UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES *ACATLÁN*

ELABORACIÓN DE MATERIAL DIDÁCTICO PARA LA ASIGNATURA DE ESTADÍSTICA APLICADA A LA COMUNICACIÓN. DESARROLLO DE CONTENIDOS Y DISEÑO INSTRUCCIONAL

T E S I S
QUE PARA OBTENER EL TÍTULO DE:
LICENCIADO EN COMUNICACIÓN
P R E S E N T A:

FRED CENTENO MORALES

ASESOR: M. en C. VÍCTOR MANUEL ULLOA ARELLANO

JUNIO DE 2006

Agradecimientos

A Dios:
Por todo lo que me ha dado y no merezco.

A mis amados padres:
María Pueblito Morales Jiménez e Ildefonso Centeno Morales, por el hermoso regalo de la existencia, por guiarme en la vida; con todo mi amor y gratitud por su infinito amor, cuidados y sacrificios para llenarme de felicidad. Los amo.

A mi amada Liliana Lejarazu:
Por compartir conmigo la vida, los buenos y los malos momentos, los sueños y las ilusiones. Por impulsarme siempre a seguir adelante y creer en mí. Por regalarme tu hermosa sonrisa y tu mirada llena de luz en todo momento. Con admiración por tu valor para enfrentar retos y cumplir metas por inalcanzables que parezcan. Gracias por darme tanta felicidad. Tu eres mi inspiración.

A mis Maestros:
Víctor Manuel Ulloa Arellano, Héctor Jesús Torres Lima, así como a (en estricto orden alfabético):

Esteban Lizama, Julio César Kala, Griselda Aguilar, Miguel Ángel Maciel, Raquél Ábrego, Verónica Quijada

Por su invaluable apoyo y entrañable amistad

A mis hermanas:
Marisol y Nancy; porque sin su vida, la mía no sería posible.

A mis amigos:

♡**Elisa**♡ Sin tu grata compañia y apoyo, el término amistad sería tan "real" como un encantacornio. TE QUIERO.

En estricto oden alfabético

Adriana Chávez♡ Aline H. Cortés♡ Bárbara Flores♡ Carolina Ramírez♡ Heriberto García ♡ Israel Piña♡ Liliana Lejarazu♡ Lupita♡ Manuel Badillo ♡ Nayhely Medel ♡ Osvaldo♡ Paty Galán♡ Salvador Jiménez♡ Suzy♡

Agradecimiento especial para:

Erica García, Liliana Cabrera y María Mercedes.

A Todos Mis Primos y Tíos:
Por su interés, confianza, apoyo y cariño, yo también los amo.
Aidé, Cheli, Edith, Lalito, los llevo en un lugar muy especial de mi alma.

Al amor ilimitado de Papá Lupe y Mamá Moye♡

A mis alegrías: Jesús (Quitush), Lucía (Mi Gushi) y Eduardo (Ponchito); siempre los querré♡♡♡♡♡

Índice general

Introducción

El presente trabajo de titulación lleva por nombre: *Elaboración de material didáctico para la asignatura de estadística aplicada a la comunicación. Desarrollo de contenidos y diseño instruccional*; dicho material, se desarrolló con el propósito de apoyar el *proceso de enseñanza-aprendizaje* en la Licenciatura de Comunicación, de la Facultad de Estudios Superiores Acatlán, en el área metodológica de su estructura curricular. Aunque el presente trabajo se apega directamente al ámbito de la *investigación en comunicación, cuya base metodológica es cuantitativa*, no excluye ni exime de su aprendizaje al estudiante interesado en áreas "apartadas" de la investigación.

Quien propone el presente material didáctico, lo hace desde las experiencias que tuvo como alumno que ya cursó la signatura en el periodo 2000 - 3, ayudante del Maestro en Comunicación Víctor Manuel Ulloa Arellano (quien impartió esta cátedra durante el periodo 2003 - 2 y que permitió retomara los apuntes, por él dictados, como base para la creación del material), así como las ayudantías en otras asignaturas del área metodológica. Lo anterior, le permitió detectar situaciones que retrasan y complican el proceso de enseñanza-aprendizaje del curso de estadística *aplicada a la investigación en comunicación*, y con su propuesta pretende coadyuvar a contrarrestar tales situaciones, así como a la potenciación del capital humano que en la UNAM se desarrolla.

Desde el momento de su formación académica hasta el de su egreso y ejercicio profesional, las herramientas formales que el presente texto contiene, *acompañarán al profesional de la comunicación en su mundo de vida cotidiana.* Ya sea como elementos constitutivos de una investigación, herramientas *para analizar-interpretar los datos duros pertinentes a un hecho de interés público*, parte sustancial de un reporte organizacional o como cultura general que permita explicar y/o informar a la sociedad lo que en su seno acontece; la estadística permitirá al comunicólogo *dar un tratamiento adecuado a los datos e información* que entre sus manos se encuentre, con lo que será posible disminuir considerablemente la superficialidad en las prácticas informativas e incrementar la precisión y exactitud de éstas.

El texto que ahora se presenta, está pensado y diseñado de acuerdo a las necesidades planteadas por el objetivo general de la licenciatura en comunicación y, específicamente, el de la asignatura en cuestión. Como se sabe, dichos objetivos son producto de un estudio *prospectivo* de carácter académico y social; elemento en que, quien suscribe, ubica la base del impacto que dicho trabajo aportará más allá de las aulas donde se pretende sirva como apoyo de profesores y estudiantes, pues el profesional de la comunicación con su trabajo sustentado formalmente, contribuirá en el análisis y la solución de las condiciones y problemas nacionales –respectivamente– relacionados con el quehacer comunicativo.

El aspecto teórico-metodológico empleado para el diseño y elaboración del material didáctico es de carácter Constructivista. Éste, se basa en el Cognoscitivismo y el Estructuralismo. A partir de lo anterior se considera, y se parte, de que el aprendizaje tiene una base psicogenética que se fundamenta en aceptar que existen estructuras abiertas a las múltiples interrelaciones de los cambios de los elementos, relaciones y códigos tanto al interior como al exterior de la estructura cognoscente, por lo que el proceso del aprendizaje nunca "parte de cero" y siempre debe ser interactivo.

De este modo, en el citado material didáctico, una condición fundamental para iniciar el abordaje del contenido programático de la asignatura, será la posesión de conocimientos básicos que permitan al estudiante comprender la nueva información que se le presenta. Por ello, en los casos que sea requerido, al inicio de cada capítulo, se presentará una *evaluación diagnóstica*.

Ésta consiste en una serie de preguntas y ejercicios planteados con la finalidad de comprobar si el estudiante cuenta con los conocimientos previos para iniciar el estudio de los contenidos en cuestión. De ser así, el proceso de aprendizaje dará comienzo en lo referente al contenido propio de la asignatura; en caso contrario, el capítulo en cuestión contendrá en el apéndice de éste, una sección con *información remedial* para solventar dicha situación.

Una vez comprendida la información remedial, se dará paso a la verificación del aprovechamiento que el estudiante consiguió. Para tal efecto, será necesaria la resolución de los *ejercicios remediales*; prueba en la que se aplicarán los conocimientos proporcionados en la sección del apéndice previamente comprendida. Una vez que el usuario del material haya cotejado sus respuestas con las que se le proporcionarán (para todos los casos en que haya ejercicios o algún tipo de evaluación habrá una sección con las respuestas pertinentes), el paso de esta sección a la del contenido capitular, será determinada –una vez más– por su desempeño; sí el aprovechamiento sigue siendo insatisfactorio, no podrá iniciar el estudio del capítulo.

En el último de los casos, los *ejercicios de asimilación* serán el recurso final para que aprenda y aplique la información remedial. En ellos, se trabajará primordialmente con *redes conceptuales* (organizadores avanzados del pensamiento) que cuentan con la información más esencial y elemental que el estudiante puede *asimilar*. Dichas redes, se encontrarán (por lo general) en pares que contengan un duplicado con espacios en blanco para ser llenados con la información faltante. En caso de fallar en éste último recurso del material, se propone al estudiante la asesoría inmediata de un profesor.

Es por todo lo anterior, que el contenido del texto se dispuso de la siguiente manera: En el *primer capítulo* se abordarán aspectos generales de la estadística, a nivel histórico y como epistemológico, que darán pauta a una reflexión que permita vincular esta rama de las ciencias formales con las disciplinas sociales y humanísticas entre las que ser encuentra la comunicación. Asimismo, se brindarán aspectos generales para introducir al estudiante en la terminología estadística que comenzará a estudiar en el capítulo tres. Así que en cierta medida, éste capitulo permitirá al estudiante la valoración de la asignatura en relación con su quehacer académico y profesional.

El capítulo dos será la *columna vertebral* del aprendizaje matemático requerido para el resto del material. En él, se recordarán elementos básicos en la teoría de conjuntos, aritmética y álgebra para que, mediante la resolución de ejercicios y comprensión de conceptos, el estudiante cuente con las herramientas matemáticas suficientes para el aprendizaje de conceptos y fórmulas estadísticas planteadas en el programa de la asignatura. Del presente capítulo, cabe destacar la inclusión de los siguientes temas: Medidas de proporción, debido a su frecuente uso en los medios, y Técnicas de conteo, por su utilidad en el diseño de muestras probabilísticas.

La organización de los datos pertenecientes a una muestra que se ha obtenido correctamente, en forma aleatoria, así como el análisis y la descripción de sus características, serán los asuntos a tratar en el capítulo número tres. En primer lugar, se expondrá la metodología para construir una *tabla de datos agrupados* (instrumento en el que se ordenan los datos de una muestra y permiten el cálculo de los estadígrafos) para dar paso a la elaboración de diferentes gráficas que posibilitan el análisis primario de los datos.

En el capítulo número cuatro, se revisarán el cálculo y análisis de los descriptores conocidos como *medidas de tendencia central*. Éstos son aquellos que cuantifican la concentración de datos respecto a un cierto valor. Se ejercitará el cálculo de la *media*, *moda* y *mediana* tanto para datos agrupados como para datos no agrupados y la *regla empírica* para calcularlas; además, se efectuará el cálculo de las medidas denominadas *cuantiles*, como una herramienta formal que permite el análisis de un conjunto de observaciones mediante su división en partes iguales.

Cabe destacar que del capítulo 4 en adelante, tanto la *evaluación diagnóstica* como la sección de *información remedia*, ya no aparecen en la estructura de los capítulos. La razón para ello, es muy sencilla. A partir de dicho capítulo, el contenido que se presenta en el material se considera nuevo para el estudiante, pues no todos durante el bachillerato cursaron alguna asignatura relacionada con la estadística y, en adelante, ya no hay secciones ni capítulos como el dos (que se enfoca en las habilidades y conocimientos matemáticos básicos), dado que se consideran cubiertos los requerimientos mínimos para el estudio de la estadística.

En el capítulo cinco, se estudiarán otro tipo de medidas para el análisis de una muestra. Éstas, son las denominadas *medidas de dispersión* o *variabilidad*, entre las que se encuentran: el *rango*, la *varianza* y la *desviación estándar*; ellas, permiten cuantificar el grado de dispersión o variación de los datos con respecto a la media. Por tanto, las medidas de dispersión se erigen como elementos complementarios de las medidas de tendencia central.

En el sexto capítulo del texto, se encuentran los elementos necesarios para calcular las magnitudes denominadas Momentos (en sus diferentes modalidades) y los elementos generales de las *medidas de asimetría y forma*. Siendo que los *momentos*, son magnitudes específicas que se pueden localizar al interior de un conjunto de datos; por su parte, los elementos generales de las medidas de asimetría, permiten determinar el tipo de sesgo que posee una distribución de datos muestrales, mientras que la *curtosis* posibilita la interpretación de la forma que presenta una gráfica. Cabe destacar que en este capítulo se encuentra el complemento de los dos anteriores, en cuanto al análisis de los datos muestrales.

En el capítulo siete, corresponde la revisión de cálculos como la *regresión* y la *correlación*. En el primer caso, se calcula la regla matemática que describe un fenómeno con relación causal mediante el *método de los mínimos cuadrados*; mientras que en el caso de la *correlación*, se cuantifica el grado de *relación lineal* que se da entre dos variables (una independiente y la otra dependiente), a través del cálculo del *coeficiente de correlación* que se traduce en la descripción matemática del comportamiento que siguen las variables al cambiar sus valores.

En el capítulo ocho, se revisarán conceptos fundamentales para el estudio de la Estadística. Esto es: una *introducción a la probabilidad*; la diferencia que existe entre los *fenómenos probabilísticos*, *deterministas*; con lo que se afianzará la pertinencia del estudio estadístico de ciertos fenómenos. Para ello se aprenderá el *concepto clásico de probabilidad*, la *ley de los grandes números* y el cálculo de la *distribución de la probabilidad*.

La *distribución normal* será revisada en el capítulo nueve. Para ello serán aprovechados temas como los de *función y relación* con sus respectivos conceptos. En este tenor, se revisará el comportamiento las curvas resultantes de una gráfica y lo que son las *variables estandarizadas*, que posibilitan al investigador, determinar la probabilidad que tiene un evento de ocurrir en la realidad.

En el décimo y penúltimo capítulo, se revisarán las técnicas apropiadas para la elección-obtención de una *muestra* (y su tamaño) que puede o no ser probabilística. Se estudiarán diferentes *tipos de muestreo* junto con las características que los diferencian, así como el *teorema del límite central* y el *nivel de confianza*.

En el último capítulo pertinente al contenido de la asignatura (decimoprimero capítulo), se revisarán diferentes medidas estadísticas que permiten cuantificar los cambios en las variables a través del tiempo. A estas medidas se les conoce como *números índice*. De éstas se revisarán algunas propiedades y diferentes formas de calcularlos. Entre otros se revisarán el *índice de Laspeyres*, de *Paasche* y de *Fischer* con sus respectivas fórmulas y aplicaciones en la economía.

Así, el material didáctico aquí presentado, obedece a necesidades muy específicas de la licenciatura para la cual se diseñó. En primer lugar, resuelve la falta de un producto comunicativo especializado en *estadística aplicada a la comunicación*; con este material, estudiantes y maestros podrán potenciar el tiempo dedicado al estudio de la asignatura y agilizar el proceso de enseñanza-aprendizaje a lo largo del semestre. Lo anterior, se traduce en el máximo aprovechamiento del tiempo consagrado al trabajo intelectual –ya sea en el aula o fuera de ésta.

Otro de los problemas que el presente material didáctico coadyuvará a solucionar, es la acreditación extraordinaria de la asignatura. En particular, un grupo de estudiantes que se favorecerá, es el de aquellos que por diversos motivos adeudan la asignatura y desean acreditarla por sus propios medios. Puesto que en el material se encuentran sistematizados los contenidos del programa, éste se constituirá como una excelente guía de estudios para quienes presentarán el examen extraordinario correspondiente, ya que además de los contenidos mencionados, contará con ejemplos, ejercicios, autoevaluaciones y elementos matemáticos básicos-necesarios para facilitar el estudio y aprendizaje de las herramientas estadísticas propuestas por el programa de la asignatura.

Ya que la asignatura, *estadística aplicada*, se encuentra en la base del *área metodológica* y su relevancia va más allá del semestre en que se encuentra situada dentro de la estructura curricular; el material propuesto también trasciende la asignatura para la que fue diseñado. La estadística es un elemento capital en el diseño y realización de investigaciones que el estudiante profesional de la comunicación será capaz de plantear y ejecutar; de tal manera que con este material, el estudiante podrá enfrentar y solucionar problemas metodológicos de otras asignaturas, trabajos como el de titulación y de su vida en el mercado de trabajo. De tal manera que las dificultades posteriores al segundo semestre serán aminoradas, e incluso solventadas, con el uso del texto en cuestión.

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

MATERIAL DIDÁCTICO PARA LA ASIGNATURA DE ESTADÍSTICA APLICADA A LA COMUNICACIÓN. DESARROLLO DE CONTENIDOS Y DISEÑO INSTRUCCIONAL

FRED CENTENO MORALES

JUNIO DE 2006

Si uno logra medir lo que está diciendo y lo puede expresar en números, es que sabe lo que dice; pero si no lo puede expresar con números, es que el conocimiento que tiene de ello es escaso e insatisfactorio.

–Lord Kelvin–

Esta Tesis forma parte del proyecto EN311004 del Programa de Apoyo a Proyectos Institucionales para el Mejoramiento de la Enseñanza (PAPIME)

Objetivos

Objetivo General

- *El alumno conocerá y aplicará las bases instrumentales y los elementos básicos de la estadística aplicados a las Ciencias Sociales y a la Investigación en Comunicación.*

Contenidos mínimos de la asignatura

1. Aspectos básicos de la estadística
2. Organización de los datos para que transmitan un significado: tablas y gráficas
3. Medidas de tendencia central
4. Medidas de variabilidad
5. Números índice
6. Regresión y correlación simples
7. Técnicas de muestreo.

Introducción al material didáctico

Como estudiante de la Licenciatura en Comunicación de la Facultad de Estudios Superiores Acatlán (FES Acatlán), la reflexión en torno a los elementos que intervienen en el proceso de formación profesional, fue una concomitante, para quien suscribe, que derivó en la detección de algunas ventajas y áreas de oportunidad que en dicho proceso intervienen, v. gr.:

- En Acatlán, la Licenciatura en Comunicación es una carrera multifacética; su núcleo de conocimientos, comprende una rica gama de áreas tales como pensamiento filosófico, ético, político, histórico, científico, tecnológico, social, teórico, metodológico y humanístico. Esta variedad de conocimientos, respalda a la carrera y a los egresados de Acatlán, en el amplio mundo académico de la comunicación social, así como en el saturado y, por tanto, competido mercado de trabajo al que los estudiantes y egresados profesionales de la Licenciatura se integran.

- Debido a las características que predominan en el campo de trabajo y el mercado laboral, las habilidades, capacidades y competencias necesarias para una exitosa carrera profesional, son cada vez mayores tanto a nivel interior como a nivel exterior. Así, la obtención de un trabajo depende de las herramientas formales con que el profesional de la comunicación se enfrenta al mundo.

- La planta docente es buena pero hace falta más, un respaldo que acompañe incondicionalmente al estudiante. Por ello, es menester llevar a cabo un constante proceso formal de evaluación-actualización del proceso enseñanza-aprendizaje y, en consecuencia, el diseño de estrategias y materiales didácticos de vanguardia que permitan la congruencia entre la formación de profesionistas y las exigencias de la sociedad, el campo laboral y el mercado de trabajo.

En respuesta a dicha reflexión y a las necesidades someramente planteadas en esta introducción, es que surge la idea de generar el material didáctico que ahora tienes en tus manos. Éste, se encuentra diseñado para la enseñanza-aprendizaje de la estadística y su aplicación en el área metodológica de la carrera en Comunicación. La finalidad del texto, es impulsar y potenciar el aprendizaje de los elementos propuestos del área metodológica, lo que implica el integral desarrollo de capital humano que en la carrera se cultiva; cuestión que, en gran medida, puede coadyuvar a mantener la consolidación de la carrera de Comunicación en Acatlán, como una de las mejores del país.

Ciertamente, el contenido temático de Estadística es extenso ya que incluye aspectos básicos de la estadística, antecedentes matemáticos, organización de los datos para que transmitan un significado (tablas y gráficas), medidas de tendencia central, medidas de variabilidad, números índice, regresión y correlación simples y técnicas de muestreo. No obstante, el temario se puede cubrir de manera suficiente con el apoyo del presente material didáctico; para ello, es necesario que conozcas sus características y consideres las recomendaciones propuestas.

El enfoque bajo el cual fue concebido este material, es de carácter Constructivista; asimismo, se basa en el Cognoscitivismo y el Estructuralismo, por ello, se parte de la convicción de que "el estudiante no es un recipiente hueco al que se debe llenar como sea y con lo que sea", sino que es un individuo con cierta historia de vida que le dota de conocimientos iniciales –más o menos iguales-diferentes de los que poseen las personas que le rodean al momento de inicial la aventura del conocimiento– que le sirven de base para incorporar a su universo simbólico nuevos conceptos con los que puede re-conocer el mundo que le rodea e interesa.

Sobre dicho entendido, se sabe que en ocasiones el conocimiento del que se parte para aprender nuevas cosas, no es igual en todas las personas aunque se encuentren en el mismo nivel de alguna estructura curricular. Por ello, la forma en que se presentan los contenidos de la asignatura en este material, es el siguiente:

- **Preliminar**. Es el preámbulo del capítulo. En él, se introduce a los contenidos que se abordarán y contiene como elementos adicionales:
    - **Evaluación diagnóstica**. En esta sección, se evalúan los conocimientos previos que el estudiante debe poseer como la base que le permitirá comprender el contenido del capítulo en cuestión. Cabe destacar, el hecho de que esta sección se encuentra presente hasta el capítulo 3, ya que del 4 en adelante, se considera al contenido del capítulo como totalmente nuevo y su base es el contenido de los capítulos 1 al 3, así como los anteriores a cada nuevo capítulo.
    - **Respuestas a la evaluación diagnóstica**. Las respuestas a cada actividad propuesta y problema planteado a lo largo del texto, se encuentran en el material, con la finalidad de posibilitar el estudio en forma autodidacta.

- **Ejercicios de recuperación**. En caso de error al contestar las preguntas planteadas en la evaluación diagnóstica, tienes que resolver los ejercicios de este apartado, para comprobar que adquiriste el conocimiento necesario para el estudio del contenido propio del capítulo (previa lectura del la sección de *Información remedial*).
- **Respuestas a los ejercicios de recuperación**. Apoyo a tu proceso de estudio, para verificar que tus respuestas son las apropiadas, evento después del cual podrás pasar al estudio del contenido planteado para el capítulo.
- **Ejercicios de asimilación**. En caso de un nuevo error (al resolver los ejercicios de recuperación), deberás revisar nuevamente la información remedial, y repasar los elementos básicos de ésta, para que se incorporen a su universo simbólico. Si te equivocas nuevamente, como estrategia de aprendizaje se te sugiere que consultes con un profesor del área o asignatura para que resuelva tus dudas.
- **Respuestas a los ejercicios de asimilación**. Cabe destacar que cada una de las secciones, cuenta con referencias que indican en qué página se encuentran las respuestas correspondientes a los ejercicios y evaluaciones como ésta.

- **Contenido del capítulo**. En adelante se revisan varios aspectos de interés para la asignatura y, por ende, tu formación. Para afirmar tus conocimientos, esta parte del texto cuenta con varias series de ejercicios y sus respectivas respuestas, que encontrarás en el apéndice del capítulo.
- **Contenido...**
- **Contenido...**

- **Apéndice del capítulo**. Esta sección es de apoyo al capítulo y cuenta con elementos tales como:

  - **Información remedial**. Este apartado se hace presente en los capítulos 1, 2 y 3; ya que su lectura, te permitirá resolver apropiadamente los ejercicios de recuperación y de asimilación en caso de ser necesario. Recuerda que del capítulo 4 en adelante, la información remedial ya no es necesaria, puesto que la base de los conocimientos necesarios para el estudio de estos capítulos se encuentra en los tres primeros y los ulteriores.
  - **Respuestas a los ejercicios del capítulo**. Para verificar que tu desempeño es el adecuado al estudiar y resolver los ejercicios planteados en el capítulo.
  - **Respuestas de la autoevaluación**. Sección en la que podrás obtener un indicador de tu aprovechamiento, al comprobar que mediante la lectura, elaboración de apuntes y resolución de ejercicios, te es posible solucionar diversos problemas planteados.
  - **Glosario**. Este se te proporciona con el fin de agilizar tu proceso de aprendizaje, ya que el acudir a buscar la definición de un concepto de difícil comprensión o desconocido, puede distraerte. En este sentido, cabe destaca que una buena lectura es aquella en que al finalizar el recorrido del texto en cuestión, todos los términos que lo conforman quedan claros al lector; situación que se puede traducir como un factor incidente en aprovechamiento terminal.

Como podrás notar, el presente material, busca fortalecer el área metodológica, específicamente la de estadística, a través del desarrollo y sistematización de los contenidos del programa, así como la propuesta de estrategias didácticas para la enseñanza-aprendizaje de la Estadística; exprofeso para alumnos de Comunicación y de otras carreras de Humanidades. Este material, se planeó haciendo énfasis en la parte afectiva, motivacional y emocional que todo proceso de estudio y aprendizaje implica. Asimismo, el material se diseñó de manera que los contenidos de la asignatura se relacionen con otras asignaturas que cursas, para que puedas aplicar el conocimiento de la Estadística en ellas.

Por todo lo anterior, es importante que sigas las instrucciones que en el material se encuentran, ya que pueden ser un gran apoyo para ti como estudiante. Asimismo, este material puede ser de gran utilidad para el profesor, que imparta la asignatura, cuya profesión no sea la de Comunicólogo. El material permitirá orientar al profesor acerca del enfoque que debe dar a la enseñanza de la Estadística para la Licenciatura en Comunicación y carreras afines; por lo que es de suma importancia la aplicación y evaluación del presente texto.

Capítulo 1

Aspectos básicos de la estadística

Objetivos específicos

- El estudiante explicará cómo la incorporación de las matemáticas en general y de la estadística en particular, a la Investigación Social, reforzó el pensamiento científico en esa área, mediante la revisión de su historia, fundamentos y objetivos.

- El estudiante reconocerá a la Estadística, como una herramienta en la Investigación de la Comunicación Social, a través del análisis de los elementos presentados en este capítulo.

1.1. Preliminar

1.1.1. Evaluación diagnóstica

El aprendizaje es un proceso en el cual se requiere que el estudiante posea conocimientos previos que le sirvan como base para comprender la nueva información que enfrentará; a continuación, se presentan una serie de ejercicios para recordar elementos indispensables con el fin de comprender el material que conforma la presente unidad.

Instrucciones:

- Señala con un círculo la respuesta correcta y/o responde brevemente y con precisión lo que se te pide, en un espacio no mayor al indicado.
- Al finalizar verifica que tus respuestas coincidan con las del subapartado **1.1.2. Rspuestas**, que se localiza en la página 7
- Si todas tus respuestas fueron correctas, pasa al apartado **1.2. Orígenes de la estadística**, en la página 17
- Si tuviste errores, en la sección **1.1.2.** encontrarás elementos instruccionales que deberás seguir para corregir tus fallas y, en consecuencia, facilitar el aprendizaje del contenido propio del capítulo y la asignatura.

1. ¿Qué es un sistema?

____

____

____

____

____

____

____

2. Elige la opción que corresponda al concepto de sistema.

   a) Las cosas suspendidas en el viento durante un tornado.

   b) El conjunto de animales salvajes que huyen de un incendio.

   c) El conjunto de todos los peces que habitan en el fondo del mar.

   d) El conjunto de los puntos parasitarios en blanco y negro en la pantalla de T. V.

3. ¿Qué es un Acto Expresivo?

4. ¿Cuál es la principal diferencia entre los actos ejecutivos y los expresivos?

5. El trabajo de un científico consiste en estudiar las cosas y sus comportamientos para:

   a) Escribir reportes, discutir sus conclusiones con la comunidad científica y divulgar los resultados obtenidos al público en general.

b) Describir, explicar y predecir, en la medida de lo posible, algunos aspectos y/o propiedades de los fenómenos que se estudian para generar conocimiento científico.

c) Descubrir, Inventar y renovar cosas para mejorar la calidad y el nivel de vida en la sociedad.

6. ¿A qué se debe la precisión que caracteriza a la matemática?

   a) A que es una ciencia que sólo estudia relaciones entre elementos abstractos (ideales) y por tanto la veracidad de éstas no se pueden someter a comprobación empírica.

   b) Son ciencias puras en las que todo mundo está de acuerdo con los resultados, sin importar a qué se refieran.

   c) A que se rige por una estructura estrictamente apegada al sentido del deber ser, así: $2 + 3 = 5$ bajo cualquier circunstancia y lugar.

7. Se llama así a las ciencias que se encargan de hechos que se pueden comprobar mediante la experiencia para contrastar sus fórmulas:

   a) Ciencias Naturales.

   b) Ciencias Factuales

   c) Ciencias Positivas

8. ¿Qué es el Método Científico?

____

____

____

____

____

____

____

____

9. ¿Cuáles son las principales características del Conocimiento Científico?

____

____

____

____

____

____

____

10. De la siguiente lista, elige un caso en el cual se esté efectuando un proceso de comunicación social.

    a) Acto proselitista en la explanada del Palacio Municipal.

    b) Transmisión del tema musical de moda.

    c) El proceso de adiestramiento de un perro policía.

    d) La elaboración de un material didáctico para Estadística.

11. Explica tu respuesta.

____

____

____

____

____

1.1.2. Respuestas

En caso de cualquier error, *no te preocupes*: Te encuentras en el proceso de *construcción del conocimiento* y el primer paso para *crecer sólidamente*, es detectar tus **áreas de oportunidad** con el fin de *no construir* sobre "arenas movedizas".

- Los recuadros que encontrarás a lo largo de esta sección, agrupan las respuestas por tema; en dichos elementos, encontrarás el tema a que corresponden las preguntas anteriores a él, así como las indicaciones que deberás seguir en caso de error.

- Finalmente:

    - Si obtuviste un número de aciertos mayor o igual que siete y menor a once, revisa las instrucciones pertinentes a las respuestas incorrectas para corregirlas.

    - En caso tener menos de siete aciertos, revisa todo el contenido de la sección **1.1.7. Información remedial** (pág. 56) y responde nuevamente la sección **1.1.1. Evaluación diagnóstica** (pág. 3).

    - Si te volvieras a equivocar, pasa a la sección **1.1.3. Ejercicios de asimilación** (pág. 11).

1. Un **sistema** es un conjunto *ordenado* de elementos. Un todo dentro del cual los *componentes* (elementos) que lo conforman, también pueden ser *subsistemas* y guardan *relaciones solidarias* entre ellos, así como con otros elementos y sistemas (*dinámica de sistemas*). Dichas relaciones, producen *afectaciones mutuas* y cambios en los sistemas.

2. **c)** El conjunto de todos los peces que habitan en el fondo del mar.

Si una o dos de tus respuestas anteriores no fue(ron) adecuada(s), atiende a las siguientes instrucciones:
Lee nuevamente la respuesta y revisa el tema: *Teoría de sistemas* (pág. 56) del apartado 1.7.1.
Luego de la revisión, intenta resolver correctamente la pregunta número 1 y razona el porqué de la 2^a respuesta.

3. Un **Acto Expresivo** es el *comportamiento* encaminado al logro de un objetivo a partir de la *producción de señales* relacionadas con *representaciones* (trabajo expresivo). Éstas son adecuadas para lograr la *interacción* con el Otro a partir del *intercambio de información* situada dentro de un contexto (marco de referencia).

4. La *energía aplicada* para el logro de objetivos *es menor* cuando se recurre a los **Actos Expresivos**, puesto que la interacción se genera a partir del intercambio de información; mientras que en los **Actos Ejecutivos** se recurre al *trabajo físico directo* para mover un objeto o influir en el Otro (coactuación).

Si entre tu(s) error(es) se encuentran las respuestas 3 y/o 4, atiende a las siguientes instrucciones:
Lee nuevamente la respuesta y revisa el tema: *Tipología de actos* (pág. 58) del apartado 1.7.1.
Luego de la revisión, intenta resolverla(s) correctamente

5. **b)** Describir, explicar y predecir, en la medida de lo posible, algunos aspectos y/o propiedades de los fenómenos que se estudian para generar conocimiento científico.

Si tu respuesta número 5 no fue adecuada, atiende a las siguientes instrucciones:
Lee nuevamente la respuesta y revisa el tema: *Objetivos de la ciencia* (pág. 59) del apartado 1.7.1.
Luego de la revisión, intenta resolverla correctamente

6. **a)** En que es una ciencia que sólo estudia relaciones entre elementos abstractos (ideales) y por tanto la veracidad de éstas no se pueden someter a comprobación empírica.

7. **b)** Factuales.

Si entre tu(s) error(es) se encuentran las respuestas a las preguntas 6 y/o 7, entonces:
Lee nuevamente la respuesta y revisa el tema: *Clasificación de las ciencias* (pág. 59) del apartado 1.7.1.
Luego de la revisión, intenta resolver correctamente la pregunta número 6 y razona el porqué de la 7

8. **El método científico** es el *procedimiento lógico*, previamente fijado, que se orienta a la consecución del conocimiento científico. Implica toda una serie de *conocimientos teóricos, conceptuales, filosóficos y técnicos* que en su conjunto, remiten a *una forma de pensar y actuar con, en y sobre el mundo*, de manera estratégica y *congruente* con las necesidades que plantean los diversos *problemas intelectuales* y se aplica al ciclo entero de la investigación de manera general.

Si tu respuesta no fue la adecuada, atiende a las siguientes instrucciones:
Lee nuevamente la respuesta y revisa los siguientes conceptos: **Método** y **Método Científico** en la sección **1.6.1. Glosario** (pág. 77)
Luego de la revisión, intenta resolver correctamente la pregunta número 8

9. El **conocimiento científico** es *racional, objetivo, sistemático, y falible*. Racional porque se ha obtenido mediante el método de la ciencia; objetivo porque es impersonal y se niega a admitir entidades o explicaciones no naturales; sistemático porque aspira a la coherencia lógica de enunciados fundados y contrastables; asimismo, es falible porque se puede volver a someter a prueba, enriquecerse y, llegado el caso, superarse mediante el mismo método.

Si tu respuesta no fue la adecuada, atiende a las siguientes instrucciones:
Lee nuevamente la respuesta y revisa el tema: **El conocimiento científico** (pág. 61) del apartado 1.7.1.
Luego de la revisión, intenta resolver correctamente la pregunta número 9

10. **a)** Acto proselitista en la explanada del Palacio Municipal.

11. El **acto proselitista** es un proceso de **comunicación social** porque es propio de la actividad social humana. En él, incurren *actores, instrumentos, expresiones y representaciones* pertenecientes al *sistema social*. Hay mutua influencia (interacción) entre los actores (que son agentes sociales), por el recurso a *actos expresivos* y con ello, le imprimen energía e influyen en la dinámica del sistema social.

Si entre tu(s) error(es) se encuentran las respuestas 10 u 11, atiende a las siguientes instrucciones:
Lee nuevamente la respuesta y revisa el tema: **Comunicación social** (pág. 65) del apartado 1.7.1.
Luego de la revisión, intenta resolver correctamente las preguntas 10 y 11

1.1.3. Ejercicios de asimilación

- *Instrucciones*: Lee una vez más la información pertinente a cada una de las respuestas del cuestionario anterior.
- Coloca sobre cada línea la palabra que complete correctamente los conceptos

1. Un sistema es conjunto ______________ de elementos. Un todo dentro del cual los __________________ (elementos) que lo conforman, también pueden ser __________________ y guardar relaciones _____________ entre ellos, así como otros elementos y sistemas (_________________________). Dichas relaciones producen ________________________ y cambios en los ____________.

sistemas	solidarias	componentes	dinámica de sistemas
ordenado	subsistemas	afectaciones mutuas	

2. Un *acto expresivo* es el ________________________ encaminado al logro de un objetivo a partir de la ___________________ *de* ________________ relacionadas con _________________ (trabajo expresivo). Éstas son adecuadas para lograr la *interacción* con el Otro a partir del _________________ *de* _____________________ situada dentro de un contexto (marco de referencia).

comportamiento	representaciones	producción
información	señales	intercambio

3. En los actos _________________ se recurre al trabajo ______________ directo para mover un objeto o __________ en el Otro (coactuación).

físico	influir	ejecutivos

4. El trabajo de un científico (la investigación científica) consiste en describir, __________ y ____________, en la medida de lo posible, algunos aspectos y/o ____________ de los fenómenos que se estudian para __________ conocimiento científico.

generar	propiedades	explicar	predecir

5. La exactitud de la matemática se debe a que pertenece a las ciencias ____________, es decir: sólo se encargan del estudio de las relaciones entre elementos ______________ (ideales) y por tanto la veracidad de éstas no se pueden someter a comprobación __________.

formales	abstractos	empírica

6. Las ciencias __________ se refieren a __________ que se supone ocurren en el mundo, y, consiguientemente, tienen que apelar a la __________________ para ____________ sus fórmulas.

experiencia	contrastar	hechos	fácticas

7. **El método científico** es el ______________________________, previamente fijado, que se orienta a la consecución del conocimiento científico. Implica todo una serie de *conocimientos* ________________, *conceptuales, filosóficos y* ________________ que en su conjunto, remiten a *una forma de pensar y* ______________ *con, en y sobre el mundo*, de manera ________________ y *congruente* con las necesidades que plantean los diversos problemas __________________ y se aplica al ciclo entero de la investigación de manera general.

estratégica	procedimiento lógico	teóricos	actuar
intelectuales	técnicos		

8. El **conocimiento científico** es *racional, objetivo, sistemático, y falible* ______________ porque es impersonal y se niega a admitir entidades o explicaciones no naturales; ______________________ porque se ha obtenido mediante el método de la ciencia; ______________ porque aspira a la coherencia lógica de enunciados fundados y contrastables; asimismo, es ______________ porque se puede volver a someter a prueba, enriquecerse y, llegado el caso, superarse mediante el mismo método.

sistemático	falible	objetivo	racional

9. El proceso de comunicación social es una actividad en la que incurren los siguientes elementos: ____________, que pueden ser individuos organizaciones o instituciones que se desenvuelven en el Sistema ____________; instrumentos, que pueden ser ____________ o tecnológicos y están al servicio de la comunicación; ______________________, que son sustancias informadas (con forma) que designan algo para alguien y las ________________ que dotan de sentido a las expresiones; dichos elementos pertenecen al ______________ social; por ende, la comunicación social es propia de la actividad relacional humana.

sistema	social	actores	expresiones	biológicos	representaciones

Ahora coteja tus respuestas con los de la sección correspondiente (véase **1.1.4. Respuestas a los ejercicios de asimilación** en la página 14), corrige los posibles errores y estudia la información contenida en las respuestas; si aún con lo anterior, no te queda claro alguno de los conceptos aquí presentados, solicita la ayuda de algún profesor.

1.1.4. Respuestas a los ejercicios de asimilación

1. Un sistema es un conjunto **ordenado** de elementos. Un todo dentro del cual los **componentes** (elementos) que lo conforman, también pueden ser **subsistemas**y guardar relaciones *solidarias* y **funcionales** entre ellos, así como otros elementos y sistemas (**dinámica de sistemas**). Dichas relaciones producen **afectaciones mutuas**y cambios en los **sistemas** y/o subsistemas

sistemas	funcionales	componentes	dinámica de sistemas
ordenado	subsistemas	afectaciones mutuas	

2. Un *acto expresivo* es el **comportamiento** encaminado al logro de un objetivo a partir de la **producción** de **señales** relacionadas con **representaciones** (trabajo expresivo). Éstas son adecuadas para lograr la *interacción* con el Otro a partir del **intercambio** de **información** situada dentro de un contexto (marco de referencia).

comportamiento	representaciones	producción
información	señales	intercambio

3. En los actos **ejecutivos** se recurre al trabajo **físico** directo para mover un objeto o **influir** en el Otro (coactuación).

físico	influir	ejecutivos

4. El trabajo de un científico (la investigación científica) consiste en describir, **explicar** y **predecir**, en la medida de lo posible, algunos aspectos y/o **propiedades** de los fenómenos que se estudian para **generar** conocimiento científico.

generar	propiedades	explicar	predecir

5. La exactitud de la matemática se debe a que pertenece a las ciencias **formales**, es decir: sólo se encargan del estudio de las relaciones entre elementos **abstractos** (ideales) y por tanto la veracidad de éstas no se pueden someter a comprobación **empírica**.

formales	abstractos	empírica

6. Las ciencias **fácticas** se refieren a **hechos** que se supone ocurren en el mundo, y, consiguientemente, tienen que apelar a la **experiencia** para contrastar sus fórmulas.

experiencia	hechos	fácticas

7. El método científico es el **procedimiento lógico**, previamente fijado, que se orienta a la consecución del conocimiento científico. Implica toda una serie de *conocimientos* **teóricos**, *conceptuales, filosóficos* y **técnicos** que en su conjunto, remiten a *una forma de pensar y* ***actuar*** *con, en y sobre el mundo*, de manera **estratégica** y *congruente* con las necesidades que plantean los diversos problemas **intelectuales** de que se encarga la ciencia, y se aplica al ciclo entero de la investigación de manera general.

Estratégica	procedimiento lógico	teóricos
actuar	intelectuales	técnicos

8. El conocimiento científico es *racional, objetivo, sistemático, y falible.* **objetivo**, porque es impersonal y se niega a admitir entidades o explicaciones no naturales; **racional**, porque se ha obtenido mediante el

método de la ciencia; **sistemático**, porque aspira a la coherencia lógica de enunciados fundados y contrastables; asimismo, es **falible**, porque se puede volver a someter a prueba, enriquecerse y, llegado el caso, superarse mediante el mismo método.

sistemático	falible	objetivo	racional

9. El proceso de comunicación social es una actividad en la que incurren los siguientes elementos: **actores**, que pueden ser individuos organizaciones o instituciones que se desenvuelven en el sistema **social**; instrumentos, que pueden ser **biológicos** o tecnológicos y están al servicio de la comunicación; **expresiones**, que son sustancias informadas (con forma) que designan algo para alguien y las **representaciones** que dotan de sentido a las expresiones; dichos elementos pertenecen al **sistema** social; por ende, la comunicación social es propia de la actividad relacional humana.

sistema	social	actores	expresiones	biológicos	representaciones

1.2. Orígenes de la estadística

1.2.1. El Conocimiento en la antigüedad

"La ciencia está constituida por tradiciones de razonamiento y, en particular, por patrones de **explicación** que se constituyen y desarrollan históricamente a lo largo de los siglos."[1] Por naturaleza, el ser humano es una especie cuya condición racional le hace *especial e inevitablemente curioso*. El hombre, a lo largo de su historia, se caracteriza por la intensa búsqueda de explicaciones adecuadas a los **fenómenos** que le rodean y afectan; esto incluye desde aquellos en los que se encuentra involucrado él mismo y sus semejantes, hasta los que ya por la distancia o por su naturaleza propia, le son momentáneamente inaccesibles. En dicha *curiosidad-racional* se encuentra el origen de la investigación y la **ciencia** (y/o la investigación científica) tal y como la conocemos hoy día.

El antecedente más remoto de la ciencia se encuentra en la magia, la hechicería, la religión y, en un sentido más amplio: en la relación del hombre con la divinidad, lo ultraterreno; pues si bien es cierto que las ciencias son una herramienta para conocer y explicar la realidad. "El mito, la religión, la magia, la superstición, también son explicaciones (a su manera), de esta realidad"[2] a lo largo de la historia. A partir de la creación de dichas cuestiones místicas, generadas, albergadas y re-creadas en el imaginario social, es que se comienza a estructurar y reflejar el **conocimiento** que el ser humano tiene de su entorno, de sí mismo –como especie e individuo– y el interés por explicarse los fenómenos que le circundan, afectan y de los que forma parte.

[1] MARTÍNEZ, Sergio F.; **De los Efectos a las Causas. Sobre la Historia de los Patrones de Explicación Científica**; Paidós; México; 1997; pp. 15 Negritas agregadas

[2] COLMENARES, Ismael; et. al. **De la Prehistoria a la Historia**; México; Ediciones Quinto Sol; 1994; pp. 20

Las anteriores formas para explicarse la realidad, no son sino respuestas infundidas por las partes volitiva y afectiva del espíritu, pues la naturaleza humana aún carecía de la experiencia necesaria para integrarlas entre sí, de manera más o menos equilibrada, junto a la razón. Así, el pensamiento mágico, místico y/o religioso, alcanzó la cúspide como una herramienta para explicar la realidad.

Poco a poco se generaron las condiciones necesarias para iniciar un lento descenso en el estatus de esta tradición explicativa de la realidad, por lo que paulatinamente ya no pudo satisfacer, *del todo*, la curiosidad natural del ser humano; y con ello, se inició la tradición explicativa sustentada por el conocimiento a partir del sentido común. Claro, el proceso no es tan sencillo como podrían dibujarlo estas líneas; tampoco obedece a uno o dos factores como causas directas: es el resultado de esa natural inquietud humana que vinculada con la racionalidad, el azar y todas las esferas en que se desenvuelve el ser humano, dan como resultado un complejo y paulatino proceso de evolución cultural.

Con el paso del tiempo y el progresivo perfeccionamiento de las facultades y habilidades que le han permitido al hombre descollar del resto de las especies, las explicaciones de los fenómenos que se presentan al espíritu humano, fueron estructuradas de una forma mucho más elaborada desde un punto de vista lógico, natural y objetivo; es decir: las inferencias aumentan su nivel de complejidad hasta transformarse en verdaderos razonamientos y las sentencias con que se pretenden explicar los fenómenos, se generan y rigen por los datos obtenidos mediante los sentidos. Resultado: El re-conocimiento de "nuevas" aristas intrínsecas en la naturaleza de los acontecimientos que interesan al ser humano, aún sesgadas por la imprecisión típica que ofrecen los sentidos cuado carecen de un apoyo teórico-conceptual y metodológico-técnico sistematizados, pero ya fundamentadas en mayor medida a partir del equilibrio entre el deseo, la voluntad y la razón.

Tales condiciones se reforzaron cuando el hombre dejó la vida nómada y su condición gregaria para convertirse en un ser sociable (que vive en sociedad y se relaciona poco más o menos armónicamente con sus semejantes); es en este momento que se generó una condición más que coadyuvó al proceso de evolución humana (en el más ámplio sentido de la palabra evolución), esto es: "la formación de la ciudad fue lo que hizo posible los progresos técnicos, y con ellos, todo un conjunto de invenciones intelectuales, económicas y políticas: los números, la escritura, el comercio, la evolución del nuevo sistema de clases y del gobierno organizado. Entonces empezó a surgir una ciencia conciente y las disciplinas distinguibles de la astronomía, la medicina, y la alquimia adquirieron sus primeas tradiciones."[3]

1.2.2. Ciencia y estadística en las primeras culturas de la antigüedad

Las primeras sociedades de que se tiene cuenta en la antigüedad son la egipcia y la babilónica; en ellas, el proceso evolutivo de la naciente ciencia, se originó debido a factores de esencial importancia para el desarrollo del capital humano. Las condiciones materiales, económicas (excedentes de producción generados, principalmente, por la agricultura y la ganadería), sociales (la coexistencia en ciudades bajo la tutela de un contrato social que resguarda al hombre de peligros tanto internos como externos) y el desarrollo técnico, le concedieron al ser humano tiempo para el ocio. Ello le permitió tener mayor libertad espiritual para contemplar, pensar, imaginar y cuestionar tanto los

[3]ROSAS, y Novelo María de Lourdes; **La Ciencia y el Método Experimental**; México; Universidad Nacional Autónoma de México; 1997; pp. 10

fenómenos concretos como los abstractos y contrastarlos sensual, cognitiva e incluso, afectivamente. Así comenzó un ciclo de rupturas, reconstrucciones, mezclas y/o sincretismos de antiguos y nuevos mitos, tabúes, leyendas, dogmas, creencias y conocimientos.

El conocimiento del entorno en dichas civilizaciones ya no se daba únicamente a partir de elementos divinos o ultraterrenos, sino que se crean y sistematizan herramientas formales (ideales, como la matemática) que aun y cuando sólo existen en la mente humana, se aplican a objetos factuales (materiales, reales: como los naturales y sociales). El saber se transforma en conocimiento al volverse complementarias estas dos esferas (la formal y la factual). Un elemento de suma importancia para esta transformación, es la matemática que se comienza a constituir como "herramienta para realizar la más precisa reconstrucción de las complejas relaciones que se encuentran entre los hechos y entre los diversos aspectos de los hechos."[4] Aunque en el caso de las primeras culturas, el término complejidad aún no resulta muy apropiado en cuanto a sus afanes.

Tal es el caso de los egipcios, quienes utilizaban sus conocimientos en matemática, para la resolución de problemas prácticos. Asentados en la orilla del Nilo, ellos, aprovecharon y lidiaron con las inundaciones anuales de éste. Cada temporada tenían que trazar las lindes desaparecidas por la invasión del limo (que era utilizado como fertilizante natural). Así nació la Geometría. En el caso de la Estadística, ésta se encuentra directamente relacionada con actividades tales como la recaudación de impuestos así como el control y mantenimiento del imperio mediante el conocimiento de los recursos como son la población, sus características y muy en especial la industria

[4]BUNGE, Mario; **La Ciencia, su Método y su Filosofía**; Buenos Aires; Ediciones Siglo Veinte; 1975; pp. 10

bélica (soldados, armamento, recursos militares). Ellos, calculaban la carga tributaria en función del índice de fertilidad que se obtenía a partir del caudal del Nilo.

Así, hemos llegado al punto en que encontramos nuestro primer referente directamente ligado a la Estadística, pues dicho vocablo "está relacionado con la palabra «estado», y originalmente la actividad llamada estadística fue una clase sistemática de ciencia política, comparada. Esta actividad, se centró gradualmente en tablas numéricas de hechos económicos, demográficos y políticos, y así, «estadística» vino a significar la recopilación y análisis de tablas numéricas"[5] cuyo propósito era estrictamente administrativo. He aquí, un caso concreto en el que se puede apreciar la utilidad de la naciente ciencia, donde al ser aplicada, el ser humano piensa y actúa *en*, *sobre* y *con* el mundo.

En el mundo por ser la cosa a la cual objetiva mediante su cuantificación, por ser el ente en que centra y al cual aplica sus pensamientos-reflexiones formales y es, al mismo tiempo, el lugar donde se desarrolla. *Sobre* el mundo, porque es la entidad que manipula, transforma y domina tanto física como abstractamente, es el lugar al cual intenta señorear (y en algunos aspectos lo consigue) mediante la aplicación de sus invenciones (la administración y control de los recursos cuantificados y registrados en las tablas numéricas). *Con* el mundo, porque dicha entidad no sólo está constituida por elementos materiales, sino que una importante porción de él, está ocupada por entes psicofísicos semejantes con los que interactúa a partir de diversas herramientas, entre las que se encuentra la estadística al servicio de la relación existente entre gobernados y gobernantes.

[5]SILLS, L. David; **Enciclopedia Internacional de las Ciencias Sociales**; Tomo 4; Madrid; 1974; Aguilar S. A. de Ediciones; pp. 392

Hacia el año 5000 a C. en la cultura babilónica ya realizaban cuantificaciones de todo lo que les pertenecía (ejemplo: el ganado), registrando en tablas de arcilla sus posesiones materiales, con el propósito de controlar y administrar los recursos de que podía disponer el Estado. De manera que, en principio, la actividad estadística es de carácter puramente descriptivo, únicamente da cuenta de los recursos que el Estado posee para subsistir. Todo parece indicar que la única técnica estadística con la que se contaba a la fecha, era el levantamiento de censos, proceso laborioso, costoso e impreciso en cuanto a los resultados que arroja y las decisiones que se pueden tomar en función de éstos, aún cuando las poblaciones era menores en comparación con las actuales.

Con los progresos culturales y técnicos[6] de la civilización, lo que antaño permanecía a nivel intuitivo (se presentía pero no era posible expresar con claridad), se pudieron representar de manera más comprensible en un soporte físico que facilitó su enjuiciamiento y reconstrucción, por lo que los avances intelectuales incrementaron. Con lo anterior, quedaron sentadas las bases que permitieron iniciar el cambio (que no la absoluta sustitución) de algunas explicaciones mágico-místicas por aquellas cuya fuente argumentativa residio en las vivencias (conocimiento de base empírico). Mas, en Babilonia y Egipto, el interés que coadyuvó a la superación científica, y técnica fue, mayoritariamente, práctico.

[6]Principalmente la invención de la escritura y los números, plasmados en las tablas de arcilla y papiro.

1.2.3. Ciencia y estadística en la época clásica

Sin embargo, ellas, son las más remotas influencias que la cultura occidental ha tenido en su historia, pues tuvieron un peso significativo en la conformación de la sapiencia helénica, que alcanzó su máximo esplendor en la época clásica (600 a. C. - 500 d. C.). Los griegos produjeron una importante cantidad de especulaciones filosóficas[7] en cuanto a la naturaleza y al ser en el mundo antiguo. Hacia el siglo VI a. C. en la parte occidente del mundo griego, los jonios dilucidaban en torno a cuestiones pertinentes a la naturaleza, absortos en los primeros intentos de una explicación científica del cosmos en cuanto a materia y forma; mientras en la parte oriental, los pitagóricos enarbolaban el ideal de la filosofía como una guía para la vida[8] al tiempo que consiguieron avances en la matemática.

Por un lado, los jonios consideraban que tras el aparente caos de la realidad, existían un orden y unidad universales que pueden ser comprendidas por el ser humano; por otra parte, los pitagóricos constituyeron un parte aguas para nuestra cultura científica cuantitativa; en ellos, se podía encontrar una fuerte tendencia intelectual religiosa y otra contemplativa y práctica basada en la matemática. Los pitagóricos estaban convencidos de que "el mundo está hecho de números,"[9] lo que permitió avances en áreas de la matemática como la aritmética, el álgebra, la geometría y el cálculo.

Finalmente, la conjunción y contraposición de esas tendencias y otras más, resultó en las primeras expresiones explicativas de carácter preponde-

[7]Finalmente *filosofía*, la madre de todas las ciencias.

[8]GUTHRIE, William, K. C; **Los Filósofos Griegos**; México; Fondo de Cultura Económica; 1995; pp. 31

[9]XIRAU, Ramón; **Introducción a la Historia de la Filosofía**; México; Universidad Nacional Autónoma de México; 1995; pp. 29

rantemente racional. En adelante, el pensamiento griego se enfila hacia el progreso impulsado por la razón."Los griegos quieren establecer un orden racional, una forma de vida que ya no dependa de los monstruos y de los sacrificios primitivos. Ante un fenómeno inexplicable, tratan de dar una explicación congruente capaz de ser entendida por todos los hombres."[10]

La estadística cumplía su función, principalmente, al servicio del Estado, pero no como en la actualidad (con sus técnicas y métodos inferenciales): en Grecia, al igual que en Egipto y Mesopotamia, se realizaban censos poblacionales que incluían información sobre las pertenencias de los habitantes y otras características sociodemográficas, ello con la finalidad de recabar impuestos y, más concretamente, para saber con qué recursos contaba el Estado, pues en sociedades donde prácticamente sólo había dos clases sociales (esclavistas y esclavos), lo que importaba era conocer la cantidad y utilidad que tenían los recursos gobernados, por lo que la estadística no requería de mayor ciencia.

Pero no todo en el mundo helénico era teoría, especulación y abstracción. Éste, realizó importantes avances en cuestiones científicas practicás como medicina, educación, astrología, etc. (lo que impactó significativamente nuestra ciencia y cultura), ni todo el mundo antiguo estaba constituido por estas culturas. En oriente, por ejemplo, hacia el año 4000 a. C. China realizó un registro de los soldados del reino, pero fuera de ello, nuestra herramienta y objeto de estudio no sufrió grandes cambios; incluso durante el imperio romano, con todo lo que implica el control administrativo por parte del Estado en cuestión.

[10]Ibidem. XIRAU, Ramón; México; 1995; pp. 15

1.2.4. Avances en la edad media e ilustración

"El periodo denominado Edad Media, fue mixto en sus actitudes intelectuales; no hubo esa tendencia retrógrada que se solía atribuir al término «medieval». Hubo, sobre todo, un florecimiento de la ciencia, tanto pura como aplicada desde el siglo XI al XIII."[11] Debido a esa "tendencia retrógrada" con la que se ha caracterizado a la Edad Media, a este periodo histórico también se le ha denominado como *el oscurantismo* y abarca los 900 años comprendidos entre 500 y 1400 d. C., pero una revisión más detallada, puede cambiar el concepto general y aportar información de gran utilidad en diversos campos. La contribución cultural que más nos importa de esta época, corresponde a las aportaciones de la Cultura Islámica. La curiosidad y aceptación de los científicos islámicos hacia la ciencia helénica, aglutinada en la filosofía que contenía disciplinas como la Astronomía, Medicina y Astrología, sirvió como la punta de lanza que produjo un significativo florecimiento de dicha cultura, y la ciencia de la época.

"El interés en la astronomía produjo un notable avance en las matemáticas, hubo progresos tanto en geometría como en cálculo, el manejo de los números tuvo un desarrollo mayor, y con *la invención de los números arábigos*, la aritmética quedó (potencialmente) al alcance de cualquier individuo."[12] El anterior progreso científico, pertenece al primer periodo de la Edad Media y se ubica en el área de las ciencias puras; la cultura islámica también fue la que prácticamente fundó la química; en Europa los avances se pueden notar en las aplicaciones de la ciencia, tal es el caso de la matemática en la arquitectura, que es una de las expresiones más características del saber medieval.

---

[11] GORDON, Scott; **Historia y Filosofía de las Ciencias Sociales**; Barcelona; Ariel Referencia; 1995; pp. 34

[12] op. cit. ROSAS, y Novelo María de Lourdes; 1997; pp. 13 Cursivas agregadas.

Hacia el siglo XVII comenzó a germinar un movimiento histórico-espiritual donde el ser humano se propone llegar a un orden de vida dictado por la razón. Tras los 900 años en que predominó la cosmovisión teológica, se dieron las condiciones necesarias para virar completamente la orientación científica del feudalismo. Dicho movimiento es de carácter laico y parte de la convicción de que el progreso y la felicidad de la especie humana son terrenales y sólo se pueden alcanzar mediante la razón.

Este proceso alcanzó su pináculo hacia el siglo XVIII con el predominio de la razón; lo que cambió la visión del universo y reinició el desarrollo de la ciencia como la conocemos en la actualidad: Un verdadero instrumento holístico al servicio de la curiosidad humana. Galileo, Newton, Descartes y los **enciclopedistas**, entre muchos otros, son figuras capitales para el desarrollo y difusión del pensamiento ilustrado y la ciencia positiva.

Tanto Galileo como Newton son fundadores del pensamiento científico moderno. Basándose en las ideas complementarias de filósofos como Platón y Aristóteles (el primero creaba una deidad: La Razón; y el segundo daba forma a los objetos sensibles del mundo: Conocimiento empírico), sentaron las bases de la experimentación científica de nuestra época. Galileo con sus observaciones y teorías respecto al movimiento, infundió en el espíritu de Newton, junto con las inquietudes del pensamiento ilustrado, el impulso para pasar de la teoría y la experimentación "por sí mismas," al uso de ellas como herramientas para la fórmulación de Leyes Científicas.

Esto es la nueva investigación científica: la búsqueda de la naturaleza misma (causa - efecto) de los fenómenos, explicado por sentencias generales que describan sus principios. Fue Newton quien *formalizó* la explicación de los fenómenos naturales relativos al movimiento. Sus observaciones fueron expresadas de tal manera que le fue posible sintetizarlas en lenguaje matemático, tal es le caso de la Ley de la Gravitación Universal, con lo que posibilitó su demostración en los planos fáctico e ideal (formal).

Por su parte, Descartes coloca la cereza en el pastel científico, al enfatizar la necesidad de la creación de un mundo nuevo. Éste, debería ser hecho por el hombre desde sus cimientos, sin otros recursos que los proporcionados por el ingenio y la razón netamente humanas. Descartes constituyó el gran cisma del hombre con el pensamiento circunstancial que envolvía las explicaciones de vida humana y su derredor; a partir de sus propuestas, el avance científico depende de la razón humana, ya no más explicaciones divinas a fenómenos reales, el hombre contiene en sí, en su razón, la capacidad de explicar, predecir y reedificar el mundo que le rodeaba, ese mundo mitológico creado por el pensamiento religioso. De aquí en adelante, todo dependerá de la inteligencia humana, pues la vocación del pensamiento cartesiano es definitiva: la de un hombre que desea llegar a la verdad absoluta, vía el uso de la razón y, dicho sea de paso, con el uso de la matemática como una de sus más importantes herramientas.

En este periodo, también se ubican los antecedentes inmediatos de la sistematización en el campo de la estadística, así como de la investigación social empírica. Éstos, "se remontan al movimiento denominado *estadística social*: un grupo de estudiosos interesados en la recogida y organización de datos económicos, demográficos y sociales."[13] que a diferencia de sus antecesores (las culturas de la antigüedad que sólo utilizaban los recuentos censales con finalidades tributarias y militares, es decir, de contabilidad y control social)

[13]CEA, D' Ancona María Ángeles; **Metodología Cuantitativa. Estrategias y Técnicas de Investigación Social** Madrid; Proyecto Editorial Síntesis Sociológica; 1998; pp. 20

pretendieron aplicar los avances de la matemática con mayor precisión a los fenómenos de la naturaleza y, en especial los de la estadística, a los procesos sociales con propósitos sí de control social, pero también científicos.

Dicha propuesta analítica no se debe únicamente a la efervescencia intelectual del pensamiento ilustrado. Con el advenimiento de la *Revolución Industrial*, un nuevo modo de vida surgió en el viejo continente: los conglomerados urbanos en derredor de los centros industriales. Con esta nueva forma de vida, el despotismo ilustrado, que antaño surgiera como una moda, se transforma en una necesidad para cualquier Estado que enfrentase los problemas de la naciente sociedad en masa.

Por lo que las cuantificaciones, la aplicación de los conocimientos matemáticos al estudio del cosmos y en especial lo referente a la estadística en el campo de lo social, adquieren paulatinamente un carácter científico para sí, y el de una herramienta de precisión al servicio de otras ciencias. Dicha transformación, en el campo de la estadística, "fue posible gracias a las aportaciones fundamentales de una serie de autores englobados en dos *escuelas estadísticas* principales: Los aritméticos políticos ingleses y la escuela estadística alemana."[14]

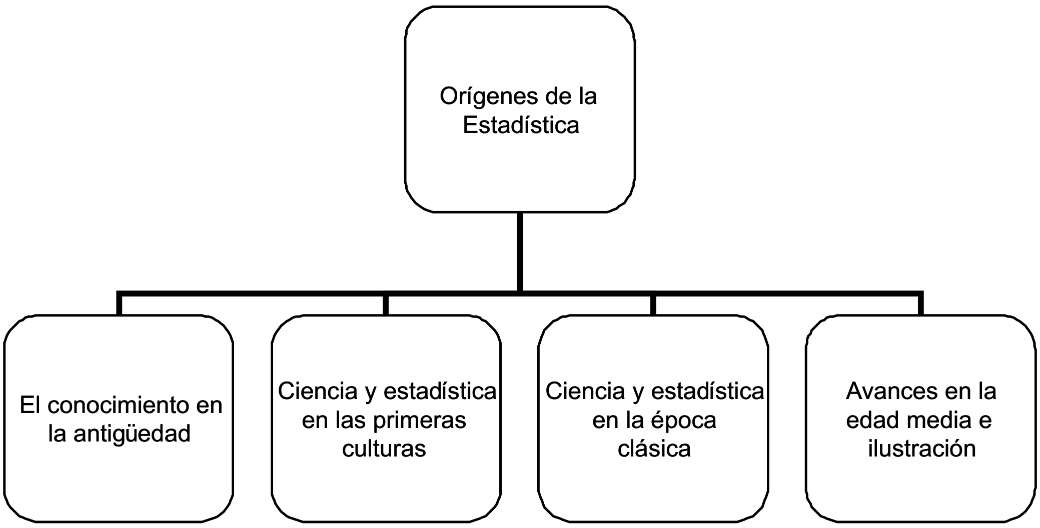
Ejercicios 1

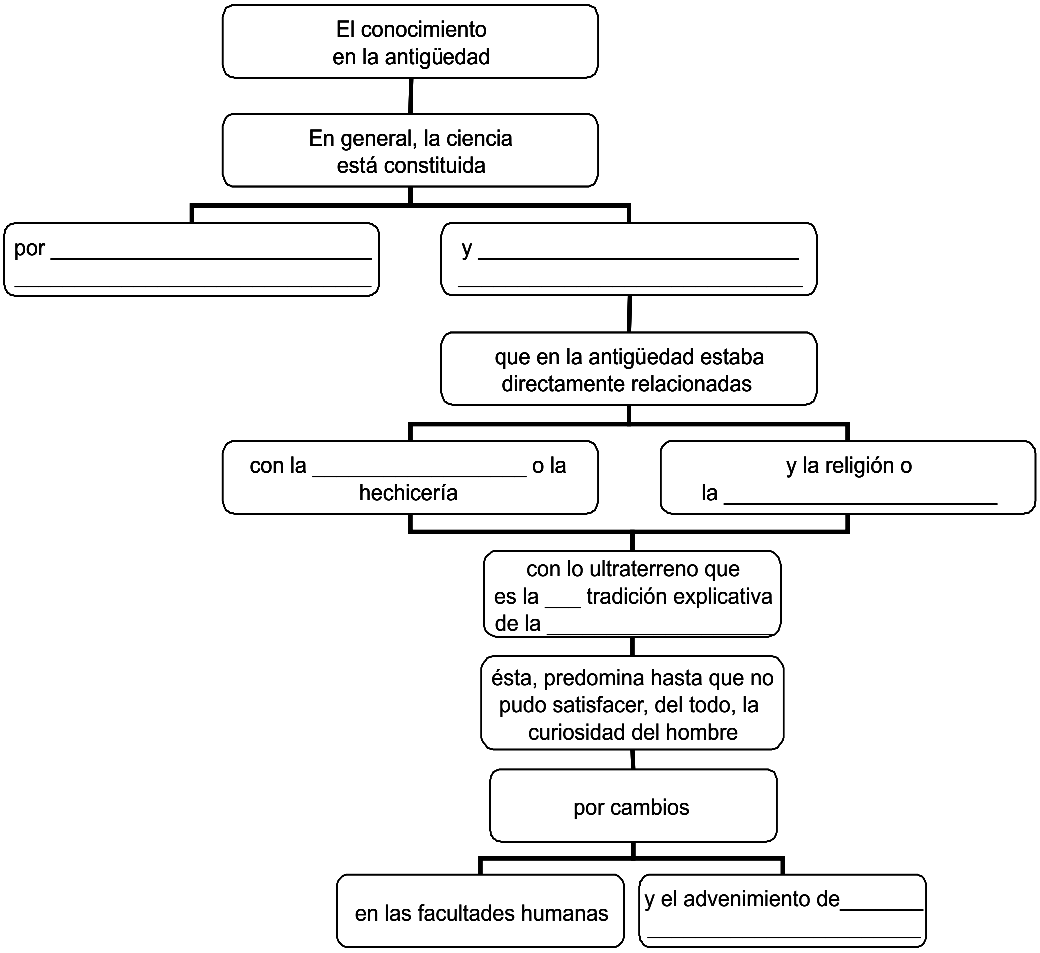
Como parte de tu perfil profesional, debes contar con el conocimiento de los referentes históricos que envuelven la construcción de las herramientas teóricas, metodológicas y técnicas que formarán parte de tu acervo; ello, te permitirá la comprensión integral de estas herramientas y por ende su mejor aprovechamiento en el ámbito laboral en que te desempeñes.

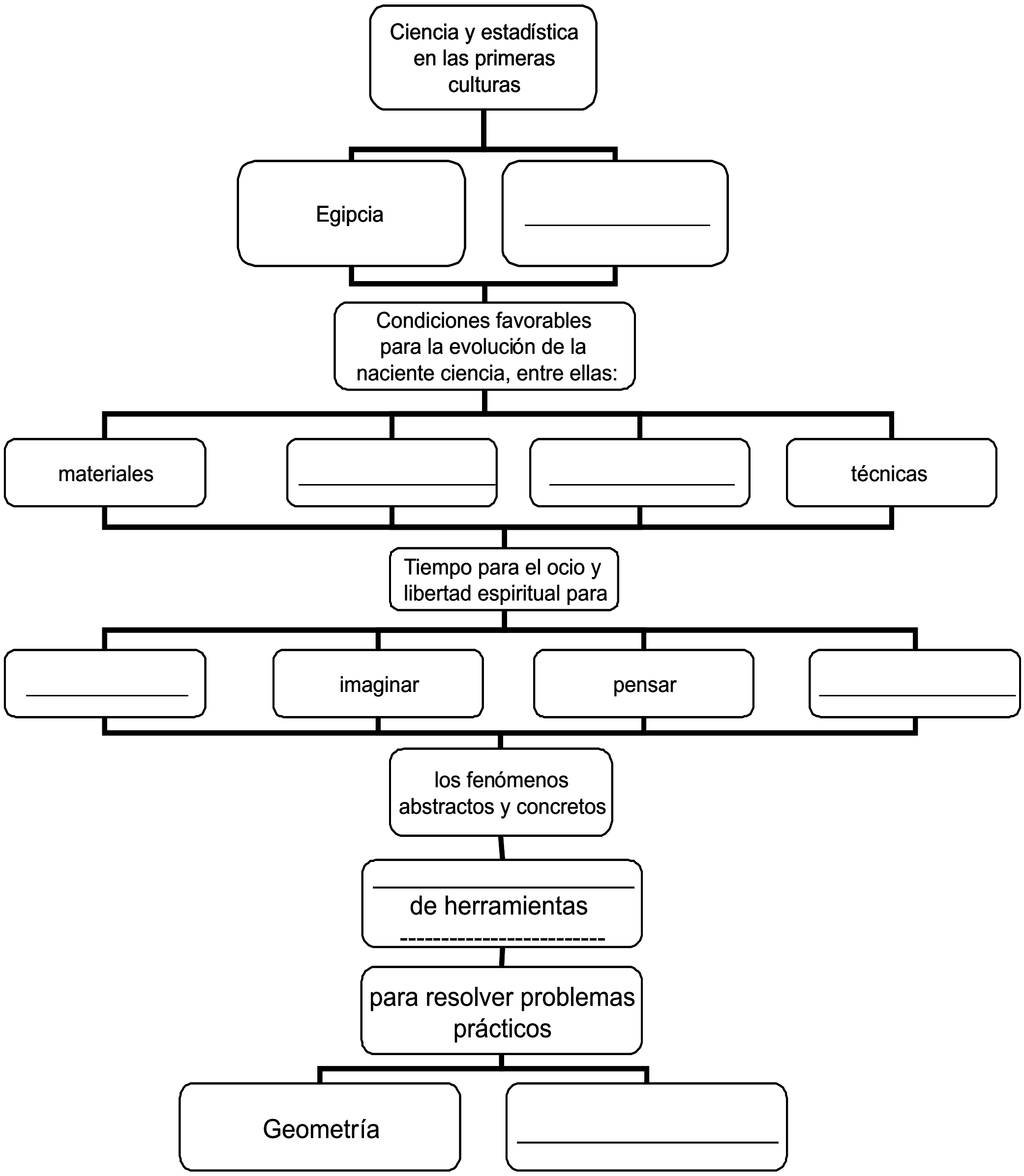
[14]Ibidem. CEA, D' Ancona María Ángeles; Madrid; 1998; pp. 20

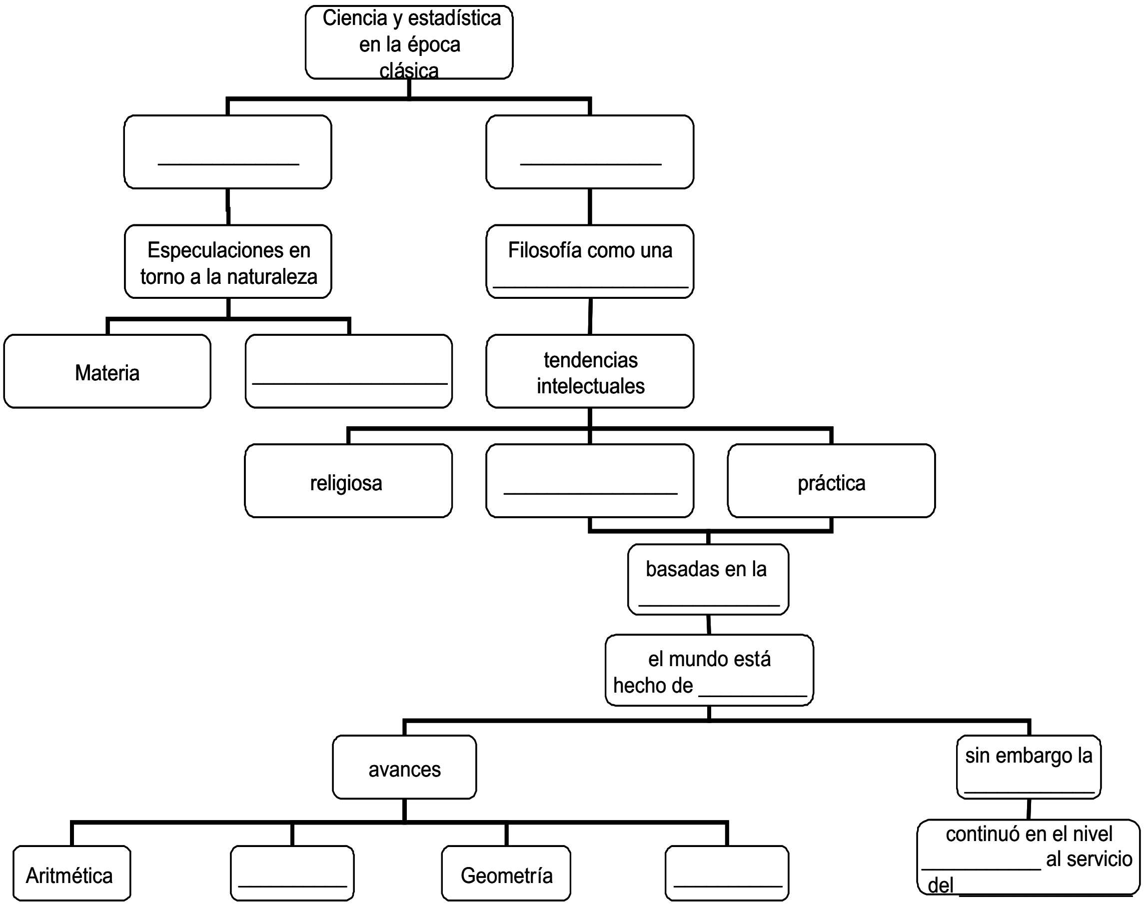
A continuación, se te presenta la primera serie de ejercicios que te permitirá reforzar el conocimiento que sobre el contenido del capítulo has conseguido. Es recomendable que antes de responder, repases tus apuntes (recuerda que una excelente técnica de estudio es la elaboración de acordeones) y acotaciones que hiciste en el texto.

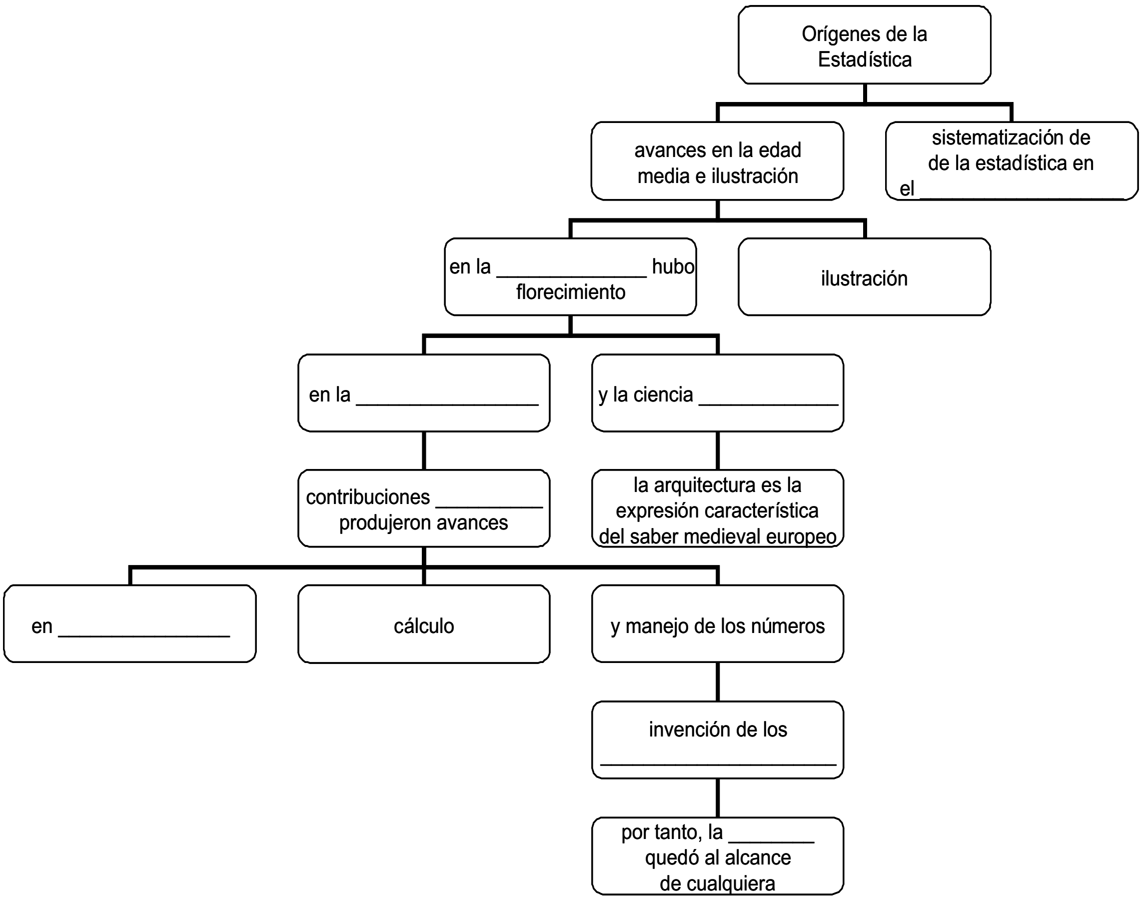
- Instrucciones: Anota en los espacios vacíos, los elementos que completen correctamente las siguientes redes conceptuales. Posteriormente, verifica que tus respuestas coincidan con las de la página 67

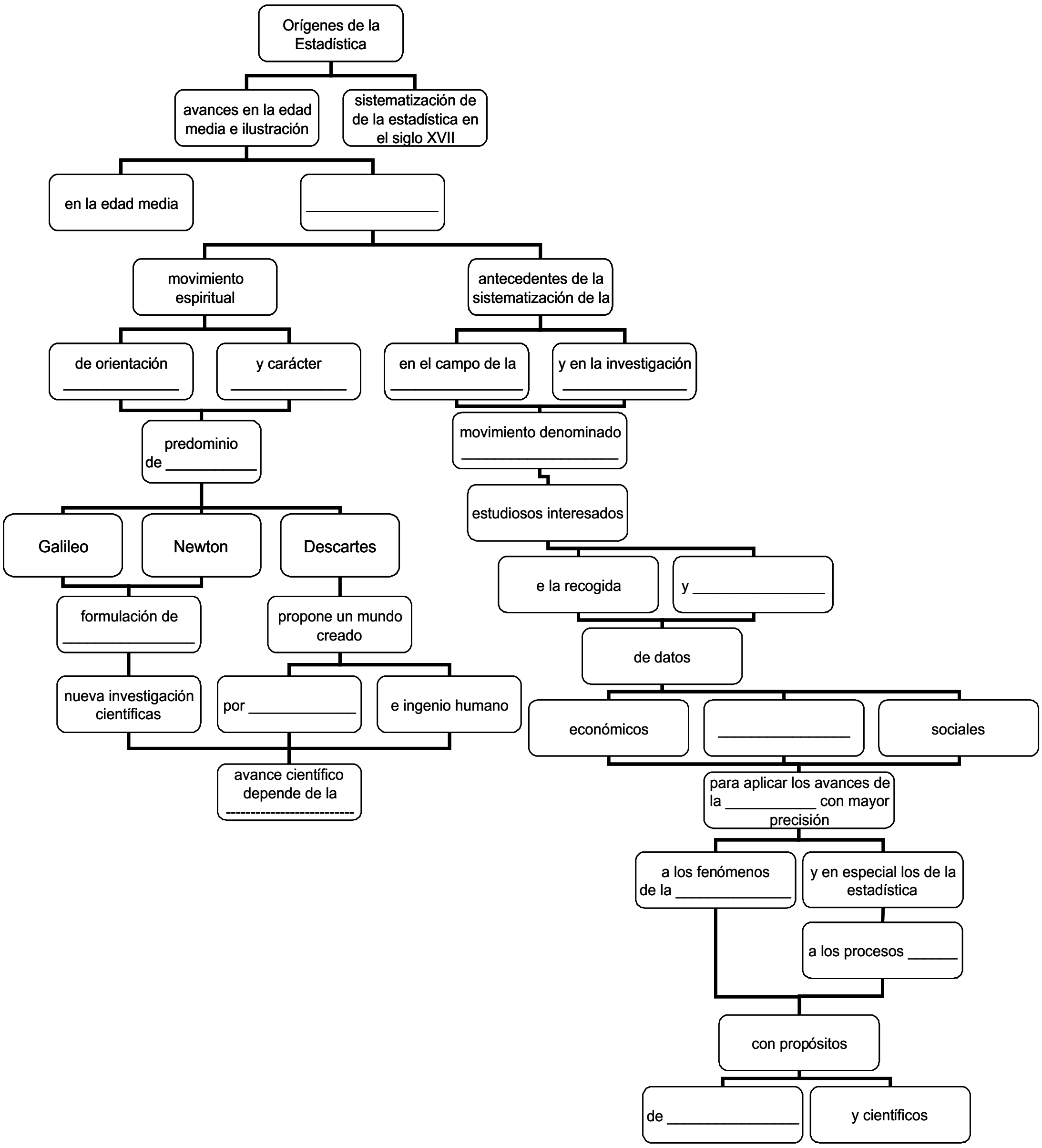



El conocimiento en la antigüedad
En general, la ciencia está constituida
por ____
y ____
que en la antigüedad estaba directamente relacionadas
con la ____ o la hechicería
y la religión o la ____
con lo ultraterreno que es la ___ tradición explicativa de la ____
ésta, predomina hasta que no pudo satisfacer, del todo, la curiosidad del hombre
por cambios
en las facultades humanas
y el advenimiento de____


Ciencia y estadística en las primeras culturas
Egipcia
____
Condiciones favorables para la evolución de la naciente ciencia, entre ellas:
materiales
____
____
técnicas
Tiempo para el ocio y libertad espiritual para
____
imaginar
pensar
____
los fenómenos abstractos y concretos
____________ de herramientas ____________
para resolver problemas prácticos
Geometría
____


Ciencia y estadística en la época clásica
Especulaciones en torno a la naturaleza
Filosofía como una ____
Materia
tendencias intelectuales
religiosa
práctica
basadas en la ____
el mundo está hecho de ____
avances
sin embargo la ____
Aritmética
Geometría
continuó en el nivel ____ al servicio del ____


Orígenes de la Estadística
avances en la edad media e ilustración
sistematización de de la estadística en el ____________
en la ____________ hubo florecimiento
ilustración
en la ____________
y la ciencia ____________
contribuciones ________ produjeron avances
la arquitectura es la expresión característica del saber medieval europeo
en ____________
cálculo
y manejo de los números
invención de los ____________
por tanto, la ________ quedó al alcance de cualquiera


Orígenes de la Estadística
avances en la edad media e ilustración
sistematización de de la estadística en el siglo XVII
en la edad media
____
movimiento espiritual
antecedentes de la sistematización de la
de orientación ______________
y carácter ______________
en el campo de la ________________
y en la investigación ______________
predominio de ___________
movimiento denominado ___________________
estudiosos interesados
Galileo
Newton
Descartes
formulación de ________________
propone un mundo creado
e la recogida
y ________________
de datos
nueva investigación científicas
por _____________
e ingenio humano
económicos
____
sociales
avance científico depende de la ------------------------
para aplicar los avances de la ___________ con mayor precisión
a los fenómenos de la _______________
y en especial los de la estadística
a los procesos ______
con propósitos
de ________________
y científicos

1.2.5. Sistematización de la estadística

El siglo XVII vio el nacimiento de la estadística moderna, en cuanto a técnicas de inferencia para el análisis de los datos sociales. Una repercusión inmediata fue que en Europa, el estudio y registro de información de los habitantes se utilizó como una herramienta de espionaje al servicio del Estado, lo que fuera del ámbito científico-académico significó el uso de la estadística (al más puro estilo de la vieja escuela) en un sentido literal con respecto a su origen, es decir: como un medio de control y administración estatal.

Por su parte, la matemática también consiguió un avance que repercutió sustancialmente en el campo de nuestro interés. A esta aleación entre la estadística y el estudio de la dinámica social, en 1660 Petty, científico naturalista inglés –autor adscrito a la corriente de los aritméticos políticos ingleses–, la denominó *aritmética política* en su obra *Political Arithmetic.* "En esta obra se insta a la *medición* de los fenómenos sociales, con los mismos instrumentos empleados en las ciencias naturales. Además, se defiende la utilización gubernamental de los datos estadísticos que se obtengan.

"En el siglo XVIII aparecen las primeras *investigaciones sociales empíricas*, propiamente dichas, de la mano de los llamados *reformadores.* Un conjunto de profesionales, de distintas disciplinas académicas (médicos, ingenieros, naturalistas), unidos por un mismo propósito: aportar evidencia empírica de los problemas sociales existentes en su época. Su finalidad era eminentemente política: movilizar a la opinión pública para exigir medidas concretas de *reforma social.*[15] De ahí les viene el nombre.

[15] Hecho en que se puede advertir uno de los beneficios aportados por la Matemática no unidireccionalmente al servicio del Estado controlador, sino como un fundamento que posibilitaría el contrapeso ejercido a la elite en el poder. Nota del Autor

“Estos investigadores constituyeron (en sus países respectivos) *sociedades estadísticas* que promovieron el desarrollo de *encuestas sociales* en la investigación social.

“En España el *el primer censo de población* (considerado Oficial y moderno) aparece en 1857.

“De las *encuestas sociales*, realizadas en Gran Bretaña destaca en lo metodológico:

Vida y trabajo de los habitantes de Londres (1889-1891), de Charles Booth por los siguientes aspectos:

1. La *observación directa* de barrios obreros.
2. La realización de *entrevistas abiertas* a hombres de negocios.
3. La elaboración de *mapas sociales* de diferentes barrios de Londres. Mediante estos *mapas*, el autor analiza la distribución de la población en la ciudad. De manera especial, la relación existente entre la ubicación de la industria y el nivel de vida de la población.

“La figura de Booth destaca, no sólo por su obra, sino también por los seguidores que tuvo a comienzos del siglo XX. Entre ellos suelen citarse los siguientes:

“a) *Rowntree*. A este autor se le reconoce un cierto perfeccionamiento de la *metodología de los sondeos*: preguntar directamente a los obreros, sin acudir a intermediarios.

“b) *Bowley* y *Burnet*. Introdujeron mejoras en el *muestreo aleatorio*. Éstas se centraron en la disminución de los *errores muestrales* y de los gastos en su realización.

“c) *El matrimonio Webb*. Principalmente por el análisis que efectuaron (mediante fuentes secundarias) de la evolución de determinadas instituciones sociales.

De las encuestas sociales efectuadas en Francia merece destacarse:

“a) *Obreros europeos (1877 -1879), de Frederic Le Play.*

“Un ingeniero de minas interesado en el análisis de la familia como núcleo de toda la sociedad. Estudia no sólo las características externas de las familias, sino también sus sentimientos, valores y deseos. Para ello compagina la *comparación sistemática* de diversas familias de países diferentes.

“Nisbet (1966: 61) califica esta investigación como ‘una obra estrictamente sociológica, el primer trabajo sociológico genuinamente científico del siglo XIX.’ Sea o no cierta esta afirmación, parece indudable la contribución de Le Play a lo que después se denominaría el *estudio de casos*.

“La gran revolución que experimenta la investigación social empírica a lo largo del siglo XIX fue pareja a los avances producidos en las técnicas de recogida y análisis de la información. Adolfo Quételet (1796 - 1874) fue uno de los más grandes impulsores de la aplicación de las mejorías introducidas en el campo de la estadística al estudio de los fenómenos sociales.

“Pero, si por algo destaca Quételet es, sobre todo, por su *teoría de las regularidades de los fenómenos sociales*: en la conducta humana, aparentemente azarosa, pueden observarse regularidades, que pueden medirse a través de técnicas estadísticas de análisis. De ahí que se le considere un ‘un pionero en la aplicación de modelos matemáticos en la sociología’ (García Ferrando, 1979: 33).

“En sus investigaciones socio-antropológicas (*Sobre la ley del crecimiento del hombre e Investigaciones sobre las tendencias criminales en distintas edades*), Quételet analiza las características físicas de las personas y las estadísticas de delincuencia. Para ello aplica el *cálculo de las probabilidades estadísticas.*

“En 1835, publica su obra más relevante: *sobre el hombre y el desarrollo de las facultades humanas: ensayo sobre física social.* Años más tarde (en 1869), reeditaría esta obra, corregida y ampliada, pero con un título más breve: *Física social.* De ella se ha dicho que ‘marcó la transición de la simple descripción estadística al empleo consciente de los datos cuantitativos empíricos, para establecer las regularidades de la vida social’

“Además de Quételet hubo otros autores que realizaron importantes aportaciones en el campo de la estadística. Entre ellos destacan:
“a) Laplace, que introduce el principio de mínimos cuadrados en su *Teoría analítica de las probabilidades.*
“b) Gauss generaliza el *método de mínimos cuadrados.*
“c) Yule aplica, por primera vez, los análisis de *regresión múltiple* y *de correlación* a datos sociales. Esto sucede a finales del siglo XIX.

"En Alemania, los enfoques de Le Play y Quételet Fueron sintetizados por tres autores principales:
"a) Ernest Enguel, en sus investigaciones sobre el presupuesto.
"b) Adolf Wagner, considerado epígono de Quételet. A este respecto es muy ilustrativo el título de uno de sus libros: *La regularidad de las acciones que parecen casuales desde el punto de vista de la estadística*, de 1864.
"c) Wilhelm Lexis, que profundiza en las ideas de Quételet sobre la cuantificación de los fenómenos sociales.

"La recogida y análisis de datos sociales poco a poco va profesionalizándose. El número de investigadores *amateurs* se reduce y aumenta, por el contrario, el de profesores universitarios. Entre estos caber citar a Gustav Schmoller, Ferdinand Tönnies y Max Weber, en Alemania; y, a Emile Durkheim, en Francia."[16]

Así, la estadística pasó de ser una actividad puramente descriptiva a una herramienta científica cuyo valor radica en la precisión, tanto de su lenguaje como de sus técnicas y métodos, que permite obtener conclusiones sustentadas por principios no contradictorios. Ésta característica es importante para el ser humano en dos aspectos:

- El primero, es que gracias a ella, el hombre puede tomar decisiones altamente racionales y con un elevado rango de certidumbre, pues al incorporar las nociones de no contradicción y aleatoriedad con las que se encuentra profundamente ligada, hasta lo "desconocido" se puede medir.

- El segundo, es que sus aportaciones en el ámbito científico, sobre todo en lo referente a las ciencias sociales y humanidades, son fundamentales

[16]Ibidem. CEA, D' Ancona María Ángeles; Madrid; 1998; pp. 20-25

en tanto que se erigen como un elemento constitutivo (fundador) de una nueva visión en disciplinas tan complejas como el hombre mismo.

1.3. Fundamentación de la matemática

La matemática *per se* es una ciencia que, como la lógica, se encarga de estudiar hechos abstractos, ideales; es decir, estudia ideas. La verificación de éstas, sólo tiene cabida en la mente humana. "En resolución: la ciencia formal es *autosuficiente* por lo que hacen al contenido y al método de prueba."[17] La exactitud de la matemática, se basa en el principio de la no contradicción. Los elementos básicos de la estructura **apodíctica** que infunde la precisión de dicha ciencia, son los siguientes:

Axioma: Es una afirmación que se considera *tan evidente* que se acepta como verdad sin que sea necesario demostrarla. De estas aseveraciones parte el conocimiento matemático para caracterizarse como irrefutable.

Postulado: Es una proposición que se obtiene a partir de los Axiomas; no es tan evidente por sí misma, pero también se acepta sin demostración y forma parte de un sistema deductivo.

Lema: Enunciado que se construye a base de Axiomas y Postulados: Sí requiere de demostración lógico-matemática. Se puede decir que es un Teorema preliminar.

Teorema: Es una proposición matemática demostrable; se construye con Axiomas, Postulados y Lemas. En este caso, es indispensable la demostración lógico-matemática para validar la conclusión a que se ha llegado; por tanto,

[17] BUNGE, Mario; México; 2000; pp. 19

es la proposición más formal de las matemáticas y permite la construcción de nuevos teoremas.

Corolario: Es una consecuencia o particularización que se desprende de un teorema, lo que significa que después de la aplicación de un teorema a un caso particular, se desprende una conclusión extra.

En la base del rigor matemático tenemos el pensamiento ordenado. Las conclusiones a que se puede llegar *son consecuencia lógica* de la unión entre elementos abstractos que van desde el axioma hasta el teorema. El conocimiento matemático está íntimamente ligado a una forma correcta de razonar; toda conclusión se debe encontrar apropiadamente sustentada mediante una demostración formal. Es por ello que la matemática se ha constituido como una herramienta sumamente confiable para el estudio de los fenómenos naturales y sociales.

Al aplicarse al estudio de fenómenos naturales, por ejemplo, nos ofrece cierta precisión e incluso exactitud. Podemos conocer dimensiones como la longitud de una autopista o de un bolígrafo, determinar el peso de una hoja o un megalito, la velocidad de un auto o la de la luz... con el "simple" hecho de elegir los instrumentos apropiados y conocer las operaciones aritméticas y algebráicas necesarias para ello, se pueden obtener resultados satisfactorios y de gran utilidad.

Pero en el terreno de los acontecimientos sociales, la relatividad y el orden que se dan en ellos, pueden (y lo hacen) mostrar una gran cantidad de facetas que difícilmente podemos describir, comprender, explicar y/o controlar con sólo la aritmética o instrumentos de medición físicos. Como se pudo observar en las secciones anteriores, las regularidades en el mundo social no son mesurables de forma directa, pues éstas no se pueden apreciar o demostrar de forma física e individualmente.

- Ejemplos:
  - ¿Qué ocurre cuando se trata de conocer el porqué de la popularidad de una serie de televisión en el Distrito Federal?
  - ¿Cómo determinamos el impacto de las expresiones contenidas en una campaña propagandística o publicitaria en el Área Metropolitana?
  - ¿Qué elementos debemos considerar para confiar o no en los resultados de una encuesta para medir la opinión pública?
  - ¿Cómo se mide la opinión pública?
  - ¿Cómo saber si el tema principal de una revista es interesante para el público al que se dirige y no la consumen por algún otro elemento?

El estudio de la comunicación social es complicado en tanto que la naturaleza de su objeto, no es tal que permita las regularidades mesurables con que se puede tratar en la ciencia factual natural o lógica. En ella sí hay regularidades, pero son de un orden diferente, como sabemos: El *objeto formal de la teoría de la comunicación*, en general, son "las interacciones entre los seres vivos, que se llevan a cabo por el recurso a actos expresivos."[18] y éstos, como su influencia en los actores sociales de la comunicación, pueden ser tan diversos y heterogéneos como la naturaleza misma del ser humano.

[18]MARTÍN, Serrano Manuel; et. al. **Teoría de la Comunicación I. Epistemología y Análisis de la Referencia**; México; Universidad Nacional Autónoma de México; 1993; pp. 65

Para ello contamos con la estadística: un instrumento matemático para el estudio de fenómenos cuya naturaleza dista mucho de ser ordenada y predecible en un sentido tradicional, es decir inmediata y particularmente. En ella, encontramos el apoyo de teorías, métodos y técnicas que nos permiten hacer frente al estudio de fenómenos naturales y sociales, en que las variaciones inherentes a sus propiedades, imposibilitarían un abordaje científico o teórico funcional y confiable dentro de sus respectivas generalidades y limitaciones.

1.4. Objetivos de la estadística

En el estudio de fenómenos sociales, entre ellos la comunicación, se encuentran circunscritos por implicaciones teórico-conceptuales, metodológico-técnicas e incluso epistemológicas que dan validez al conocimiento generado por las disciplinas que de ellos se encargan. Dichas implicaciones van desde la elección del paradigma y todos sus porqués, hasta la determinación de los elementos a ser estudiados y la pertinencia de las técnicas e instrumentos para la recolección, análisis e interpretación de los datos con que se trabajará, pues como ya se mencionó, no es lo mismo medir un objeto físico que uno social.

Lo anterior, se debe a factores pertinentes a las propiedades de los fenómenos estudiados y a las particularidades del sistema en que se encuentran: En la actualidad, es impensable realizar constantemente censos poblacionales para recolectar toda la información social que se requiere por diversas instituciones, organizaciones e incluso particulares. La sociedad en masa que se comenzó a configurar durante la Revolución Industrial, cobró dimensiones exorbitantes en todo el mundo y en múltiples aspectos.

Casi cualquier núcleo social o sociedad en que nos corresponda estudiar al sistema comunicativo, es tan complejo y heterogéneo que los aspectos y dimensiones posibles de análisis se antojan imposibles de abordar apropiadamente, dadas sus dimensiones y variabilidad. Ante la aparente imposibilidad espacial, temporal y económica de estudiar en su totalidad a las interacciones que se dan entre los *actores sociales de la comunicación* (que se llevan a cabo por el recurso a actos expresivos) en el sistema social, la estadística sigue la estrategia de seleccionar, mediante los mecanismos adecuados, sólo una parte representativa del sistema; pues como ya vimos, en los fenómenos sociales, las regularidades se pueden ubicar con mayor facilidad de forma grupal que individualmente.

Así, la imposibilidad para estudiar a la totalidad, no es más que aparente gracias a las técnicas de muestreo. A la parte representativa se le denomina *muestra* (denotada por la literal "**n**"), mientras que al total de elementos del sistema se les conoce como *población objetivo* (representada por la literal "**N**"). Dicho de otro modo, la población constituye el universo que se quiere estudiar, es un todo compuesto por elementos que guardan relaciones solidarias y éstas no son muy claras a simple vista, por que la estadística es de gran utilidad para su análisis; por otra parte, la muestra es un micro universo, la copia a escala –que aspira a ser perfecta representativa – de ese todo que es objeto de atención. Lo que ubica a la lógica de investigación en el campo de la *metodología cuantitativa*; de ella se, pueden obtener técnicas de análisis social científicamente válidas, de gran precisión, bajo costo y comprobables, lo que se traduce en eficiencia, competitividad y rentabilidad, al ser aplicadas apropiadamente.

A través de técnicas específicas se obtienen *descriptores y características* de la muestra (n) que mediante procedimientos formales pueden ser generalizados a la población (N). “La teoría del muestreo es un estudio de las relaciones existentes entre una población y muestras extraídas de ellas.”[19]

“Si una muestra es representativa de una población, se pueden deducir importantes conclusiones acerca de ésta, a partir del análisis de la misma. La parte de la estadística que trata de las condiciones bajo las cuales tales inferencias son válidas se llama *estadística inductiva* o *estadística inferencial.*

“La parte de la estadística que trata solamente de describir y analizar a un grupo dado sin sacar conclusiones o inferencias de un grupo mayor, se llama *estadística descriptiva* o *deductiva.*”[20] Como se puede observar, éstos son ámbitos diferentes (pero complementarios) de la estadística; ellos, pueden ser aplicados al estudio cuantitativo de fenómenos cuya naturaleza se presenta, en primera instancia, como algo que no se pueden medir (tal es el caso de una gran cantidad de fenómenos sociales).

Los resultados obtenidos a través de estudios sociales apoyados en la estadística, se emplean para la toma de decisiones y resoluciones razonables de problemas intelectuales y prácticos, cuyos orígenes y naturaleza “obligan” a su análisis holístico mesurable. Desde luego, no es la única herramienta ni es infalible; pero sí es una de las más confiables y su aplicación corresponde al tipo de pregunta bajo la cual surge la necesidad y/o problema en cuestión.

---

[19]SPIEGEL, Murray R.; **Estadística**; México; McGrawg Gil, Serie Schaum; 1970; pp. 141

[20]Ibidem. SPIEGEL, Murray R.; México; 1970; pp. 1

Ejercicios 2

- Instrucciones: Responde breve y suficientemente las siguientes preguntas; al terminar, compáralas con las que se te proporcionan en la página 72

1. ¿Por qué se dice que en el siglo XVII se dio el nacimiento de la estadística moderna?

   ____

   ____

   ____

   ____

2. ¿Cuál es el propósito de las primeras investigaciones sociales realizadas por los denominados reformadores?

   ____

   ____

   ____

3. ¿Por qué Adolfo Quételet es considerado como un pionero en la aplicación de modelos matemáticos en la sociología?

   ____

   ____

   ____

   ____

4. ¿Cuál es el principio que sustenta la exactitud de la matemática como una ciencia formal?

   ____

   ____

   ____

   ____

- Coloca en las lineas correspondientes los términos que completen correctamente los siguientes conceptos:

5. _____________: Es una _________________ que se considera tan ___________ que se acepta como ___________ sin que sea necesario _____________. De estas aseveraciones parte el conocimiento __________ para caracterizarse como __________.

demostrarlas	evidente	irrefutables	Axioma
verdadera	matemático	afirmación	

6. _________________: Es una _________________ matemática demostrable; se construye con _________________, Postulados y _________________. En este caso, es indispensable la _________________ lógico-matemática para validar la _________________ a que se ha llegado; por tanto, es la proposición más _________________ de las matemáticas y permite la construcción de nuevos _________________.

Proposición	demostración	formal	teoremas
aximas	Teorema	lemas	conclusión

7. ¿Cómo son las regularidades que se pueden encontrar en los sistemas sociales mediante el uso de la estadística?

____

____

____

____

____

____

____

8. ¿Qué es una muestra?

1.5. Reflexiones

Al llegar a este punto, debemos hacer una pausa para ahondar en los siguientes aspectos propuestos:

- La ciencia es falible. No obstante su perfeccionamiento y confiabilidad, una de sus características más trascendentales es que ella misma, con su dinamismo, muestra que no existen los conocimientos terminales, absolutos y definitivos; tras cada determinación científica, viene la revelación de una nueva incógnita.

- A pesar del alto grado de racionalidad que hemos alcanzado en nuestra cosmovisión, los elementos ultraterrenos siguen siendo un paradigma explicativo para gran parte de la humanidad. Incluso, hay quienes afirman que la ciencia misma, se ha transformado en un importante metarrelato sagrado semejante al de las antiguas deidades, pues en ella perciben a un dios providente y/o a un demonio cruel.

- La evolución de la matemática y su precisión, en combinación con los avances teóricos producidos en las ciencias sociales, dieron una nueva orientación y alcance a los estudios sociales. Con las técnicas del muestreo (por ejemplo), es posible analizar fenómenos cuyas dimensiones suelen rebasar, sin éstas, las capacidades ejecutivas de cualquier instituto o agencia de investigación.

- Los beneficios de dicha fusión se pueden enumerar tanto en la producción del conocimiento por sí mismo, como en la reducción de los costos que generan las investigaciones sociales, esto es: precisión y economía; el máximo aprovechamiento de los recursos. Desde luego no son las únicas mercedes obtenidas, pues no sólo a nivel intelectual repercuten los avances científicos, sino que alcanzan múltiples dimensiones que van desde lo político hasta lo lúdico.

- La matemática le permite a las ciencias sociales manipular objetos y propiedades de los fenómenos que sin ellas, serían incontrolables (intelectualmente) para su estudio. Es por ello que términos como *variable*, *factor*, *rango*, *parámetro*, y otros, se utilizan frecuentemente en los estudios sociales y humanísticos. Respecto a ello, se reflexionará en el siguiente capítulo.

1.6. Autoevaluación

- La sección que ahora inicias, fundamentalmente obedece a dos propósitos:

    - El primero y más importante es: Averiguar, en cierta medida, el resultado de tus esfuerzos de lectura y comprensión, relativos a los temas tratados durante el capítulo. Tener *conciencia* de tus propios progresos, debe ser factor de *motivación personal* que se traduce en éxito. De igual manera, detectar respuestas inexactas, te permitirá iniciar un proceso de reconocimiento personal, encaminado a encontrar los agentes internos y externos que interfieren con la eficacia de tu estudio.

- En segundo lugar, te permitirá *ejercitar tus capacidades cognoscitivas.* La práctica constante y el trabajo conciente, son reflejo del interés que por tu desarrollo tienes. Aprehende esto: la llave del triunfo no es una memoria superior, sino una autoestima creciente, suficiente para la automotivación y la constancia en el trabajo. Si en algún momento consideras (o lo has hecho) que tienes problemas con la memoria, piensa en lo siguiente: "Lo que se usa evoluciona y lo que no, se atrofia."

- Instrucciones:
  - Antes de responder las preguntas, léelas con atención y reflexiona sobre lo que se te pide.
  - Responde exhaustivamente cada cuestionamiento. Para ello, se te proporcionan espacios suficientes que debes aprovechar al máximo.
  - Evita, en lo posible, las respuestas memorísticas, es decir, fundamenta tus afirmaciones con argumentos basados en las lecturas realizadas.
  - Al concluir, acude a la sección **1.7.2. Respuestas de la autoevaluación** (pág. 73) y coteja tus razonamientos con las respuestas que ahí se proponen. En esta sección se te proporcionarán algunos criterios para el cotejo.

1. ¿En qué característica del ser humano se puede ubicar el origen de la ciencia y por qué?

   ____

   ____

   ____

   ____

   ____

2. Explica por qué el antecedente más remoto de la ciencia se encuentra en la relación del hombre con lo ultraterreno.

3. ¿Cuáles son los elementos que permitieron el paso de las explicaciones ultraterrenas al dominio de las que se basan en el sentido común?

4. ¿Cuál es el común denominador para la estadística en las culturas más antiguas del mundo?

5. ¿Cuál es la importancia de la cultura helénica para el desarrollo de la ciencia moderna y la estadística?

   ____

   ____

   ____

   ____

   ____

   ____

   ____

6. ¿En qué campo de la ciencia, se dio el más notable invento durante la edad media y en qué consistió?

   ____

   ____

   ____

   ____

   ____

   ____

7. ¿Qué acontecimiento definió el inicio de una nueva era en el campo de la investigación científica?

   ____

   ____

   ____

   ____

   ____

   ____

   ____

   ____

   ____

8. ¿Cuál es la importancia del pensamiento cartesiano para el desarrollo de la ciencia moderna?

   ____
   ____
   ____
   ____
   ____
   ____
   ____
   ____

9. ¿De qué manera incidio el pensamiento ilustrado en el campo de la estadística?

   ____
   ____
   ____
   ____
   ____
   ____
   ____
   ____

10. ¿Cuál es la importancia del trabajo de Adolfo Quételet para la constitución de la investigación social y la estadística moderna?

    ____
    ____
    ____
    ____
    ____
    ____

____

____

____

11. ¿Cuál es la importancia de la incorporación de la estadística a la investigación social?

____

____

____

____

____

____

____

____

____

____

____

____

____

____

____

12. ¿Qué característica de la comunicación social, hace necesario el uso de las herramientas estadísticas para su estudio?

____

____

____

____

____

____

13. Explica qué es una población y una muestra

14. Sugiere un ejemplo de población y, a partir de ésta, un ejemplo de muestra

15. ¿Cuál es la diferencia que existe entre la estadística descriptiva y la inferencial?

Consulta las respuestas de los ejercicios de autoevaluación en la página 73

1.7. Apéndice del capítulo

1.7.1. Información remedial

Teoría de sistemas

A medida que la ciencia avanza, se ha caído en la cuenta de que el universo que habitamos es un todo en el que cada parte contenida en él, guarda una serie, incluso inimaginable, de relaciones con otros elementos que lo conforman. "En 1933, en una obra titulada *Modern Theories of Development*, el biólogo Luwing von Bertalanffy establecía las bases de lo que formalizaría en la posguerra como la «teoría de los sistemas» [...] Bertalanffy usa el término «función» relacionándolo con los «procesos vitales u orgánicos en la medida en que contribuyen al mantenimiento del organismo»"[21] con lo que pone de manifiesto la interacción que existe entre diferentes partes de un todo.

A ese "todo" se le denomina **sistema**. Un sistema es un conjunto de elementos. Dichos elementos se encuentran ordenados y cumplen con diferentes funciones al interior de éste; lo que implica que entre ellos, guarden una serie de relaciones que permiten el mantenimiento de ese conjunto. Asimismo, los componentes del sistema (los elementos que conforman al conjunto) y el mismo sistema, se relaciona solidariamente con otros sistemas, lo que en consecuencia suele producir cambios en las estructuras de éstos.

Dichas relaciones no son siempre claras, por lo menos de forma evidente. La concepción determinista del universo planteada por la visión positiva del

[21]MATTELART, Armand; et. al.; **Historia de las Teorías de la Comunicación**; Barcelona; Paidós; 1997; pp. 44

pensamiento ilustrado, es posible bajo condiciones "controladas" por el ser humano en sus laboratorios; pero en la naturaleza, las condiciones bajo las que ocurren los fenómenos distan mucho de ser controladas y predecibles en su totalidad.

La *teoría de sistemas*, es un paradigma planteado por la necesidad de un estudio más completo de los fenómenos que interesan a las ciencias. Desde esta perspectiva, resulta evidente que para cumplir con las principales encomiendas de la ciencia, es necesario dejar el pensamiento parcial y mecanicista para analizar tanto los fenómenos naturales como los sociales. Se considera la existencia de un orden universal que podría ser comprendido bajo un espectro de luz más amplio que el anteriormente explotado.

El concepto de sistema, implica la interacción de los elementos que componen un todo organizado. Asimismo, envuelve la conciencia de que en ese todo organizado (el sistema) donde se dan relaciones solidarias, se generan afectaciones internas y externas a considerar con respecto de él y sus componentes (que pueden ser otros sistemas de menor extensión a los que conocemos como subsistemas). En consecuencia, al dejar de lado el reduccionismo por el cual sus propiedades se abordaban como unidades elementales investigables independientemente, se puede ampliar y profundizar el conocimiento del fenómeno.

"El sistemismo y el funcionalismo comparten por tanto un mismo concepto fundamental: el de función, que denota la primicia del todo sobre las partes"[22] Cuestión que en gran medida compagina con las herramientas y leyes proporcionadas por la estadística, puesto que ella se encarga de regularidades que se pueden apreciar cuando se toma un grupo de datos y no a un individuo o evento en particular.

[22]Ibidem. MATTELART, Armand; et. al.; Barcelona; 1997; pp. 44

Tipología de actos

"Cualquiera que sea la tipología del comportamiento, el logro de un objetivo determinado, pone en juego una secuencia de actos.

"Los actos expresivos movilizan igualmente energía. En el caso de los actos expresivos, esa energía, aplicada sobre los órganos fonológicos, produce una secuencia de señales relacionadas con una representación; Las señales son adecuadas para lograr una interacción con el Otro, pero la eficacia no procede de la cantidad de energía que comportan, sino de la información que poseen. [...] En el caso de los actos ejecutivos, la energía que se moviliza suele ser mayor; se aplica inmediatamente como gasto energético destinado a producir un cambio en el objeto."[23]

La diferencia radica en la naturaleza de las acciones; en el primer caso, la interacción que se da entre los actores de la comunicación se da a partir del intercambio de información, mientras que el acto de naturaleza ejecutiva, implica el trabajo físico directo; el desplazamiento de un objeto o la persona mediante la fuerza aplicada directamente. Tomemos como ejemplo la siguiente situación:

Cuando un niño desea que su hermano cambie de lugar, puesto que le obstruye su campo de visión, tiene dos opciones para lograr su objetivo: Si el niño es demasiado pequeño y aún no ha desarrollado su habilidad expresiva, lo más seguro es que intente satisfacer su necesidad por medio de actos ejecutivos: cambiándose de lugar, empujando, jalando o golpeando a su hermanito para que no le "estorbe"; en cambio, si su capacidad expresiva es suficiente, seguramente recurrirá al uso de la información para expresar su deseo y procurar que su hermano le permita ver lo que desea.

[23]op. cit. MARTÍN, Serrano Manuel; et. al.; México; 1993; pp. 47-48

Objetivos de la ciencia

El objetivo principal de toda ciencia es la formulación de leyes científicas. Estos enunciados generales que describen el comportamiento de los fenómenos a que se refieren, se pueden formular a partir de la experimentación dirigida por las teorías, métodos y técnicas propios de la ciencia en cuestión. Gracias a ellas, el ser humano se explica, predice y controla algunos aspectos del **kosmos** que le rodea.

Quizá al llegar a este punto, cabría entrar en la discusión respecto de que si el Universo es determinista, aleatorio o caótico, pues la naturaleza a que nos remite el término kosmos, ciertamente conlleva a una tendencia; pero bástese, por el momento, con saber que el ideal de la ciencia es descubrir el orden que rige las interacciones entre los componentes que integran al Universo, sea cual fuere. Esta discusión será abordada en el **Capítulo 8 Introducción a la probabilidad** (pág. 407).

Clasificación de las ciencias

Según Mario Bunge "La diferencia primera y más notable entre las varias ciencias es la que se presenta entre *ciencias formales* y *ciencias fácticas*, o sea, entre las ciencias que estudian ideas y las ciencias que estudian hechos. La lógica y la matemática son ciencias formales: no se refieren a nada que se encuentre en la realidad, y, por tanto, no pueden utilizar nuestros contactos con la realidad para convalidar sus fórmulas. La física y al psicología se encuentran en cambio entre las ciencias fácticas: se refieren a hechos que se supone ocurren en el mundo, y, consiguientemente, tienen que apelar a la experiencia para contrastar sus fórmulas."[24]

[24]BUNGE, Mario; **La Investigación Científica**; México; Siglo XXI Editores; 2000; pp. 19

En otras palabras, las ciencias formales trabajan con símbolos que representan entes no reales. La lógica, es la ciencia que estudia las estructuras del pensamiento en busca del razonamiento correcto, y, ni el pensamiento ni sus estructuras se pueden ubicar –físicamente– en el cerebro, de tal manera que puedan ser captadas para su estudio empírico; por otra parte, la matemática trabaja con símbolos que no representan, originalmente, cosa u objeto alguno de la realidad. Los números y el pensamiento no se pueden comprobar en el campo físico donde nos desenvolvemos; sólo tienen cabida –hasta donde se sabe– en la mente humana y ésta, incluso, ni siquiera se puede ubicar con precisión en el individuo. Las ciencias formales, organizan –dan forma– a esos entrs que estudian.

Cuando la matemática sólo estudia relaciones entre símbolos se trata de una ciencia pura y cuando se encuentra involucrada en el estudio y/o resolución de fenómenos comprobables empíricamente, se trata de matemática aplicada; así la matemática es "ama y esclava de todas las ciencias. Tomemos por caso el siguiente:

$$4$$

Dicho número, por sí mismo no expresa nada. Pero si a éste se le dota de un contexto, entonces cobra sentido: "En la presente investigación, serán 4 las variables sociodemográficas a considerar. . . " En tal caso, ya no se trata de fórmulas puramente analíticas (fórmulas que pueden convalidarse únicamente por medio del análisis racional como es el caso de las relaciones numéricas), sino de la aplicación del producto de ellas, a la realidad para *establecer relaciones no contradictorias* entre los elementos a considerar, que es la propiedad de la matemática que más atractiva la hace para otras ciencias.

El conocimiento científico

El universo en que habitamos se ha revelado como un ente del que apenas conocemos algunos aspectos. Nuestra desarrollada capacidad de observación es limitada en comparación con todas las variables que intervienen en los procesos que nos interesan. En las ciencias naturales, por ejemplo, la aplicación de los principios provenientes de la matemática, permitió cierta precisión en áreas como la Física y la Química entre otras; no obstante, ésta es relativa.

Por ello, casi todo conocimiento que se ha obtenido a través del método general de la ciencia, es conocimiento relativo, esto es: conocimiento perfectible, dado que un momento determinado las limitaciones de las herramientas teóricas, metodológicas y técnicas con que se abordó la problemática planteada por la investigación científica, no permiten un acercamiento más preciso al fenómeno en cuestión, así, el conocimiento generado por ésta "puede volver a someterse a prueba, enriquecerse y, llegado el caso, superarse mediante el mismo método." [25]

Tales limitaciones suelen implicar la omisión deliberada (sesgo de la investigación debido al objetivo con que se realiza y a la elección de la(s) teoría(s) con que se aborda el objeto de estudio) o circunstancial de cierta(s) variable(s) (errores propios de los instrumentos de medición e incluso elecciones inapropiadas de teoría(s), metodología y técnicas de estudio), que influyen de manera importante en la construcción del conocimiento científico. De hecho, la misma lógica de la investigación considera y se encuentra con dichas limitaciones, por lo que una vez edificado un nuevo cuerpo de conocimientos, éste apunta hacia otros problemas que pueden, incluso, referirse a la misma investigación.

[25] Op. cit. BUNGE, Mario; México; Siglo XXI Editores; 2000; pp. 3

Otro aspecto que incide es la necesidad de contrastar el conocimiento con la realidad. “Lo único que puede probarse hasta quedar más allá de toda duda razonable, son o bien teoremas de la lógica y la matemática o bien, enunciados fácticos triviales (particulares y de observación)”[26] como “este líquido está congelado”, fuera de ello, la única certeza que existe es el **falibilismo**, la certidumbre de que nuestro conocimiento del universo es provisional e incierto. Tal “relatividad” podría encontrarse dada tanto por las relaciones no evidentes o consideradas que los elementos de un sistema guardan entre sí, como por las de naturaleza semejante que los componentes del sistema establecen de manera irregular o aleatoria con los componentes de otros sistemas.

Los hechos que se encuentran sujetos a la comprobación empírica, se encuentran afectados por estas limitantes humanas para su estudio; consideremos el siguiente experimento:

Galileo definió: *El Movimiento Rectilíneo Uniforme (M. R. U.), es aquel en el cual se recorren distancias iguales en tiempos iguales. En este caso, el desplazamiento es directamente proporcional al tiempo.*

Supongamos que un alumno, es capaz de correr a una velocidad de 8,25 m/s durante un lapso de 12 segundos. Como tenemos la velocidad y el tiempo, se puede determinar la distancia que recorre, y ésta sería igual a 99 m

Para corroborar el principio enunciado por Galileo, un grupo de físicos hizo correr al mismo muchacho, en varias ocasiones durante periodos de 36 segundos, obteniendo “el mismo resultado” con pequeñas, “insignificantes” variaciones.

[26]Op. cit. BUNGE, Mario; México; Siglo XXI Editores; 2000; pp. 5

Ante tal situación se plantea la siguiente disyuntiva: ¿Dichas variaciones son tan insignificantes como para omitirlas o tan importantes como par descalificar el principio del M. R. U.? Ocurre que si se repite el mismo experimento un número indefinido de veces en forma teórica, el resultado siempre será el mismo, pero si se hace en la vida real, existirán un importante número de variables (componentes del sistema) que al no estar contempladas por los investigadores, producirán efectos que incidirán en el resultado de los eventos (cada uno de los recorrido hechos por el sujeto) y por ende en la comprobación de dicha definición científica.

Dichas variables podrían ser: Las variaciones de temperatura que afectan al corredor en los diferentes eventos que conforman el experimento; las diferentes corrientes de aire que causan resistencia al avance del corredor; el desgaste que causa al corredor un número considerable de repeticiones del evento; el incremento en la fuerza muscular como producto de dichas repeticiones; la predisposición, disposición, y/o actitud, del sujeto; las posibles distracciones causadas por situaciones "ajenas" a la carrera; las diferentes trayectorias seguidas por el corredor; la amplitud y la potencia de cada zancada y sus diferencias; el ritmo cardiorrespiratorio, las diferencias e irregularidades en la pista donde se desarrolla; el desgaste de sus zapatos, la altitud y latitud en donde se desarrolla el experimento, presión atmosférica, la hora en que se realizan los eventos del experimento ...

Todos esos elementos y otros que no se pueden cuantificar con respecto de su impacto en cada evento, hacen que la ciencia fáctica sea relativamente precisa. Quizá para este punto ya surgió la inquietud con respecto a que se trata de un sistema de experimentación donde uno de los elementos es un subsistema vivo muy inestable (el corredor es un subsistema que difícilmente se puede comportar de la misma manera más de una vez), con lo que una vez más se puede cuestionar dicho principio o la misma investigación.

Cosas similares ocurrirían si en lugar de un corredor, el objeto fuera un balín que corre sobre una canaleta; debemos recodar que por cada evento, las condiciones cambian y no podemos medir con la más exacerbada de las exactitudes los errores de medición (los humanos, como los propios de los instrumentos), ni el número de partículas que se transfieren y modifican las superficies donde se realizan los eventos, si ocurrirá un sismo durante el desarrollo de algún evento, si la presión y temperatura... variables imprevistas que condicionan la exactitud del conocimiento científico.

Lo anterior no es otra cosa que hacer notar algunas limitaciones a que se enfrenta, en ciertos momentos, la investigación científica (herramienta en la producción del conocimiento científico). Estas limitaciones, ciertamente, no son insuperables y en algunos casos se pueden obviar. El conocimiento científico es conciente de estas limitaciones y por ello, se asume como falible. Bajo esta lógica, todo conocimiento es perfectible y opera mientras sea útil; "parte del sentido común de hoy día, es resultado de la investigación científica de ayer."[27]

Finalmente, cabe destacar que el ejemplo remite al ámbito natural en que las condiciones pueden ser (relativamente) controladas. En el caso de las Ciencias Sociales, el asunto suele ser más complejo. Al retomar una de las variables del anterior ejemplo, ello quedaría más claro: ¿Qué ocurriría si lo que nos interesa es determinar la predisposición o actitud del corredor? ¿Cómo obtener indicadores que nos permitan calcular la veracidad de la respuesta si aplicamos un cuestionario? Ciertamente hay técnicas que nos

[27]Op. cit. BUNGE, Mario; México; Siglo XXI Editores; 2000; pp. 3

permitirían hacerlo desde la estructura del cuestionario, pero nada garantiza que la habilidad mental del sujeto le impida detectarla y manipularla a su antojo.

Comunicación social

El término comunicación, sin un contexto que lo determine, es tan amplio y complicado como el que más. Éste tiene tantos sentidos y significados como la polisemia misma de la palabra. En sí, la comunicación no es un proceso privativo de la especie humana, e incluso se podría decir que ni siquiera de los seres vivos.

Se habla de comunicación al referir la posibilidad de poner en común dos localidades o más mediante el aprovechamiento de una nueva carretera (comunicación como transporte), a la conexión de dos habitaciones por medio de una puerta o ventana (comunicación como unión o puesta en común), también hay quien sostienen que si una planta varía su aspecto se está comunicando, e incluso cuando se habla de la emisión-recepción de impulsos electromagnéticos entre dos máquinas (telecomunicaciones).

Desde este punto de vista, se estará compartiendo uno de posprincipios fundamentales expresados por El Colegio Invisible (La escuela de Palo Alto) que dice Lo siguiente: "No se puede no comunicar", que es igual a decir "Todo comunica" y siendo así, nos encontramos ante una problemática concreta que podemos definir mediante la siguiente pregunta: ¿Cuál es el objeto de estudio de la comunicación?

De seguir la propuesta teórica de Palo Alto, nos correspondería, como estudiosos de la comunicación, abordar las variaciones físicas de una planta cuyo hábitat es el centro de una sala, la forma en que nos mira un perro, el sentido y/o significado de un peinado, la ruta que sigue un automóvil y más. Así, no hay un objeto preciso que tratar y por tanto, las teorías, métodos y técnicas con que cuenta este campo de estudio, de muy poco o nada servirían.

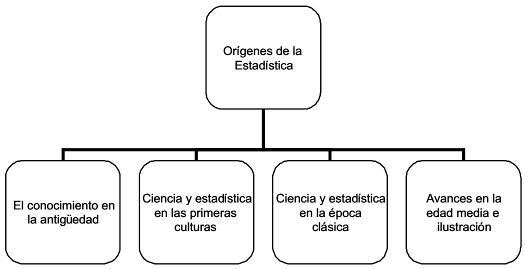
La visión que se tiene de la comunicación es el hilo conductor que guía su quehacer. En lo que concierne a la FES Acatlán, dicha visión ha sido sesgada de manera intencional por una dirigida hacia el antropocentrismo, donde la *interacción comunicativa* entre los *actores sociales de la comunicación* es la parte medular del trabajo que se realiza en dicha área del conocimiento. Los actores –que al menos deben ser dos– pueden ser entes particulares, pequeños grupos, organizaciones, o instituciones.

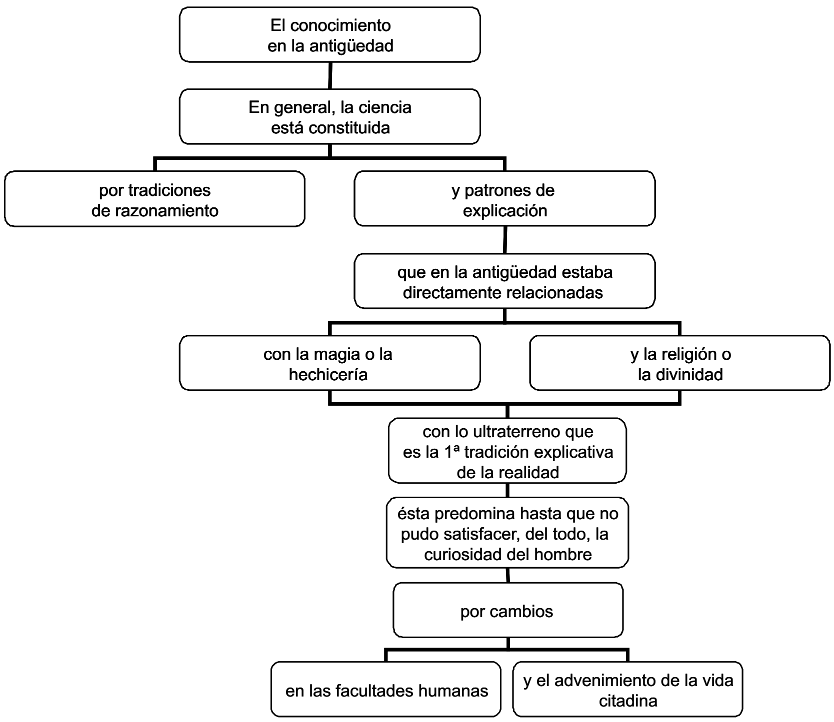
Así, el centro de atención son *aquellas interacciones por el recurso a actos expresivos que se dan al interior del sistema social*, es decir: Los procesos comunicativos de la actividad social humana; lo que incluye las relaciones pertinentes a los siguientes elementos: los *actores*, los *instrumentos*, las *expresiones* y las *representaciones*.

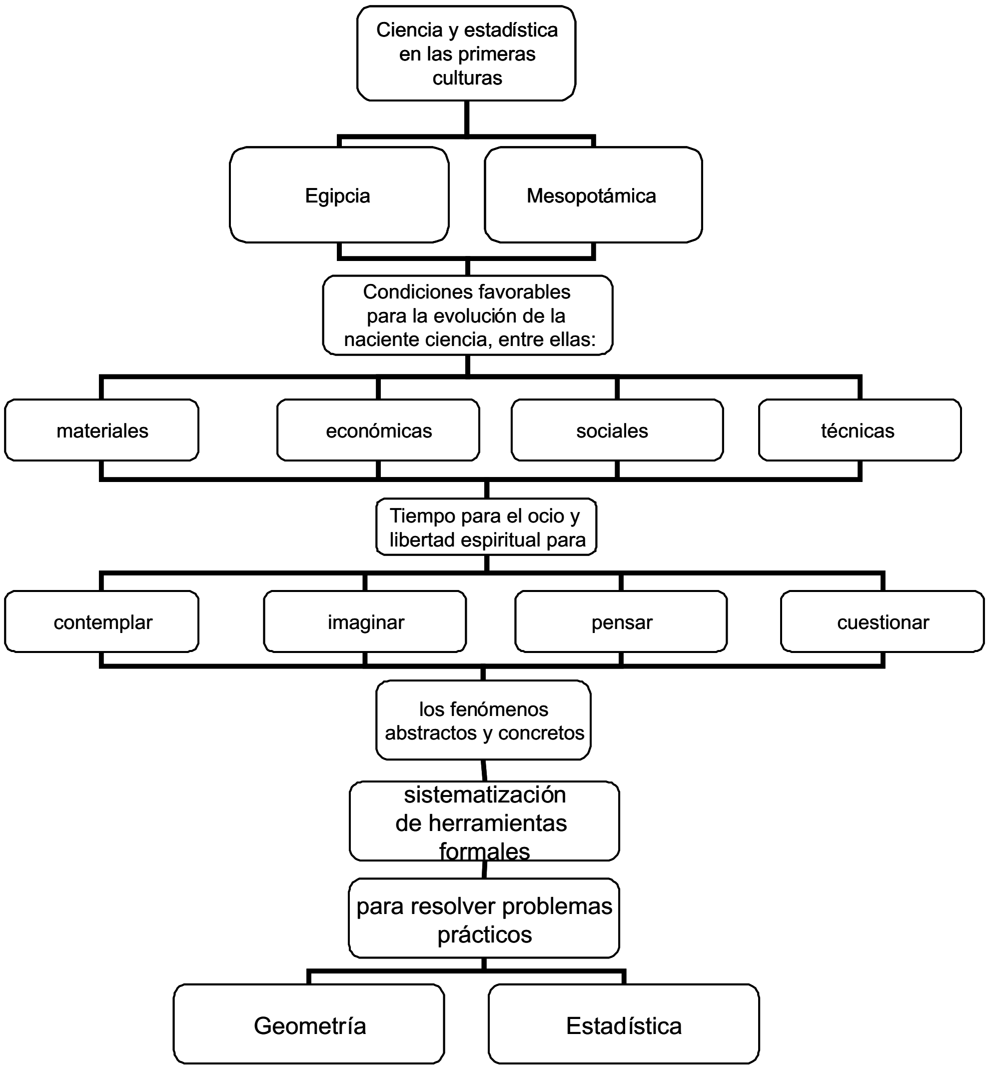
La *comunicación social*, es aquella que se da entre los elementos que conforman el sistema social, pero más aún es aquella en que la intencionalidad del trabajo expresivo tiene un valor de uso concreto y éste es el de la socialización. Por medio de la comunicación social, los actores le encuentran y dan forma al mundo que los rodea; se integran al Sistema Social aprehendiendo roles, adquieren su propia identidad y establecen las relaciones grupales con que imprimen la dirección a la dinámica del sistema social.

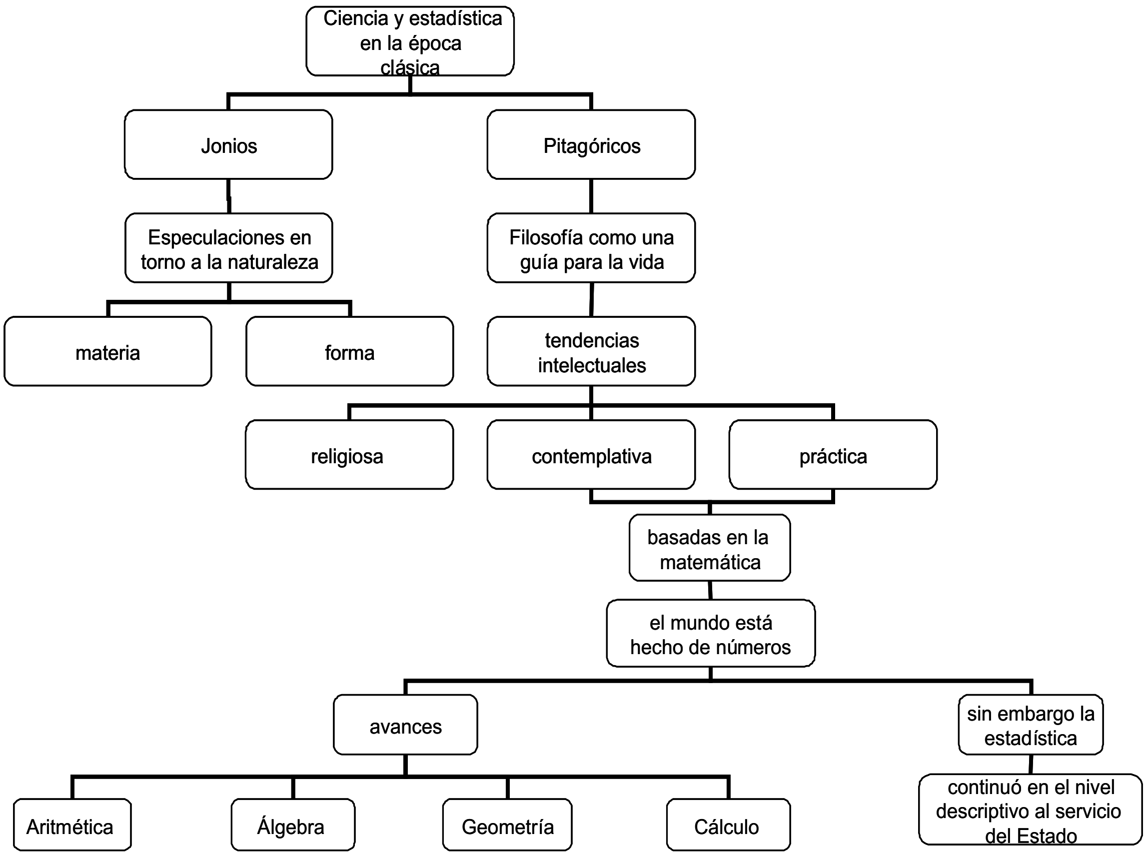
1.7.2. Respuestas a los ejercicios de la unidad

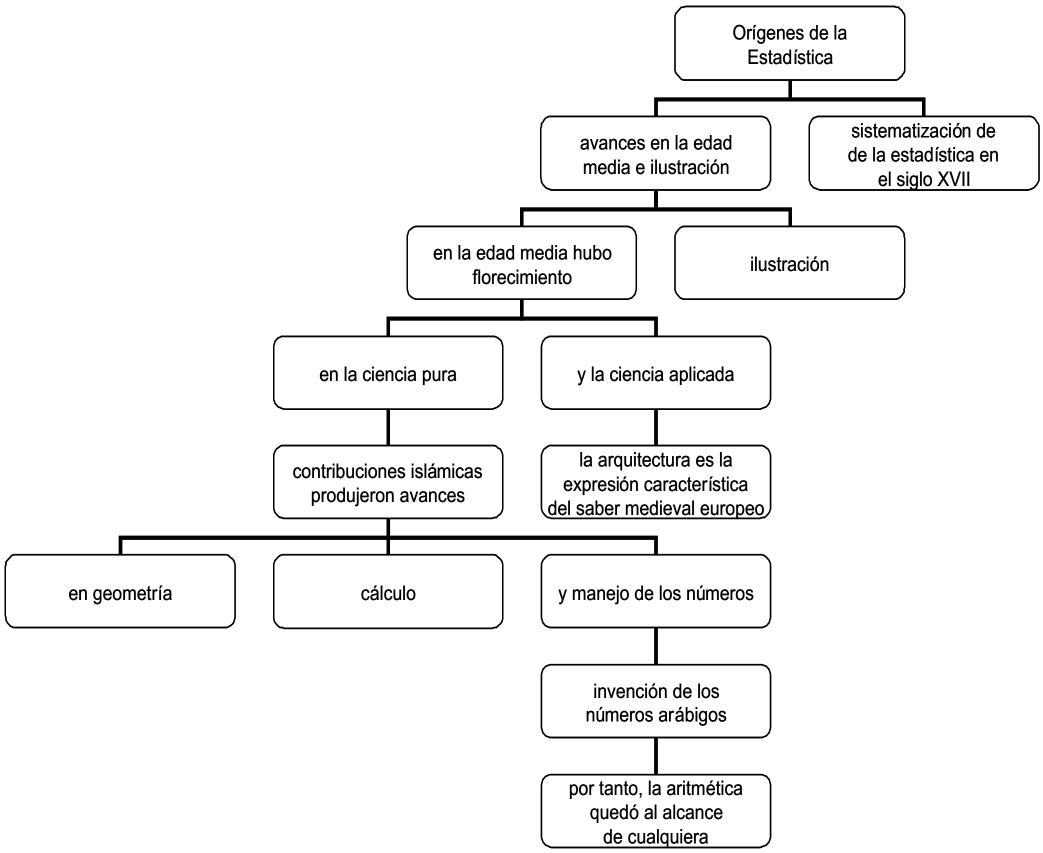
Ejercicios 1

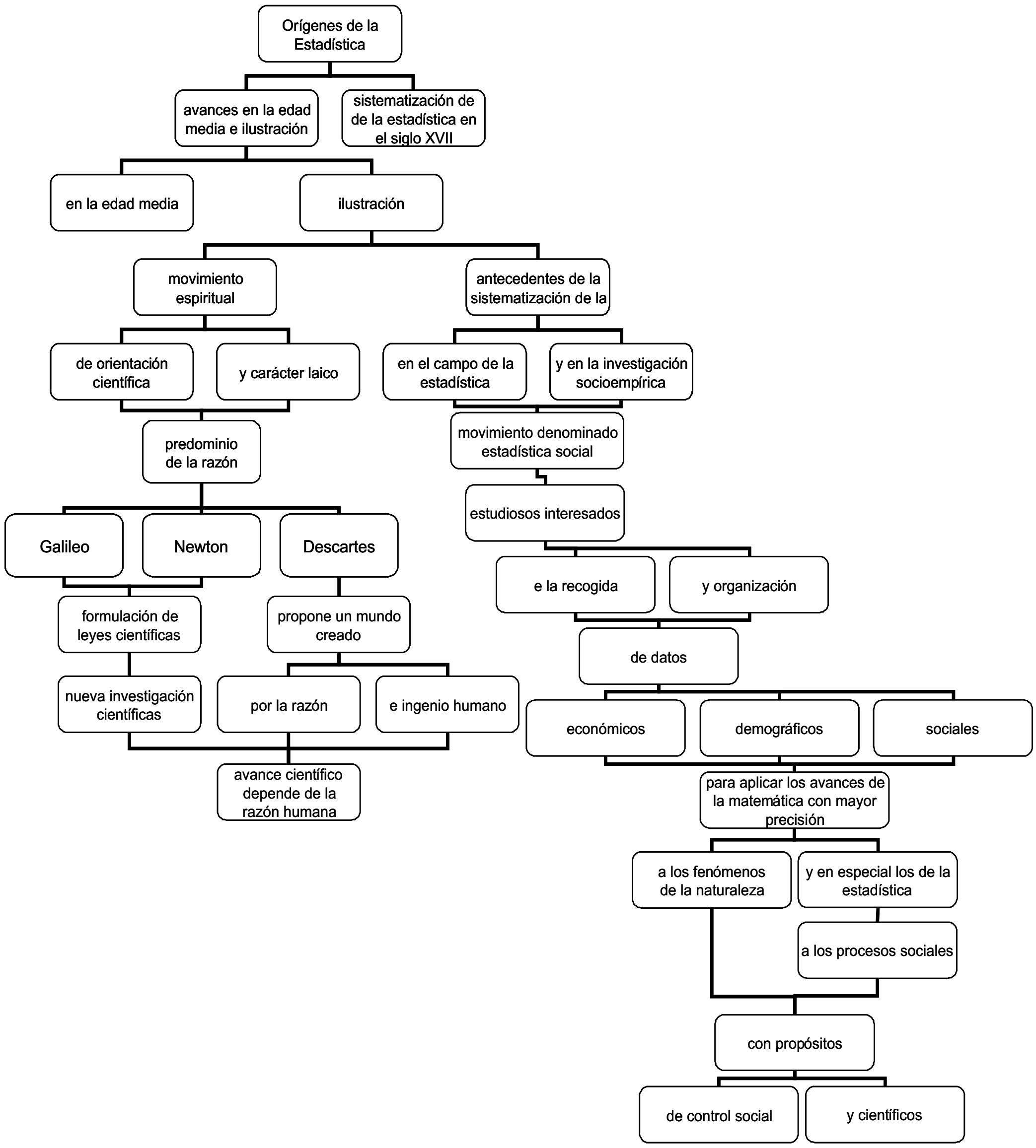





Ciencia y estadística en las primeras culturas
Egipcia
Mesopotámica
Condiciones favorables para la evolución de la naciente ciencia, entre ellas:
materiales
económicas
sociales
técnicas
Tiempo para el ocio y libertad espiritual para
contemplar
imaginar
pensar
cuestionar
los fenómenos abstractos y concretos
sistematización de herramientas formales
para resolver problemas prácticos
Geometría
Estadística


Ciencia y estadística en la época clásica
Jonios
Pitagóricos
Especulaciones en torno a la naturaleza
Filosofía como una guía para la vida
materia
forma
tendencias intelectuales
religiosa
contemplativa
práctica
basadas en la matemática
el mundo está hecho de números
avances
sin embargo la estadística
continuó en el nivel descriptivo al servicio del Estado
Aritmética
Álgebra
Geometría
Cálculo


Orígenes de la Estadística
avances en la edad media e ilustración
sistematización de de la estadística en el siglo XVII
en la edad media hubo florecimiento
ilustración
en la ciencia pura
y la ciencia aplicada
contribuciones islámicas produjeron avances
la arquitectura es la expresión característica del saber medieval europeo
en geometría
cálculo
y manejo de los números
invención de los números arábigos
por tanto, la aritmética quedó al alcance de cualquiera


Orígenes de la Estadística
avances en la edad media e ilustración
sistematización de de la estadística en el siglo XVII
en la edad media
ilustración
movimiento espiritual
antecedentes de la sistematización de la
de orientación científica
y carácter laico
en el campo de la estadística
y en la investigación socioempírica
predominio de la razón
movimiento denominado estadística social
estudiosos interesados
Galileo
Newton
Descartes
e la recogida
y organización
formulación de leyes científicas
propone un mundo creado
de datos
nueva investigación científicas
por la razón
e ingenio humano
económicos
demográficos
sociales
avance científico depende de la razón humana
para aplicar los avances de la matemática con mayor precisión
a los fenómenos de la naturaleza
y en especial los de la estadística
a los procesos sociales
con propósitos
de control social
y científicos

Ejercicios 2

1. ¿Por qué se dice que en el siglo XVII se dio el nacimiento de la estadística moderna?
   Porque a principios de este siglo, los avances de la matemática incidieron sustancialmente en el campo de la estadística, de tal manera que mejoraron las técnicas de inferencia para el análisis de los datos sociales.

2. ¿Cuál es el propósito de las primeras investigaciones sociales realizadas por los denominados reformadores?
   Aportar evidencia empírica de los problemas sociales existentes en su época. Su finalidad era eminentemente política y consistía en movilizar a la opinión pública para exigir medidas concretas de reforma social.

3. ¿Por qué Adolfo Quételet es considerado como un pionero en la aplicación de modelos matemáticos en la sociología?
   Porque con su texto Física social, marcó la transición de la simple descripción estadística al empleo consciente de los datos cuantitativos empíricos para establecer, mediante técnicas estadísticas de análisis, las regularidades de la vida social.

4. ¿Cuál es el principio que sustenta la exactitud de la matemática como una ciencia formal?
   El principio de no contradicción, es decir: Los elementos básicos de la estructura matemática (axiomas, postulados, lemas, teoremas y corolarios), se encuentran encadenados de tal manera que no se podría admitir lo contrario de lo que se afirma lógicamente y, por tanto, se admite como verdadero.

5. *Axioma*: Es una afirmación que se considera *tan evidente* que se acepta como verdad sin que sea necesario demostrarla. De estas aseveraciones

parte el conocimiento matemático para caracterizarse como irrefutable.

6. *Teorema*: Es una proposición matemática demostrable; se construye con Axiomas, Postulados y Lemas. En este caso, es indispensable la demostración lógico-matemática para validar la conclusión a que se ha llegado; por tanto, es la proposición más formal de las matemáticas y permite la construcción de nuevos teoremas.

7. ¿Cómo son las regularidades que se pueden encontrar en los sistemas sociales mediante el uso de la estadística?
Las regularidades en el mundo social no son mesurables de forma directa, pues éstas no se pueden apreciar o demostrar de forma física e individualmente. La naturaleza de los fenómenos sociales dista mucho de ser ordenada y predecible en un sentido tradicional, es decir inmediata y particularmente, pues en ellas intervienen múltiples variables y propiedades que no se pueden medir físicamente, pero sí formalmente.

8. ¿Qué es una muestra?
La parte representativa (extraída mediante mecanismos adecuados) de una población o universo al que se desea estudiar, por lo que se dice que la muestra es un microuniverso o copia a escala, del universo en cuestión.

1.7.3. Respuestas de la autoevaluación

1. En la curiosidad-racional inherente a su naturaleza; pues ella le obliga a buscar explicaciones, congruentes y sustentadas por pruebas empíricas, de los fenómenos que le rodean, afectan y de los que forma parte.

2. Porque en su momento, como la ciencia actual, fue la herramienta abstracta con que el ser humano explicaba la realidad e intentaba controlarla. Asimismo, era reflejo del conocimiento y la cosmovisión que el

ser humano tenía de su entorno y de sí mismo. Finalmente, en ella, se comenzaban a ordenar los componentes que conformaban el acervo de conocimientos (estructurar) que se estaban generando.

3. Son varios los elementos que intervienen en el proceso. Entre ellos se encuentran la incesante curiosidad-racional humana (como una constante); la evolución de las capacidades que permitieron realizar inferencias lógicas que a la postre llevaron a la elaboración de verdaderos razonamientos lógicos; la formación de la ciudad y sus beneficios materiales y sociales, así como otros elementos combinados, todos ellos, con la casualidad e incluso el azar.

4. La estadística era una actividad puramente descriptiva; se realizaban censos para conocer y administrar los recursos con que contaba el Estado para subsistir y no incluía – como actividad cuantitativa– ninguna técnica de inferencia.

5. En que fue la primera cultura del mundo antiguo que se ocupó de establecer un orden universal mediante explicaciones preponderantemente racionales, científicas; cuestión que impulsó el avance múltiples aspectos del conocimiento. Entre esos aspectos, se encuentran los que se dieron en el campo de la matemática, que si en ese momento no afectaron de manera directa a la estadística, sí lo hicieron a la postre.

6. En la primera etapa de la edad media, la cultura islámica hizo una importante contribución al campo de las ciencias puras (en la matemática) al inventar los números arábigos; lo que prácticamente puso al alcance de cualquier persona la aritmética.

7. La conjunción de elementos formales con factuales para indagar el origen (causa - efecto) de fenómenos naturales. Dicha conjunción derivó en

la formulación de la primera ley científica que permitió explicar y predecir, con base en la matemática, un fenómeno natural. Newton sintetizó y completó las ideas sostenidas por Galileo, entre otros, en su primera ley de movimiento, y formalizó la explicación de la gravitación universal (Ley de la gravitación universal), es decir: la enunció de tal manera que le fue posible sintetizarla en una fórmula matemática que permite la demostración de su afirmación.

8. El pensamiento cartesiano constituyó el gran cisma con la cosmovisión teológica que predominó durante la edad media; lo que implica el rompimiento con el pensamiento circunstancial y/o divina que envolvía la razón de los fenómenos acaecidos en la realidad y cuyas causas no eran evidentes. Con ello, responsabiliza al ser humano por sus actos y reconoce en él, la capacidad que tiene para alcanzar la felicidad y el progreso por sus propios medios, principalmente por la razón.

9. Es la base para la sistematización de la estadística. Esta nueva forma de pensar, infundió en la comunidad científica el ánimo para aplicar los conocimientos de las ciencias formales, particularmente los de la matemática, a los fenómenos de la naturaleza y, posteriormente, a la medición de los fenómenos sociales con los mismos instrumentos empleados en las ciencias naturales.

10. Quételet vislumbra la posibilidad de entender el orden intrínseco de los hechos sociales. Con su "*teoría de las regularidades de los fenómenos sociales*" aplica los avances en el campo de la estadística, como la incorporación del muestreo aleatorio y la disminución de los errores muestrales, para elaborar conclusiones que van más allá de la simple descripción y observar las regularidades de los fenómenos sociales. En su texto *Física Social* se puede ver, por primera vez, el empleo consciente

de los datos cuantitativos empíricos para establecer las regularidades de la vida social.

11. La estadística se constituyó como una herramienta científica cuyo valor radica en la precisión, tanto de su lenguaje como de sus técnicas y métodos, que permite obtener conclusiones sustentadas por principios no contradictorios. Ello, permite reconstruir, expresar y analizar, incluso, las relaciones más complejas entre los hechos y diversos aspectos de ellos. Esta característica es importante para el ser humano en dos aspectos. El primero, es que gracias a ella, el hombre puede tomar decisiones altamente racionales y con un elevado rango de certidumbre, pues al incorporar las nociones de no contradicción y aleatoriedad con las que se encuentra profundamente ligada, hasta lo “desconocido” se puede medir; en segundo lugar, sus aportaciones en el ámbito científico, sobre todo en lo referente a las ciencias sociales y humanidades, son fundamentales en tanto que se erigen como un elemento constitutivo (fundador) de una nueva visión en disciplinas tan complejas como el hombre mismo.

12. La complejidad propia de la naturaleza de su objeto de estudio -las interacciones entre los seres vivos, que se llevan a cabo por el recurso a actos expresivos. Las relaciones que se dan entre los componentes del Sistema de Comunicación, son tan diversos y heterogéneos que en apariencia distan mucho de ser ordenados o sistemáticos en el sentido tradicional, es decir inmediata y particularmente. Por lo que las nociones de aleatoriedad y regularidad estadística, son apropiadas para el estudio de la comunicación.

13. Una población es el total de los elementos del sistema que se desea estudiar, en otras palabras: Es el universo del que se pretende averiguar algo.

La muestra, es una parte representativa de la población, es como una copia a escala, un micro universo, lo que significa que contiene todas las características que constituyen a la población como un objeto de estudio.

14. Población: La comunidad estudiantil de la UNAM. Muestra: un grupo de estudiantes que reproduzca lo más fielmente posible las características que posee la comunidad universitaria, es decir estudiantes de todas las carreras en sus diferentes semestres y de ambos sexos.

15. Que tiene objetivos diferentes. La estadística descriptiva se encarga, principalmente, de analizar y describir sólo una parte representativa del sistema o universo que se desea estudiar. A ella se le llama muestra, y se obtiene, al igual que los descriptores y características que posee, mediante procedimientos formales. Por otra parte, la estadística inferencial, también utiliza procedimientos formales, mas lo hace con el propósito de generalizar las propiedades y/o resultados obtenidos de la muestra, al sistema o universo (también llamado población objetivo) del que fue extraída la muestra. Ambas ramas de la estadística son diferentes pero complementarias.

1.7.4. Glosario

Apodíctica: "El encadenamiento de proposiciones de tal manera que no se podría admitir lo contrario de lo que se afirma por lo que equivaldría a contradecir lo que previamente ya se había admitido."[28]

Ciencia: "Una ciencia es una disciplina que utiliza el método científico

[28]GUTIERREZ, Saenz Raúl; **Introducción a la Lógica**; México; Editorial Esfinge; 1997; pp. 275

con la finalidad de hallar estructuras generales (leyes)."[29] Ésta, es una actividad social que se encarga de cuestiones intelectuales y con la cual se genera conocimiento caracterizado como objetivo, racional, sistemático, exacto, verificable –por tanto falible– y, en éste tenor, creciente. Por otra parte, el vocablo *Ciencia*, en general, equivale literalment al de conocimiento.

Conocimiento:"El conocimiento es un proceso histórico, y la manera como ha sido constituido históricamente no puede desligarse de la manera como se formulan las preguntas y evalúan las respuestas."[30] En dicho proceso, se da la identificación de algunas propiedades pertinentes a los fenómenos objetos de estudio y las relaciones reciprocas que guardan dichas propiedades. Es una situación objetiva "y da lugar, una vez debidamente comprobado y sistematizado, a la ciencia"[31]

Despotismo ilustrado: Tendencia política que surgió en Europa hacia el siglo XVIII, a raíz de movimiento denominado la ilustración. Bajo esta perspectiva, algunos monarcas del continente, intentaron instaurar una forma de gobierno en la que el monarca se asumía como amigo de la ciencia, impulsador del progreso, tolerante, fiel a las leyes naturales y hasta defensor de los derechos humanos. En varios casos, sólo era pose, una moda.

Enciclopedistas: Grupo de pensadores europeos que se dedicaron a difundir las ideas del pensamiento ilustrado del siglo XVIII a partir de la elaboración de una enciclopedia publicada entre los años 1751 y 1756 en Francia. Compartieron la dirección de la enciclopedia D. Diderot y DÁlembert. Entre los pensadores que colaboraron con sus textos para la conformación de La

[29]Op. cit. BUNGE, Mario; México; 2000; Siglo XXI Editores; pp. 14

[30]Op. cit. MARTÍNEZ, Sergio F.; México; 1997; pp. 16

[31]FERRATER, Mora José; **Diccionario de Filosofía**; Argentina; Editorial Sudamericana Buenos Aires; 1975

Enciclopedia destacan Voltaire, Russeau y Montesquieu. Dicha obra es el resultado y el medio de difusión de las teorías del progreso, la fuerza liberadora del saber y de la convicción, éste es el mejor medio de lograr la perfección y la felicidad de la especie humana.

Estadística: Es el siguiente paso en la construcción del conocimiento científico. La estadística, nos da un conjunto de reglas para apoyar las investigaciones elaboradas a partir del método científico.

Explicación: "El concepto de explicación se ha desarrollado y ha sido examinado filosóficamente en dos direcciones que pueden detectarse desde la Grecia antigua: por un lado, se ha tratado de ver la teoría de la explicación desde el punto de vista de una teoría de la argumentación y, por otro, se ha visto como parte de la búsqueda de criterios filosóficos que nos permitan distinguir las creencias que constituyen conocimiento objetivo del mundo (creencias que desde la Grecia antigua se han se han identificado con la ciencia), de las que sólo reflejan la ilusoria realidad de nuestras percepciones y la falta de claridad de nuestros conceptos de uso cotidiano."[32]

Fenómeno: "Suele emplearse en el sentido de *apariencia*. Lo que se da a los sentidos y en la percepción"[33]

Holístico: Relativo al Holismo.

Holismo: El vocablo 'holismo' ha sido empleado para designar un modo de considerar ciertas realidades -y a veces todas las realidades en cuanto tales-primariamente como totalidades o 'todos', Y secundariamente compuestas de

[32]Op. cit. MARTÍNEZ, Sergio F.; Paidós; México; 1997; pp. 15

[33]Op. cit. XIRAU, Ramón; México; 1995; pp. 460

ciertos elementos o miembros. El holismo afirma que las realidades de que trata son prominentemente estructuras. Los miembros de todas las estructuras se hallan funcionalmente relacionados entre sí, de suerte que cuando se habla de dichos miembros se habla de relaciones funcionales más bien que de disposición u orden"[34]

Kosmos: "Palabra intraducible que combina las ideas de orden, correspondencia y belleza. Se dice que fue Pitágoras el primero que llamó al mundo de esa manera."[35]

Lógica natural:"Es una aptitud para razonar que todo hombre posee en mayor o menor grado... es capaz de desarrollo y perfeccionamiento... *Es una capacidad para razonar correctamente.*"[36]

Método: "La palabra *método* viene del griego (*metá*, al lado; *odos* camino) y significa: al lado del camino. *Es el camino o procedimiento adecuado para conseguir una finalidad.*"[37] Si se aborda este concepto únicamente desde su estructura verbal, se corre el riesgo de reducir su sentido, e incluso su significado, a una esfera de acción ejecutiva. Pero si se repara en la libertad misma que proponen los términos con que se define, es preciso enfatizar que la meta o finalidad perseguida, implica la conjunción y coordinación de las esferas ejecutiva y abstracta (las acciones y el pensamiento).

Método científico: Es el procedimiento lógico, previamente fijado, que se orienta a la consecución del conocimiento científico. Surge para cuestionar y criticar el conocimiento generado a partir del sentido común y propone,

[34]Op. cit. FERATER, Mora José; Argentina; 1975
[35]op. cit. GUTHRIE; Wiliam K. C.; México; 1995; pp. 48
[36]Op. cit.; GUTIERREZ, Saenz Raúl; México; 1997; pp. 12-13
[37]Op. cit. GUTIERREZ, Saenz Raúl; México; 1997; pp. 271

para ello, un conjunto de pasos ordenados. No se restringe a términos puramente operativos (ejecutivos), lo que de ninguna manera los excluye, sino que implica todo una serie de conocimientos teóricos, conceptuales, filosóficos y técnicos que en su conjunto, remiten a *una forma de pensar y actuar con, en y sobre el mundo*, de manera estratégica y congruente con las necesidades que plantean los diversas problemas intelectuales de que se encarga la ciencia en sus diferentes ramas."Cada método especial de la ciencia es, pues, relevante para algún estadio particular de la investigación científica de problemas de cierto tipo. En cambio, el *método general* de la ciencia es un procedimiento que se aplica al ciclo entero de la investigación en el marco de cada problema de conocimiento."[38]

Mitología: La acepción que del mito nos interesa, no es la que se relaciona con la literatura, sino aquella que se gestó en la antigüedad clásica, es decir, como a una forma atenuada de intelectualidad; en esta época, "el Mito fue considerado como un producto inferior o deformado de la actividad intelectual. Al M. se le atribuyó, a lo sumo, la 'verosimilitud' frente a la 'verdad', propia de los productos genuinos del entendimiento."[39] Aparejado con este concepto de lo que es el mito, se encuentra su significación religiosa, que se manifiesta en el uso de la mitología como un elemento de explicación a eventos tales como la formación del universo (cosmogonía) y según algunas teorías naturalistas alemanas el mito, en este sentido, "es un producto de la misma actitud teórica o contemplativa que luego dará lugar a la ciencia."[40] Primera interpretación de la realidad, cuya base es la imaginación, el pensamiento mágico. Es decir, todas aquellas tradiciones del pensamiento con que se pretende explicar la realidad a partir de elementos ultraterrenos

---

[38]Op. cit. BUNGE, Mario; México; Siglo XXI Editores; 2000; pp. 7

[39]ABAGNANO, Nicolo; **Diccionario de Filosofía** México; Fondo de Cultura Económica; 1985

[40]Ibidem. ABAGNANO, Nicolo; México; 1985

y fantásticos o fantasiosos, asociados a los fenómenos observados por el ser humano. Tradición caracterizada por la mezcla de elementos ficticios con situaciones factuales.

Positivismo: Corriente filosófica inaugurada por Augusto Comte, quien ve en la sociología una auténtica disciplina científica para el estudio de los hechos sociales, de los cuales afirma que "están regidos por un principio determinista, que debido a la complejidad del comportamiento social todavía no era conocido."[41] Lo que implica a la postre, la posibilidad de conocer en su totalidad al mundo a partir de la experiencia inmediata de la realidad que es objetiva.

Realidad: "Lo real es dado, como sugiere Kant, en el marco de la experiencia posible y por eso lo que concuerda con las condiciones materiales de la experiencia (de la sensación) es real. Algo que se presenta o puede presentarse a una conciencia"[42]

Sistema: "El término «sistema» cuando se utiliza para designar entidades reales, se opone al término «agregado». Un sistema y un agregado son igualmente conjuntos, es decir, entidades que se constituyen por la concurrencia de más de un elemento; la diferencia entre ambos, consiste en que el conjunto de los elementos de un sistema, muestra una organización de la que carecen los elementos del agregado."[43]

Sentido común: Después de la utilización de la mitología se apela al

[41]PEÑA de la, Ricardo; et. al.; **Cómo Acercarse a la Sociología**; México; Limusa Noriega; 1991; pp. 36

[42]Op. cit. FERRATER, Mora José; Argentina; 1975

[43]Op. cit. MARTÍN, Serrano Manuel; México; 1993; pp. 94 - 95

sentido común para la explicación de los fenómenos; el conocimiento que se genera a partir del sentido común aspira a ser racional, objetivo y coherente; sin embargo éste suele ser engañoso, puesto que su base es de carácter sensual, es decir: "El sentido común no puede conseguir más que una objetividad limitada porque está demasiado estrechamente vinculado a la percepción y a la acción."[44] En otras palabras, se apoya en las apreciaciones recogidas por los sentidos.

Técnica: "La técnica representa las etapas de operaciones limitadas, unidas a unos elementos prácticos, concretos, adaptados a un fin definido."[45] En otras palabras, las técnicas son la concreción de la parte metodológica, en ellas se encuentran fuertemente marcadas por los actos ejecutivos (lo práctico-operativo) que se planearon y diseñaron en una etapa anterior (diseño metodológico).

[44] Op. cit. BUNGE, Mario; México; Siglo XXI Editores; 2000; pp. 4

[45] ANDER-EGG, Ezequiel, **Técnicas de Investigación Social** Argentina; Editorial Hvmanitas, 21; 1990; pp. 41

Capítulo 2

Antecedentes matemáticos

Objetivos específicos

- El estudiante resolverá ejercicios con base en la teoría de conjuntos para la posterior aplicación de ésta, principalmente, en los temas relacionados con la teoría de la probabilidad.

- El estudiante reafirmará conocimientos básicos de aritmética y álgebra, mediante la resolución de ejercicios, para resolver los cálculos matemáticos necesarios en el estudio de la estadística.

2.1. Preliminar

Los conocimientos matemáticos fundamentales para entender, aprehender y aplicar la estadística, se encuentran en los principios de la teoría de conjuntos, aritmética y álgebra. Algunos de éstos, se practican con relativa frecuencia en la cotidianeidad de la vida y es por ello que su ejecución suele resultar sencilla; sin embargo, existen otros conocimientos, de tales áreas, que luego de ser explorados en la academia, se excluyen casi por completo de la práctica conciente, situación que puede entorpecer el aprendizaje de la materia en cuestión.

En este capítulo, se abordarán elementos matemáticos de las áreas antes mencionadas, con el fin de que su recuperación agilice el aprendizaje de la teoría estadística. Para ello, es necesario confirmar el conocimiento de aspectos mínimos que permitan dicha recuperación. De igual manera, es preciso que atiendas puntualmente las instrucciones de la siguiente evaluación, pues por medio de ésta, podrás diagnósticar y subsanar carencias teórico-matemáticas que dificultarían el proceso.

Finalmente, cabe aclarar que para denotar la separación de los enteros y fracciones, se prescindió del uso del *punto decimal*; en cambio se adoptó el sistema internacional, donde se utiliza la coma con el mismo propósito.

2.1.1. Evaluación diagnóstica

- *Instrucciones*: Lee con atención las preguntas e indicaciones que anteceden a cada reactivo y resuelve lo que se te pide con cuidado. Recuerda que no es un ejercicio contra reloj, así que tómate el tiempo suficiente para razonar tu respuesta y hacerlo correctamente.
- En el caso de las preguntas con opción múltiple, marca con una "X" el inciso de la respuesta correcta.

1. Realiza, *sin el uso de calculadora*, las siguientes sumas algebraicas.

   a) $(-7) + (-8) =$

   b) $(15) + (-5) =$

   c) $(10 - 17 - 15 - 8) - (11 - 2 + 6) =$

2. ¿Qué nombre recibe el resultado de una multiplicación?

   a) Cociente

   b) Término

   c) Producto

   d) Factor

3. ¿Cuál es el nombre que reciben los números que al multiplicarse permiten la obtención de un producto?

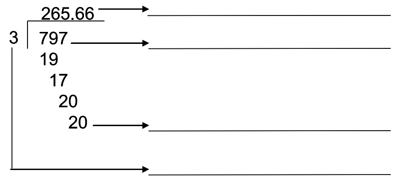
   a) Términos

   b) Multiplicación

   c) Factores

   d) Elementos

4. Coloca el nombre de cada elemento perteneciente a la división en la línea correspondiente.



5. Observa las siguientes operaciones matemáticas que se te presentan:

$$1)\ 25 \div \_\_\_\_ = 5 \qquad\qquad 2)\ 5 \times 5 = 25$$

A partir de éstas, se puede afirmar que una división es:

a) Encontrar el factor desconocido de una multiplicación.

b) Un producto notable de dos elementos numéricos.

c) El resultado de un cociente matemático.

d) Una operación aritmética proporcional a la multiplicación.

6. ¿En qué consiste la potenciación de un número.

a) Multiplicar la base de la potencia por el exponente que lo afecta naturalmente (sin cambiar de signo a la base).

b) Multiplicar sucesivamente un número base, por sí mismo, cuantas veces lo indica el exponente.

c) Multiplicar al número base por su valor absoluto indicado en el exponente.

d) Exponer variablemente la proporción del número natural mediante una multiplicación de productos cruzados.

7. En álgebra ¿qué representan las literales?

   a) Variables probabilísticas

   b) Incógnitas racionales

   c) Números reales

   d) Cantidades conocidas y desconocidas

8. Calcula el resultado de las siguientes operaciones, *sin el uso de calculadora.*

   a) $-5(-369) =$

   b) $-14 \div 7 =$

   c) $(-15) - 2(-3) =$

   d) $-81 \div -9 =$

9. Dibuja un plano cartesiano y anota en él, el nombre de *todas* sus partes.

Ahora comprueba que tus respuestas hayan sido las correctas con ayuda de la sección **1.1.2. Respuestas de la evaluación diagnóstica** (pág 91) y sigue las indicaciones que ahí se te brindan.

2.1.2. Respuestas de la evaluación diagnóstica

En esta sección, como lo indica el título, se encuentran las respuestas correspondientes a los reactivos de la evaluación diagnóstica. En general, sólo se presentan la pregunta y su respuesta correcta; esto es: En el caso de las preguntas de opción múltiple, únicamente aparece el inciso que se debía marcar y en el de las operaciones, el resultado correcto.

Compara tus respuestas con las de esta sección, cada una tiene un porcentaje de valor indicado entre corchetes []. En los casos donde el valor aparece como una fracción, éste se divide entre el número de respuestas y/o incisos que posee el reactivo; en caso de un error se pierde el total del valor asignado. Luego del cotejo, atiende a las instrucciones ulteriores; en ellas, encontrarás actividades y lecturas para resolver tus dudas y corregir los errores de cada reactivo.

1. Realiza, *sin el uso de calculadora*, las siguientes sumas algebraicas. [20 %] En caso de error, véase la *información remedial* en la pág. 181

   a) **-15**

   b) **10**

   c) **-45**

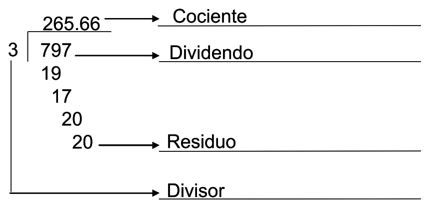
2. ¿Qué nombre recibe el resultado de una multiplicación? [20 %] *Información remedial* en la pág. 185

   c) Producto

3. ¿Cuál es el nombre que reciben los números que al multiplicarse permiten la obtención de un producto? [20 %] *Información remedial* en la pág. 185

   c) Factores

4. Coloca el nombre de cada elemento perteneciente a la división en la línea correspondiente. [$\frac{20\,\%}{4}$ Por tanto, cada acierto suma 0,5 %] *Información remedial* en la pág. 189



5. Observa las siguientes operaciones matemáticas que se te presentan:

$$1)\ 25 \div \_\_\_\_ = 5 \qquad\qquad 2)\ 5 \times 5 = 25$$

   A partir de éstas, se puede afirmar que una división es: [20 %] *Información remedial* en la pág. 189

   a) Encontrar el factor desconocido de una multiplicación.

6. ¿En qué consiste la potenciación de un número. [20 %] *Información remedial* en la pág. 190

   b) Multiplicar sucesivamente un número base, por sí mismo, cuantas veces lo indica el exponente.

7. En álgebra ¿qué representan las literales? [20 %] *Información remedial* en la pág. 194
   d)Cantidades conocidas y desconocidas

8. Calcula el resultado de las siguientes operaciones, *sin el uso de calculadora.* [$\frac{20\,\%}{4}$ Por tanto, cada acierto suma 0,5 %] *Información remedial* en las páginas 185 y 189

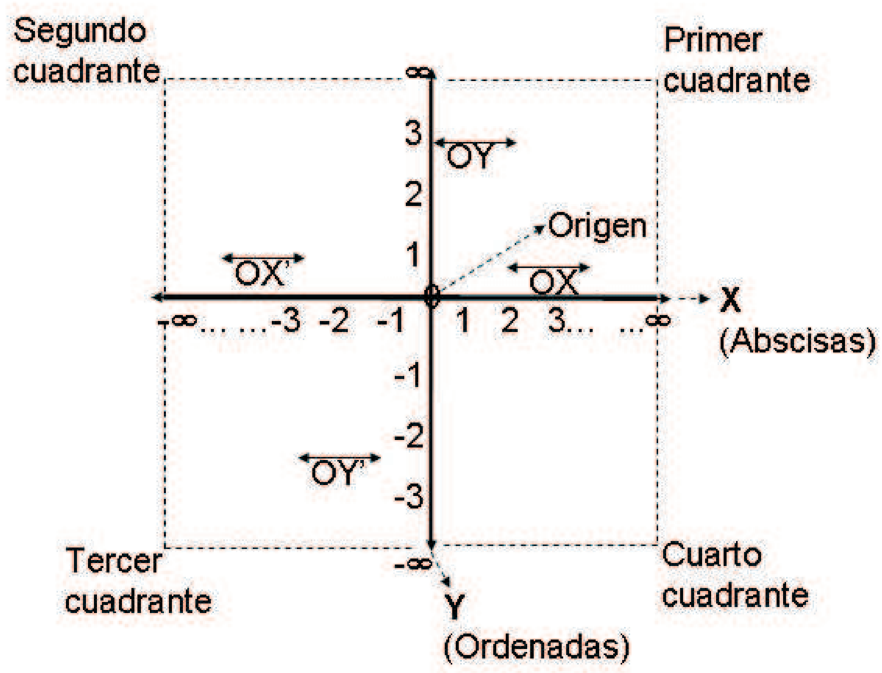
   a) 1845

b) -2

c) -90

d) 9

9. Dibuja un plano cartesiano y anota en él, el nombre de *todas* sus partes. [$\frac{20\,\%}{4}$ Por tanto, cada acierto suma 0,5 % (ejemplo: +0,5 por nombrar los cuatro cuadrantes). Para este caso, existe un elemento opcional que podría sumar 0,5 % adicional, éste es: La correcta señalización de los semiejes $\overleftrightarrow{OX}$, $\overleftrightarrow{OX'}$, $\overleftrightarrow{OY}$, y $\overleftrightarrow{OY'}$ con sus respectivas formas de graduar. El criterio para determinar si el elemento de la respuesta es correcto, consiste en verificar si está completo y coincide totalmente con la respuesta proporcionada en este apartado.] *Información remedial* en la pág. 195



Una vez realizada la verificación de tus respuestas y elaborada la suma del porcentaje, *debes felicitarte por tu esfuerzo.* **Si todas tus respuestas fueron correctas**, pasa a la sección **2.2. Simbología** (pág. 114) para iniciar con la revisión del contenido propuesto para este capítulo. **En caso de alguna duda**, te puedes apoyar en la sección *2.15.1. Información remedial de la evaluación diagnóstica* pág. (181), donde encontrarás algunos elementos matemáticos con los que podrías resolverlas.

Como habrás notado, cada respuesta cuenta con la referencia propia a la *información remedial* que deberás leer en caso de error. Si al finalizar la suma de tus aciertos, el **resultado fue mayor a** 75 % y sólo olvidaste algunas cosas, para recordarlas lee la *información remedial* indicada para cada reactivo y resuelve únicamente las actividades del apartado **2.1.3.** (pág. 95) que tienen el místmo número que aquellas en que fallaste; posteriormente podrás pasar a la revisión del apartado **2.2.** (pág. 114), con el cual se inicia el contenido de ésta sección.

Pero **si los resultados del ejercicio no fueron óptimos** (menor o igual que 75 %), para cada reactivo de la *evaluación diagnóstica*, se diseñó una *actividad de recuperación* con el objetivo de que, *si tu respuesta fue errónea o tuviste dudas* al contestar, reconsideres tus razonamientos y te sirva –dicha actividad- como incentivo para recordar y/o reafirmar tus conocimientos.

En este sentido, cabe destacar que un factor de capital importancia para tu desarrollo integral es la *honestidad intrapersonal.* Por tanto, **si tu resultado fue menor o igual que** 75 % y/o tuviste dudas en reactivos "bien" contestados, debes leer la *información remedial* (ubicada en el *apéndice del capítulo* pág. 181) para *resolver todos* los **Ejercicios de recuperación** (pág. 95).

2.1.3. Ejercicios de recuperación

- Instrucciones:
  - Antes leer los reactivos de esta sección, debes *estudiar con cuidado y atención* la Información Remedial propuesta de acuerdo los resultados de la *evaluación diagnóstica*.
  - En los reactivos de opción múltiple, señala con una "X" el inciso de la respuesta correcta; en el caso de los reactivos que contienen operaciones a realizar, no omitas ningún paso para calcular el resultado.

1. Elige la opción que defina correctamente el siguiente término: **Número**

   a) Ente imaginario que determina la cantidad de objetos (unidades) acumulados(as) unitariamente.

   b) Signo matemático cuyas características indican acumulación y/o disminución de unidades.

   c) **Símbolos** que fungen como medios auxiliares para representar cantidades de objetos.

   d) Elemento abstracto con que se realizan acumulaciones unitarias calculadas por el sentido de unidad.

2. En general: ¿qué es una suma algebraica?

   a) En general, la *suma algebraica*, es la adición en donde se puede obtener como resultado una cantidad representada por números con signo negativo.

   b) Es la adición en donde se puede obtener como resultado un signo negativo, gracias a la adición de números positivos y negativos en una misma operación.

c) La operación matemática donde intervienen números y letras (llamadas literales) que representan cantidades.

d) es la acumulación unitaria de unidades a un símbolo original cuyo resultado es la sucesión de un nuevo número real.

3. Realiza las siguientes sumas algebraicas:

   a) $(-15) + (-12) =$

   b) $11 + (-7) =$

   c) $25 - (-35) =$

   d) $(8 - 13 + 12 - 17 + 5) - (-9 - 6 - 5 + 15) =$

4. Observa la tabla de las leyes de los signos para la multiplicación y sintetízala en dos enunciados afirmativos.

$$\boxed{\begin{array}{c} (-)\cdot(+) = (-) \\ (-)\cdot(-) = (+) \\ (+)\cdot(-) = (-) \\ (+)\cdot(+) = (+) \end{array}}$$

____

____

____

____

____

5. Complementa la igualdad del siguiente modelo matemático para la ley asociativa de la multiplicación:

$$\boxed{(a \cdot b)c = \_\_\_\_\_\_\_\_}$$

6. Calcula el producto de las siguientes expresiones:

    a) (2)(9-4)=

    b) $(-15) \cdot (-7 + 2 + 5 - 10 + 12) =$

7. Coloca las palabras faltantes sobre cada línea para completar las expresiones:

    a) En principio, la división es una ______________________ a la multiplicación. Dividir, en otras palabras, consiste en ____________ el factor __________ de una multiplicación; esto, matemáticamente se expresa de la siguiente manera:

    b) Sean a, b dos números reales tal que $a \neq$ __, se dice que a es ___________ por b (a/b) si existe un número real r al que se denomina ___________, tal que: $\boxed{\frac{a}{b} = r \Rightarrow a = b \cdot r}$

    c) Donde al número a se le conoce como ___________ y al número b como ___________; cuando el cociente de $\frac{a}{b}$ no es un número ___________ (no es exacto) y los dígitos posteriores al punto decimal son infinitos, existe un número que posibilita la secuencia infinita de éstos, a tal número se le denomina ___________ y se caracteriza por ser menor que el ___________.

divisor	desconocido	cociente	entero
dividido	residuo	0	calcular
dividendo	operación inversa	divisor	

8. Resuelve las siguientes operaciones con potencias de la misma base.

    a) $10^3 \cdot 10^5 =$

    b) $\frac{10^5}{10^8} =$

    c) $(5^2)^3 =$

d) $-3^5 =$

9. Localiza los siguientes elementos en el plano cartesiano:

   a) Semieje $\overleftrightarrow{OX'}$

   b) Eje de las ordenadas

   c) Cuadrante donde se pueden localizar los pares ordenados de números cuyos signos son $(-,-)$

   d) Graduación del semieje $\overleftrightarrow{OX}$

Compara tus respuestas con las del apartado *Respuestas a los Ejercicios de Recuperación* ubicado en la página 99

2.1.4. Respuestas a los ejercicios de recuperación

- Instrucciones:
  - Verifica que tus respuestas sean correctas; para ello, deberán poseer los mismos elementos que las presentadas en esta sección. En caso de que el reactivo posea varios incisos, con un solo error la actividad completa estará mal.
  - Para los reactivos en que se realizaron operaciones, cuentan tanto el procedimiento como el resultado, si uno de los dos no corresponde con los presentados en esta sección, la actividad completa estará mal. Finalmente es importante se cuide el aspecto de los signos para cada número.

1. Elige la opción que defina correctamente el siguiente término: *Número*
   b) Símbolos que fungen como medios auxiliares para representar cantidades de objetos.

2. En general: ¿qué es una suma algebraica?

   a) En general, la *Suma Algebraica*, es la adición en donde se puede obtener como resultado una cantidad representada por números con signo negativo.

3. Realiza las siguientes sumas algebraicas:

   a) $(-15) + (-12) = -(15 + 12) = -27$

   b) $11 + (-7) = 11 - 7 = 4$

   c) $25 - (-35) = 25 + 35 = 60$

   d) $(8-13+12-17+5)-(-9-6-5+15) = (-5)-(-5) = -5+5 = 0$

4. Observa la tabla de las leyes de los signos para la multiplicación y sintetízala en dos enunciados afirmativos.

   - **La multiplicación de *factores con signos iguales* dará como resultado un signo *positivo.***
   - **La multiplicación de *factores con signos diferentes* dará como resultado un signo *negativo.***

5. Complementa la igualdad del siguiente modelo matemático para la ley asociativa de la multiplicación:

$$\boxed{(a \cdot b)c = a(b \cdot c)}$$

6. Calcula el producto de las siguientes expresiones:

   a) (2)(9-4)= 18 - 8 = 10 ó (2) (5) = 10 Cualquiera de los dos procedimientos está bien

   b) $(-15) \cdot (-7 + 2 + 5 - 10 + 12) = 105 - 30 - 75 + 150 - 180 = -30$ ó $(-15) \cdot (2) = -30$ ambos desarrollos están bien

7. Coloca las palabras faltantes sobre cada línea para completar las expresiones:

   a) En principio, la división es una **OPERACIÓN INVERSA** a la multiplicación. Dividir, en otras palabras, consiste en **CALCULAR** el factor **DESCONOCIDO** de una multiplicación; esto, matemáticamente se expresa de la siguiente manera:

   b) Sean a, b dos números reales tal que $a \neq 0$, se dice que a es **DIVIDIDO** por b (a/b) si existe un número real r al que se denomina **COCIENTE**, tal que: $\boxed{\frac{a}{b} = r \Rightarrow a = b \cdot r}$

c) Donde al número a se le conoce como **DIVIDENDO** y al número b como **DIVISOR**; cuando el cociente de $\frac{a}{b}$ no es un número **ENTERO** (no es exacto) y los dígitos posteriores al punto decimal son infinitos, existe un número que posibilita la secuencia infinita de éstos, a tal número se le denomina **RESIDUO** y se caracteriza por ser menor que el **DIVISOR**.

8. Resuelve las siguientes operaciones con potencias de la misma base.

a) $10^3 \cdot 10^5 = 10^{3+5} = 10^8 = 100000000$

b) $\frac{10^5}{10^8} = 10^{5-8} = 10^{-3} = 0{,}0001$

c) $(5^2)^3 = 5^{2 \cdot 3} = 5^6 = 15625$

d) $-3^5 = -243$

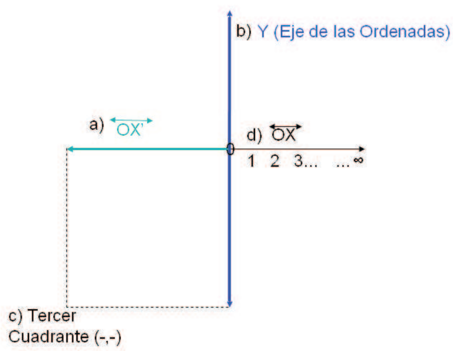
9. Localiza los siguientes elementos en el plano cartesiano:

a) Semieje $\overleftrightarrow{OX'}$

b) Eje de las ordenadas

c) Cuadrante donde se pueden localizar los pares ordenados de números cuyos signos son $(-,-)$

d) Graduación del semieje $\overleftrightarrow{OX}$



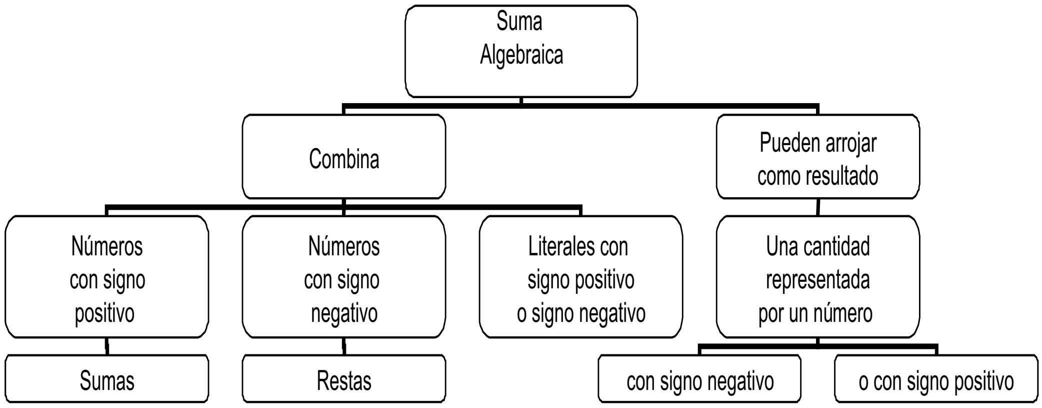
Para cada actividad resuelta en los *ejercicios de recuperación*, hay una que te permitirá *asimilar* la información que hasta el momento no haya quedado lo suficientemente clara. Sólo en caso de haber fallado en alguno de los ejercicios anteriores, pasa al siguiente apartado y resuelve el ejercicio que corresponda al número del reactivo en que te equivocaste.

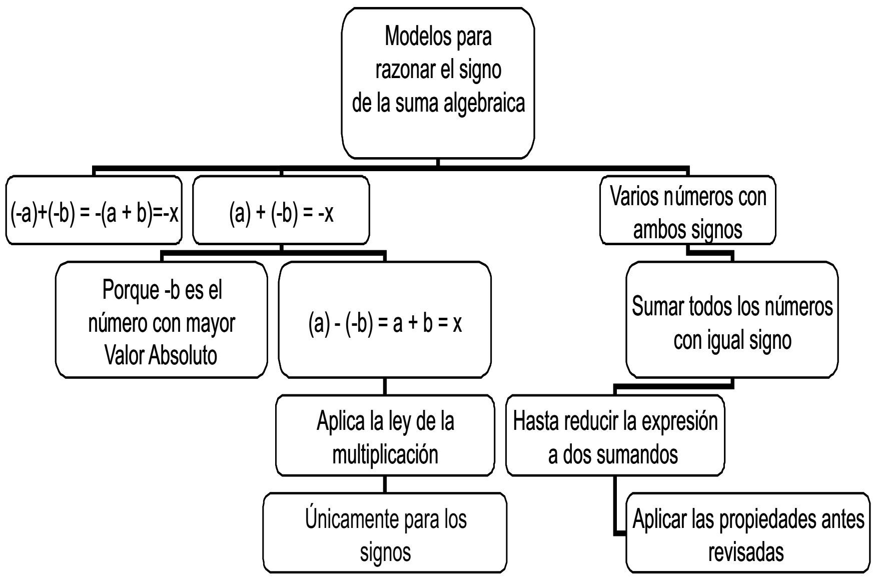
2.1.5. Ejercicios de asimilación

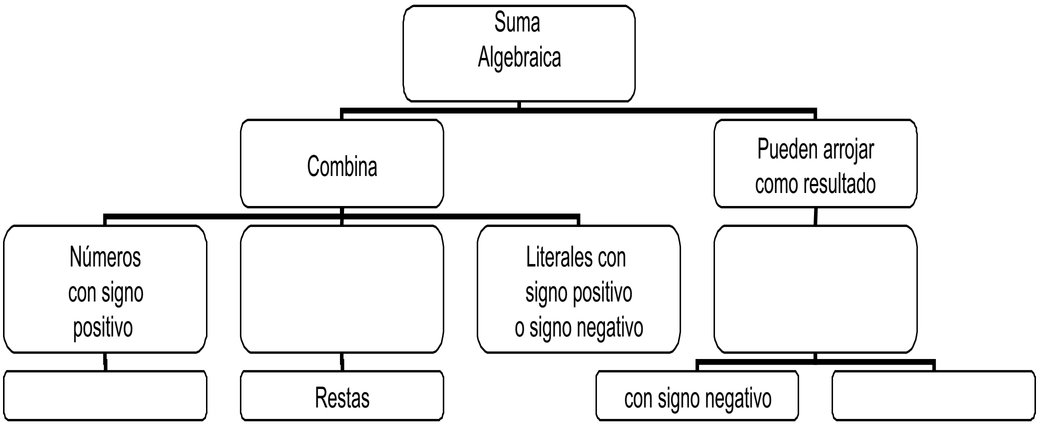
- *Instrucciones*: Lee con atención cada una de las redes conceptuales que a continuación se presentan. En ellas hay información básica referente a los conocimientos previos al abordaje del contenido de la unidad.

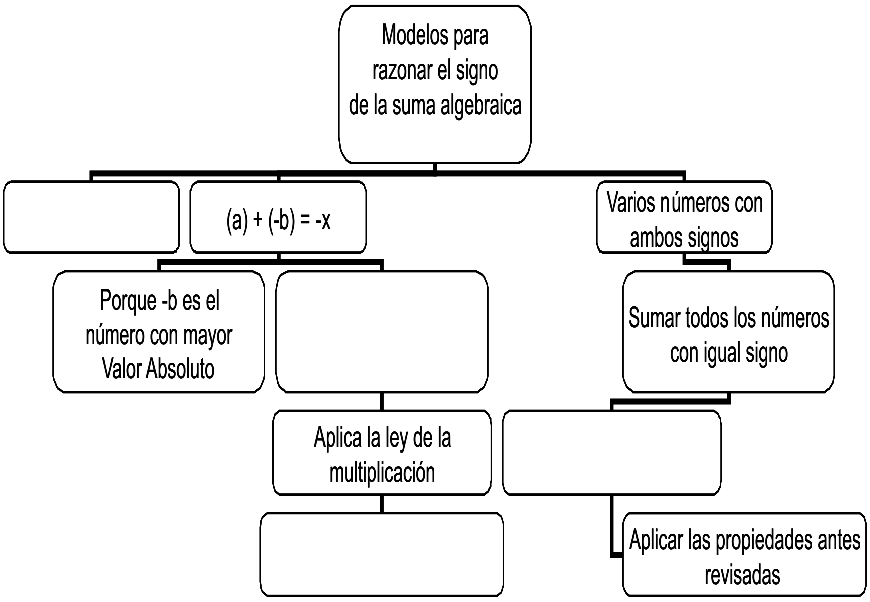
- Luego de su lectura, completa los espacios presentados en las redes posteriores; para verificar tu respuesta, regresa a la página anterior y corrobora la información que allí se proporciona.

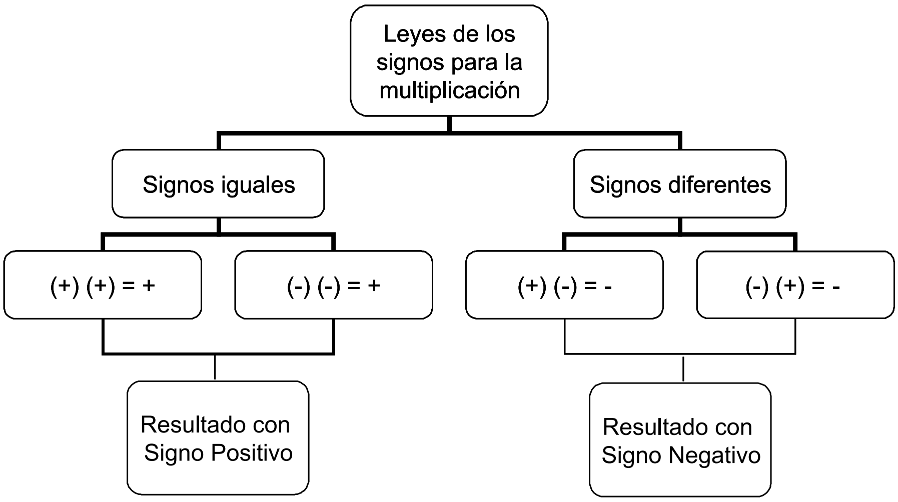
- Finalmetne, repite los ejercicios de recuperación para aplicar el conocimiento asimilado; en caso de un nuevo error, consulta tus dudas con un profesor del área.

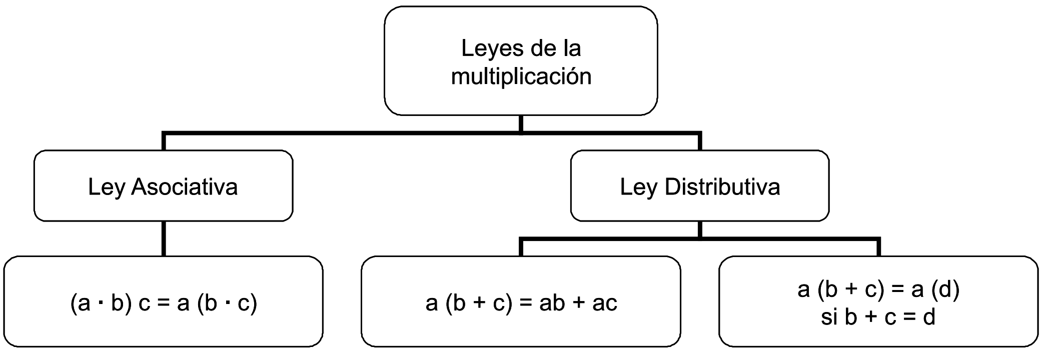

Suma Algebraica
Combina
Pueden arrojar como resultado
Números con signo positivo
Números con signo negativo
Literales con signo positivo o signo negativo
Una cantidad representada por un número
Sumas
Restas
con signo negativo
o con signo positivo

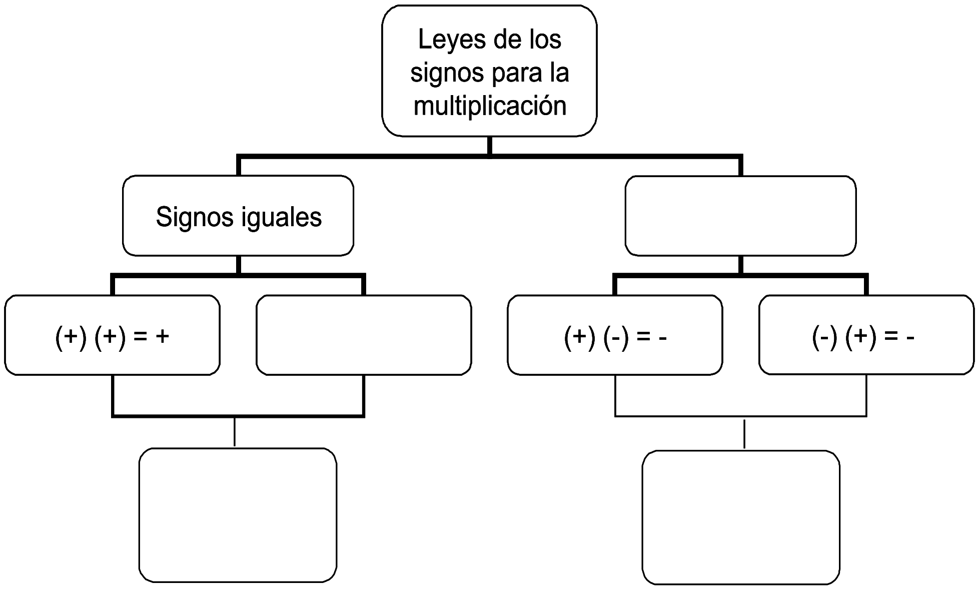

Modelos para razonar el signo de la suma algebraica
(-a)+(-b) = -(a + b)=-x
(a) + (-b) = -x
Varios números con ambos signos
Porque -b es el número con mayor Valor Absoluto
(a) - (-b) = a + b = x
Sumar todos los números con igual signo
Aplica la ley de la multiplicación
Hasta reducir la expresión a dos sumandos
Únicamente para los signos
Aplicar las propiedades antes revisadas

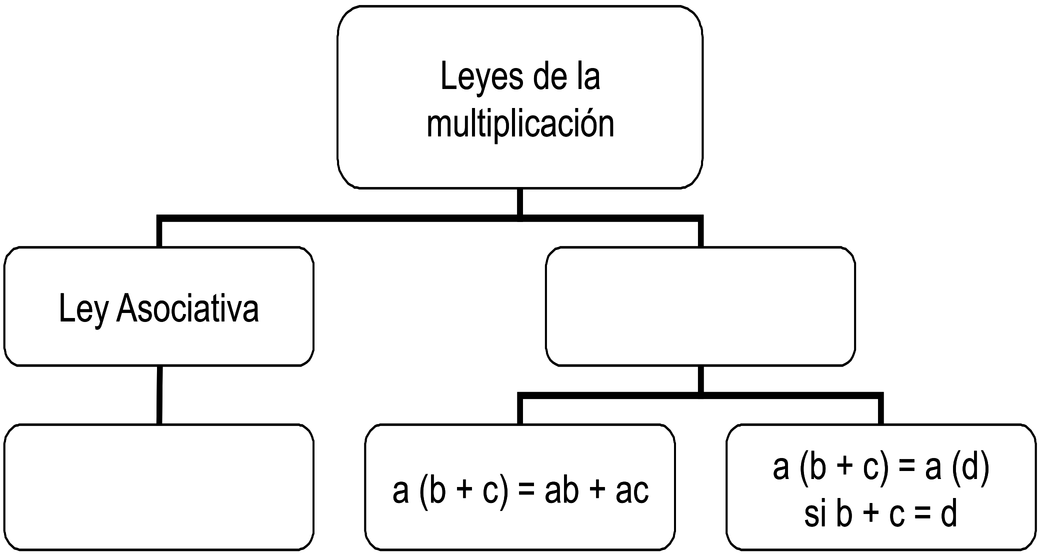

Suma
Algebraica
Combina
Pueden arrojar
como resultado
Números
con signo
positivo
Literales con
signo positivo
o signo negativo
Restas
con signo negativo

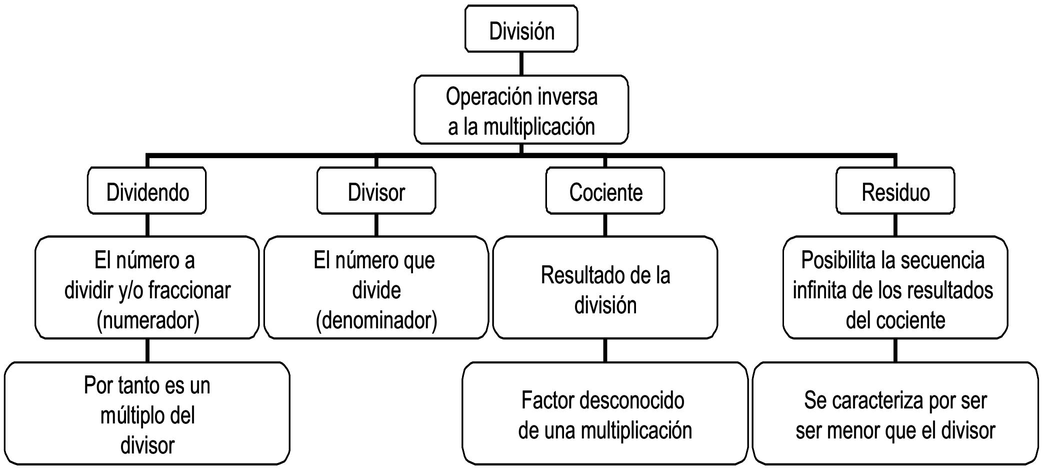

Modelos para
razonar el signo
de la suma algebraica
(a) + (-b) = -x
Varios números con
ambos signos
Porque -b es el
número con mayor
Valor Absoluto
Sumar todos los números
con igual signo
Aplica la ley de la
multiplicación
Aplicar las propiedades antes
revisadas

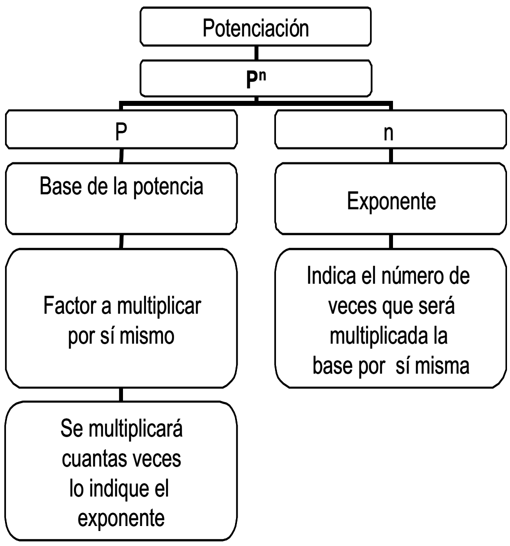

Leyes de los signos para la multiplicación
Signos iguales
Signos diferentes
(+) (+) = +
(-) (-) = +
(+) (-) = -
(-) (+) = -
Resultado con Signo Positivo
Resultado con Signo Negativo

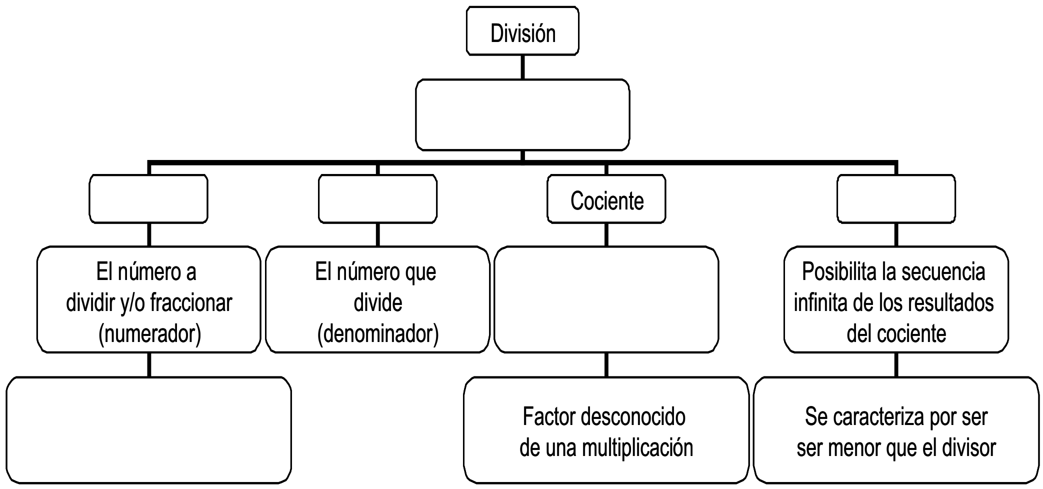

Leyes de la multiplicación
Ley Asociativa
Ley Distributiva
(a · b) c = a (b · c)
a (b + c) = ab + ac
a (b + c) = a (d)
si b + c = d

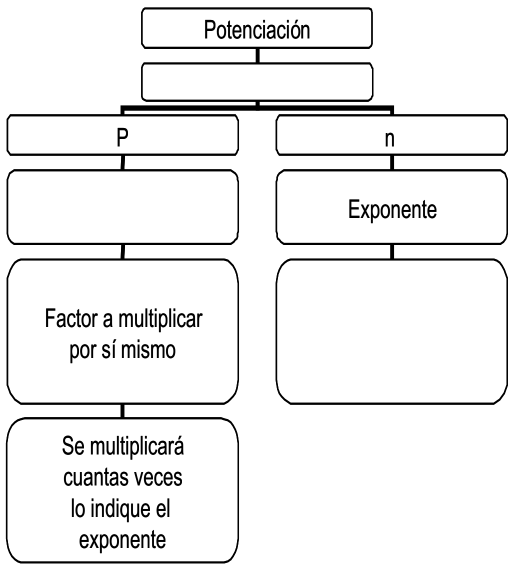

Leyes de los signos para la multiplicación
Signos iguales
(+) (+) = +
(+) (-) = -
(-) (+) = -

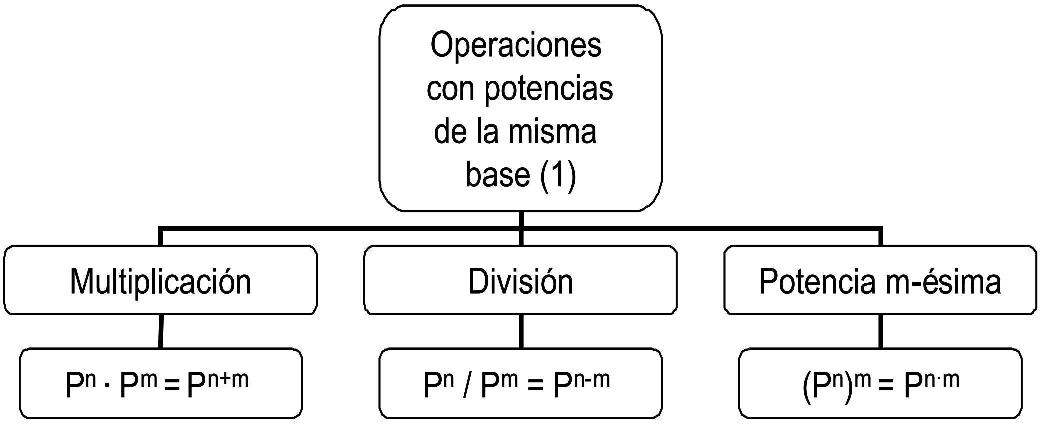

Leyes de la multiplicación
Ley Asociativa
a (b + c) = ab + ac
a (b + c) = a (d)
si b + c = d

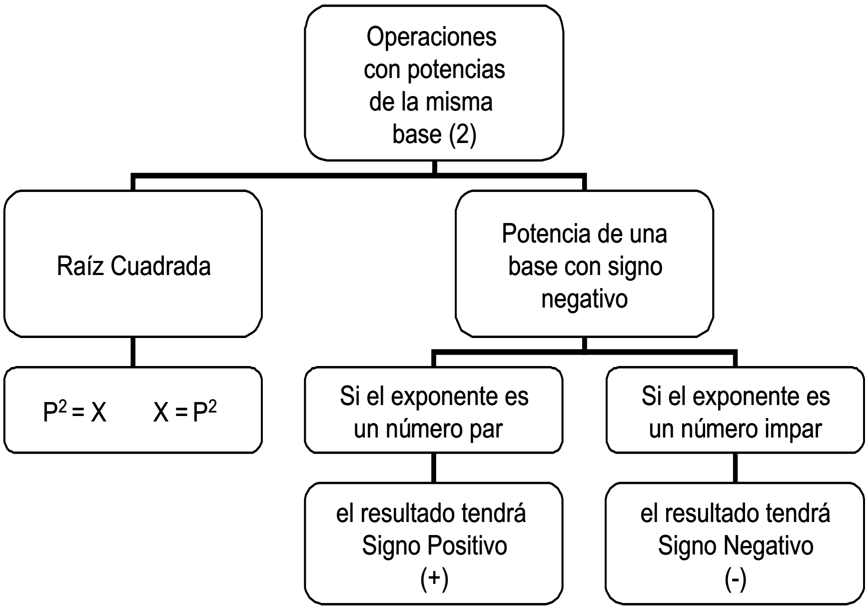

División
Operación inversa a la multiplicación
Dividendo
Divisor
Cociente
Residuo
El número a dividir y/o fraccionar (numerador)
El número que divide (denominador)
Resultado de la división
Posibilita la secuencia infinita de los resultados del cociente
Por tanto es un múltiplo del divisor
Factor desconocido de una multiplicación
Se caracteriza por ser ser menor que el divisor

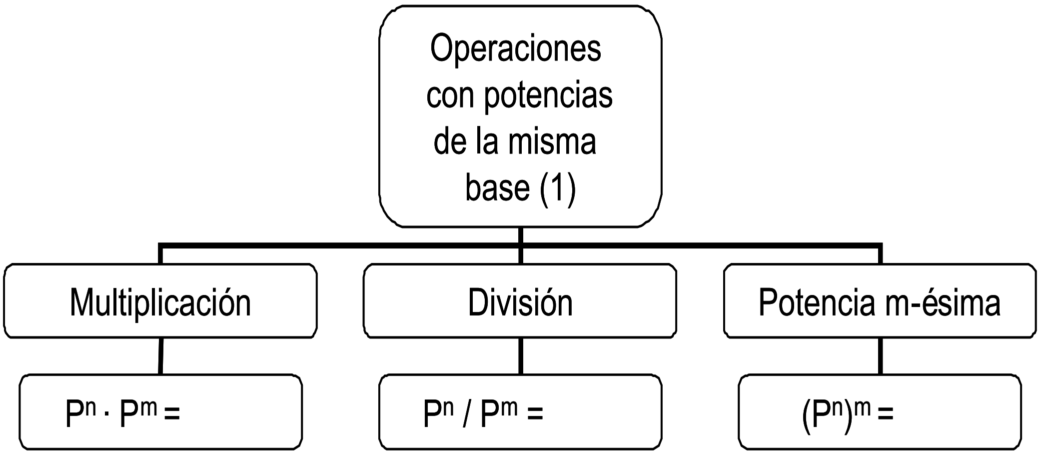

Potenciación
P^n
P
n
Base de la potencia
Exponente
Factor a multiplicar por sí mismo
Indica el número de veces que será multiplicada la base por sí misma
Se multiplicará cuantas veces lo indique el exponente

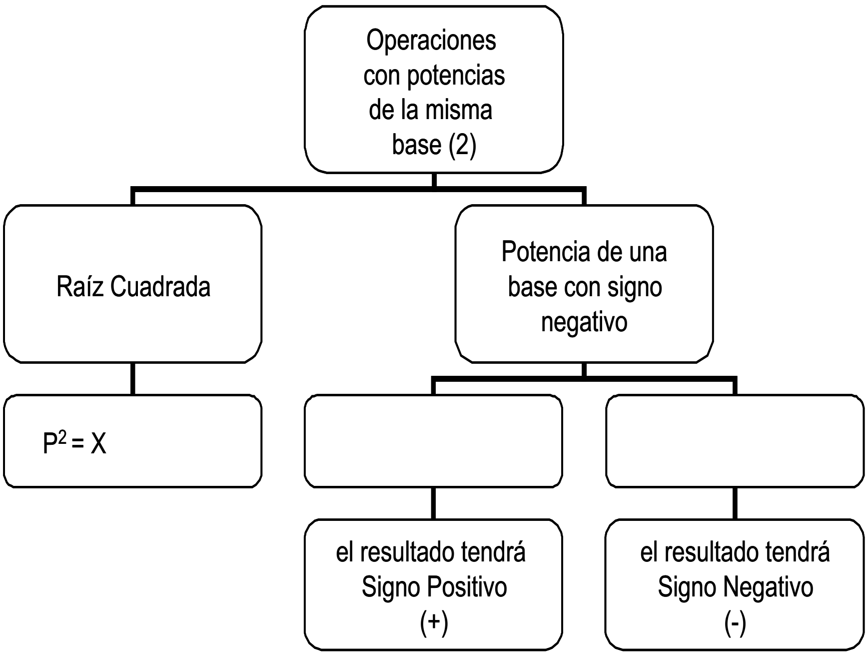

División
Cociente
El número a dividir y/o fraccionar (numerador)
El número que divide (denominador)
Posibilita la secuencia infinita de los resultados del cociente
Factor desconocido de una multiplicación
Se caracteriza por ser ser menor que el divisor

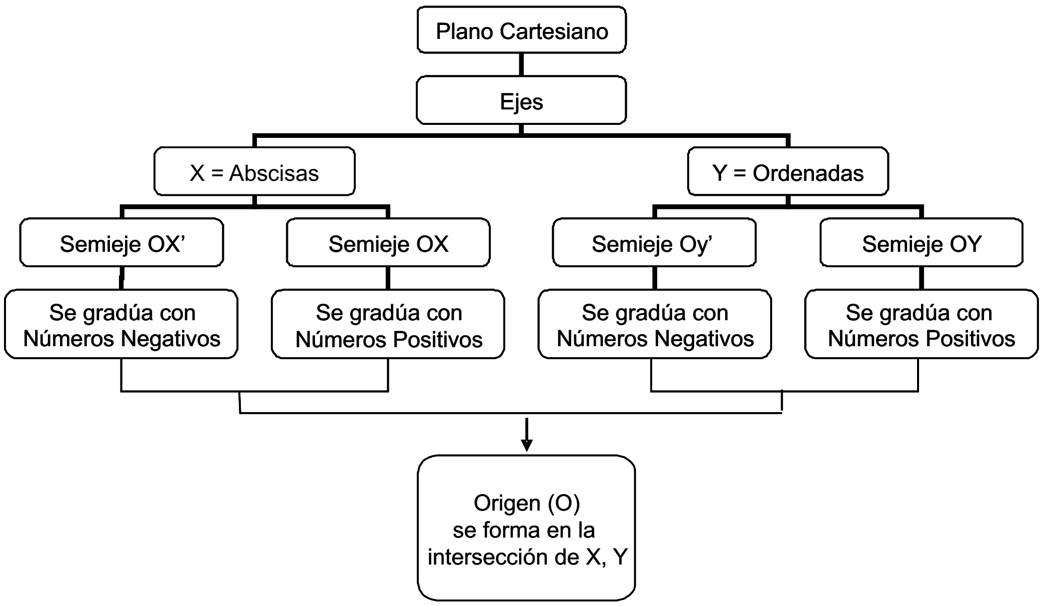

Potenciación
P
n
Exponente
Factor a multiplicar por sí mismo
Se multiplicará cuantas veces lo indique el exponente

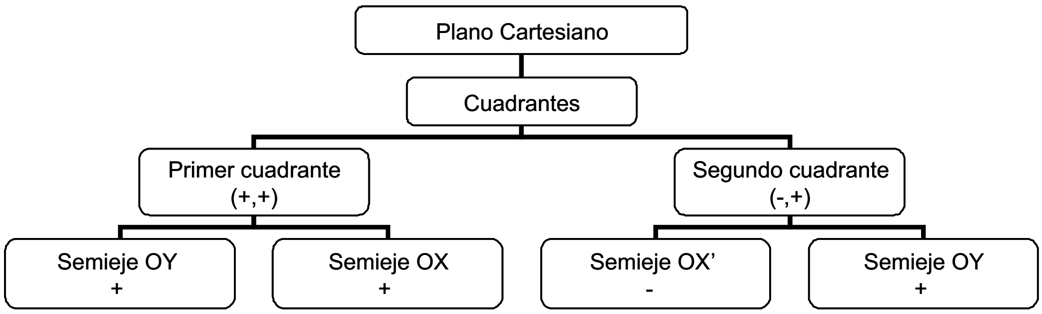

Operaciones con potencias de la misma base (1)
Multiplicación
División
Potencia m-ésima
$P^n \cdot P^m = P^{n+m}$
$P^n / P^m = P^{n-m}$
$(P^n)^m = P^{n \cdot m}$

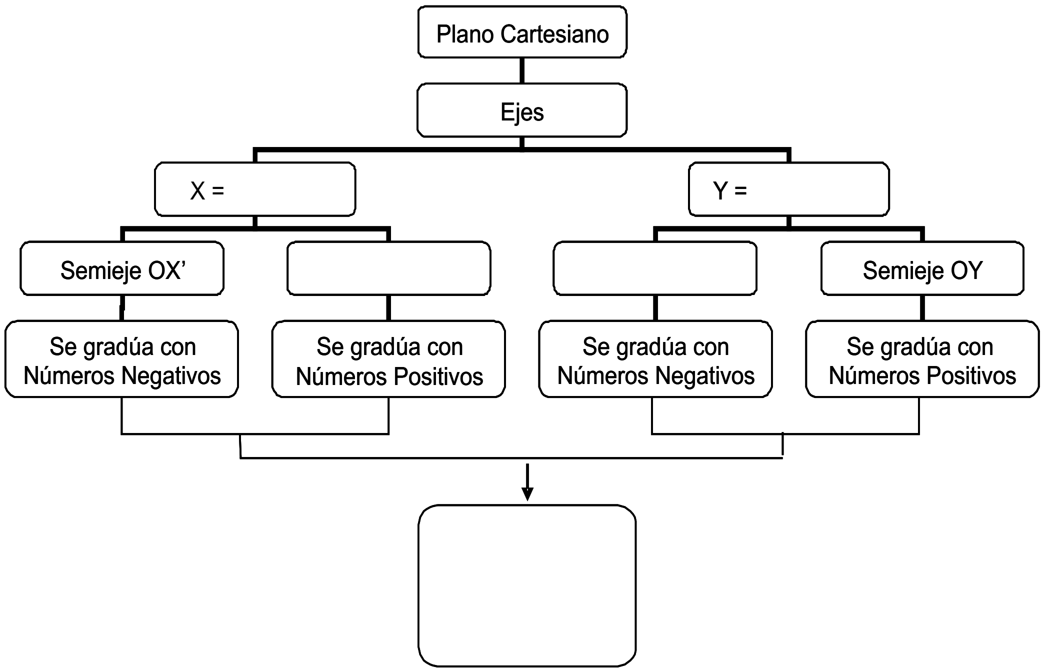

Operaciones con potencias de la misma base (2)
Raíz Cuadrada
Potencia de una base con signo negativo
$P^2 = X$ $X = P^2$
Si el exponente es un número par
Si el exponente es un número impar
el resultado tendrá Signo Positivo (+)
el resultado tendrá Signo Negativo (-)

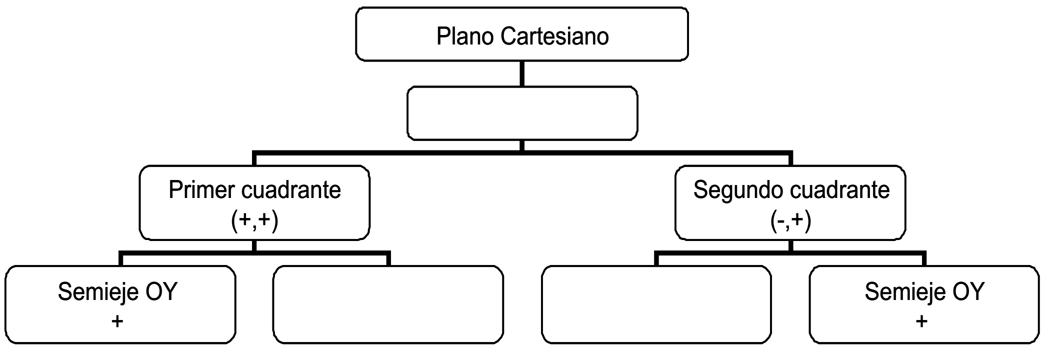

Operaciones con potencias de la misma base (1)
Multiplicación
División
Potencia m-ésima
$P^n \cdot P^m =$
$P^n / P^m =$
$(P^n)^m =$

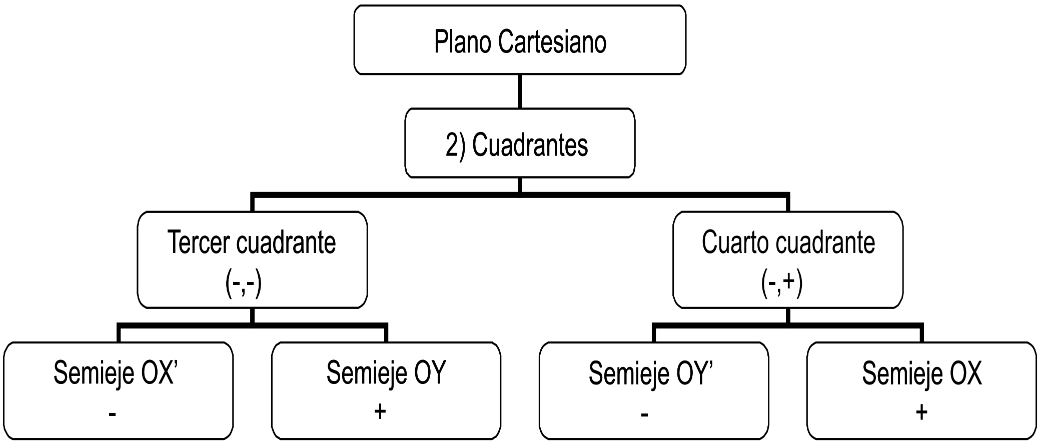

Operaciones con potencias de la misma base (2)
Raíz Cuadrada
Potencia de una base con signo negativo
$P^2 = X$
el resultado tendrá Signo Positivo (+)
el resultado tendrá Signo Negativo (-)

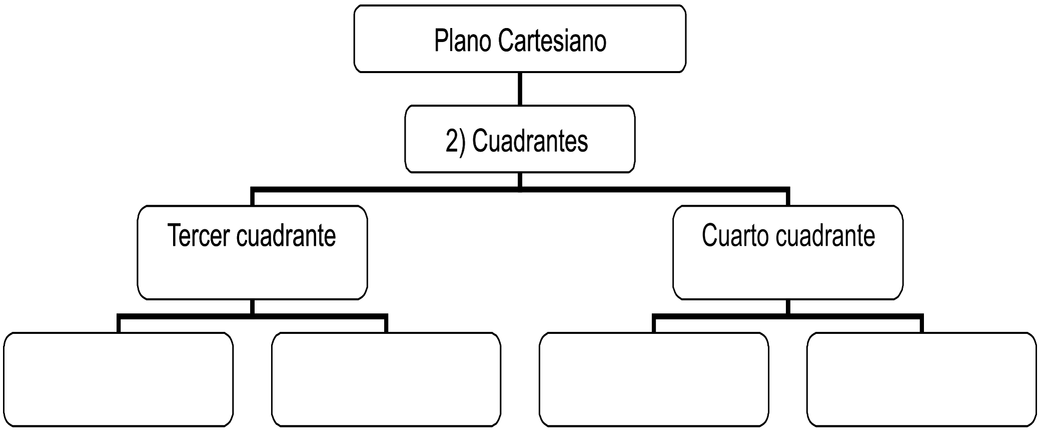

Plano Cartesiano
Ejes
X = Abscisas
Y = Ordenadas
Semieje OX'
Semieje OX
Semieje Oy'
Semieje OY
Se gradúa con Números Negativos
Se gradúa con Números Positivos
Se gradúa con Números Negativos
Se gradúa con Números Positivos
Origen (O) se forma en la intersección de X, Y


Plano Cartesiano
Cuadrantes
Primer cuadrante (+,+)
Segundo cuadrante (-,+)
Semieje OY +
Semieje OX +
Semieje OX' -
Semieje OY +


Plano Cartesiano
Ejes
X =
Y =
Semieje OX'
Semieje OY
Se gradúa con Números Negativos
Se gradúa con Números Positivos
Se gradúa con Números Negativos
Se gradúa con Números Positivos


Plano Cartesiano
Primer cuadrante (+,+)
Segundo cuadrante (-,+)
Semieje OY +
Semieje OY +


Plano Cartesiano
2) Cuadrantes
Tercer cuadrante
(-,-)
Cuarto cuadrante
(-,+)
Semieje OX'
-
Semieje OY
+
Semieje OY'
-
Semieje OX
+


Plano Cartesiano
2) Cuadrantes
Tercer cuadrante
Cuarto cuadrante

2.2. Simbología

La siguiente tabla, muestra los símbolos empleados a lo largo del presente trabajo, con el objetivo de que el estudiante se familiarice con su significado y su uso.

Símbolo	Significado	Símbolo	Significado
$\in$	Es elemento de, o pertenece a	$\Rightarrow$	Si. . . , entonces. . . o Implica que
$\notin$	No es elemento de	∞	Infinito
$\subset$	Subconjunto propio de	$n!$	Factorial de n
$\subseteq$	Subconjunto impropio de	$\aleph$	Aleph
$\neq$	No es igual que	$\int$	Integral
$\leq$	Menor o igual que	α	Alfa
$\geq$	Mayor o igual que	β	Beta
$\emptyset$ ó $\{\}$	Conjunto vacío	μ	Mu
$\equiv$	Idéntico a	π	Pi
$\approx$	Aproximadamente igual	$\sim$	Se distribuye como
$\rightarrow$	Tiende a	σ	Sigma
$\Leftrightarrow$	Si y solo si	Ω	Omega
$\{\ldots\}$	Conjunto	$\pm$	Mas menos
n(A)	Cardinalidad del conjunto A	$\sum$	Suma
$\mid$	Tal que	$\surd$	Raíz Cuadrada

2.3. Conjuntos

Al nombrar las **cosas**, el hombre construye y da forma a lo que él mismo ha denominado *realidad* (su realidad). En el proceso de **significación**, el ser humano **caracteriza** las cosas (ahora **objetos**), re-conoce lo dado en el marco de la experiencia posible, es decir, hace la traducción lingüística de "lo ahí", de lo que se presenta a la conciencia, lo **informa**, clasifica y manipula; a través del **signo lingüístico**, y su natural arbitrariedad, hace posible el conocimiento cada vez más preciso de su entorno, de sí.

El conocimiento que se tiene de los objetos, parte del **análisis** y clasificación de éstos y sus características más evidentes. Analizar, conocer, reconocer y clasificar, son los principios básicos para la conformación de *conjuntos*. Esta actividad, habilita la posibilidad de abordar los fenómenos que se constituyen como problemas intelectuales, por separado (interior y exteriormente), y analizarlos con más cuidado y perspicacia; a la postre, ello permite generar un conocimiento científico más profundo y ceñido a la esencia de los fenómenos objeto de estudio.

A continuación, se repasarán algunos elementos de *Teoría de Conjuntos* y las características de ciertos *conjuntos especiales*, con la finalidad de que los recuerdes y aproveches en estadística e investigación. Considera la siguiente definición:

Conjunto: En general, es una colección de elementos que tienen una característica en común. Pero desde un punto de vista formal, se dice que un conjunto es una *agrupación de elementos perfectamente definidos*, es decir, un grupo de objetos con, por lo menos, una característica en común que permite saber, con exactitud, lo que pertenece a él y lo que no.

Generalmente, los conjuntos se **denotan** mediante las primeras letras del alfabeto, en mayúsculas, seguidas por el signo de igualdad (=) y un par de llaves ({}) que contienen al conjunto y/o su descripción; de igual manera, los elementos del conjunto, pueden ser representado por las primeras letras minúsculas del alfabeto. *Los conjuntos se pueden expresar por comprensión o por extensión.* En el primer caso, los elementos del conjunto se encuentran definidos por una **proposición abierta**, mientras que en el segundo (por extensión), se enumeran todos y cada uno de los elementos (separados por comas) que componen dicha agrupación; ejemplo:

- Por comprensión:
    - A = {las diez primeras letras del alfabeto}
      ≡ A = {x|x es de las primeras diez letras del alfabeto} que se lee: el conjunto A, está formado por todos los elementos "x", tal que "x" es de las primeras diez letras del alfabeto.
- Por extensión (numeración):
    - A = {a, b, c, d, e, f, g, h, i, j}

Como se sabe, existen diversos tipos de conjuntos en el universo. Los hay finitos e infinitos. Un tipo especial de conjunto es *el conjunto vacío*, éste no cuenta con elemento alguno y se denota con los símbolos $\emptyset$ ó {}. Otro conjunto especial es el *conjunto universal*: "En el análisis de una situación particular, hay un conjunto o colección fija de elementos que se denomina *conjunto universal* y se denota por la letra griega Ω (omega). Dicho conjunto Ω, consta de *todos* los elementos a los que se pueda referir esa situación. Es algo así como la fuente de todos los elementos que forman parte de los conjuntos sobre los que vamos a trabajar.

"Hay dos circunstancias que se deben tener en cuenta cuando se trata de elegir el conjunto universal:

"i) el conjunto universal no es único; depende del problema que se esté considerando y puede cambiar según la situación particular de que se trate;

"ii) aún para un mismo problema el conjunto universal no está definido en forma única; podemos elegirlo a nuestra conveniencia con relativa libertad."[1]Ejemplos:

- Conjuntos vacíos:
  - A={x|x es actor de telenovelas mexicanas menor de 30 años que escribe libros de matemáticas} $\equiv$ A $= \emptyset$
  - B={x|x es un ingeniero que estudia la secundaria} $\equiv$ B $= \{\}$
  - C={x|x es una estudiante de comunicación sin información} $\equiv$ C $= \emptyset$
  - D={x|x es un hombre sano que no conoce el agua} $\equiv$ D$= \{\}$
  - E={x|x es un ser humano no engendrado} $\equiv$ C $= \emptyset$
- Conjuntos universales.
  - Ω = {x|x es hombre mexicano que nació en 1980}
  - Ω ={x|x es un teatro ubicado en la ciudad de Guanajuato}
  - Ω = {x|x es un libro publicado por la UNAM}
  - Ω = {x|x es habitante del Estado de México}
  - Ω = {x|x es estudiante de la UNAM}

[1]KLEIMAN, Ariel; Et. al. **Conjuntos. Aplicaciones Matemáticas a la Administración**; México; Editorial Limusa, Biblioteca Didáctica de Matemáticas; 1975; pp. 25

En el campo de la investigación social, el abordaje riguroso de un fenómeno requiere cierta **delimitación**. Para ello, la formación de conjuntos es de suma utilidad, ya que un tema-objeto de estudio en particular, tiene diferentes implicaciones y efectos en función de, por ejemplo, algunas **variables sociodemográficas** con que se asocia.

En un primer acercamiento, podemos pensar en el siguiente caso: Se desea conocer la **opinión pública** de los mexicanos, con respecto al Presidente de la República. Si se parte de la anterior propuesta de investigación, sin aplicar criterios formales para realizarla, el investigador se enfrentará a las siguientes dificultades (entre otras):

1. Opinión pública, es decir ¿lo que se le ocurra al sujeto con respecto de la figura presidencial y cualquier tema?
2. ¿A qué mexicanos se refiere?
   a) ¿niños y adultos por igual?
   b) ¿Adultos con y sin derecho al voto?
   c) ¿Empresarios y desempleados indistintamente?
   d) ¿Yucatecos, regiomontanos, capitalinos?
   e) ¿Adultos que radican fuera de México?
3. ¿Al presidente de cuál República? ¿De qué periodo?

Como se podrá observar, la conformación de un conjunto, puede evitar las anteriores ambigüedades. Desde luego, la selección de las características que determinarán la pertenencia o exclusión de los elementos, obedecerá a los criterios que dictan tanto la delimitación como el **planteamiento del problema** que dan origen a la investigación.

El anterior razonamiento, apunta en dirección al concepto de *pertenencia a un conjunto*. Se puede afirmar que un objeto pertenece a cierto conjunto, cuando posee la(s) característica(s) y/o atributo(s) que definen al conjunto; esta relación se denota simbólicamente con la ayuda del siguiente símbolo: $\in$, que significa *es elemento de* o *pertenece a*; mientras que la exclusión (no pertenencia) se simboliza con $\notin$, que significa *no pertenece a*.

Recuérdese que las primeras letras del abecedario, en mayúscula, se pueden utilizar para representar conjuntos y que las minúsculas pueden representar a sus elementos. Así, se puede expresar formalmente que "a" *es elemento del conjunto* "A", de la siguiente manera:

$$\mathrm{a} \in \mathrm{A}$$

Por el contrario, si un elemento "b" *no pertenece al conjunto* "A", esto se representa por la expresión:

$$\mathrm{b} \notin \mathrm{A}$$

Ejemplos: Considere los siguientes conjuntos:

$\mathrm{A} = \{2, 4, 6, 8, 10\}$

$\mathrm{B} = \{1, 3, 5, 7, 9\}$

$\mathrm{C} = \{p, q, r, s\}$

$\mathrm{D} = \{w, x, y, z\}$

Se tiene, entre otras afirmaciones, que:

- $2 \in A$
- $5 \in B$
- $s \in C$
- $x \in D$
- $2 \notin B$
- $7 \notin A$
- $q \notin D$
- $z \notin C$
- $z \notin A$

Ejercicios 1

1. Propón 3 ejemplos de conjuntos, denótalos por comprensión.

   a) ____

   b) ____

   c) ____

2. Retoma tus anteriores 3 ejemplos y exprésalos por extensión.

   a) ____

   b) ____

   c) ____

3. Asigna el símbolo que corresponda (Ω, $\{\}$ ó $\emptyset$) a las siguientes proposiciones, de acuerdo con la naturaleza del conjunto que expresan.

   a) A={x|x es un signo del zodiaco}................................ ≡ A = ______

   b) B={x|x es un ave de corral}.................................... ≡ B = ______

   c) C={x|x es un número positivo menor que cero}........ ≡ C = ______

   d) D={x|x es un ser mitológico}................................... ≡ D = ______

   e) E={x|x es un soldado mexicano de 5 años}............. ≡ E = ______

   f) F={x|x es un extranjero en su país}.......................... ≡ F = ______

4. Considera los siguientes conjuntos y complementa las afirmaciones en que se expresan las relaciones de pertenencia; sean:

   Ω = {x|x es una variable sociodemográfica}
   A = {x|x es una vocal}

   a) "b" Representa la Edad ⇒... b ___ Ω

   b) "c" Representa el Estado civil ⇒....................................... c ___ Ω

   c) "d" Representa el Color de cabello ⇒............................... d ____ Ω

   d) "f" Representa los Libros que conoce ⇒........................... f ____ Ω

   e) "g" Representa el Color de piel ⇒..................................... g ___ Ω

   f) e ___ A

   g) f ___ A

   h) a ___ A

   i) i ___ A

   j) x ___ A

5. Propón 3 ejemplos de Conjunto Vacío y 3 de Conjunto Universal.

    a) ____

    b) ____

    c) ____

    d) ____

    e) ____

    f) ____

- Ahora verifica que tus respuestas sean las correctas; para dicho efecto, comparalas con las que se te porporcionan en la sección **3.15.2** página 197

La *teoría de conjuntos* es la base para comprender los elementos matemáticos de la *teoría de la probabilidad.* Esta última, es una valiosa herramienta en la realización de investigaciones cuya metodología es **cuantitativa**, puesto que su aplicación al diseño de muestras para el estudio de fenómenos naturales y sociales, permite, como se ha señalando, conocer con mayor precisión el fenómeno que se estudia. Por ello, es importante que aproveches el tiempo que inviertes en el estudio de este capítulo. Considera los siguientes elementos de *teoría de conjuntos*:

Cardinalidad de un conjunto

El número de elementos que contiene un conjunto, se denomina *cardinalidad*, y se denota con el símbolo: n( ). Por ejemplo, observa los siguientes conjuntos:

- $A = \{-2, -4, -6, 7, 8, 9\}$ Se tiene entonces que: n(A) = 6

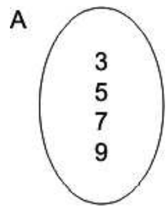
- $B = \{0, 1, 2, 3, 4, 5, 6, 7\}$ Entonces: n(B) = 8

- $C = \{\}$. (Conjunto vacío) Implica que: n(C)= 0

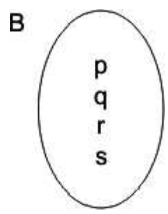
- $D = \{1, 2, 3, 4, 2\}$ En este caso, se observa que el conjunto incluye un elemento repetido (2), el cual sólo debe ser considerado una vez, por lo que: n(D)=4

Diagramas de Venn

Johon Venn (1834 - 1923), fue un matemático inglés que desarrolló una técnica para trabajar con los conjuntos; ésta es conocida como *diagrama de Venn* en honor a su creador. Un diagrama de Venn, es la *representación gráfica* de un conjunto. Por ejemplo, el conjunto A = {3, 5, 7, 9} se representa como diagrama de Venn de la siguiente manera:



De igual manera, el conjunto B = {p, q, r, s} tiene la siguiente representación en diagrama de Venn:

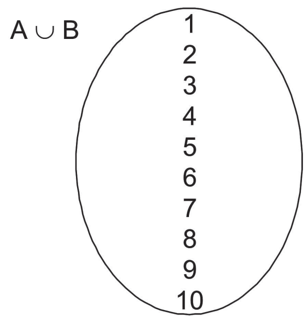


Operaciones con conjuntos

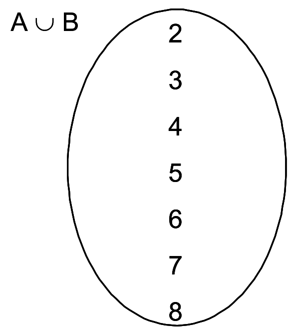
Sean “A” y “B” dos conjuntos. *La unión* de los conjuntos “A” y “B” es el conjunto de elementos que pertenecen al conjunto “A” o al conjunto “B” o a ambos, y se representa por el símbolo: ∪

- Por comprensión: A∪B={x|x ∈ A o x ∈ B}

Ejemplos: Sea el conjunto A = {2, 4, 6, 8, 9, 10} y B = {1, 3, 5, 7, 9, 10}. Se tiene entonces que A∪B = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}. En diagrama de Venn, se tiene:



Ahora bien, considera los siguientes conjuntos: A={2, 3, 4, 5} y B={5, 6, 7, 8}. Entonces, A ∪ B = {2, 3, 4, 5, 6, 7, 8}, operación que representada mediante el diagrama de Venn da como resultado la siguiente figura:

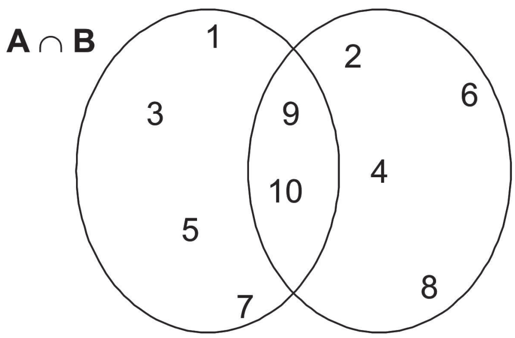


Nótese que ambos conjuntos incluyen como a uno de sus elementos al número 5, no obstante, al realizar la unión, dicho elemento sólo se considera una vez en el resultado de la operación.

Por otra parte, *la intersección de conjuntos* se define como el conjunto de elementos que pertenecen al conjunto "A" y al conjunto "B", y se representa por el símbolo $\cap$

- Por comprensión: A$\cap$B=$\{$x$|$x $\in$ A y x $\in$ B$\}$

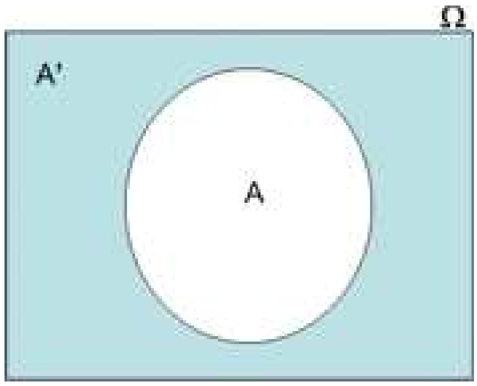
Considerando los conjuntos A = $\{2, 4, 6, 8, 9, 10\}$ y B = $\{1, 3, 5, 7, 9, 10\}$, la intersección se representa por el siguiente diagrama de Venn:



El *complemento de un conjunto* "A", es el conjunto de elementos que no pertenecen a éste, es decir, los elementos no contenidos en el conjunto "A", pero que son parte del conjunto referencial, el conjunto universal (Ω); el complemento de un conjunto se denota con el símbolo A', que por comprensión se define de la siguiente manera:

- Por comprensión: A'=$\{$x$\in$ $\Omega$$|$x $\notin$ A$\}$

Su representación en diagrama de Venn se muestra en la siguiente figura:



Subconjuntos de un conjunto

Otra de las posibles relaciones entre conjuntos, se da a partir de la similitud entre las características que definen a sus elementos y el tamaño de cada conjunto. En general, si se tienen dos conjuntos cuales quiera "A" y "B" tales que todos los elementos del conjunto "A" pertenezcan al conjunto "B", se dice que "A" *es un subconjunto* de "B", y se simboliza por:

$$A \subseteq B$$

La definición anterior considera el caso especial de aquellos conjuntos que *pertenecen a sí mismos*: $A \subseteq A$, es decir, el conjunto "A" es un *subconjunto impropio* del conjunto "A". Por otra parte, si todos los elementos de "A" pertenecen a "B" pero no todos los elementos de "B" pertenecen a "A", se dice que el conjunto "A" es un *subconjunto propio* del conjunto "B" y la pertenencia se expresa de la siguiente forma:

$$A \subset B$$

Ejemplos: Observe los siguientes conjuntos:

$A = \{1, 2\}$, $B = \{1, 2, 3, 4, 5\}$, $C = \{1, 2\}$

Se tiene que:

- A $\subset$ B, es decir: El conjunto "A" *es subconjunto propio* del conjunto "B"
- A $\subseteq$ A, es decir: El conjunto "A" *es subconjunto impropio* del conjunto "A"
- Se puede observar que A$\subseteq$ C, y asimismo, C $\subseteq$ A. Por tanto ambos conjuntos son iguales: A = C

Ejercicios 2

1. ¿Qué es la cardinalidad de un conjunto?

   ----

2. Estima y desarrolla (sin importar el tipo de elementos que propongas) la cardinalidad de los siguientes conjuntos.

   a) D = {El Universal, La Jornada, Reforma, Uno más uno, La prensa, Esto, Ovaciones}, por tanto...................................... n(D) = ______

   b) E = {1, 3, 5, 7, 9, 11, 13, 15, 17, 19}, por tanto......... n(E) = ______

   c) F = {x|x es un mes del año}, por tanto.................. n(F) = ______

   d) n(G) = 9 ⇒ D = {   }

   e) n(H) = 6 ⇒ D = {   }

   f) n(I) = 10 ⇒ D = {   }

3. Mediante el diagrama de Venn, realiza las operaciones que se indican con los siguientes conjuntos:

A = {1, 3, 5, 7, 9, 11, 13, 15, 17, 19}
B = {1, 2, 3, 5, 7, 11, 13, 17, 19}
C = {2, 4, 6, 8, 10, 12, 14, 16, 18}

a) B ∩ A

b) A ∪ B

c) A'

4. Con los conjuntos del anterior ejercicio, realiza las siguientes operaciones y denota sus resultados por comprensión.

    a) $B' = \{$ $\}$

    b) $B \cap C = \{$ $\}$

    c) $C \cup A = \{$ $\}$

5. Complementa las siguientes definiciones:

    a) En general, si se tienen dos conjuntos cuales quiera "A" y "B" tales que todos los ______________ del conjunto "A" pertenezcan al conjunto "B", se dice que "A" __________________________ de "B", y se simboliza por $A\_\_B$. La definición anterior considera el caso especial de aquellos conjuntos que *pertenecen a* ______________________: $A \subseteq$ ___, es decir, el conjunto "A" es un subconjunto ______________ del conjunto "A".

b) Si todos los elementos de un conjunto "A" ____________________ a "B" pero ____________________ los elementos de "B" pertenecen a "A", se dice que el conjunto "A" es ____ del conjunto "B" y la pertenencia se expresa de la siguiente forma: A_B

sí mismos	un subconjunto propio	$\subset$	B	A
elementos	pertenecen	impropio	no todos	subconjunto

- Ahora compara tus **respuestas** con las que se te presentan el la página 199

Número de subconjuntos de un conjunto

Como ya se ha expresado, una vez definido el Conjunto Universal Ω correspondiente a un problema específico, todos los conjuntos construidos en él, son necesariamente subconjuntos de Ω. Asimismo, sobre cualquier conjunto puede calcularse el *número posible de subconjuntos que lo componen.* El número de subconjuntos de un conjunto "A", se denota con la expresión $n_s(A)$; este número se calcula mediante la siguiente fórmula:

$$\boxed{n_s(A) = 2^{n(A)}}$$

Es decir: El número de subconjuntos del conjunto "A" es igual a dos elevado al número de elementos que contiene el conjunto "A"; ejemplos:
Analiza los siguientes conjuntos y determina su número de subconjuntos: A={1, 2} y B={p, q, r}

- Como n(A)= 2, entonces $n^{s(A)} = 2^2 = 4$ subconjuntos posibles.
- Como n(B)= 3, entonces $n^{s(B)} = 2^3 = 8$ subconjuntos posibles.

Conjunto potencia

Al conjunto de todos los posibles subconjuntos de "A", se le denomina conjunto potencia. Éste, se representa con la siguiente notación:

$$\boxed{2^A = \{X | X \subseteq A\}}$$

Ejemplo: Considera los conjuntos del ejemplo anterior; su conjunto potencia se expresa por:

- A={1, 2}. Se tiene que $2^A = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$
- B={p, q, r}. Se tiene que $2^B = \{\emptyset, \{p\}, \{q\}, \{r\}, \{p, q\}, \{p, r\}, \{q, r\}, \{p, q, r\} \}$

Como podrás notar, en ambas potencias, el conjunto vacío aparece como subconjunto; de igual manera, el mismo conjunto, es un subconjunto de sí, cuestión que ocurre para todos los casos en que se quiere expresar la potencia de un conjunto.

Finalmente, cabe resaltar que la diferencia entre el número de subconjuntos de un conjunto y un conjunto potencia, es que en este último, el resultado se debe expresar por extensión, mientras que en el primero, el resultado se expresa numéricamente.

Conjuntos especiales de números

En matemáticas, se manejan conjuntos especiales de números, clasificados de acuerdo a sus características propias. Entre dichos conjuntos y los que se pueden formar a partir de las características sociodemográficas de una población, existe una estrecha relación. La aplicación de los primeros, facilita el manejo de los datos recogidos. Dicha clasificación se presenta a continuación.

Cabe agregar que por las características de sus elementos y dimensiones propias, los conjuntos pueden guardar (y lo hacen) una serie de relaciones que es posible expresar formalmente. El caso de los diferentes conjuntos de números existentes, no es la excepción; conocer sus características y propiedades es de gran utilidad al momento de utilizarlos y/o aplicar sus propiedades en el terreno de la investigación factual.

En el contexto de la matemática, los conjuntos numéricos se denotan en función de las características de los elementos que los componen. El primer caso, corresponde a los *números naturales*. Desde la antigüedad, su utilidad se encuentra profundamente vinculada con las nociones de unidad y pertenencia.

Con los números naturales, el hombre pudo representar sus herramientas, ganado, las presas de caza, prisioneros tomados en la batalla y en general sus posesiones; mediante la creación de los números naturales, el ser humano configuró un medio formal para reconocer y cotejar los objetos que le interesaban y de los que debía apropiarse (una vez poseídos) simbólicamente, dicho de otra manera: formaba conjuntos y realizaba operaciones con los elementos que poseían cada uno de ellos para no recurrir al “tanteo” en la conservación y administración de sus pertenencias.

El conjunto de los números naturales se **denota** por la literal **N**, y es aquel cuyos elementos son los números *enteros positivos* que van desde el uno (1) hasta el infinito (∞). Este conjunto está asociado al concepto de cardinalidad o conteo de objetos. Formalmente se expresa por:

$$\boxed{\mathbf{N}=\{1,2,3,4,5\ldots\infty\} \equiv \mathbf{N}=\{x|x \text{ es un número entero positivo}\}}$$

La segunda expresión se lee de la siguiente manera: “El conjunto **N** está formado por los elementos ‘x’ *Tal que* ‘x’ es un número entero positivo”; así que cualquier número que posea dichas características, será un elemento del conjunto **N**, por ejemplo:

- Considérese el número 3 con respecto del conjunto **N**; éste, es un número positivo y entero, por tanto se puede decir que: $3 \in \mathbf{N}$. Lo que significa que el número 3 *es elemento de* **N**

Las **variables** y los resultados obtenidos en una investigación, pueden ser representadas numéricamente. La elección del conjunto de números mediante los cuales serán representadas dichas variables, dependerá de sus características y las del conjunto de números. El conjunto de los **N**, sirve para representar variables y resultados enteros no negativos, por ejemplo, número de hijos en una familia, grupos políticos a los que está afiliada una persona, entre otras, ya que es claro que nadie tiene 2.3 ó -2 hijos, por citar un caso.

El primer conjunto especial de números (**N**), es de gran utilidad hasta cierto punto, pero al rebasar éste, por el motivo que sea, es necesario recurrir a un conjunto distinto, puesto que no siempre se puede trabajar con el conjunto de los **N**. Hay ocasiones en que al realizar una medición, balance o comparación, los resultados quizá sean enteros pero no positivos. En tal caso se debe recurrir a un conjunto que permita representar el déficit de unidades, ejemplo:

- Considérese el número -2 (menos dos); si aplicamos la definición del conjunto **N**, desde un punto de vista formal, se tiene que:

  $$-2 \notin \mathbf{N}$$

  Puesto que es entero, pero no es positivo.

Una vez revisado el conjunto de los números naturales, se define el conjunto de los *Números Enteros.* Este conjunto abarca a todos los números enteros, tanto los negativos como los positivos, y se denota con la literal **Z**. Por comprensión:

$$\boxed{\mathbf{Z}= \{-\infty \ldots -3,-2,-1,0,1,2,3\ldots+\infty\} \equiv \mathbf{Z}= \{x|x \text{ es número entero}\}}$$

El conjunto de los naturales un subconjunto de los enteros, es decir:

$$\mathbf{N}\subset\mathbf{Z}$$

Si definimos a $\mathbf{Z}^+$ como el conjunto de los números enteros positivos, se tiene que: $\mathbf{Z}^+ = \{0,1,2,3\ldots+\infty\}$, por lo que se puede afirmar que $\mathbf{N}\equiv \mathbf{Z}^+$ y/o $\mathbf{N}\subseteq\mathbf{Z}^+$; Ejemplos:

- $3 \in \mathbf{Z}$
- $12 \in \mathbf{Z}$
- $2{,}5 \notin \mathbf{Z}$
- $2{,}9 \notin \mathbf{Z}$

Por otra parte, **Q** es el conjunto de los *Números Racionales*, agrupación formada por aquellos números que *pueden representarse como una fracción*; ejemplos:

- El número cuatro (4) puede expresarse con las fracciones: $\frac{8}{2}$,$\frac{16}{4}$, etc. En consecuencia, se puede afirmar que $4 \in Q$
- El -2, es un número que puede expresarse con las fracciones: $\frac{4}{-2}$,$\frac{-4}{2}$, entre otras. En consecuencia $-2 \in Q$

- $\sqrt{2} = 1{,}415\ldots$ tiene un desarrollo decimal infinito y no periódico (es decir tiene un número infinito de decimales que no siguen una secuencia). $\sqrt{2} \notin \mathbf{Q}$ ya que no existe una fracción que pueda expresar dicho número de forma exacta, como pudo hacerse en los dos casos anteriores.

- $\frac{1}{3} = 0{,}333333\ldots$ Pertenece a los racionales, ya que aunque tenga un desarrollo decimal infinito, éste es periódico, es decir, sus decimales siguen una secuencia.

A partir de la definición anterior puede observarse que:

$$\mathbf{N}\subset\mathbf{Z}\subset\mathbf{Q}$$

Por otra parte, los números que no pueden ser representados de forma exacta a través de una fracción, como es el caso de $\sqrt{2}$ y π, pertenecen al conjunto de los *números irracionales*, el cual se denota con la literal **I**. Los números irracionales tienen un **desarrollo decimal** infinito no periódico, el cual no es posible expresar a través de una fracción.

Todos los conjuntos numéricos anteriormente descritos, se encuentran contenidos en un conjunto mayor: *El conjunto de los números reales*, que se representa con el símbolo **R**, es decir:

$$\mathbf{N}\subset\mathbf{Z}\subset\mathbf{Q} \subset \mathbf{R} \text{ asimismo: } \mathbf{I}\subset\mathbf{R}$$

El concepto de cardinalidad puede extenderse a los conjuntos numéricos **N**, **Z**, **Q**, **I** y **R**, sin embargo, aunque es claro que la cardinalidad de cada uno de estos conjuntos es infinito, cada uno tiene su propia *medida del infinito*, de acuerdo a los estudios realizados por el matemático alemán George Cantor en 1847. Cantor demostró que existen *diversos tamaños de infinitos*, los cuales son expresados por números denominados *números transfinitos* y que se representan por la primera letra del alfabeto hebreo: Aleph ($\aleph$). El *menor de los infinitos se representa con el símbolo*: $\aleph_0$. El siguiente infinito, mayor al anterior se representa por: $\aleph_1$ y así respectivamente. Para el caso de los conjuntos numéricos se tiene que:

- n(**N**)= $\aleph_0$
- n(**Z**)= $\aleph_0$
- n(**Q**)= $\aleph_0$
- n(**R**)= $\aleph_1$, aceptando la hipótesis de Cantor[2]

Ejercicios 3

1. Observa los siguientes conjuntos y realiza lo que se te pide.

   B = {b, c, d, f, g, h, j, k}
   C = {a, e, i, o, u}
   D = {a, b, c, d, e, f, g, h, i, j, k}

   a) Determina el n_s(B)= ______________________________

   b) Determina el n_s(D)= ______________________________

   c) Denota la potencia del conjunto C por comprensión.

   ____

[2]George Cantor demostró que la cardinalidad de los números reales es mayor que la de los naturales, enteros, racionales e irracionales; y la denominó con el nombre de *"Continuo"* y se representa con la letra **C**. Asimismo, formuló una conjetura conocida como *"Hipótesis del Continuo"*, en la cual, sostiene que no existe un conjunto con una cardinalidad mayor a la de $\aleph_0$ y menor a la de **C**. En este sentido, se asume que la cardinalidad de los números reales, es igual a $\aleph_1$, es decir el continuo.

d) Desarrolla la siguiente expresión: $2^C =\{$______________________________

____

____

____

____

____

____

____

____$\}$

2. Responde las siguientes cuestiones:

a) ¿Qué elementos componen el conjunto de los números naturales? Denótalo por comprensión:

____

b) ¿Cómo se expresa por extensión el conjunto de los números enteros?

____

c) Cuál es el significado de la siguiente expresión? $\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q}$

____

____

____

____

d) ¿Cómo se puede distinguir a un número racional de uno irracional cuando ambos poseen un desarrollo decimal infinito?

____

____

____

3. Anota si el elemento pertenece ($\in$) o no ($\notin$) a los conjuntos de nuúmeros y explica por qué

   a) -14 ____ $\mathbf{N}$ porque ____

   b) $\frac{5}{3}$ ____ $\mathbf{Q}$ porque ____

   c) $\sqrt{7} = 2{,}6457513$ ____ $\mathbf{I}$ porque ________________________________

   d) ____$\in \mathbf{Z}^+$ porque ____

   e) $\frac{10}{1}$____ $\mathbf{Q}$ porque ____

- Compara tus **respuestas** con las de la página 202 y en caso de cualquier error, repite el ejercicio.

Para concluir esta sección, es pertinente resaltar que los ejercicios propuestos, coadyuvan a reforzar el aprendizaje de los contenidos, pero su dominio, sólo será posible con la práctica. En este sentido, la estrategia propuesta es el autoconocimiento; los ejercicios pueden ser un modelo para el diseño de nuevas prácticas, pero cada quien tiene características y necesidades diferentes al estudiar-aprender, así que la mejor estrategia para ello, es ir más allá de lo propuesto tanto en el presente material como en cualquier otro.

En adelante, el contenido del capítulo estará centrado en el ejercicio y desarrollo de habilidades matemáticas específicas. Éstas, son necesarias para facilitar el aprendizaje de la Estadística Aplicada a la Investigación en Comunicación. La "parquedad" en las referencias de aplicación, tiene como objetivo agilizar el proceso de ejercitación y aprehensión de las mismas; dicha utilidad, será advertida paulatinamente en el resto del material.

2.4. Prioridad de los operadores

En estadística, como en cualquier rama de las matemáticas, la exactitud en el resultado de las operaciones, depende tanto de la correcta ejecución de los cálculos, como del adecuado orden en su realización. El segundo aspecto, es fuente frecuente de errores, ya que en determinados casos el orden en que se han de realizar las operaciones no es tan evidente.

A este orden se le llama *prioridad de operadores* o *jerarquía de operadores*. A continuación se indica la prioridad, es decir, el orden en que se deben realizar los cálculos que incluyen diversos tipos de operaciones:

1. Paréntesis
2. Potencias y raíces
3. Producto y cociente
4. Adición y sustracción

Ejemplo, evaluar las siguientes expresiones:

- La expresión $8 + 6/2 + 5$. Primero se realiza el cociente y después las sumas. Así, la expresión se evalúa como $8 + 3 + 5 = 16$ y no como $8 + 6 = 14$, $14/2 = 7$ y finalmente $7 + 5 = 12$, que claramente es un resultado incorrecto.
- Sin embargo, si se introduce paréntesis en la expresión, cambia totalmente el sentido de ésta, de manera que: $(8 + 6)/2 + 5$ si corresponde a realizar primero $8 + 6 = 14$, dividir este resultado entre 2: $14/2 = 7$ y finalmente sumarle 5: $7 + 5 = 12$

A continuación se presentan más ejemplos acerca de la prioridad de operadores:

- $12/4 + 2 \cdot 5$
  $= 3 + 2 \cdot 5$
  $= 3 + 10$
  $= 13$

- $12/(4 + 2) \cdot 5$
  $= 12/6 \cdot 5$
  $= 10$

- $(12/4 + 2) \cdot 5$
  $3 + 2 \cdot 5$
  $5 \cdot 5$
  $= 25$

- $2 + 30/\sqrt{25} - 4(18 - 3^2)$
  $= 2 + 30/5 - 4(18 - 9)$
  $= 2 + 30/5 - 4(9)$
  $= 2 + 6 - 36$
  $= 8 - 36$
  $= -28$

2.5. Propiedades de la suma

Otra forma de expresar una adición es mediante esta expresión: $\sum_{i=1}^{n}$ Con ella, se *indica la suma* ($\sum$) de varios elementos en forma abreviada. El subíndice (i=1), señala que la suma inicia desde un elemento (i), que ordinalmente es el primero (i es igual a 1) del conjunto de valores a sumar; el superíndice (n), indica que el número de elementos a sumar, puede ser

prácticamente cualquiera y frente a una situación concreta, se sustituye por el número de elementos que conforman al conjunto de datos en cuestión.
La suma tiene ciertas propiedades que permiten simplificar sus cálculos. Las más importantes se exponen a continuación.

- *Propiedad I*

  Considera la siguiente proposición: Sea $C \in \mathbf{R}$, es decir un número real, al cual se le denomina *constante*. Este número se puede sumar cuantas veces se quiera o necesite y tal operación se puede expresar como sigue: $\sum_{i=1}^{n} C$, término que desarrollado expresa lo siguiente: $C + C + C + \ldots + C$. Es decir, el número C se suma n veces. Por lo tanto, esta suma equivale en realidad al producto $n \cdot C$, o sea:

  $$\boxed{\sum_{i=1}^{n} C = n \cdot C}$$

  Que se lee: La suma de todas las C, cuando sus elementos van desde uno hasta n, es igual al producto de n por C; ejemplos:

  - Calcular $\sum_{i=1}^{12} 3$. Para este caso, se tiene que $n = 12$ y $C = 3$; por lo que su desarrollo sería el siguiente: $3 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3$
    Si se aplica la anterior propiedad, entonces el resultado de la suma se calcula mediante el producto de $12 \cdot 3 = 36$
  - Calcular $\sum_{i=1}^{9} 7$. Para este caso, se tiene que $n = 9$ y $C = 7$; por lo que su desarrollo sería el siguiente: $7+7+7+7+7+7+7+7+7$
    Si se aplica la anterior propiedad, la suma se calcula por $9 \cdot 7 = 63$
  - Calcular $\sum_{i=1}^{123} 38$. Para este caso, se tiene que $n = 123$ y $C = 38$; desarrollar la expresión sería muy tardado y laborioso, pues como

se habrá notado la suma del 38 se tendría que repetir en 123 ocasiones. En cambio, al aplicar la anterior propiedad, el proceso se simplifica al calcular la suma por $123 \cdot 38 = 4674$

- Calcular $\sum_{i=1}^{13} \ 84 \equiv 13 \cdot 84 = 1092$, por tanto $\sum_{i=1}^{13} \ 84 = 1092$

- *Propiedad II* Sea $C \in \mathbf{R}$. La expresión $\sum_{i=1}^{n} \ (x_i + C)$, simboliza la suma de todas las x de i más C, cuando i va desde 1 hasta n, y puede desarrollarse de la siguiente manera:

$$\sum_{i=1}^{n} \ (x_i + C) = (x_1 + C) + (x_2 + C) + \ldots + (x_n + C)$$

Asimismo, ésta, puede ser denotada como dos sumas separadas:

$$\sum_{i=1}^{n} \ x_i + \sum_{i=1}^{n} \ C$$

Y de acuerdo a la propiedad I, la segunda suma de la expresión anterior, puede reordenarse y se obtiene:

$$\sum_{i=1}^{n} \ x_i + (n \cdot C)$$

Ejemplos:

- Sean $X = \{2, 4, 6, 8\}$ y $C = 5$
  La suma $\sum_{i=1}^{4} \ (x_i + 5)$ se tendría que determinar como sigue:
  $\sum_{i=1}^{4} \ (x_i + C) = (2 + 5) + (4 + 5) + (6 + 5) + (8 + 5)$
  Sin embargo, para simplificar el cálculo, se puede aplicar la *Propiedad II*.

$$\boxed{\sum_{i=1}^{n} \ (x_i + C) = \sum_{i=1}^{n} \ x_i + (n \cdot C)}$$

Así que:
$\sum_{i=1}^{4}\ (x_i+5)=\sum_{i=1}^{4}\ x_i+(4\cdot 5)=$
$\sum_{i=1}^{4}\ x_i+(4\cdot 5)=2+4+6+8+(20)=$
$\sum_{i=1}^{4}\ x_i+(4\cdot 5)=20+20=$
$\sum_{i=1}^{4}\ x_i+(4\cdot 5)=40$

Se debe recordar que n es la cardinalidad del conjunto X y la x_i representa a cada uno de los elementos que lo conforman, de tal manera que al desarrollar la expresión y sustituir a x_i por x_1, este representa el primer elemento del conjunto (2), x_2 al segundo (4) y así hasta x_4, que es el último elemento del conjunto (8).

- Sean $X=\{3,5,7,9,11\}$ y $C=12$
  La suma $\sum_{i=1}^{5}\ (xi+12)$ se determina como sigue:
  $\sum_{i=1}^{5}\ (x_i+12)=\sum_{i=1}^{5}\ x_i+(5\cdot 12)=$
  $\sum_{i=1}^{5}\ x_i+(5\cdot 12)=3+5+7+9+11+(60)=$
  $\sum_{i=1}^{5}\ x_i+(5\cdot 12)=35+60=$
  $\sum_{i=1}^{5}\ x_i+(5\cdot 12)=95$
- Finalmente, como se verá en este ejemplo, con la práctica se pueden omitir algunas especificaciones:
  Sean $X=\{3,6,9\}$ y $C=171$.
  Calcular el resultado de la siguiente operación: $\sum_{i=1}^{3}\ (xi+171)$
  $\sum_{x_i}^{3}\ x_i+(3\cdot 171)=18+(513)$
  $\sum_{x_i}^{3}\ x_i+(3\cdot 171)=531$

- *Propiedad III*

  Sea $C\in\mathbf{R}$. La suma $\sum_{i=1}^{n}\ (C\cdot x_i)$, puede desarrollarse de la siguiente manera:
  $\sum_{i=1}^{n}\ C\cdot x_i=(C\cdot x_1)+(C\cdot x_2)+\ldots+(C\cdot x_n)$, que equivale a:

  $$C\cdot\sum_{i=1}^{n}\ x_i$$

Ejemplos:

- Sea $X = \{8, 4, 6, -1\}$, y $C = 3$. Calcular $\sum_{i=1}^{4} \ (3 \cdot x_i)$. Como

$$\boxed{\sum_{i=1}^{n} \ (C \cdot x_i) = C \cdot \sum_{i=1}^{n} \ x_i}$$

  La suma se calcula así $3 \cdot \sum_{i=1}^{4} \ x_i$, puesto que $C = 3$, y se desarrolla de la siguiente manera: $3 \cdot (8 + 4 + 6 + (-1)) = 51$

- Sean $X = \{9, 12, 15, 18, 21\}$ y $C = 7$. Calcular $\sum_{i=1}^{5} \ (7 \cdot x_i)$
  $\sum_{i=1}^{5} \ (7 \cdot x_i) = 7 \cdot \sum_{i=1}^{5} x_i$
  $7 \cdot \sum_{i=1}^{5} x_i = (7) \cdot (9 + 12 + 15 + 18 + 21)$
  $7 \cdot \sum_{i=1}^{5} x_i = (7) \cdot (75) = 525$

2.6. Redondeo de datos

En estadística se requiere, en muchos casos, el manejo de cantidades enteras, como puede ser el número de personas que cumplen con cierta característica, de mujeres que cursan estudios de licenciatura, y de autos que produce una fábrica, entre otros. No obstante en la gran mayoría de los casos, los cálculos estadísticos arrojan resultados con decimales, por lo que se debe llevar a cabo un proceso de redondeo para poder expresar dichos resultados.

La norma a seguir para el redondeo, consiste en, dependiendo a cuántos decimales se desee redondear, considerar el último de ellos y observar el número que tiene a su derecha. Si éste es menor o igual a 5, el último dígito permanece sin cambio. En caso contrario, a ese último dígito se le suma la unidad, ejemplos:

- Redondear 72,024 a dos decimales: $72{,}024 \approx 72{,}02$. Donde el símbolo $\approx$, indica que se aproxima.

- Redondear 3,01 a un decimal: $3{,}01 \approx 3{,}0$
- Redondear 25,8146 a tres decimales $25{,}8146 \approx 25{,}815$
- Redondear 183,5769 a tres decimales: $183{,}5769 \approx 183{,}577$
- Redondear 12,9153 a dos decimales: $12{,}12{,}9153 \approx 12{,}92$

Se adopta como criterio la numeración del 0 al 9 y a partir de 5 sube o baja la cantidad, esto es: según la proximidad que se tiene con respecto de 0 y de 9; cuando el 5 es el dígito que antecede al número que se redondeará, se agrega la unidad a éste, ya que el cinco se encuentra situado más a menos distancia del 9 que del 0.

2.7. Notación científica o exponencial

Este tipo de notación permite manejar, de forma compacta, números de gran magnitud. En el sistema decimal, ésta se encuentra relacionada con la notación desarrollada.

Se tiene que:

- $10^0 = 1$
- $10^1 = 10$
- $10^2 = 100\ldots$
- $10^n = 1$ seguido de n ceros

Por ejemplo: el número 8000000 se puede expresar con notación científica de la siguientes formas:

- $0{,}008 \times 10^9$
- $0{,}08 \times 10^8$

- $0{,}8 \times 10^7$
- 8×10^6
- 80×10^5
- 800×10^4
- 8000×10^3
- 80000×10^2
- 800000×10^1

Por otra parte, el exponente negativo implica el uso de números decimales:

- $10^{-1} = \frac{1}{10} = 0{,}1$
- $10^{-2} = \frac{1}{10^2} = 0{,}01$
- $10^{-3} = \frac{1}{10^3} = 0{,}001\ldots$
- $10^{-n} = \frac{1}{10^n} =$ (a partir del punto decimal, $n - 1$ ceros seguidos del número 1)

Por ejemplo, el número 0,000003 puede representarse de las siguientes maneras:

- 3×10^{-6}
- $0{,}3 \times 10^{-5}$
- $0{,}03 \times 10^{-4}$
- $0{,}003 \times 10^{-3}$
- $0{,}0003 \times 10^{-2}$

- $0{,}00003 \times 10^{-1}$

Las siguientes propiedades, permiten la realización simplificada de operaciones con números de gran magnitud (ya sean enteros o decimales):

Sean p y q dos números enteros. Se tiene que:

- $(10^p)(10^q){=}10^{(p+q)}$
- $\frac{10^p}{10^q} = 10^{(p-q)}$

Por ejemplo:

- Si se tiene (1000)(10000) Puede expresarse como: $(10^3)(10^5)$, que corresponde a $10^{(3+5)}$, es decir: 10^8, o sea 100000000
- Sea (5000000)(4000). Esto puede expresase como: $(5{\times}10^6)(4{\times}10^3)$, que equivale a $(5)(4) \times 10^{(6+3)}$, es decir 20×10^9. Lo que da como resultado, 20000000000
- (40000000)(0,00002) puede representarse como: $(4{\times}10^7)(2{\times}10^{-5})$, que es igual a $(4)(2) \times 10^{(7-5)}$, que es igual a 8×10^2. Cuyo resultado es 800
- $\frac{120000000}{300000}$. Puede rescribirse, $\frac{(12\times10^7)}{(3\times10^5)}$ que es igual a $\frac{12}{3} \times 10^{(7-5)}$, es decir $\frac{12}{3} \times 10^2$, que da como resultado 400

Ejercicios 4

1. Aplica la Prioridad de los operadores para realizar los siguientes cálculos:

   a) $(1 + 2^3)^2 + 8/\sqrt{16} - 2(5 - 3) =$

   ----

   b) $(1 + 2^3)^2 + 8/\sqrt{16} - 2(5) - 3 =$

   ----

c) $(1+2)^3 + 8/\sqrt{16} - 2(5) - 3 =$

____

d) $(1+2^3) + 8/\sqrt{16} - (2 \cdot 5 - 3) =$

____

2. Mediante los principios expuestos en la sección "Propiedades de la Suma", resuelve lo que se te pide:

   a) Sea $C = 9$.
   Calcular $\sum_{i=1}^{5} \ C =$ ________________________

   b) Sea $X = \{-4, -3, -2, 0, 1, 4\}$ y $C = 7$.
   Hallar: $\sum_{i=1}^{5} \ (xi + C)$ ______________________

   c) $X = \{8, 4, 6, -1\}$, y $C = 3$.
   Hallar: $\sum_{i=1}^{4} \ C \cdot xi$ ________________________

3. Redondeo

   a) Redondear 52,8859013 a tres decimales...... 52,8859013 $\approx$ _________

   b) Redondear 6,3384247 a tres decimales......... 6,3384247 $\approx$ _________

   c) Redondear 0,7003589 a tres decimales......... 0,7003589 $\approx$ _________

   d) Redondear 25,9907088 a tres decimales..... 25,9907088 $\approx$ _________

   e) Redondear 14,0846061 a tres decimales..... 14,0846061 $\approx$ _________

4. Expresa y desarrolla los siguientes números mediante notación científica.

   a) $11 \times 10^{-3} =$

   ____

   b) $7 \times 10^{6} =$

   ____

c) 0,000052 =

----

d) (0.2)(6000) =

----

e) $\frac{250000}{500000} =$

- Localiza las **respuestas** a estos ejercicios en la página 203 y corrige cualquier error que hayas tenido.

Dígitos significativos

Este concepto se ha desarrollado para designar formalmente a la confiabilidad de un valor numérico. Los dígitos significativos de un número, son aquellos que pueden ser empleados en forma confiable. Cualquier instrumento de medición tiene un número limitado de dígitos significativos, por ejemplo: en un reloj digital se pueden expresar con certeza la hora, los minutos y los segundos –o sea, maneja seis dígitos significativos–, las unidades de tiempo menores no pueden ser descritas por este dispositivo.

Otros ejemplos pueden ser: Un termómetro, éste, maneja tres dígitos significativos; una lectura de 36,7° de temperatura es muy razonable, sin embargo, afirmar que una persona lee en ese instrumento una temperatura de 36,70001°, no es verosímil; el velocímetro de un automóvil puede manejar cuatro dígitos significativos, es decir, una lectura de velocidad de 123.5 Km./Hr. es convincente, no así la afirmación de que el velocímetro arroja una velocidad de 123.5000017 Km./Hr. En este sentido, el número de dígitos significativos es una medida de la precisión de una cantidad; ejemplos:

- 87324,45 tiene siete dígitos significativos.

- 87324,450 tiene ocho dígitos significativos.
- 87324,4500 tiene nueve dígitos significativos.
- 0,00001845 tiene cuatro dígitos significativos. Es importante observar en este ejemplo, que los ceros que se encuentran inmediatamente a la derecha del signo decimal, no se consideran dígitos significativos.
- 0,0001845 tiene cuatro dígitos significativos.
- 0,001845 tiene cuatro dígitos significativos.
- 0,01845 tiene cuatro dígitos significativos.
- 0,1845 tiene cuatro dígitos significativos.

Por otro lado, el número 45300 puede tener tres, cuatro o cinco dígitos significativos, dependiendo de los ceros que se conozcan con exactitud. Para evitar la incertidumbre, se usa la notación científica:

- $4{,}53 \times 10^4$ tiene tres dígitos significativos.
- $4{,}530 \times 10^4$ tiene cuatro dígitos significativos.
- $4{,}5300 \times 10^4$ tiene cinco dígitos significativos.

2.8. Medidas de proporción

En las Ciencias Sociales, algunas características del objeto de estudio pueden describirse a través de un número, que, las más de las veces, por sí mismo, es insuficiente para brindar información que fundamente las conclusiones y toma de decisiones. No obstante, cuando este número se compara con otros, se dispone de un sistema de referencia que arroja información más completa sobre el objeto de estudio. Este tipo de descriptores numéricos son conocidos como *medidas de proporción*, y entre las más importantes se encuentran las razones, las tasas, los índices, las proporciones y los porcentajes.

Razones

Formalmente, se dice que una *razón* es el cociente de dos cantidades. Ésta se considera como la principal operación de transformación o normalización[3] estadística. El concepto de *razón*, se emplea cuando se quiere definir una medida o norma para expresar **datos primarios**. *Para calcularla, se divide la cantidad que se quiere normalizar (estudiar), por la cantidad normalizadora (referente).*[4]

$$\boxed{\frac{a}{b}}$$

Donde el numerador de la fracción (a) es la cantidad a normalizar y el denominador de ésta (b), es la *cantidad normalizadora*; ejemplo: Si en la carrera de Comunicación se encuentran inscritos 800 estudiantes de los cuales 600 son mujeres y 200 hombres, se puede estimar cuál es la razón de mujeres tomando como referente al número de hombres.

En este caso $a = 600$ y $b = 200$, entonces la razón de mujeres viene dada por: $\frac{600}{200} = 3$; esta cantidad indica que *hay tres mujeres inscritas por cada hombre*. Por otro lado, si se desea calcular la razón de hombres se tiene que $a = 200$ y $b = 600$ y el cociente $\frac{200}{600} = \frac{1}{3}$ indica que por cada tres mujeres inscritas hay un hombre.

[3]Aunque hay cierta similitud, no debe confundirse el concepto *normalización* asociado a una razón, con el que se maneja en la *distribución normal de probabilidad*, que se revisará en el noveno capítulo sección **9.3. Normalización y cálculo de probabilidad** (pág. 460)

[4]CRUCIANELLI, Sandra. **Curso Matemáticas para Periodistas. Apuntes Electrónicos**; Texas University; 2005

Tasas

Una *tasa* es la medida que cuantifica la *frecuencia* con que ocurren ciertos fenómenos en relación con una determinada variable. Ésta, siempre toma como base otras medidas conocidas como *números referentes*. Las referencias (números referentes) más comunes son: Per cápita, cada 1000, cada 10000, cada 100000 o cada 1000000. En Ciencias Sociales, las tasas se emplean para comparar situaciones, por ejemplo, en distintas ciudades o países, ya que permite usar el mismo número de referencia que es fijo e independiente de la población total, he aquí un ejemplo:

Supóngase que en la ciudad de México se cometen, al día, 25 robos por cada 10000 habitantes y en el la ciudad de Los Ángeles (California) se cometen 32 robos por cada 10000 habitantes. Claramente se puede apreciar que Los Ángeles es más insegura, ya que la cantidad empleada como referente para ambas ciudades es la misma (10000), independiente de la cantidad total de pobladores que tiene cada ciudad. Una tasa se calcula aplicando la *regla de proporción* o *regla de tres simple*:

$$\begin{gathered} a \text{ --- } b \\ c \text{ --- } x \end{gathered}$$

Para x se resuelve de la siguiente manera:

$$\boxed{x = \tfrac{c \cdot b}{a}}$$

En donde c es el referente, b la frecuencia de casos y a el número total de elementos del objeto de estudio. Por ejemplo, si en la ciudad A existen 500000 habitantes y de ellos fuman 85000, calcular la tasa de fumadores por cada 10000 habitantes. Planteando el problema como una regla de tres simple, se tiene que: $a = 500000$, $b = 85000$, $c = 10000$ por lo que la tasa buscada se calcula por:

$$x = \frac{10000{\cdot}85000}{500000} = 1700$$

Es decir, se tiene una tasa de 1700 fumadores por cada 10000 habitantes de la ciudad A.

Proporciones

La proporción es la *razón* que existe entre la frecuencia, o número de casos de una categoría, y el número total de categorías; ésta se calcula por:

$$\boxed{\frac{a}{n}}$$

Donde a es la frecuencia o número de casos de una categoría y n es el número total de categorías. Es importante mencionar que el resultado siempre es un número entre 0 y 1; ejemplo: Supóngase que en cierta Universidad, de 16000 alumnos, 7200 son mujeres y 8800 hombres.

- Calcular la proporción de mujeres de dicha universidad. Para este caso se tiene que a =7200 y n =16000. Luego $\frac{7200}{16000} = 0{,}45$. Es decir, la proporción de mujeres en esa Universidad es de 0,45

- Calcular la proporción de hombres. Para este caso se tiene que a =8800 y n =16000. Luego $\frac{8800}{16000} = 0{,}55$. Es decir, la proporción de hombres en esa Universidad es de 0,55

Porcentajes

Expresa una cantidad como un número de partes por cada 100 unidades. Se obtiene al multiplicar la proporción por cien:

$$\boxed{\frac{a}{n} \cdot 100}$$

Donde a es la frecuencia o número de casos de una categoría y n es el número total de categorías. En los dos ejemplos anteriores, se tiene una proporción de mujeres de 0,45. Al multiplicar por 100, se obtiene: 45, es decir, las mujeres conforman el 45 % de la matrícula total de la Universidad del ejemplo anterior. Asimismo, si la proporción de hombres 0,55 se multiplica por 100, se obtiene 55, o sea: el 55 % de los alumnos de la Universidad del ejemplo anterior son hombres.

2.9. Ecuaciones e identidades

Una ecuación es una expresión abierta que puede tener *una o más variables* (literales) que reciben el nombre de *incógnitas*. Ejemplos de ecuaciones:

- $x + 7 = 16$

- $3k - 1 = 5$

- $-x + y = 0$

En las ecuaciones anteriores, las incógnitas están elevadas a la primera potencia, por ello se dice que son *ecuaciones de primer grado*. Las dos primeras, sólo incluyen una incógnita, y se conocen como ecuaciones de *primer grado con una incógnita*. El tercer ejemplo es una *ecuación de primer grado con dos incógnitas*. En esta sección, se revisarán únicamente las ecuaciones de primer grado con una incógnita.

Ecuaciones de primer grado con una incógnita

Una ecuación de primer grado con una incógnita, es una expresión constituida por *dos miembros* relacionados mediante el signo de *igual*. La resolución de este tipo de ecuaciones, consiste en determinar el valor de la incógnita, tal que la ecuación se convierte en *identidad*. Por ejemplo, sea la ecuación:

$$x + 5 = 12$$

en este caso, es claro que el valor de x que hace cierta la expresión anterior es $x = 7$, por lo que al sustituirlo en la ecuación original se tiene:

$$7 + 5 = 12, \text{ es decir:}$$
$$12 = 12$$

Esta última expresión, que pareciera un tanto absurda, es una *identidad*, misma que se obtiene al resolver una ecuación.
Las ecuaciones de primer grado, se resuelven por transposición de términos para obtener identidades; ejemplo:

1. $2x = 7$
   Transponiendo el 7 al segundo miembro:
   $x = \frac{7}{2} = 3{,}5$

2. $7x - 5 = 23$. En este caso, primero se transpone el término que no multiplica a la incógnita (término independiente), es decir, el -5 se transpone al lado derecho, con lo que cambia su signo de negativo a positivo: $7x = 23 + 5 \Longrightarrow 7x = 28$. Posteriormente, se transpone el término que multiplica a la variable, es decir, el 7 se transpone al lado derecho de manera que ahora divida al segundo miembro:
   $\frac{7x}{7} = \frac{28}{7}$
   $x = \frac{28}{7}$
   $x = 4$

Ejercicios 5

1. Anota en la línea correspondiente, la cantidad de dígitos significativos que posee cada número.

   a) 1,0000230

   ____

   b) 0,00053

   ____

   c) $5{,}030 \times 10^2$

   ____

   d) 6000000

   ____

   e) 5,0300

   ____

2. ¿Qué son las medidas de proporción?

   ____
   ____
   ____
   ____
   ____

3. Según datos del Instituto Nacional de Estadística Geografía e Informática (INEGI), hacia el año 2000 la población total de México era de 97483412 habitantes en total. Ésta, se componía por 47592253 personas son de sexo masculino y 49891159 de sexo femenino.

   a) Calcula la *razón* de mujeres en función de los hombres que habitan el país, hacia el 2000.

   ____

b) Tomando como referente el número 10000, calcular cuál es la *tasa* de hombres que conformaban hacia el año 2000 la población total de mexicanos.

____

c) Calcula cuál es el *porcentaje* de ciudadanos mexicanos que habitaban la República Mexicana en el 2000, oriundos del Estado de México, si su número era de13096686 individuos de ambos sexos.

____

d) ¿A cuántos habitantes haciende la población total de mexicanos que habitan el Estado de México y Distrito Federal si entre ambos si entre ambas constituyen el 22,26525 %?

____

4. ¿Qué es una *proporción* y cómo se calcula?

____

____

____

____

____

5. Resuelve las siguientes ecuaciones.

a) $5x + 3 = 28$

b) $41 - 1x = -20$

c) $12x - 7x = 35$

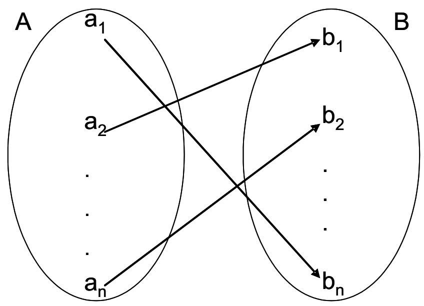
d) $\frac{15x}{22} = 11$

e) $(10 - 5)^2 + x = 30$

Respuestas en la página 205

Funciones y relaciones

Considera un conjunto A=$\{a_1, a_2, a_3 \ldots a_n\}$ y un conjunto B=$\{b_1, b_2, b_3 \ldots b_n\}$. Si se define una regla que permita asociar los elementos de A con los elementos de B, se dice que se ha establecido una relación.



Relación

Si a cada elemento del conjunto A se le asocia uno y sólo un elemento del conjunto B, se dice que la *relación* es una *función*. En términos generales, al primer conjunto se le conoce como *dominio de la función*, mientras que al segundo conjunto se nombra *codominio o contradominio*. Asimismo, la regla que permite asociar entre sí a estos conjuntos, se denomina *regla de correspondencia*.

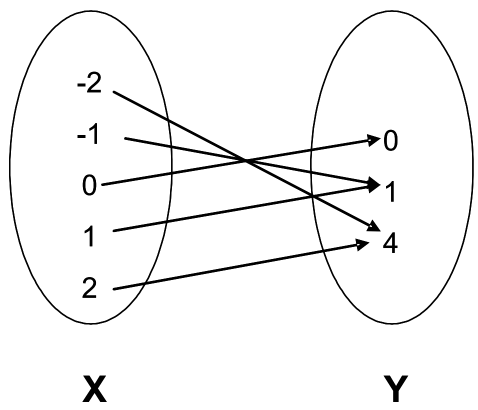
Ejemplo: Sea X={-2, -1, 0, 1, 2}. Y considera la regla que consiste en elevar a cada miembro del conjunto X, al cuadrado. Esto se denota con la siguiente expresión:

$$f(x) = x^2$$

De tal manera que se aplica a cada elemento de X, para obtener:
$f(-2) = -2^2 = 4$
$f(-1) = -1^2 = 1$

$f(0) = 0^2 = 0$
$f(1) = 1^2 = 1$
$f(2) = 2^2 = 4$

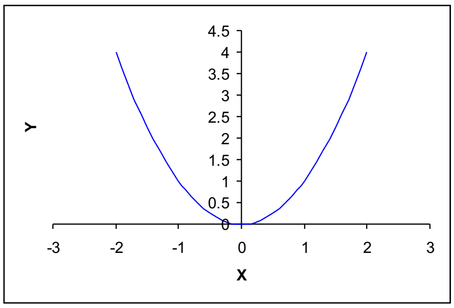
Estos valores conforman el conjunto Y, o codominio de la función, la cual puede representarse con el siguiente diagrama de Venn:



Gráficas de funciones

Una vez establecidos un conjunto X, una regla de correspondencia para éste y el conjunto Y que se genera de la aplicación de la regla al conjunto X, se obtiene una colección de parejas, denominadas *parejas ordenadas*, las cuales al ser representadas como puntos con coordenadas (x,y) sobre el plano cartesiano, permiten la construcción de la gráfica de la función.

En símbolos, dado un conjunto X=$\{x_1, x_2, x_3 \ldots x_n\}$ y una regla de correspondencia $f(x)$, se obtiene un conjunto Y=$\{f(x_1), f(x_2), f(x_3) \ldots f(x_n)\}$. De tal manera que se generan las siguientes parejas ordenadas: $(x_1, f(x_1))$, $(x_2, f(x_2)) \ldots (x_n, f(x_n))$. Éstas pueden representarse como puntos en el plano cartesiano. El ejemplo anterior, definido sobre el conjunto X={-2, -1, 0, 1, 2} y con la regla de correspondencia $f(x) = x^2$, genera el conjunto Y= Sea X={4,1,0}, con lo que pueden formarse las parejas: (-2,4), (-1,1), (0,0), (1,1), (2,4) que generan la siguiente gráfica:



2.10. Logaritmos

Todo número positivo N puede expresarse como una potencia de 10. Es decir, es posible hallar un número p tal que $10^p = N$. A este número p; se le conoce como el *logaritmo en base 10* del número N y se denota por:

$$\boxed{Log_{10}N = p}$$

Ejemplos:

- Sea N = 100. Determine el valor de p tal que $10^p = 100$. Claramente puede observarse que el número $p = 2$, es decir, $10^2 = 100$. Dicho de otra manera, *2 es el logaritmo de 100 en base 10.* Lo que se denota por:

$$log_{10}100 = 2$$

- Sea N = 1000. Determine el valor de p tal que $10^p = 1000$. Claramente puede observarse que el número $p = 3$, es decir, $10^3 = 1000$. Dicho de otro modo, *3 es el logaritmo de 1000 en base 10*. Lo que se denota por:

$$log_{10}1000 = 3$$

- Sea N = 10000. Determine el valor de p tal que $10^p = 10000$. Claramente puede observarse que el número $p = 4$, es decir, $10^4 = 10000$. En otras palabras, *4 es el logaritmo de 10 000 en base 10*. Lo que se denota por:

$$log_{10}10000 = 4$$

- Sea N = 100 000. Determine el valor de p tal que $10^p = 100000$. Claramente puede observarse que el número $p = 5$, es decir, $10^5 = 1000$. Por tanto, *5 es el logaritmo de 100 000 en base 10*. Lo que se denota de la siguiente manera:

$$log_{10}100000 = 5$$

En los ejemplos anteriores, los logaritmos obtenidos corresponden a números enteros, ya que se calcularon sobre números que son potencias de 10 (100, 1000, 10000 y 100000). Sin embargo, cuando se calcula el logaritmo de números que no son potencia de 10, el resultado es un número que no es entero. Este tipo de logaritmos se determinan con calculadora o con una tabla de logaritmos; ejemplo:

- Calcular: $Log_{10}7$ Para estimar en la calculadora, primero introduzca el número y cantidad asignada a N (en este caso 7) y después oprima la tecla *log*, que también puede estar rotulada con la expresión 10^x. Como resultado se obtiene, redondeando a cuatro decimales, 0.8451, esto es: $Log_{10}7 = 0{,}8451$.

- Por otro lado, cuando se calcula el logaritmo en base 10 de un número positivo menor que 1, se tiene como resultado un valor negativo no entero. Por ejemplo: Calcular $log_{10}0{,}5$. Utilizando la calculadora se obtiene $-0{,}30103$ redondeado a 5 decimales.

Operaciones con logaritmos

Como con la suma (Σ), es posible realizar operaciones de forma simplificada entre Logaritmos. Para poder operar con ellos, se debe cumplir la siguiente condición: *Éstos, deben tener igual base (que para el presente caso, es 10).* A continuación se presentan algunas expresiones generales que indican la lógica a en algunas operaciones con logaritmos de base 10.

1. *Logaritmo de un producto*

   $\boxed{Log_{10}(N \cdot M) = Log_{10}M + Log_{10}N}$

   Ejemplo: Calcular $Log_{10}(10000 \cdot 1000)$
   $Log_{10}(10000 \cdot 1000) = Log_{10}10000 + Log_{10}1000$
   $=4 + 3$
   $= 7$

2. *Logaritmo de un cociente*

   $\boxed{Log_{10}\frac{N}{M} = Log_{10}M - Log10N}$

   Ejemplo, calcular: $Log_{10}\frac{10000}{1000}$
   $Log_{10}\frac{10000}{1000} = Log_{10}10000 - Log_{10}1000$
   $= 4 - 3$
   $= 1$

3. *Logaritmo de una potencia*

   $\boxed{Log_{10}M^N = N \cdot Log_{10}M}$

Ejemplo, calcular $Log_{10}10000^{1000}$

$Log_{10}10000^{1000} = 1000 \cdot Log_{10}10000$

$= 1000 \cdot 4$

$= 4000$

Logaritmo natural

Todos los conceptos sobre logaritmos vistos anteriormente, se han considerado sobre el sistema decimal, empero, en la práctica de la estadística es necesario el uso de logaritmos con base diferente de 10. Específicamente, la base empleada en diversos cálculos estadísticos es el número: 2,718281828 (conocido como número de Euler), cuyo símbolo es e. Por ejemplo; calcular $log_{2,718281828}100$ equivale a encontrar el número p tal que $2{,}718281828^p = 100$.

En este caso, cuando se utiliza la base 2,718281828, se dice que se calcula el *logaritmo natural* de una cantidad, lo que se denota con el símbolo ln. En el ejemplo anterior, se está calculando, entonces, $ln100$. Utilizando al calculadora, primeramente se introduce el valor del cual se calculará el logaritmo natural (100), posteriormente se presiona la tecla ln, con lo que se obtiene 4.60517 redondeado a 5 decimales.

En general, si N es un número mayor que cero, calcular su logaritmo en base 2,718281828 corresponde a calcular su logaritmo natural, o sea:

$$log_{2,718281828}N \equiv lnN$$

Ejercicios 6

1. Dados los conjuntos A y B ¿bajo qué circunstancias se dice que la relación de sus elementos es una función?

   ____

   ____

   ____

   ____

2. Dados el conjunto X = $\{5, 10, 15, 20, 25\}$ y la regla $f(x) = -2x/2$; obtener el contradominio (Y) y representar dicha relación mediante el diagrama de Venn.

3. Calcula el logaritmo en base 10 de los siguientes números

   a) $Log_{10}9 =$

   b) $Log_{10}0{,}3$

   c) $Log_{10}100000000$

4. Efectúa las siguientes operaciones con logaritmos

   a) $Log_{10}(10 \cdot 10000) =$

   b) $Log_{10}10000000 - Log_{2{,}718281828}100 =$

   c) $Log_{10}100^{100} =$

5. Estima el ln de las siguientes cantidades

   a) $ln 1000 =$

   b) $ln 5 =$

   c) $ln 0{,}5 =$

6. Considera el conjunto X={-3, -2, -1, 0, 1, 2, 3}. Construye el contradominio del mismo para las siguientes reglas de correspondencia y traza su respectiva gráfica:

   a) $f(x) = x$

   b) $f(x) = -x$

c) $f(x) = 5x$

d) $f(x) = -5x$

e) $f(x) = 5x + 1$

Compara tus **respuestas** con las de la página 207

2.11. Técnicas de conteo

Las *técnicas de conteo* son aplicaciones de una rama de la matemática, denominada *Combinatoria*, que consiste en un conjunto de procedimientos formales para *realizar conteos de forma eficaz y abreviada*; por ello, a la combinatoria se le considera como "El arte de contar sin contar". Esta variante de la matemática, se encuentra directamente relacionada con otra denominada como *Probabilidad* y, en su conjunto (Combinatoria y Probabilidad), son una importante herramienta para el estudio y aplicación de la *Estadística*.

2.11.1. Factorial de n

El *factorial* de n, expresado por $n!$, permite *determinar el número total de las posibles ordenaciones* de n elementos.

- Ejemplo: Supóngase que en una fila se encuentran formadas tres personas, denotadas por las literales A, B y C; y se desea conocer todas las posibles ordenaciones y/o posiciones que podrían ocupar en la fila. Desarrollando de forma exahustiva todas las formas posibles en que pueden ordenarse, se tiene:

    1. A,B,C
    2. A,C,B
    3. B,A,C
    4. B,C,A
    5. C,A,B
    6. C,B,A

Para este caso, se tienen seis posibles ordenaciones de las personas A, B y C en una fila. En este ejemplo el desarrollo exhaustivo de todas las posibles ordenaciones de los elementos no es muy laborioso. Sin embargo, si se intenta conocer el número de posibles ordencaciones de, siete y diez personas en la fila, el problema se complica, pues el número de posibles ordenaciones crece dramáticamente. *El factorial de n, permite determinar el número total de posibles ordenaciones de n elementos*, sin tener que hacerlo de forma exahustiva; su definición matemática es la siguiente:

Sea $n \in \mathbf{N}$, es decir n es un número natural. El factorial de n se define por:

$$n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n.$$

Por definición:

$$0! = 1 \text{ y } 1! = 1$$

- A partir de lo anterior se tiene que:
  - $2! = 1 \cdot 2 = 2$
  - $3! = 1 \cdot 2 \cdot 3 = 6$
- Ahora bien, supongase que para el caso de las personas A, B y C formadas en una fila, se tiene que $n = 3$, por lo que para obtener el número de posibles ordenaciones medinte la fórmula $n!$ se debe realizar la siguiente operación:
  $3! = 1 \cdot 2 \cdot 3 = 6$; en consecuencia, se pueden formar de $3! = 6$ formas distintas.
- Por otra parte, si se tienen 5 personas, éstas pueden formarse en un total de $5! = 120$ formas diferentes, dado que $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$

- Concecuentemente, si se tienen 7 personas formadas en una fila, éstas pueden formarse de $7! = 5040$ formas diferentes, ya que $7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$

- Finalmente si se tiene que conocer las posibles ordenaciones de 10 personas en la fila, el resultado será $10! = 3628800$

Como puede observarse, el conteo de todas las ordenaciones posibles desarrollando exhaustivamente todos los casos se vuelve prácticamente imposible. Por ello, el uso de factorial es un método que evita tener que recurrir al desarrollo exhaustivo de todos los casos posibles.

2.12. Combinaciones

Ahora, supongamos que de un grupo de tres personas, denominadas A, B y C, se deben seleccionar únicamente a dos para darles una plaza a cada una en una compañía. Las posibles parejas que pueden conformarse se muestran a continuación:

1. A,B (que es la misma pareja que B,A)

2. A,C (que es la misma pareja que C,A)

3. B,C (que es la misma pareja que C,B)

Como puede observarse, pueden conformarse tres parejas tomadas de un total de tres personas. Por otro lado, si se desean seleccionar dos personas tomadas de un grupo de cuatro, denominadas A, B, C y D, puede desarrollarse de forma exhaustiva todas las posibles parejas y determinar cuántas son:

1. A,B = B,A

2. A,C = C,A
3. A,D = D,A
4. A,E = E,A
5. B,C = C,B
6. B,D = B,D
7. B,E = E,B
8. C,D = D,C
9. C,E = E,C
10. D,E = E,D

Se tiene un total de diez posibles parejas tomadas de un total de cuatro personas. En este contexto, se denomina *combinaciones* a los conjuntos de k elementos tomados de un total de n. Al igual que en el caso de las ordenaciones, determinar el número total de las posibles combinaciones de k elementos tomados de un total de n, se complica cuando crecen k y n. En las combinaciones no importa el orden de los elementos, por ejemplo, como ya se comentó, la pareja A,B es idéntica a la pareja B,A y por lo tanto se cuentan como una sóla. La siguiente expresión permite calcular el número de posibles combinaciones:

$$\boxed{C_k^n = \frac{n!}{k!(n-k)!}}$$

Donde:
n = número total de individuos
k = Tamaño de la muestra.

- Ejemplo: para el caso de conformar parejas tomadas de un total de cuatro individuos se tiene que:
  $n = 4$
  $k = 2$
  $C_2^4 = \frac{4!}{2!(4-2)!} = \frac{4!}{2!(2!)} = \frac{24}{2(2)} = \frac{24}{4} = 6$
  es decir, se pueden conformar seis posibles parejas de un total de cuatro individuos.

- Para el caso determinar el número de posibles parejas tomadas de un total de cinco personas, se tiene que:
  $n = 5$
  $k = 2$
  Entonces el número de posibles combinaciones, viene dado por:
  $C_2^5 = \frac{5!}{2!(5-2)!} = \frac{5!}{2!(3!)} = \frac{120}{2(6)} = \frac{120}{12} = 10$
  Lo que indica que se tienen diez posibles parejas

2.13. Permutaciones

Considérese ahora que de un total de tres personas, A, B y C, se deben seleccionar una de ellas para ser el jefe de un departamento y a otra de ellas para el puesto de asistente. Como puede observarse, y a diferencia de las combinaciones, aquí sí importa el orden de los elementos y por lo tanto, la pareja A, B es diferente de la pareja B, A, luego entonces, cada una cuenta en el cálculo. A este tipo de agrupamientos se le denomina *permutaciones*.

- **Ejemplo**: Para el caso propuesto, puede determinarse de forma exhaustiva el número total de permutaciones:

Jefe	Asistente
A	B
B	A
A	C
C	A
B	C
C	B

Se observa que se tiene un total de seis posibles permutaciones.

- Ahora supón que se tienen cinco personas, A, B, C, D, y E, de las cuales se debe seleccionar una para jefe de departamento y otra para asistente.

Desarrollando de forma exhaustiva:

Jefe	Asistente
A	B
B	A
A	C
C	A
A	D
D	A
A	E
E	A
B	C
C	B
B	D
D	B
B	E
E	B
C	D
D	C
C	E
E	C
D	E
E	D

se tiene un total de 20 permutaciones.

La expresión general para el cálculo de permutaciones es:

$$\boxed{P_k^n = \frac{n!}{(n-k)!}}$$

Donde:

$n =$ número total de individuos.

$k =$ tamaño de la muestra.

Retomando el caso de seleccionar de un grupo de tres personas, a una para jefe de departamento y a otra para asistente, puede determinarse el número de posibles parejas mediante el cálculo de las permutacion para $n = 3$ y $k = 2$, es decir:

$P_2^3 = \frac{3!}{(3-2)!}$
$= \frac{3!}{1!}$
$= \frac{6}{1}$
$= 6$ posibles permutaciones.

Ahora en lo que respecta al caso de seleccionar de un grupo de cinco personas, a una para jefe de departamento y a otra para asistente, puede determinarse el número de posibles parejas mediante el cálculo de las permutacion para $n = 5$ y $k = 2$:

$P_2^5 = \frac{5!}{(5-2)!}$
$= \frac{5!}{3!}$
$= \frac{120}{6}$
$= 20$ posibles permutaciones.

Ejercicios 7

- *Instrucciones*: Considera las situaciones propuestas y con base en las fórmulas revisadas, calcula lo que se te pide.

  Supóngase que cierto grupo de jugadores de naipes, se reunirán para hacer apuestas en la casa de uno de estos. Ante tal situación, el anfitrión se encuentra en la necesidad de acomodarlos en una mesa para desarrollar la partida. En total el grupo es de 20 individuos, pero llegarán por grupos diferidos.

1. Calcular, las posibles ordenaciones de los comensales para los momentos en que a la mesa se encuentren los siguientes números de jugadores:

   a) 4

   b) 9

   c) 13

   d) 16

   e) 20

2. Para el mismo grupo de jugadores. Mientras se desarrolla la partida de naipes, en diferentes momentos se decide que éstos jugarán por pequeños grupos, para tal situación, *calcular el número total de posibles de combinaciones* cuando...

   a) de un total de 9 individuos, se pretenden conformar parejas.

   b) si los jugadores son 13 y desean jugar en grupos de tres.

   c) Cuando los 20 jugadores están en la mesa y desean jugar en grupos de 5.

3. Finalemente, se decide (durante la partida) elegir 2 parejas para que una reparta las cartas y la otra cuantifique el dinero de los depósitos para las apuestas. Mediante la técnica de permutaciones, calcular cuántas posibles combinaciones se pueden generar cuando en la mesa hay 16 personas y cuántas se pueden dar al estar los 20 individuos jugando.

   a)

   b)

Confronta tus **respuestas** con las que se te proporcionan en la página 210

2.14. Autoevaluación

El avance obtenido con la realización de los ejercicios, es importante para tu formación en el área metodológica dentro de la carrera y para tu crecimiento integral en general. Dicho avance, puede ser comprobado en la sección que ahora inicia; los resultados que obtengas, son exclusivamente el fruto de tu trabajo, si ellos te son satisfactorios ¡felicidades! En caso contrario, reconsidera tu proceder e incrementa tus esfuerzos.

- Responde correctamente las siguientes preguntas. Cuida que tus respuestas expresen correctamente el concepto que se te solicita. Considera que no es necesario *repetir pie juntillas* los conceptos, sino la expresión de su sentido, cuidando que en tus respuestas se encuentren las palabras clave que los conforman.

1. ¿Qué es un conjunto?
____
____
____
____
____
____

2. ¿Qué implica denotar a un conjunto por comprensión?
____
____
____

3. ¿Qué situaciones se deben considerar al momento de proponer un conjunto universal?
____

____

____

____

4. Denota la cardinalidad del siguiente conjunto: G={x|x es una letra del alfabeto griego}

   ____

5. Define por comprensión las siguientes operaciones con conjuntos: unión, intersección y complemento.

   a) Unión: ____

   b) Intersección: ____

   c) Diferencia: ____

6. ¿Cuántos conjuntos especiales de números existen entre el menor de los infinitos y $\aleph_1$? Descríbelos.

   ____

   ____

   ____

   ____

   ____

   ____

7. Realiza las siguientes operaciones siguiendo el orden dictado por la prioridad de los operadores que contienen.

   a) $7^4 + (40/8)^2 - (3)(9)$

   b) $((5^3) \cdot (-50))^2/4 \cdot (\sqrt{25} - 3)$

   c) $(12 + (3 \cdot 6)/(49/7 + 5))^2$

8. Complementa las expresión para denotar la *propiedad II* de la suma:

Sea $C \in$ ____. La expresión $\sum_{i=1}^{n}$ $(x_i +$ ____$)$, simboliza la suma de todas las ___ de i más C, cuando i va desde ____ hasta n, y puede desarrollarse de la siguiente manera:

$$\sum_{i=1}^{\_\_}(x_i + C) = (\_\_\_\_\_ + C) \_\_\_ (x_2 + C) + \ldots + (\_\_\_\_ + C)$$

Asimismo, ésta, puede ser denotada como ______ sumas separadas:

$$\sum_{i=1}^{n} \_\_\_\_ + \sum_{i=\_\_\_\_}^{n} C$$

Y de acuerdo a la propiedad ____ de la suma, la segunda suma de la expresión anterior, puede reordenarse y se obtiene:

$$\sum_{i=1}^{n} x_i + (n \cdot C)$$

9. Supóngase que para una investigación de mercado en el Distrito Federal, se utilizó la técnica de encuesta. El tamaño de la muestra (n) fue de 384032 individuos distribuidos en 6 delegaciones diferentes. Calcula el porcentaje de la muestra que correspondió a cada delegación y redondea el resultado a tres dígitos después del punto decimal; anota tus resultados en a línea correspondiente.

Número de cuestionarios aplicados por cada Delegación:

1^a 30500⇒......_____________%, que redondeado a tres dígitos ≈ ___________
2^a 108752⇒....____________%, que redondeado a tres dígitos ≈ ___________
3^a 935⇒.........____________%, que redondeado a tres dígitos ≈ ___________
4^a 158025 ⇒...____________%, que redondeado a tres dígitos ≈ ___________
5^a 15301 ⇒..______________%, que redondeado a tres dígitos ≈ ___________
6^a 70519 ⇒.....____________%, que redondeado a tres dígitos ≈ ___________

10. Efectúa las siguientes operaciones con notación científica.

    a) $(0{,}005)(500000) =$

    ____

    b) $\frac{15000}{2000000} =$

    ____

11. ¿Cómo se calcula una tasa?

    ____

    ____

    ____

    ____

    ____

    ____

12. Considera el siguiente conjunto y su respectiva regla de correspondencia: Sea X = $\{5,\ 10{,}15,\ 20,\ 25\}$ y $f(x) = -x + 10$; ahora construye el codominio de la función y represéntala mediante un diagrama de Venn.

13. Considera el conjunto X={-3, -2, -1, 0, 1, 2, 3}. Construye el contradominio del mismo para las siguientes reglas de correspondencia y traza su respectiva gráfica:

    a) $f(x) = 5x - 1$

    b) $f(x) = x^2$

    c) $f(x) = -x^2$

d) $f(x) = 4x^2$

e) $f(x) = 4x^2 + 1$

f) $f(x) = 4x^2 - 1$

14. Resuelve las operaciones con logaritmos.

a) $Log_{10}(15000000 \cdot 200) =$

b) $Log_{10}100^{10000} =$

c) $ln500 =$

15. Calcula el valor de x para cada ecuación.

    a) $7 + x = 21$

    b) $5x + \frac{x}{3} = 20$

    c) $9 + 4x \cdot 8 = 25$

16. Anota las fórmulas de:

    a) Permutaciones

    b) Combinaciones

Asegurate de que las respuestas que anotaste, sean las correctas comparándolas con las de la página 211

2.15. Apéndice del capítulo

2.15.1. Información remedial de la evaluación diagnóstica

Suma algebraica

Originalmente para el ser humano como especie e individuo, los sentidos de unidad y sucesión, al ser relacionados formalmente, posibilitan la construcción formal de los números (**Símbolos** que fungen como medios auxiliares para representar cantidades de objetos) y la adición,[5] esto es: un número que sucede a otro como resultado propuesto de la acumulación unitaria (la

[5]Según Juan Enrique Pestalozzi (1718 -1827), la intuición juega un papel fundamental en la elaboración de las representaciones con que el hombre reconoce el mundo. Pero éstas, no son suficientes para la naturaleza humana. Quiere elevar a conceptos precisos las *representaciones* puestas en claro sensorialmente; combinar con fuerza independiente los objetos de su intuición, separarlos y compararlos, es decir, elaborarlos lógicamente. Así,

suma de unidades). Sin embargo (en la realidad), dicho proceso no siempre aumenta la cantidad o unidad original.

La ausencia y pérdida, como hechos reales, también tuvieron que ser representados simbólicamente. El cero y el signo de la sustracción, se añadieron al universo simbólico en que se realizan las primeras operaciones matemáticas (las adiciones y/o sumas). Finalmente, la importancia de tales acontecimientos radica en la posibilidad, que con ellos se genera (histórica e individualmente hablando), de representar la pérdida mediante el uso de números con signo negativo.

En general, la *Suma Algebraica*, es la adición en donde se puede obtener como resultado una cantidad representada por números con signo negativo.[6] Para razonar el signo de su resultante, considérese lo siguiente:

Sean: a y b dos números positivos; así como: $-a$ y $-b$ dos números negativos, en general se puede afirmar que las expresiones:

1. $a + b = x$

2. $(-a) + (-b) = -x$

Indican la suma de dos elementos con sus respectivos productos, y el signo resultante de la segunda expresión, se resuelve como sigue:

$$\boxed{(-a) + (-b) = -(a + b) = -x}$$

Pestalozzi, ubica en el ser humano la *capacidad* para contar y medir, como una fuerza *inicial* (de su esencia intelectual) para elaborar lógicamente lo recogido por la intuición. PESTALOZZI; Juan Enrique; **El Canto del Cisne**; México; 1996; Editorial Porrua, Colección Sepan Cuantos; pp. 12, 42 - 44

[6]Son todos los números menores que cero; verbi gratia: $-1, -2, -3 \ldots \ \ldots -\infty$

Es preciso enfatizar el apoyo de los paréntesis que, en este caso, *únicamente fungen como signos de agrupación*; éstos, sólo indican la afectación de los signos con respecto a los números que intervienen en la suma, por lo que no se deben confundir como notación para la multiplicación; ejemplos:

- Sean $(-c) = (-5)$ y $(-d) = (-8)$, realizar la suma de éstos:
  $(-5) + (-8) = -(8 + 5) =$ -13

- Sean $(-c) = (-1)$ y $(-d) = (-0{,}5)$, realizar la suma de éstos:
  $(-1) + (-0{,}5) = -(1 + 0{,}5) =$ -1,5

- Sean $(-c) = (-120)$ y $(-d) = (-150)$, realizar la suma de éstos:
  $(-120) + (-150) = -(120 + 150) = -270$

En la suma algebraica pueden intervenir números positivos y negativos indistintamente, en tal caso, *el proceso de suma* se resuelve de manera general como se indica a continuación:

- Sean (a) y $(-b)$ dos números de los cuales $(-b)$ tiene un *Valor absoluto*[7]mayor que (a), la resolución de $(a) + (-b)$ se da como:

$$\boxed{(a) + (-b) = a - b = -x}$$

[7] "El valor absoluto de una cantidad es el número que representa a la cantidad prescindiendo del signo o sentido de la cantidad y el **valor relativo** es el sentido de la cantidad, representada por el signo. Así, el valor absoluto de +\$8 es \$8, y el valor relativo, **haber,** expresado por el signo +; el valor absoluto de −\$20 es \$20, y el valor relativo, **deuda,** expresado por el signo -."
BALDOR, Aurelio; **Álgebra**; Madrid; 1981; Ediciones y Distribuciones Códice, S. A. Madrid; pp. 13

Dónde, únicamente para los signos, aplica la ley de la multiplicación y ésta, se resuelve –finalmente– como una diferencia; para tal caso, el signo de su resultado es igual al del número cuyo valor absoluto es mayor; ejemplos:

- Sean $(a) = (5)$ y $(-b) = (-3)$ realizar la suma algebraica de $(a)+(-b)$: $(5) + (-3) = 5 - 3 = 2$

- Sean $(a) = (5)$ y $(-b) = (-8)$ realizar la suma algebraica de $(a)+(-b)$: $(5)+(-8) = 5-8 = -3$, donde el resultado es de signo negativo puesto que el número de mayor valor absoluto es el sustraendo (-8).

- Sean $(a) = (9)$ y $(-b) = (-16)$ realizar la suma algebraica de $(a)-(-b)$: $(9)-(-16) = 9+16 = 25$, puesto que se aplica la ley de los signos para resolverla, ésta se transforma en una suma donde el signo de ambos números es positivo y éste será el signo del resultado.

En los casos donde intervienen varios números con ambos signos (también se define a la suma algebraca como aquella en la que se combinan sumas y restas) se recomienda sumar, en primer lugar, a todos los números de igual signo hasta reducir la expresión a la intervención de dos sumandos y aplicar las propiedades antes revisadas, ejemplos:

- Resolver las siguientes sumas algebraicas:

    - $1 + 2 + (-5) + 3 + (-2) =$
      $(6) + (-7) = 6 - 7 = -1$

    - $(14 + 3 + 5 + 8) - (-12 - 11 - 7 - 6) =$
      $(30) + (-36) = 30 - 36 = -6$

- $(-5+8-3+2)+(3-8+5-1)=$
  $(-8+10)+(8-9)=(2)+(-1)=2-1=1$

- $(5-7-9-15+2)-(8+2-7+5-9)=$
  $(7+(-31))-(15-16)=$
  $(-24)-(-1)=-24+1=23$

- $(35-2-3-5)-(14-4+2)=$
  $(35-10)-(16-4)=$
  $(25)-(12)=13$

Finalmente, cabe agregar que a la suma algebraica no se le conoce así únicamente por la intervención de números con signos positivos y negativos que deriva en la combinación de sumas y restas, sino que en ella también pueden intervenir letras (literales)para generalizar las cantidades.[8]

Multiplicación algebraica

En general, se sabe que la multiplicación es una especie de suma abreviada. Multiplicar, implica encontrar la "suma total" de una cantidad acumulada a sí misma n cantidad de veces, es decir: Dada la suma de los números $2+2+2+2+2+2+2+2+2+2=20$ se puede observar que 20 representa la *suma total* (acumulación) del número 2 sumado a sí mismo 10 veces. Dicha operación, se puede expresar como:

1. $2 \times 10 = 20$
2. $(2) \cdot (10) = 20$
3. $2 \cdot 10 = 20$

[8]Véase el tema **Álgebra. Los Números y las Literales**, pág. 194

4. $(2)(10) = 20$
5. $2(10) = 20$

A los *factores* que intervienen en dicha operación, se les conoce así: Son *multiplicando y multiplicador* (respectivamente) los números que intervienen de tal manera que su relación permite la obtención de un resultado al que se conoce como *producto*, de tal manera que a la multiplicación también se le llama *producto*. Como en la suma algebraica, en la multiplicación, el producto de la operación también puede ser negativo, dada la intervención de factores con este signo, ejemplos:

1. $(-20) \cdot (5) = -100$
2. $(-20) \cdot (-5) = 100$
3. $(20) \cdot (-5) = -100$
4. $(20) \cdot (5) = 100$

De los anteriores ejemplos, cabe destacar el hecho de que en los casos donde intervienen términos o se obtiene un producto con signo positivo que no se denota, en tanto que el signo negativo sí se debe especificar. El signo del producto, dependerá de los signos que tengan los factores; por regla general se puede afirmar que:

- La multiplicación de *factores con signos iguales* dará como resultado un signo *positivo*.
- La multiplicación de *factores con signos diferentes* dará como resultado un signo *negativo*.

* Gráficamente:

$$\boxed{\begin{array}{l}(-)\cdot(+)=(-)\\(-)\cdot(-)=(+)\\(+)\cdot(-)=(-)\\(+)\cdot(+)=(+)\end{array}}$$

Para el caso de la multiplicación existen, por lo menos, 6 leyes que se cumplen al momento de obtener un producto, entre ellas se encuentran la *ley conmutativa*, la *ley de la identidad*, *ley de la cancelación* y otras. Éstas, aunque no siempre de manera explícita, se conocen y utilizan sin mayor problema por el público en general (pocas personas con el mínimo de escolaridad ignoran que $2 \times 9 = 9 \times 2$ ley conmutativa, o que $5 \times 1 = 5$ expresión en la que se verifica la ley de la identidad, por citar algunos casos).

Por tal motivo, de esas leyes, se considera pertinente recordar, por lo menos, dos que a juicio de quien suscribe el presente trabajo, podrían ser las menos recordadas y son de gran utilidad en el estudio de las bases para el aprendizaje de la estadística y por ende de ésta.

- *Ley asociativa.*

  Sean a, b y c tres factores relacionados en el modelo matemático $(a \cdot b)c$; el producto de tal expresión no se altera, si se efectúa el producto $(a \cdot b)$ y posteriormente se multiplica por c, como tampoco se altera al cambiar la asosiación de los factores $(b \cdot c)$ para multiplicarlos por c, es decir:

  $$\boxed{(a \cdot b)c = a(b \cdot c)}$$

  * Ejemplos:

  - $(5 \cdot 8)(9) = 360$
    $\equiv (40)9 = 360$

♣ $5(8 \cdot 9) = 360$
$\equiv 5(72) = 360$
♣ $(5) \cdot (8) \cdot (9) = 360$

- $(9 \cdot 7)5(-2) = -630$
$\equiv (63)5(-2) = -630$
♣ $9 \cdot (7 \cdot 5) \cdot (-2) = -630$
$\equiv 9 \cdot (35) \cdot (-2) = -630$
♣ $(9 \cdot 7)(5(-2)) = -630$
$\equiv 63(-10) = -630$
♣ $(9 \cdot 7)(5(-2)) = -630$
$\equiv (63)(-10) = -630$
♣ $9 \cdot 7 \cdot 5(-2) = -630$
$\equiv 315(-2) = -630$

- *Ley distributiva de la multiplicación con respecto a la suma algebraica.*

La multiplicación algebraica, también se puede combinar con operaciones entre las que se encuentra la suma algebraica. Para tales casos, el producto se puede calcular realizando, en primer lugar, la multiplicación del término por cada uno de los sumandos (*aplicando la ley de los signos*) para, posteriormente, realizar la suma algebraica; en lenguaje formal se tiene que:

- Sean a, b, c, tres números reales[9] implicados en el producto algebraico dado por $a(b + c)$, se tiene que:

$$\boxed{a(b + c) = ab + ac}$$

* Ejemplos:

◦ Sean $a = 2$, $b = 7$ y $c = 5$, calcular el producto algebraico de

[9]Véase la definición de Números Reales en la página 104.

$a(b+c)$

$2(7+5) = 14 + 10 = 24$

- Calcular el siguiente producto algebraico:

  $(-8) \cdot (3 - 5 + 2 + 4 - 6) =$

  $= (-8) \cdot (3 - 5 + 2 + 4 - 6) = -24 + 40 - 16 - 32 + 48 = 16$

- Finalmente, si $a(b+c) = ab + ac$, se puede afirmar que:
  Si $b + c = d$ entonces:

$$\boxed{a(b+c) = a(d)}$$

* Ejemplos: Considérense los dos anteriores productos algebraicos y resuélvanse simplificando la suma algebraica:

  - $2(7+5) = 2(12) = 24$

  - $(-8) \cdot (3 - 5 + 2 + 4 - 6) = (-8) \cdot (-2) = 16$

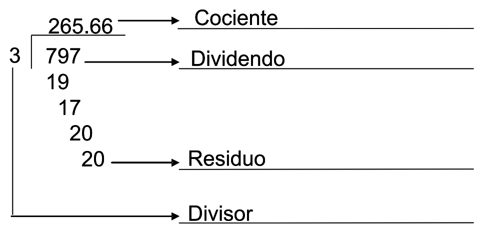
División

En principio, la división es una operación inversa a la multiplicación. Dividir, en otras palabras, consiste en calcular el factor desconocido de una multiplicación; esto, matemáticamente se expresa de la siguiente manera:

- Sean a, b dos números reales tal que $a \neq 0$, se dice que a es dividida por b (a/b) si existe un número real r al que se denomina *Cociente*, tal que:

$$\boxed{\tfrac{a}{b} = r \Rightarrow a = b \cdot r}$$

Donde al número a se le conoce como *Dividendo* y al número b como *Divisor*; cuando el cociente de $\frac{a}{b}$ no es un número entero (no es exacto) y los dígitos posteriores al punto decimal son infinitos, existe un número que posibilita la secuencia infinita de éstos, a tal número se le denomina *Residuo* y se caracteriza por ser menor que el divisor.



Finalmente, de la relación que guardan los elementos que componen a una división, se puede afirmar que el dividendo es un múltiplo del divisor. Así cuando tratase de resolver una división cuyo divisor es el diez, o uno de sus tantos múltiplos, basta con recorrer el punto decimal un dígito (cuando se trata del 10) hacia la izquierda; en el caso de la potencias de 10, el punto decimal se debe recorrer tantas veces como ceros tenga el múltiplo, ejemplos:

- $\frac{1200}{10} = 120$
- $\frac{1200}{100} = 12$
- $\frac{1200}{1000} = 1{,}2$
- $\frac{1200}{10000} = 0{,}12$
- $\frac{1200}{100000} = 0{,}012$
- $\frac{1200}{100000} = 0{,}0012$

Potenciación

Si es posible describir a la multiplicación como una forma abreviada de la suma, análogamente se puede hacer lo mismo con la *potenciación de un* *número natural.* Calcular la potencia de un número natural, implica determinar el "producto" de éste multiplicado por sí mismo n cantidad de veces, es decir:

- Dado el producto de los números

$$2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 1024$$

  se puede observar que el número 1024 es el *producto* del número dos multiplicado por sí mismo 10 veces. Dicha operación se puede expresar como potencia de la siguiente manera: 2^{10} donde al número 2 se le denomina *base de la potencia* (P) e indica el número que se multiplicará por sí mismo tantas veces como lo indica el *exponente* $(^n)$ que en este caso es el número 10. Genéricamente, la potenciación de un número se representa como:

$$\boxed{P^n = P \cdot P \cdot P \cdot P \ldots \ \ldots \cdot n \text{ veces}}$$

Para concluir con las potencias, se proporcionarán modelos generales de lo que son las operaciones básicas con éstas y la manera en que se resuelven *cuando tienen la misma base*:

- La *multiplicación* de potencias con la misma base, puede resolverse al sumar las potencias para y elevar la base de éstas según el exponente resultante de la suma, formalmente se denota por:

$$\boxed{P^n \cdot P^m = P^{n+m}}$$

  - Ejemplos:

◦ $2^5 \cdot 2^2 = 2^{5+2}$
$= 2^7 = 128$

◦ $10^5 \cdot 10^2 = 10^{5+2}$
$= 10^7 = 10000000$, nótese que en el caso de las potencias de base diez, el número de ceros corresponde al número del exponente.

◦ $10^3 \cdot 10^{-5} = 10^{3+(-5)}$
$= 10^{-2} = 0{,}001$ para un caso similar al anterior, los exponentes de signo negativo, implican colocar tantos ceros como lo indica el exponente luego del punto decimal y antes del número 1 (*potencias de base 10*).

- El resultado de la *división* entre potencias con la misma base, se calcula realizando una resta entre los exponentes para elevar la base de acuerdo al resultado de la resta:

$$\boxed{\frac{P^n}{P^m} = P^{n-m}}$$

- Ejemplos:

◦ $\frac{5^5}{5^2} = 5^{5-2} =$
$= 5^3 = 75$

◦ $\frac{10^5}{10^7} = 10^{5-7} =$
$= 10^{-2} = 0{,}001$

◦ $\frac{10^5}{10^{-7}} = 10^{5-(-7)}$
$= 10^{5+7} = 10^{12} = 1000000000000$

- La potencia de una potencia se conoce como *potencia m-ésima de una base elevada a una potencia* n, esto es $(P^n)^m$ y su resultado se calcula multiplicando las potencias:

$$\boxed{(P^n)^m = P^{n \cdot m}}$$

  - Ejemplos:
    - $(2^2)^2 = 2^{2 \cdot 2} =$
      $2^4 = 16$
    - $(10^3)^{-2} = 10^{3(-2)} =$
      $10^{-6} = 0{,}0000001$
    - $(10^3)^2 = 10^6 =$
      $= 10^6 = 1000000$

- La *raíz cuadrada* es es una operación en la que se determina la base que fue elevada al cuadrado, luego entonces, se puede afirmar que esta operación es la inversa de una potencia elevada al cuadrado.

$$\boxed{P^2 = x \Rightarrow \sqrt[2]{x} = P}$$

  - Ejemplos:
    - $\sqrt{100} = 10$, como $10^2 = 100$
    - $\sqrt{81} = 9$, como $9^2 = 81$
    - $\sqrt{64} = 8$, como $8^2 = 64$

- Cuando se calcula una *potencia cuya base es de signo negativo*$(-P^n)$, el signo del resultado se determina aplicando la siguiente regla:

  - Si el exponente es un número par, el signo del resultado será positivo, en tanto que si es impar, el exponente será negativo, ejemplo:

- $-15^2 = 225$
- $-15^3 = -3375$

Álgebra. Los números y las literales

1. **“Álgebra** es la rama de la matemática que estudia la cantidad considerada del modo más general posible.

2. “**CARÁCTER DEL ÁLGEBRA Y SU DIFERENCIA CON LA ARITMÉTICA**

   “El concepto de la cantidad en álgebra es mucho más amplio que en aritmética.

   “En Aritmética las cantidades se representan por **números** y estos expresan valores determinados. Así, 20 expresa un solo valor: **veinte**; para expresar un valor mayor o menor que éste habrá que escribir un número distinto de 20.

   “En álgebra, para lograr la generalización, las cantidades se representan por medio de **letras**, las cuales pueden **representar todos los valores.** Así, a representa el *valor que nosotros le asignemos*, y por tanto puede representar 20 o más de 20 o menos de 20, a nuestra elección, aunque conviene advertir que cuando en un problema asignamos a una letra un valor determinado, esa letra no puede representar en el mismo problema a otro valor distinto del que le hemos asignado.

3. “**NOTACIÓN ALGEBRAICA**

   “Los **símbolos** usados en Álgebra para representar las cantidades son los **números** y las **letras**.

   “Los *números* se emplean para representar cantidades conocidas y determinadas.

"Las *letras* se emplean para representar toda clase de cantidades, ya sean desconocidas o conocidas.

"Las *cantidades* conocidas se expresan por las primeras letras del alfabeto: $a, b, c, d \ldots$

"Las *cantidades* desconocidas se representan por las últimas letras del alfabeto: u, v, w, x, y, z.

"Una misma letra puede representar distintos valores diferenciándolos por medio de comillas ; por ejemplo: a'', a''', que se leen **a prima, a segunda, a tercera** o también por medio de subíndices; por ejemplo: a_1, a_2, a_3 que se leen **a subuno, a subdos, a subtres.**"[10]

Plano cartesiano

René Descartes (1596-1650) filósofo y matemático francés, "aprendió en La Flèche los fundamentos de las matemáticas y, con ellas, el deseo de encontrar una ciencia exacta aplicable a todos los terrenos [...] tuvo la visión de un nuevo método que permitiría aplicar las matemáticas, la "ciencia admirable", a todos los campos de la física [...] su vocación era definitiva: la de un matemático que quiere precisar las matemáticas y afinar un nuevo método para llegar a la verdad absoluta y necesaria. [...] Las preocupaciones de Descartes fueron sobretodo de orden teórico. Su *Geometría* donde expone el descubrimiento de la geometría a analítica permite, mediante los sistemas de coordenadas unir la geometría y el álgebra."[11]

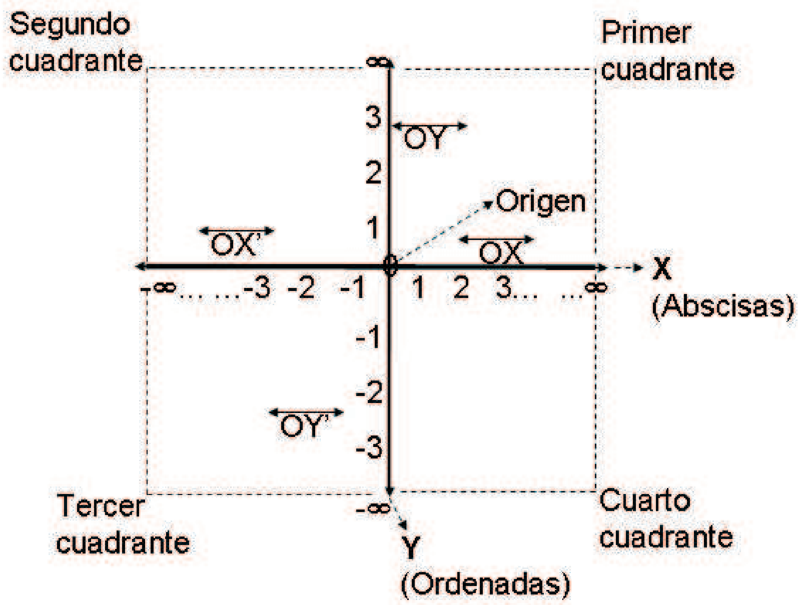
En este sentido, uno de los tantos elementos legados por Descartes fue el sistema coordenado de ejes, en cuya intersección (llamada Origen) se forma un ángulo de 90°, mejor conocido como *plano cartesiano* en honor a su

[10] op. cit. BALDOR, Aurelio; Madrid; 1981; pp. 5, 6
[11] op. cit. XIRAU, Ramón; México; 1995; pp. 183, 184

creador. "Este sistema, [...] consta de dos rectas dirigidas X', X y Y', Y llamadas *ejes de coordenadas* perpendiculares entre sí."[12]La primera recta enunciada se llama *eje de X y/o eje de las abscisas*; en tanto que a la segunda se le denomina *eje de Y* (*también conocido como eje de las ordenadas*). Ambos se pueden dividir, para su graduación, en semiejes con respecto del Origen (O), para su graduación. Partiendo de derecha a izquierda, el primer semieje, $\overleftrightarrow{OX}$, se gradúa positivamente; el segundo $\overleftrightarrow{OY}$ también se gradúa con números positivos; y los semiejes tres y cuatro, $\overleftrightarrow{OX'}$, $\overleftrightarrow{OY'}$, respectivamente, se gradúan negativamente.

Finalmente, a cada sector determinado por el área iniciada en la intersección, se le llama cuadrante. Éstos también obedecen a un orden y sobre la base de éste, es que se les denomina; el *Primer Cuadrante* se encuentra en la parte superior derecha del plano cartesiano y en él se pueden localizar las parejas ordenadas de números cuyos signos sean positivos (+,+), en el *Segundo Cuadrante* se pueden localizar las parejas ordenadas de números cuyos valores sean negativo y positivo (-,+), dicho cuadrante se encuentra a la izquierda del primero; en tercer lugar se encuentra el cuadrante donde se localizan las parejas ordenadas de números cuyos signos, en ambos casos, son negativos (-,-); Finalmente el *Cuarto Cuadrante* se encuentra justo bajo el primero y en él se encuentran las parejas ordenadas de números con signos negativo y positivo (-,+).

[12]LEHMANN, H. Charles; **Geometría Analítica**; México; Editorial Limusa; 1980; pp. 5



2.15.2. Respuestas a los ejercicios del capítulo

Ejercicios 1

1. Propón 3 ejemplos de conjuntos, denótalos por comprensión.

   - Para los tres casos, la manera en que se debe representar cada conjunto, obedecerá a la forma A={x|x es...}; los únicos aspectos que importan respecto a la naturaleza de los elementos propuestos, es que con ellos sea posible conformar un conjunto, es decir, que posean –al menos– una característica en común que permita determinar con exactitud lo que pertenece a él y lo que no.

2. Retoma tus anteriores 3 ejemplos y exprésalos por extensión.

   - Para este ejercicio, la manera correcta en que se deben representar los conjuntos, es colocando todos y cada uno de los elementos que pertenecen al conjunto dentro de las llaves con que se representa al conjunto, estos deben ir separados por comas y en caso de que el conjunto sea infinito, se tendrán que colocar algunos elementos precedidos por tres puntos suspensivos y el símbolo de infinito antes de cerrar las llaves.

3. Asigna el símbolo que corresponda Ω, $\{\}$ ó $\emptyset$ a las siguientes proposiciones, de acuerdo con la naturaleza del conjunto que expresan.

   a) A={x|x es un signo del zodiaco}................................ $\equiv A = \Omega$

   b) B={x|x es un ave de corral}...................................... $\equiv B = \Omega$

   c) C={x|x es un número positivo menor que cero}........... $\equiv C = \emptyset$

   d) D={x|x es un ser mitológico}...................................... $\equiv D = \Omega$

   e) E={x|x es un soldado mexicano de 5 años}.................. $\equiv E = \emptyset$

   f) F={x|x es un extranjero en su país}............................. $\equiv F = \emptyset$

4. Considera los siguientes conjuntos y complementa las afirmaciones en que se expresan las relaciones de pertenencia; sean:

   Ω = {x|x es una variable sociodemográfica}
   A = {x|x es una vocal}

   a) "b" Representa la Edad $\Rightarrow$... $b \in \Omega$

   b) "c" Representa el Estado civil $\Rightarrow$...................................... $c \in \Omega$

   c) "d" Representa el color de cabello $\Rightarrow$............................... $d \notin \Omega$

   d) "f" Representa los Libros que conoce $\Rightarrow$............................ $f \notin \Omega$

   e) "g" Representa el Color de piel $\Rightarrow$..................................... $g \notin \Omega$

   f) $e \in A$

   g) $f \notin A$

h) a $\in$ A

i) i $\in$ A

j) x $\notin$ A

5. Propón 3 ejemplos de conjunto vacío y 3 de conjunto Universal.

   - Para el primer caso, los conjuntos se denotarán de la siguiente manera: A={...} $\equiv$ A = $\emptyset$; puede ser por extensión o comprensión.
   - La proposición del conjunto universal, se debe apegar a la forma y criterios establecidos en las páginas 78-79 del presente capítulo, para este caso, se recomienda la notación por comprensión.

Ejercicios 2

1. ¿Qué es la cardinalidad de un conjunto?
   Es el número de elementos que contiene un conjunto y se denota con el símbolo: n( ).

2. Estima y desarrolla (sin importar el tipo de elementos que propongas) la cardinalidad de los siguientes conjuntos.

   a) D = {El Universal, La jornada, Reforma, Uno más uno, La prensa, Esto, Ovaciones}, por tanto.. n(D) = 7

   b) E = {1, 3, 5, 7, 9, 11, 13, 15, 17, 19}, por tanto........... n(E) = 10

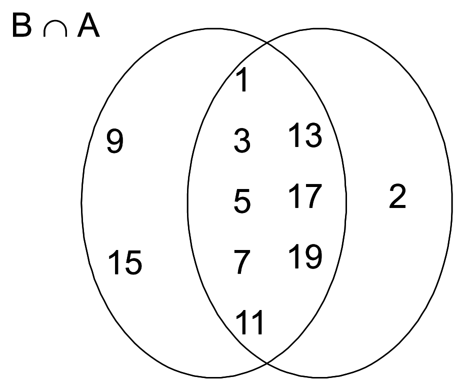
   c) F = {x|x es un mes del año}, por tanto...................... n(F) = 12

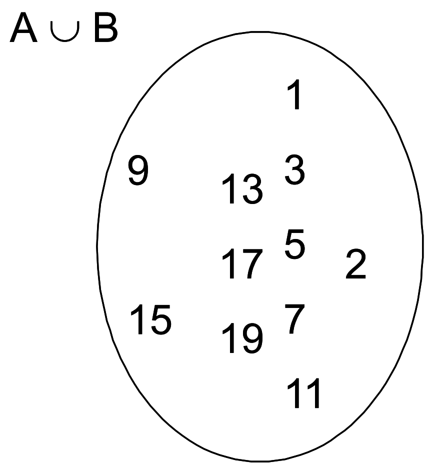
   d) n(G) = 9 $\Rightarrow$ D = {$\heartsuit$, $\heartsuit$, $\heartsuit$, $\heartsuit$, $\heartsuit$, $\heartsuit$, $\heartsuit$, $\heartsuit$, $\heartsuit$}

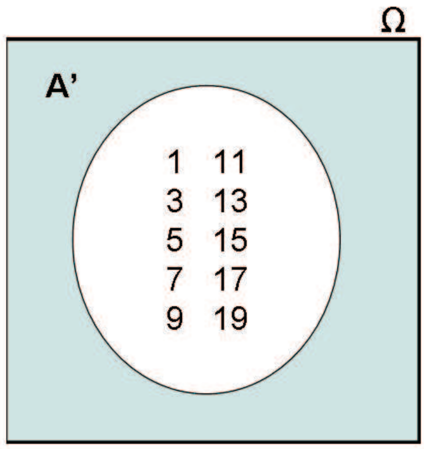
   e) n(H) = 6 $\Rightarrow$ D = {$\clubsuit$, $\clubsuit$, $\clubsuit$, $\clubsuit$, $\clubsuit$, }

   f) n(I) = 10 $\Rightarrow$ D = { $\diamondsuit$, $\diamondsuit$, $\diamondsuit$, $\diamondsuit$, $\diamondsuit$, $\diamondsuit$, $\diamondsuit$, $\diamondsuit$, $\diamondsuit$, $\diamondsuit$ }

3. Mediante el diagrama de Venn, realiza las operaciones que se indican con los siguientes conjuntos:
A = {1, 3, 5, 7, 9, 11, 13, 15, 17, 19}
B = {1, 2, 3, 5, 7, 11, 13, 17, 19}
C = {2, 4, 6, 8, 10, 12, 14, 16, 18}







4. Con los conjuntos del anterior ejercicio, realiza las siguientes operaciones y denota sus resultados por comprensión.

   a) B'= { x∈ Ω|x∈B}

   b) B ∩ C = { x| x ∈B y x∈ C}

   c) C ∪ A = { x|x ∈ C o x∈A}

5. Complementa las siguientes definiciones:

   a) En general, si se tienen dos conjuntos cuales quiera "A" y "B" tales que todos los **elementos** del conjunto "A" pertenezcan al conjunto "B", se dice que "A" **es un subconjunto** de "B", y se simboliza por $A \subseteq B$. La definición anterior considera el caso especial de aquellos conjuntos que *pertenecen a* **sí mismos**: A ⊆ A, es decir, el conjunto "A" es un subconjunto **impropio** del conjunto "A".

   b) Si todos los elementos de un conjunto "A" **pertenecen** a "B" pero **no todos** los elementos de "B" pertenecen a "A", se dice que el conjunto “A” es **un subconjunto propio** del conjunto “B” y la pertenencia se expresa de la siguiente forma: $A \subset B$

Ejercicios 3

1. Observa los siguientes conjuntos y realiza lo que se te pide.
   B = {b, c, d, f, g, h, j, k}
   C = {a, e, i, o, u}
   D = {a, b, c, d, e, f, g, h, i, j, k}

   a) n_s(B)= 256 posibles subconjuntos del conjunto B.

   b) n_s(D)= 2048 posibles subconjuntos del conjunto D.

   c) 2^C={X|X$\subseteq C$}

   d) 2^C = {∅, {a }, { a, e}, { a, i}, {a, o}, {a, u}, {a, e, i }, {a, e, o }, {a ,e ,u }, {a, e, i, o}, {a, e, i, u}, {a, e, i, o, u }, { e }, { e, i }, {e, o}, {e, u}, {e, i, o}, {e, i, u}, {e, i, o, u }, { i }, { i, o}, {i, u}, {i, o, a}, {i, o, u}, {i, o, u, a}, { o}, { o,u }, {o, u, a}, {o, u, e}, {o, u, a, e}, {u}, {u, a, i}}

2. Responde las siguientes cuestiones:

   a) ¿Qué elementos componen el conjunto de los números naturales? Denótalo por comprensión:
   N={x|x es número entero positivo}

   b) ¿Cómo se expresa por extensión el conjunto de los números enteros?
   Z=$\{-\infty \ldots \ \ldots -3, -2, -1, 0, 1, 2, 3 \ldots \ \ldots \ \infty\}$

   c) Cuál es el significado de la siguiente expresión? **N**⊂**Z**⊂**Q**
   El conjunto de los números naturales (**N**), es un subconjunto propio del conjunto de los números enteros (**Z**), y este último, es un subconjunto propio de los números racionales (**Q**).

d) ¿Cómo se puede distinguir a un número racional de uno irracional cuando ambos poseen un desarrollo decimal infinito?
Estos se distinguen entre sí, ya que el desarrollo decimal de los racionales es *periódico*, lo que indica una secuencia de números repetidos hasta el infinito; mientras que los decimales del número irracional, no son periódicos (no siguen una secuencia).

3. Complementa las siguientes proposiciones de relación entre los números y los conjuntos especiales que de números que conoces.

a) $-14 \notin \mathbf{N}$ porque es número entero pero no es positivo.

b) $\frac{5}{3} \in \mathbf{Q}$ porque su desarrollo decimal es periódico

c) $\sqrt{7} = 2{,}6457513 \in \mathbf{I}$ porque tiene un desarrollo decimal infinito no periódico.

d) $102\,005 \in \mathbf{Z}^+$ porque es entero y es positivo (el reactivo estará bien contestado sin importar qué numero se haya colocado, siempre y cuando comparta estas características).

e) $\frac{10}{1} \in \mathbf{Q}$ porque se puede expresar, de distintas maneras, como una fracción cuyo cociente es exacto.

Ejercicios 4

1. Aplica la Prioridad de los operadores para resolver las siguientes operaciones:

a) $(1+2^3)^2 + 8/\sqrt{16} - 2(5-3) = 79$

b) $(1+2^3)^2 + 8/\sqrt{16} - 2(5) - 3 = 70$

c) $(1+2)^3 + 8/\sqrt{16} - 2(5) - 3 = 16$

d) $(1+2^3) + 8/\sqrt{16} - (2 \cdot 5 - 3) = 22$

2. Mediante los principios expuestos en la sección “propiedades de la suma”, resuelve lo que se te pide:

   a) Sea $C = 9$.
   Calcular $\sum_{i=1}^{5} \ C =$
   Para facilitar el cálculo se aplica la *Propiedad I* de la suma y el resultado es 45

   b) Sea $X = \{-4, -3, -2, 0, 1, 4\}$ y $C = 7$.
   Hallar: $\sum_{i=1}^{5} \ (xi + C)$
   Se aplica la *Propiedad II* de la suma; el resultado es 168

   c) $X = \{8, 4, 6, -1\}$, y $C = 3$.
   Hallar: $\sum_{i=1}^{4} \ C \cdot xi$
   Se aplica la *Propiedad III* de la suma y el resultado es 51

3. Redondeo

   a) Redondear 52,8859013 a tres decimales........ $52{,}8859013 \approx 52{,}886$

   b) Redondear 6,3384247 a tres decimales............ $6{,}3384247 \approx 6{,}338$

   c) Redondear 0,7003589 a tres decimales............ $0{,}7003589 \approx 0{,}700$

   d) Redondear 25,9907088 a tres decimales........ $25{,}9907088 \approx 25{,}991$

   e) Redondear 14,0846061 a tres decimales........ $14{,}0846061 \approx 14{,}085$

4. Expresa y desarrolla los siguientes números mediante la notación científica.

   a) $11 \times 10^{-3} = 0{,}00011$

   b) $7 \times 10^{6} = 7000000$

   c) $0{,}000052 = 52 \times 10^{-5}$

d) $(0.2)(6000) =$
$(2 \times 10^{-1})(6 \times 10^{3})$
$= (2)(6)(10^{-1+3})$
$= 12 \times 10^{2}$
$= 1200$

e) $\frac{250000}{500000} =$
$\frac{25 \times 10^{4}}{5 \times 10^{5}}$
$= \frac{25}{5} \times 10^{4-5}$
$= \frac{25}{5} \times 10^{-1}$
$= 0{,}5$

Ejercicios 5

1. Anota en la línea correspondiente, la cantidad de dígitos significativos que posee cada número.

   a) 1,0000230
   Tiene siete dígitos significativos

   b) 0,00053
   Tiene dos dígitos significativos.

   c) $5{,}030 \times 10^{2}$
   Tiene cuatro dígitos significativos.

   d) 6000000
   Tiene seis dígitos significativos.

   e) 5,0300
   Tiene cinco dígitos significativos.

2. ¿Qué son las medidas de proporción?
   Son descriptores numéricos (números) con los que se puede constituir un sistema de referencias para comparar aquellas cantidades, con que se

han descrito ciertas características de un fenómeno objeto de estudio, que por sí mismas son insuficientes para elaborar conclusiones y tomar decisiones a partir de la investigación social.

3. Según datos del Instituto Nacional de Estadística Geografía e Informática (INEGI), hacia el año 2000 la población total de México era de 97483412 habitantes en total. Ésta, se componía por 47592253 personas son de sexo masculino y 49891159 de sexo femenino.

    a) 1,05 lo que indica una ligera superioridad en la proporción que hay de mujeres en comparación con los hombres (5 centésimas).

    b) Hay 4882,09 hombres por cada 10000 habitantes en México hacia el año 2000

    c) 13,435 % redondeado a tres dígitos.

    d) 21704925,39; pero como no puede haber "0,39 persona," la cantidad se debe redondear a números enteros y el total de habitantes es de 21704925.

4. ¿Qué es una *proporción* y cómo se calcula?
   La proporción es la *razón* que existe entre la frecuencia, o número de casos de una categoría, y el número total de categorías; ésta se calcula por: $\frac{a}{n}$, donde a es la frecuencia o número de casos de una categoría y n es el número total de categorías.

5. Resuelve las siguientes ecuaciones.

    a) $x = 5$

    b) $x = 61$

    c) $x = 7$ Para este caso en particular, nótese que la incógnita sólo es una (x) y se resuelve simplificando la expresión al sumar $12x - 7x$, lo que da como resultado $5x = 35$

d) $x = 16{,}133...$

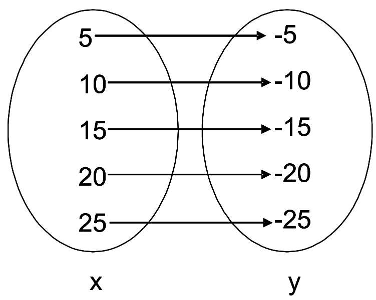
e) $x = 5$

Ejercicios 6

1. Dados los conjuntos A y B, ¿bajo qué circunstancias se dice que la relación de sus elementos es una función?
   Cuando se ha definido una regla que permite asociar a uno y sólo un elemento del conjunto A con uno de los elementos pertenecientes al conjunto B

2. Dados el conjunto X = $\{5,\ 10,\ 15,\ 20,\ 25\}$ y la regla $f(x) = -2x/2$; obtener el contradominio (Y) y representar dicha relación mediante el diagrama de Venn.

x	f (x) = -2x/2
5	-5
10	-10
15	-15
20	-20
25	-25



3. Calcula el logaritmo en base 10 de los siguientes números:

   a) $Log_{10}9 = 2{,}1972$ Redondeado a cuatro dígitos.

   b) $Log_{10}0{,}3 = -0{,}5229$ Redondeado a cuatro dígitos.

   c) $Log_{10}100000000 = 8$

4. Efectúa las siguientes operaciones con logaritmos

a) $Log_{10}(10 \cdot 10000) = 5$

b) $Log_{10}10000000 - Log_{2,718281828}100 =$ Es Imposible realizar esta operación, porque la base de ambos logaritmos no es la igual.

c) $Log_{10}100^{100} = 200$

5. Estima el ln de las siguientes cantidades

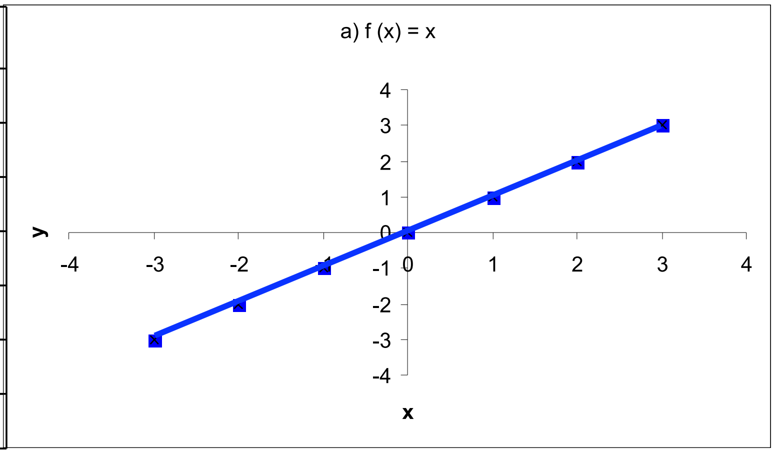
   a) $ln1000 = 6{,}9077$ Redondeado a tres dígitos.

   b) $ln5 = 1{,}6094$ Redondeado a tres dígitos.

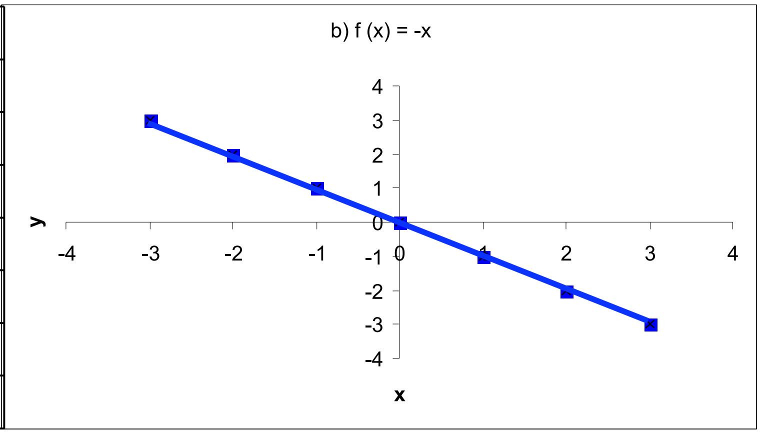
   c) $ln0{,}5 = -0{,}6931$ Redondeado a tres dígitos.

6. Considera el conjunto X={-3, -2, -1, 0, 1, 2, 3}. Construye el contradominio del mismo para las siguientes reglas de correspondencia y traza su respectiva gráfica:

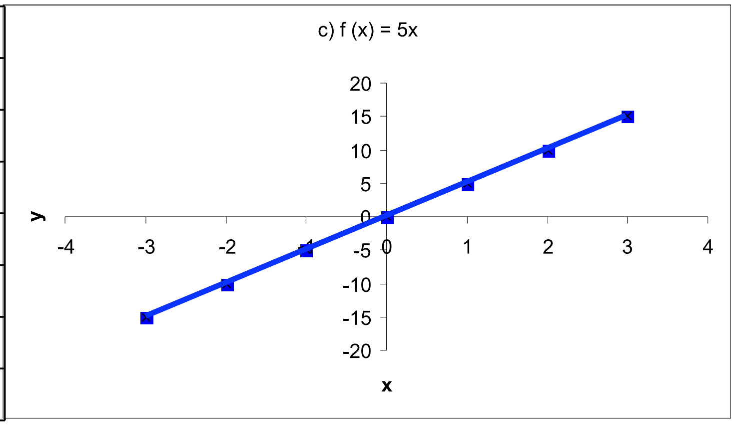
x	f (x) = x
-3	-3
-2	-2
-1	-1
0	0
1	1
2	2
3	3



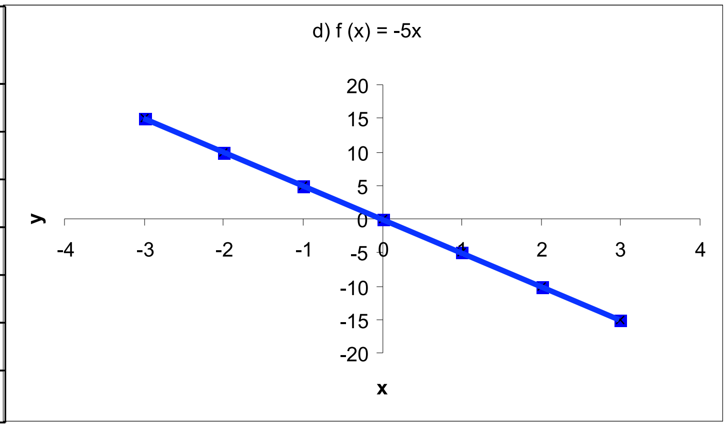
x	b) f (x) = -x
-3	3
-2	2
-1	1
0	0
1	-1
2	-2
3	-3



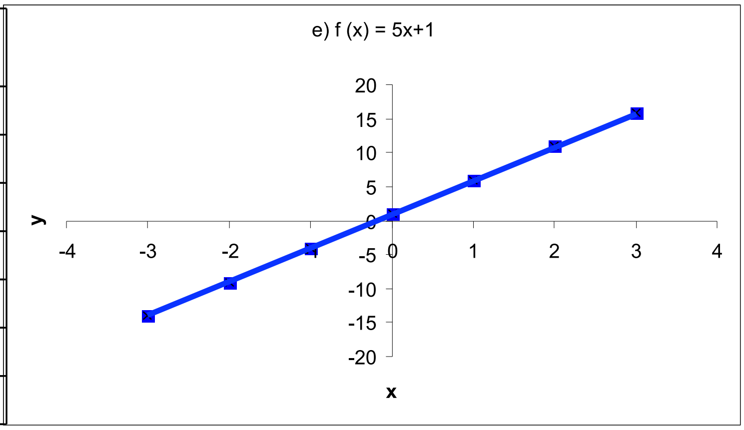
x	c) f (x) = 5x
-3	-15
-2	-10
-1	-5
0	0
1	5
2	10
3	15



x	d) f (x) = -5x
-3	15
-2	10
-1	5
0	0
1	-5
2	-10
3	-15



x	e) f (x) = 5x+1
-3	-14
-2	-9
-1	-4
0	1
1	6
2	11
3	16



Ejercicios 7

- *Instrucciones*: Considera las situaciones propuestas y con base en las fórmulas revisadas, calcula lo que se te pide.

  Supóngase que cierto grupo de jugadores de naipes, se reunirán para hacer apuestas en la casa de uno de estos. Ante tal situación, el anfitrión se encuentra en la necesidad de acomodarlos en una mesa para desarrollar la partida. En total el grupo es de 20 individuos, pero llegarán por grupos diferidos.

1. Calcular, las posibles ordenaciones de los comensales para los momentos en que a la mesa se encuentren los siguientes números de jugadores:

   a) $4! = 24$

   b) $9! = 362880$

   c) $13! = 6227020800$

   d) $16! = 2{,}092278989 \times 10^{13}$

   e) $20! = 2{,}432902008 \times 10^{18}$

2. Para el mismo grupo de jugadores. Mientras se desarrolla la partida de naipes, en diferentes momentos se decide que éstos jugarán por pequeños grupos, para tal situación, *calcular el número total de posibles de combinaciones* cuando...

    a) de un total de 9 individuos, se pueden conformar 36 parejas.

    b) si los jugadores son 13 y desean jugar en grupos de tres, entonces el número total de posibles tercias es de 286.

    c) Cuando los 20 jugadores están en la mesa y desean jugar en grupos de 5, las posibles combinaciones para formar dichas quintas, son 15504.

3. Finalemente, se decide (durante la partida) elegir 2 parejas para que una reparta las cartas y la otra cuantifique el dinero de los depósitos para las apuestas. Mediante la técnica de permutaciones, calcular cuántas posibles combinaciones se pueden generar cuando en la mesa hay 16 personas y cuántas se pueden dar al estar los 20 individuos jugando.

    a) 240 posibles combinaciones cuando hay 16 personas

    b) 380 posibles combinaciones cuando hay 20 jugadores

2.15.3. Respuestas de la autoevaluación

1. ¿Qué es un conjunto?

    En general, es una colección de elementos que tienen una característica en común. Pero desde un punto de vista formal, se dice que un conjunto es una *agrupación de elementos perfectamente definidos*, es decir, un grupo de objetos con, por lo menos, una característica en común que permite saber, con exactitud, lo que pertenece a él y lo que no.

2. ¿Qué implica denotar a un conjunto por comprensión?

   Cuando se denota a un conjunto por comprensión, los elementos de éste, se encuentran *definidos por una proposición abierta*, es decir, no aparecen explícitamente; su pertenencia está determinada por un enunciado que al describir las características de ciertos objetos, los incluye al tiempo que excluye al resto.

3. ¿Qué situaciones se deben considerar al momento de proponer un conjunto universal?

   En primer lugar, que un Conjunto Universal (Ω), consta de *todos los elementos a los que se pueda referir una situación en particular*. Dicho lo anterior, se sabe que cuando se trata de elegir (determinar) un conjunto de esta naturaleza, se debe tener conciencia, en segundo lugar, de que *el conjunto universal no es único*; depende del problema que se esté considerando y puede cambiar según la situación particular de que se trate; finalmente, cabe señalar que *para un mismo problema el conjunto universal no está definido en forma única*, se puede elegir a conveniencia, con relativa libertad (de acuerdo con el propósito para el cual se sugiere).

4. Denota la cardinalidad del siguiente conjunto: G=$\{$x$|$x es una letra del alfabeto griego$\}$
   n(G)=24

5. Define por comprensión las siguientes operaciones con conjuntos:

   a) Unión: A$\cup$B=$\{$x$|$x $\in$ A o x $\in$ B$\}$

   b) Intersección: A$\cap$B=$\{$x$|$x $\in$ A y x $\in$ B$\}$

   c) Complemento: A'=$\{$x$\in$ $\Omega|$x $\notin$ A$\}$

6. ¿Cuántos conjuntos especiales de números existen entre el menor de los infinitos y $\aleph_1$? Descríbelos.

   Existen cuatro grupos de números especiales. El primero es el conjunto de los **N**, comprendido por todos los números enteros positivos; el segundo conjunto se denota por **Z** y está conformado por todos los números enteros positivos y negativos, consecuentemente se encuentra el conjunto **Q**, sus elementos son todos los números que se pueden expresar como una fracción, finalmente, está el conjunto **I** al que pertenecen todos los números que poseen un desarrollo decimal infinito y no periódico. La cardinalidad de los conjuntos **N, Z y Q**, es $\aleph_0$, mientras que la del conjunto **I**, es $\aleph_1$.

7. Realiza las siguientes operaciones siguiendo el orden dictado por la prioridad de los operadores que contienen.

   a) $7^4 + (40/8)^2 - (3)(9) = 2399$

   b) $((5^3) \cdot (-50))^2/4 \cdot (\sqrt{25} - 3) = 19531250$

   c) $(12 + (3 \cdot 6)/(49/7 + 5))^2 = 182{,}25$

8. Complementa las expresión para denotar la *Propiedad II* de la suma:

   Sea $C \in \Re$. La expresión $\sum_{i=1}^{n}\ (x_i + C)$, simboliza la suma de todas las x de i más C, cuando i va desde 1 hasta n, y puede desarrollarse de la siguiente manera:

$$\sum_{i=1}^{n}\ (x_i + C) = (x_1 + C) + (x_2 + C) + \ldots + (x_n + C)$$

   Asimismo, ésta, puede ser denotada como dos sumas separadas:

$$\sum_{i=1}^{n}\ x_i + \sum_{i=1}^{n}\ C$$

Y de acuerdo a la propiedad I, la segunda suma de la expresión anterior, puede reordenarse y se obtiene:

$$\sum_{i=1}^{n} x_i + (n \cdot C)$$

9. Supóngase que para una investigación de mercado en el Distrito Federal, se utilizó la técnica de encuesta. El tamaño de la muestra (n) fue de 384032 individuos distribuidos en 6 delegaciones diferentes. Calcula el porcentaje de la muestra que correspondió a cada delegación y redondea el resultado a tres dígitos después del punto decimal; anota tus resultados en a línea correspondiente.
   Número de cuestionarios aplicados por cada Delegación:
   1^a 30500⇒......07,9420464 %, que redondeado a tres dígitos ≈ 07,942 %
   2^a 108752⇒....28,3184734 %, que redondeado a tres dígitos ≈ 28,319 %
   3^a 935⇒..........00,2434692 %, que redondeado a tres dígitos ≈ 00,244 %
   4^a 158025 ⇒...41,1489146 %, que redondeado a tres dígitos ≈ 41,149 %
   5^a 15301 ⇒....03,9843033 %, que redondeado a tres dígitos ≈ 03,984 %
   6^a 70519 ⇒.....18,3627926 %, que redondeado a tres dígitos ≈ 18,363 %
   NOTA: Con frecuencia, la presentación de resultados es apoyada, visualmente, con la elaboración de gráficas. Para el caso de las gráficas circulares (también conocidas como "gráficas de pastel"), es *indispensable realizar con precisión el proceso de redondeo*, pues para que éstas sean generadas, la suma de los porcentajes redondeados debe ser igual a cien por ciento; de lo contrario, será imposible efectuar la representación.

10. Efectúa las siguientes operaciones con Notación Científica.

    a) $(0{,}005)(500000) = (5)(5) \times 10^{-3+5} = 25 \times 10^2 = 2500$

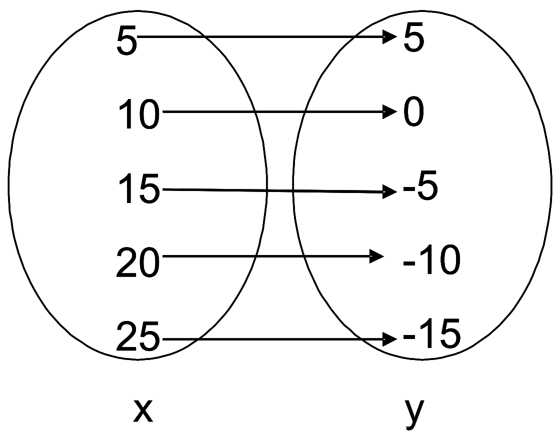
    b) $\frac{15000}{2000000} = \frac{15\times10^3}{2\times10^6} = \frac{15}{2} \times 10^{(3-6)} = \frac{15}{2} \times 10^{-3} = 0{,}0075$

11. ¿Cómo se calcula una tasa?

    Para medir la frecuencia con que ocurren ciertos fenómenos (calcular una tasa), primero se toma como base a un número referente (Per cápita, cada 1000, 10000...); posteriormente, se aplica una *regla de tres simple*: $\frac{a \cdots\cdots b}{c \cdots\cdots x}$ donde para x la expresión se resuelve de la siguiente manera: $\boxed{x = \frac{c \cdot b}{a}}$ En donde c es el referente, b la frecuencia de casos y a el número total de elementos del objeto de estudio.

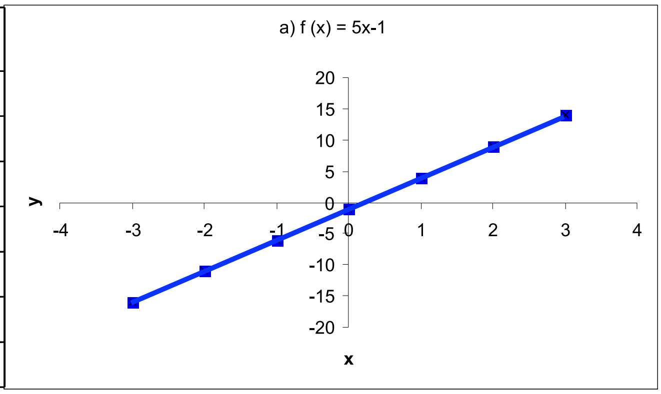
12. Considera el siguiente conjunto y su respectiva regla de correspondencia: Sea X = {5, 10,15, 20, 25} y $f(x) = -x + 10$; ahora construye el codominio de la función y represéntala mediante un diagrama de Venn.

x	f (x)=-x +10
5	5
10	0
15	-5
20	-10
25	-15

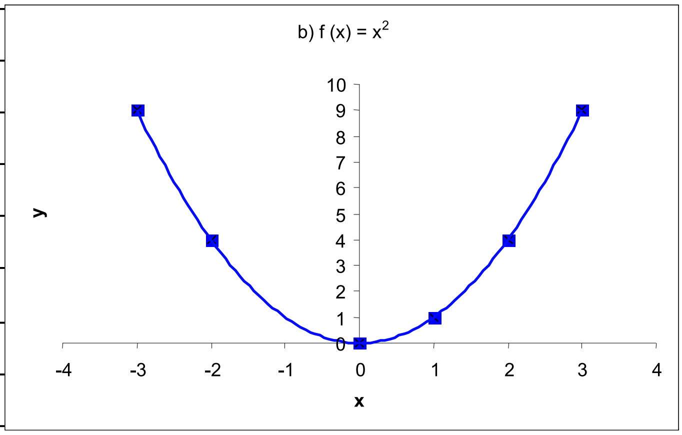


13. Considera el conjunto X={-3, -2, -1, 0, 1, 2, 3}. Construye el contradominio del mismo para las siguientes reglas de correspondencia y traza su respectiva gráfica:

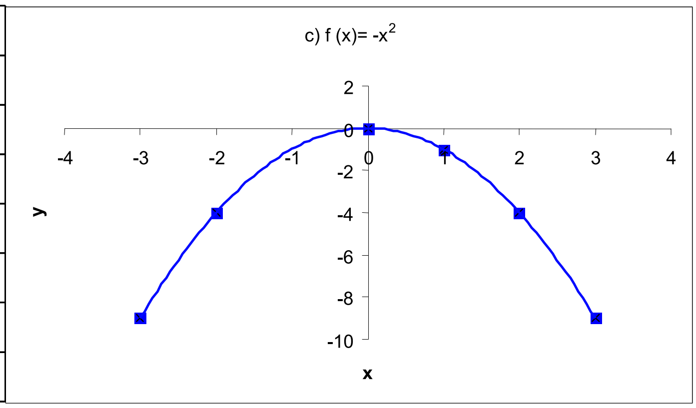
x	a) f (x) = 5x-1
-3	-16
-2	-11
-1	-6
0	-1
1	4
2	9
3	14



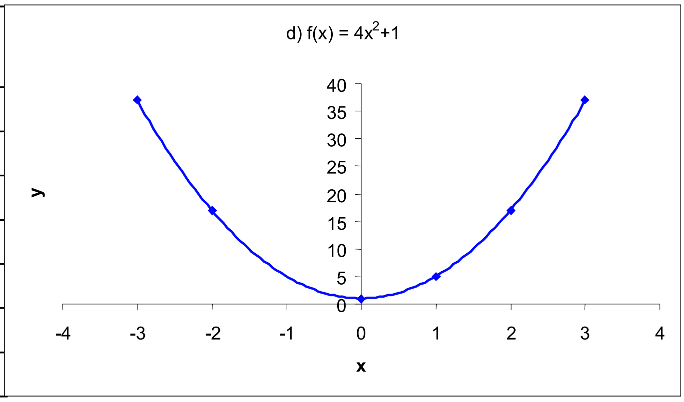
x	b) f (x) = x^2
-3	9
-2	4
1	1
0	0
1	1
2	4
3	9



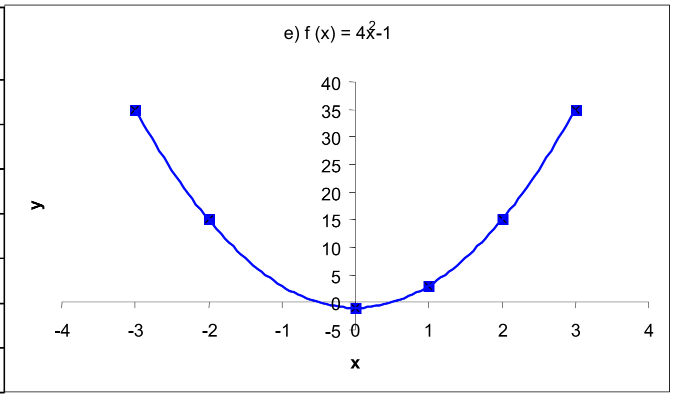
x	c) f (x) = $-x^2$
-3	-9
-2	-4
1	-1
0	0
1	-1
2	-4
3	-9



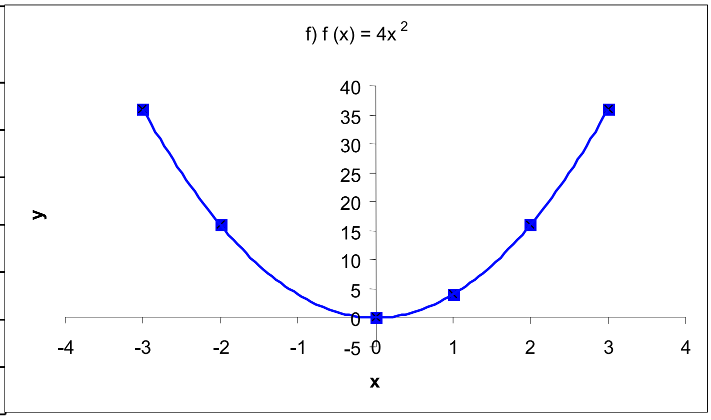
x	d) f (x) = $4x^2+1$
-3	37
-2	17
1	5
0	1
1	5
2	17
3	37



x	e) f (x) = $4x^2-1$
-3	35
-2	15
1	3
0	-1
1	3
2	15
3	35



x	f) f (x) = $4x^2$
-3	36
-2	16
1	4
0	0
1	4
2	16
3	36



14. Resuelve las operaciones con logaritmos.

    a) $Log_{10}(15000000 \cdot 200) = 9{,}4771$

    b) $Log_{10} 100^{10000} = 20000$

    c) $ln500 = 6{,}2146$

15. Calcula el valor de x para cada ecuación.

    a) $x = 7$

    b) $x = 45$

    c) $x = 0{,}5$

16. Anota las fórmulas de:

    a) Permutaciones $P_n^k = \frac{n!}{(n-k}$

    b) Combinaciones $C_k^n = \frac{n!}{k!(n-k)!}$
    Para ambos casos: k = Tamaño de la muestra y n = Número de individuos.

2.15.4. Glosario

Análisis: Del verbo *analizar*, que es descomponer un todo en sus partes elementales para conocer aspectos pertinentes a su naturaleza interior, tales como cuantificar o cualificar los componentes y sus características, las relaciones que estos guardan entre sí o con agentes externos, por mencionar algunas.

Categoría(s): "Del griego categoria, cualidad atribuida aun objeto.

"En ciencias sociales: Con el fin de establecer generalizaciones válidas, las ciencias sociales operan con categorías, no con personas. Sin embargo, las ciencias sociales no necesitan una definición precisa de tal noción, ya que existe un acuerdo tácito en concebir las categorías como meros conceptos generales clasificatorios: en las ciencias y técnicas sociales, una categoría es simplemente una clase y su existencia depende de la definición del observador.

"En la actualidad, se ha visto amalgamado al considerar también como categorías las estructuras más generales, entendidas no tanto como conceptos generales, sino como más bien como tipos diversos."[13]

Cosa: Es lo que se puede presentar a la conciencia humana pero que aún no es blanco de ella, es decir, todo aquel ente cognoscible que está fuera de los actos del pensamiento, de los actos, de la percepción o de la imaginación, al contrario de los objetos.

Cuantitativa: En el campo de la investigación social, existen dos tendencias metodológicas mediante las cuales se puede abordar un fenómeno-objeto de estudio; una de ellas, es la metodología cuantitativa. Ésta se basa en el uso de la matemática para el manejo de las variables consideradas en cierta investigación. Las características del fenómeno-objeto de estudio, se pueden representar mediante números para su estudio; así, tanto la hipótesis como los resultados de la investigación se pueden someter a comprobación formal.

Datos primarios: Magnitudes descriptivas de un fenómeno cuyo origen es formal pero no estadístico, por lo que sólo aportan una idea general de la situación que guarda el fenómeno.

[13]DEL CAMPO, Salustiano (Director); **Diccionario de las Ciencias Sociales. Vol. 1**; Madrid; Instituto de Estudios Políticos; 1975; pp. 349, 350

Delimitación: Implica circunscribirse a un campo en especifico del saber. "En este rubro requerimos situar un objeto de estudio, además de ubicar la condición teórico, metodológica y/o técnica desde la cual se abordará y por supuesto establecer la condición espacio-temporal."[14]

Denotar: "Singularizar un algo mediante signos."[15]

Desarrollo decimal: Es la serie de dígitos ulteriores al punto decimal de un número; ésta puede ser finita y/o periódica (Números Racionales), en cuyo caso la serie se puede presentar como una constante (dígito que se repite) o infinita (Números Reales), cifras decimales que por lo general son irracionales –una secuencia cualquiera de dígitos no periódicos e interminables.

Habilidades: "Una habilidad humana puede definirse como la unión de un proceso (o procesos) y un contenido (o contenidos) derivado de cambios relativamente permanentes de la conducta. La mayoría de las habilidades son del tipo cognoscitivo o del tipo psicomotor; no obstante, muchas de ellas tienen elementos de ambos tipos."[16]En otras palabras, las habilidades son capacidades humanas de adaptación al cambio (formas de aprendizaje) que implican un correcto manejo de cierta actividad intelectual, conductual y/o de ambas, que pueden, o no, permanecer en el individuo.

---

[14]AGUILAR, Vieyra Griselda; **Cuadernos de Educación y Comunicación. Un Modelo General de Diseño de Investigación, en Comunicación (Propuesta)**; México; Junio de 1999, número 4 Volumen 1 año 2; pp. 40

[15]PONTÓN, Gonzalo (Director); **Grijalbo Diccionario Enciclopédico**; Madrid; Grijalbo, 1986

[16]KLAUSMEIER, Herbert J.; et. al.; **Psicología Educativa**; México; Editorial Harla; 1971; pp. 11

Informa: Del verbo informar, dar forma, establecer un orden, "informar es ordenar, lo informado es lo ordenado. La información es **un principio de orden** que rige la organización de cualquier sistema."[17]

"**Objeto**: (lat. objectum, echado en frente) suele designar aquello hacia lo cual se dirigen los actos del pensamiento, de la imaginación o de la percepción"[18] humana.

Opinión pública: Fenómeno comunicativo (humano) en el que la interacción simbólica se genera a propósito de hechos que interesan a la sociedad, o a algunos sectores de ésta, donde las afirmaciones vertidas no son, del todo, ciertas o falsas, pues la opinión no es conocimeiento.

Planteamiento del problema: Se trata de explicar los aspectos, factores o elementos relevantes relacionados con el problema de investigación. La formulación de un problema debe concentrarse en la enunciación que responda de manera clara al qué y para qué de la investigación. Aquí a nivel descriptivo se trata de desdoblar el tema delimitado. De esta manera, plantear significa mostrar o proponer la forma de abordar el tema de investigación. Este punto comprende tres partes: 1) Descripción, 2) Contextualización y 3) Preguntas de investigación.[19]

Proposición: "Se define como proposición o enunciado a una oración declarativa carente de ambigüedad, que es verdadera o falsa, pero nunca las dos cosas simultáneamente. *Proposición **Abierta***: Una proposición abierta (o función proposicional) es una expresión que contiene una variable y que al

[17] REVILLA, Basurto Mario A. **Introducción a la Teoría de la Comunicación**; México ; S y G Editores; 1997; pp. 19

[18] op. cit. XIRAU, Ramón; México; 1995; pp. 466

[19] op. cit. AGUILAR, Vieyra Griselda; México; Junio de 1999; pp. 41

ser sustituida dicha variable por un valor determinado, hace que la expresión se convierta en una proposición. ((V) ó (F)). p(x); p(x,y,z,)."[20]

Símbolos: "Se llama símbolo a un signo sin semejanza ni contigüidad, sino solamente con un vínculo convencional entre su significante y su denotado, además de con una clase intencional para su designado."[21]

Signo Lingüístico: En general, un signo es un hecho real (por tanto perceptible) que aporta información sobre algo distinto de sí;[22] en tanto que la lengua es un sistema de signos que los seres humanos utilizan para comunicarse oralmente y/o por medio de la escritura; así, todos los hechos perceptibles de que se conforma la lengua son *signos lingüísticos*: Invenciones humanas creadas con el propósito de referir al mundo, de hacerlo explícito mediante la modificación de la materia (hecho también denominado acto expresivo por Manuel Martín Serrano); tales modificaciones, aunque arbitrarias, obedecen a ciertos acuerdos previos (convencionalismos culturales) que permiten el entendimiento intersubjetivo. Cabe resaltar el hecho de que la lengua es un sistema de signos con que se puede aportar información sobre las propiedades del mismo signo o sistema.

Variables: "Se nombra variable a cualquier característica o propiedad que permita hacer distinción entre individuos o colectivos. La distinción debe entenderse como posesión o no posesión de la característica... en todo caso, las variables representan simbólicamente –cuantitativa o cualitativamente–

[20]ROMO, Martínez Socorro Leticia; **Matemáticas 1**; en *http : //www.universidadabierta.edu.mx/SerEst/Apuntes/RomoSocorro_MateI.doc*; México Ciudad Bugambilias, Zapopan, Jal.; 2003; pp. 4; Consultada: 26/08/2005; 17:23 PM

[21]SEBEOK, Thomas A. **Signos: una Introducción a la Semiótica**; México; Paidós; 1996; pp. 49

[22]ÁVILA, Raúl; **La Lengua y los Hablantes**; México; Editorial Trillas; 1997; pp. 11

distintas dimensiones de las unidades de investigación. En tal sentido, las variables se diferenciarán del concepto por su referencia más empírica y particularizada a la realidad."[23]

Variables sociodemográficas: Elementos que sirven para distinguir a una poblacion o muestra, designados en función de situaciuones sociales y del territorio o sección de éste que ocupa.

[23]op. cit. DEL CAMPO, Salustiano (Director); Madrid; 1975; Vol. II; pp. 1148

Capítulo 3

Organización de los datos

Objetivos específicos

- El estudiante será capaz de construir Tablas de Distribución de Frecuencias y, a partir de ellas, gráficas como el Histograma y los Polígonos de frecuencia, que describan las características de una muestra aleatoria y posibiliten el posterior cálculo de los estadígrafos (descriptores) de ésta.

3.1. Preliminar

"La estadística está ligada con los métodos científicos en la toma, organización, recopilación, presentación y análisis de datos tanto para la **deducción** de conclusiones como para la toma de decisiones razonables de acuerdo con tales análisis."[1] El objetivo final de la estadística (la toma de decisiones razonables), se sustenta en un trabajo sistemático propio de la matemática.

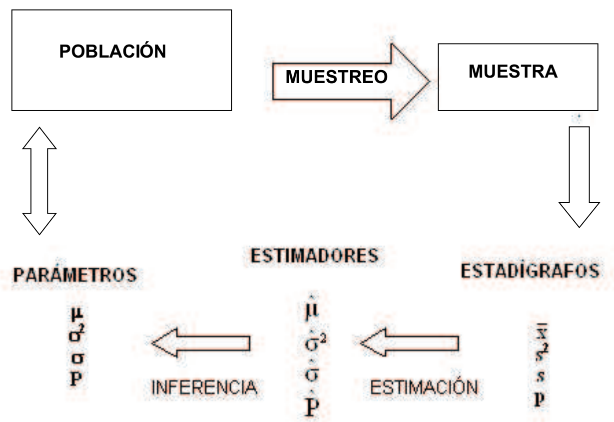
Como se explicó en el capítulo 1, ante las dificultades propias que ofrecen para su estudio los fenómenos sociales, las **disciplinas** o ciencias que se encargan de este trabajo, se apoyan en *el método estadístico conocido como muestreo; éste, sigue la estrategia de seleccionar una parte representativa del universo a estudiar*, con el propósito de hacer viable el reconocimiento formal del objeto u objetos de estudio.

Se sabe que al universo también se le llama población y ésta es el conjunto total de elementos (pueden ser individuos u objetos con características en común) que se desea conocer por medio de un tratamiento cuantitativo; de igual manera, se debe recordar que la *muestra, o parte representativa del universo*, es un subconjunto de la población; es decir, un micro universo **semejante** a la población objetivo.

Para que una muestra sea representativa de la población (*Muestra Probabilística*), se debe obtener mediante un *proceso aleatorio*. Dicha cuestión será detallada en el Capítulo 10 (**Técnicas de muestreo** página 487); por el momento bástese con saber que para la conformación de ésta, los elementos que la integran tuvieron las mimas posibilidades de ser elegidos que cualquier otro de la población.

[1] op. cit. SPIEGEL, Murray R.; México; 1970; pp. 1

Las variables en estudio tanto para la población como para las muestras se basan en las *medidas de tendencia central, de dispersión y de la proporción*. Tales medidas son rasgos esenciales que permiten el reconocimiento y estudio de la población y la muestra. Las características que describen a una población (cuantitativamente hablando) se conocen como *Parámetros* y los datos característicos obtenidos del estudio de una muestra se llaman *Estadígrafos o Estimadores*.



Si la muestra se ha obtenido a través del procedimiento correcto, los descriptores se podrán generalizar a la población, con lo que será posible el proceso de toma de decisiones. Previo al análisis de datos para la deducción de conclusiones y toma de decisiones, existe un tratamiento de éstos que permite procesar los datos que así se constituyen como información.

En esta sección se explican los procesos de organización y descripción de las características de una muestra que, se asume, ha sido seleccionada de forma aleatoria y con el tamaño adecuado, por lo que se considera representativa de la población sujeta a estudio. Pero antes, debes resolver lo siguiente:

3.1.1. Evaluación diagnóstica

- *Instrucciones*: Responde brevemente las siguientes preguntas y coteja tus respuestas con las del siguiente apartado (página 230)

1. En lógica matemática: ¿Qué son los término de enlace?

   ____

   ____

   ____

   ____

2. ¿Qué significa el signo $\Leftrightarrow$ en castellano?

   ____

3. ¿A qué término(s) de enlace equivale el signo $\Leftrightarrow$ y qué indica?

   ____

   ____

   ____

   ____

4. ¿Qué indica el signo $\geq$?

3.1.2. Respuestas de la evaluación diagnóstica

- *Instrucciones*: Verifica que el *sentido* de tus respuestas esté de acuerdo con el de las planteadas en este apartado. Pare ello, considera las palabras resaltadas y sus posibles sinónimos (expresados en tus respuestas).

  1. En lógica matemática: ¿Qué son los términos de enlace?
     Palabras que sirven para formar *proposiciones compuestas* (también llamadas moleculares) mediante la unión de *proposiciones simples* (o atómicas) e, incluso, a partir de su simple adición (como es el *caso específico de la negación*). También se denominan *conectivos lógicos*.
  2. ¿Qué significa el signo $\Leftrightarrow$ en castellano?
     Si y sólo si (término de enlace conocido como bicndicional).
  3. ¿A qué término(s) de enlace equivale el signo $\Leftrightarrow$ y qué indica?
     A dos condicionales ($\Rightarrow$) que van en sentido opuesto, e indica que se cumplirá el antecedente únicamente si se efectúa el consecuente, siendo que las proposiciones involucradas cumplen con ambas funciones (son antecedente y consecuente).
  4. ¿Qué indica el signo $\geq$?
     Dados los números p, q relacionados por éste signo ($p \geq q$), el valor relativo de p puede ser mayor o igual que el valor relativo de q

- Cada respuesta acertada, tiene un valor de 2 puntos. Si al sumar tus aciertos el resultado es 8 ¡felicidades! puedes pasar al apartado 3.2.

(pág. 237) para iniciar el estudio del capítulo.

Si al sumar tus aciertos el resultado es igual a 6, debes leer del apartado **3.6.1. Información remedial** (página 267) los temas correspondientes a las preguntas en que te equivocaste y resolver los **Ejercicios de recuperación** (apartado 3.1.3. en la página 231) pertinentes al tema.

Pero si tu resultado fue inferior a 6, debes estudiar toda la *Información remedial* (apartado 3.6.1. página 267) correspondiente al *Apéndice del capítulo* y luego resolver los *Ejercicios de recuperación* (apartado 3.1.3. pág. 231), para poder iniciar el estudio del contenido propio de este capítulo.

Las primeras tres preguntas corresponden al tema denominado *Términos de enlace* (pág. 267) en el Apéndice del capítulo y la número cuatro, al tema *Relación de orden* $\leq, \geq$ del mismo apartado, que se encuentra en la página 270

3.1.3. Ejercicios de recuperación

- *Instrucciones*: Antes de responder el cuestionario, debes leer la Información remedial del tema(s) en que tuviste errores.
- Responde brevemente las siguientes preguntas y compara tus respuestas con las del apartado 3.1.4. (pág. 232)

1. ¿Cómo se denomina a las proposiciones compuestas cuyo término de enlace es $\Leftrightarrow$?

   ----

2. Describe lo que implica para dos proposiciones simples, su unión mediante el término de enlace: $\Leftrightarrow$

   ____

   ____

   ____

   ____

3. ¿Qué indica el signo $\leq$?

   ____

   ____

4. Anota en cada una de las líneas, la palabra *Verdadera* si la proposición es correcta y *Falsa* si no lo es.

   - $q \geq p$ es ________ cuando $14 \geq 11$
   - $q \leq p$ es ________ cuando $15 \leq \frac{15}{15}$
   - $q \geq p$ es ________ cuando $13 \geq 51$
   - $q \leq p$ es ________ cuando $7 \leq 5$
   - $q \leq p$ es ________ cuando $\frac{1}{4} \leq \frac{1}{5}$
   - $q \geq p$ es ________ cuando $\frac{9}{3} \geq 2$

3.1.4. Respuestas a los ejercicios de recuperación

- *Instrucciones*: Para este caso, las respuestas del ejercicio, deberán ser correctas en su totalidad, así que luego de la comparación, si tu desempeño fue el esperado, pasa a la sección 3.2. página 237; de lo contrario, deberás realizar todos los *Ejercicios de asimilación* (apartado 3.1.5. página 233) propuestos para este capítulo.

1. ¿Cómo se denomina a las proposiciones compuestas cuyo término de enlace es $\Leftrightarrow$?
   Bicondicional

2. Describe lo que implica para dos proposiciones simples, su unión mediante el término de enlace $\Leftrightarrow$:
   Puesto que la bicondicional es el *equivalente* a dos proposiciones condicionales que van en sentido opuesto, ambas proposiciones cumplen la función de antecedente y consecuente en la proposición molecular.

3. ¿Qué indica el signo $\leq$?
   Dados los números p, q relacionados por éste signo ($p \leq q$), el valor relativo de p puede ser menor o igual que el valor relativo de q

4. Anota en cada una de las líneas, la palabra *Verdadera* si la proposición es correcta y *Falsa* si no lo es.

   - $q \geq p$ es **Verdadera** cuando $14 \geq 11$
   - $q \leq p$ es **Verdadera** cuando $15 \leq \frac{15}{15}$
   - $q \geq p$ es **Falsa** cuando $13 \geq 51$
   - $q \leq p$ es **Falsa** cuando $7 \leq 5$
   - $q \leq p$ es **Falsa** cuando $\frac{1}{4} \leq \frac{1}{5}$
   - $q \geq p$ es **Verdadera** cuando $\frac{9}{3} \geq 2$

3.1.5. Ejercicios de asimilación

- *Instrucciones*: Antes de resolver el ejercicio, lee todos los reactivos y regresa a estudiar la sección de *Información remedial* (página 267); posteriormente, realiza lo que se te indica.

- Completa las expresiones colocando la(s) palabra(a) necesaria(s) en donde lo indica la línea. Compara tus respuestas en la sección 3.1.6. (pág. 235)

1. Los términos de enlace o ____________________ son *palabras* que sirven para formar *proposiciones* __________ (también llamadas moleculares) mediante la __________ de *proposiciones simples* (o __________) e, incluso, a partir de su simple adición (como es el *caso específico de la* __________).

2. Las proposiciones que utilizan el término de enlace *Si y sólo si* se denominan ____________________ y el símbolo que lo denota es _____. Una proposición ________ se parece a dos proposiciones condicionales; éstas tienen la forma $\mathbf{P} \Leftrightarrow \mathbf{Q}$ y se pueden leer de esta manera: __________________. La proposición bicondicional $\mathbf{P}\Leftrightarrow\mathbf{Q}$ tiene la misma fuerza que __________ __________________________

3. En la proposición molecular $\mathbf{P}\Leftrightarrow\mathbf{Q}$ las proposiciones atómicas **P** y **Q** cumplen una __________________ . Éstas, son al mismo tiempo, _________ _________ y _________________; de tal manera que para determinar su valor de Verdad, la regla es la siguiente: Para que la proposición bicondicional sea _________________, el valor de verdad en las proposiciones atómicas debe ser igual (V y V o _________________), en caso contrario, su valor de Verdad será falso.

4. El lenguaje puro de la matemática, está constituido por ____________ que permiten enunciar ______________ y ______________ con toda exactitud y un mínimo de símbolos. Éste, excluye el uso de las letras en un sentido convencional, y las adscribe a él, sin la intencionalidad de formar palabras, de tal manera que __________________ (sea letra o no), __________________.

5. Uno de estos símbolos es el que expresa la idea de relación entre dos números donde uno de ellos puede ser *mayor o igual que* el otro, este es ______. Dicho símbolo aparece principalmente en ________________________, es decir, ciertas expresiones que contienen una ____________ que al ser sustituida por un valor determinado, hace que la expresión se convierta en una proposición ____________ o Falsa.

6. Dados los números q, C relacionados así: $q \geq C$, el valor relativo de q debe ser ____________________ que el valor relativo de C para que la proposición sea ____________________.

3.1.6. Respuestas a los ejercicios de asimilación

- Si alguna de tus respuestas es errónea, consulta las dudas que tengas con un profesor.

1. Los términos de enlace o **CONECTIVOS LÓGICOS**, son *palabras* que sirven para formar *proposiciones* **COMPUESTAS** (también llamadas moleculares) mediante la **UNIÓN** de *proposiciones simples* (o **ATÓMICAS**) e, incluso, a partir de su simple adición (como es el *caso específico de la* **NEGACIÓN**).

2. Las proposiciones que utilizan el término de enlace *Si y sólo si* se denominan **BICONDICIONAL** y el símbolo que lo denota es $\Leftrightarrow$. Una proposición **BICONDICIONAL** se parece a dos proposiciones condicionales; éstas tienen la forma $\mathbf{P} \Leftrightarrow \mathbf{Q}$ y se pueden leer de esta manera: ***P* SI Y SÓLO SI *Q***. La proposición bicondicional $\mathbf{P}\Leftrightarrow\mathbf{Q}$ tiene la misma fuerza que **DOS PROPOSICIONES CONDICIONALES**

3. En la proposición molecular **P⇔Q** las proposiciones atómicas **P** y **Q** cumplen una **DOBLE FUNCIÓN**. Éstas, son al mismo tiempo, **ANTECEDENTE** y **CONSECUENTE**; de tal manera que para determinar su valor de Verdad, la regla es la siguiente: Para que la proposición bicondicional sea **VERDADERA**, el valor de verdad en las proposiciones atómicas debe ser igual (V y V o **F y F**), en caso contrario, su valor de Verdad será falso.

4. El lenguaje puro de la matemática, está constituido por **SÍMBOLOS** que permiten enunciar **DEFINICIONES** y **PROPIEDADES** con toda exactitud y un mínimo de símbolos. Éste, excluye el uso de las letras en un sentido convencional, y las adscribe a él, sin la intencionalidad de formar palabras, de tal manera que **UN SÓLO SÍMBOLO** (sea letra o no), **CONNOTA UNA IDEA**.

5. Uno de estos símbolos es el que expresa la idea de relación entre dos números donde uno de ellos puede ser *mayor o igual que* el otro, este es $\geq$. Dicho símbolo aparece principalmente en **PROPOSICIONES ABIERTAS**, es decir, en ciertas expresiones que contienen una **VARIABLE** que al ser sustituida por un valor determinado, hace que la expresión se convierta en una proposición **VERDADERA** o Falsa.

6. Dados los números q, C relacionados así: $q \geq C$, el valor relativo de q debe ser **MAYOR O IGUAL** que el valor relativo de C para que la proposición sea **VERDADERA**.

3.2. Organización de los datos de una muestra

Una vez obtenida una muestra aleatoria, se procede a realizar los siguientes pasos:

1. Ordenar los datos de la muestra en forma ascendente.

2. Determinar el *rango* de los datos. Para ello, hay que restar el valor del dato menor al del dato mayor de la muestra.

3. Determinar el número k de *categorías* en las que se organizarán los datos; esto es, proponer x número de subconjuntos en los que se puedan distribuir los elementos de la muestra.

4. Definir el *ancho o longitud* de las categorías, de tal manera que *todos* los datos de la muestra se puedan concentrar en ellas; se debe cuidar que cada uno de los datos, se ubiquen en una categoría y *sólo en una*, en otras palabras: establecer el número de elementos que conformarán a cada subconjunto, de forma que ningún dato (elemento) pertenezca a más de una categoría o quede fuera de éstas. En este sentido, se dice que *las categorías deben ser mutuamente excluyentes.* A las categorías se les denomina, más apropiadamente, ***intervalos*** *de clase.*

5. Una vez definidas las categorías, se contabiliza cuántos elementos de la muestra pertenecen a cada intervalo de clase, es decir, se determinan las *frecuencias de clase*. Con estos datos se construye una tabla, que se denomina *tabla de datos agrupados.*

3.2.1. Presentación de los datos

"Frecuentemente, es posible resumir toda la información importante, que se tiene de una gran cantidad de datos, en un dibujo sencillo. Así, uno de los métodos más ampliamente utilizados para presentar datos es mediante gráficas."[2] Éstas, en general, son de gran utilidad para la interpretación de datos.

6. Con los datos de la tabla anterior se construyen las siguientes gráficos: El Histograma, Polígono de frecuencias y Ojiva porcentual.

7. Finalmente, para efectos del análisis de los datos de la muestra, se calculan los estadígrafos de la muestra (datos característicos de la muestra, también conocidos como *descriptores*, cuyo cálculo permite conocer los elementos que posee una muestra y la caracterizan) como son las *medidas de tendencia central y de variación*, que se describirán en los capítulos 4 y 5.

Ejercicios 1

- Instrucciones. Selecciona con una X el inciso que contenga la respuesta correcta. Al finalizar, verifica que tus respuestas coincidan con las que se te proporcionan en la página 271

1. ¿Cuál es el nombre que recibe el método estadístico que sigue la estrategia de seleccionar una parte representativa del universo a estudiar?

    a) Cuantitativo.

    b) Muestreo.

    c) Aleatorio.

[2]WILLOUGHBY, Stephen S; **Probabilidad y Estadística** México; Publicaciones Cultural S. A.; 1977; pp. 69

d) Probabilidad.

2. ¿Cómo se llama al conjunto total de elementos (individuos u objetos) que se desea conocer por medio de un tratamiento cuantitativo?

   a) N Población Objetivo.

   b) n Muestra Probabilística.

   c) P Proporción.

   d) Objetivo de Estudio.

3. ¿Cuál es el nombre que recibe la parte representativa del conjunto total que se conocer por medio de un tratamiento cuantitativo?

   a) N Población Objetivo.

   b) n Muestra Probabilística.

   c) P Proporción.

   d) Objetivo de estudio.

4. ¿Qué son los Estadígrafos o Estimadores?

   a) Medidas proporcionales al tamaño de la muestra.

   b) Los individuos que componen a una muestra.

   c) Las conclusiones generadas para una muestra.

   d) Datos característicos obtenidos del estudio de una muestra.

5. ¿Qué son los parámetros?

   a) Medidas proporcionales al tamaño de la población.

   b) El número total de individuos que componen a una población.

   c) Características que describen a la población cuantitativamente.
   d) Las conclusiones estadísticas generadas para una población objetivo.

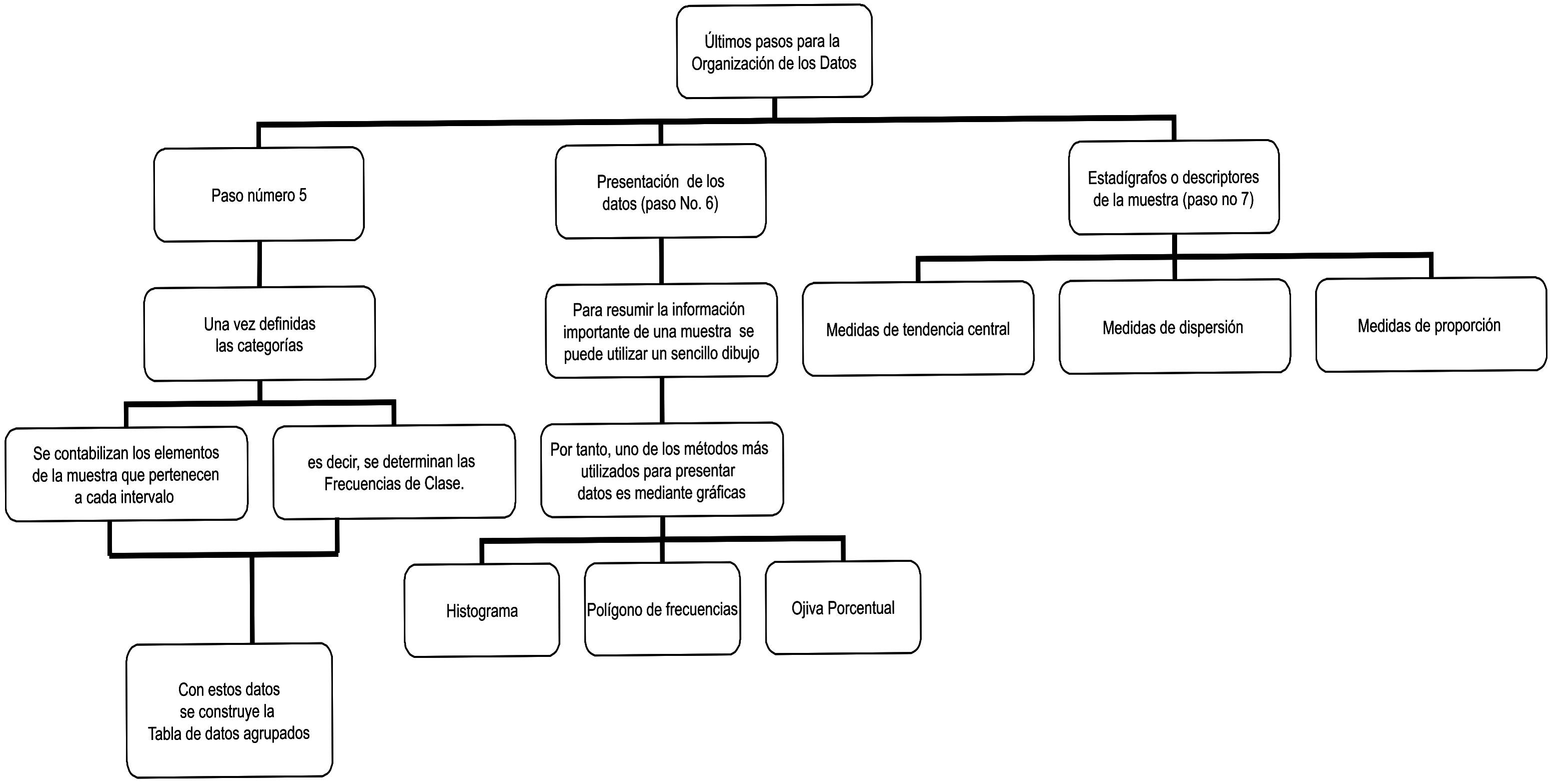
6. Anota los primeros tres pasos que se realizan en el proceso de organización de los datos de una muestra, esto es, luego de la obtención de una muestra aleatoria.

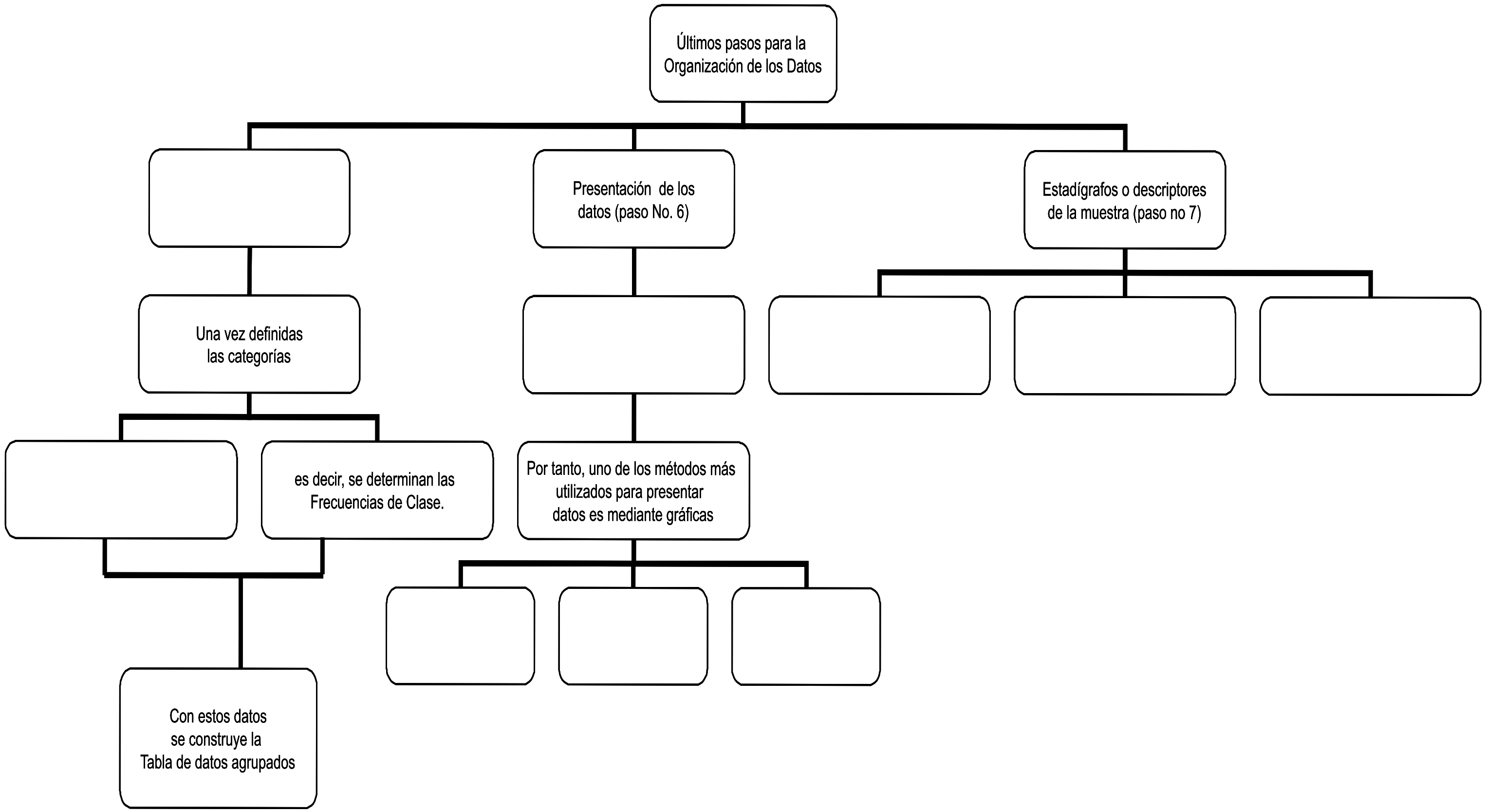
   1.- ____
   2.- ____
   ____
   3.- ____
   ____
   ____

7. Anota en cada línea, la palabra o palabras que completen de forma correcta la descripción del cuarto paso para organizar los datos de una muestra probabilística.

   4.- Definir el ____________ de las categorías, de tal manera que _____ los datos de la muestra se puedan concentrar en ellas; se debe cuidar que cada uno de los datos, se ubiquen en una categoría y ____________, en otras palabras: establecer el número de elementos que conformarán a cada subconjunto, de forma que ____________ (elemento) pertenezca a más de una categoría o ________________ de éstas. En este sentido, se dice que *las categorías deben ser* ________________. A las categorías se les denomina, más apropiadamente, *intervalos de clase*.

8. Lee con atención la siguiente red conceptual y con la información allí presentada, llena los espacios de la posterior estructura.


Últimos pasos para la Organización de los Datos
Paso número 5
Presentación de los datos (paso No. 6)
Estadígrafos o descriptores de la muestra (paso no 7)
Una vez definidas las categorías
Para resumir la información importante de una muestra se puede utilizar un sencillo dibujo
Medidas de tendencia central
Medidas de dispersión
Medidas de proporción
Se contabilizan los elementos de la muestra que pertenecen a cada intervalo
es decir, se determinan las Frecuencias de Clase.
Por tanto, uno de los métodos más utilizados para presentar datos es mediante gráficas
Histograma
Polígono de frecuencias
Ojiva Porcentual
Con estos datos se construye la Tabla de datos agrupados


Últimos pasos para la Organización de los Datos
Presentación de los datos (paso No. 6)
Estadígrafos o descriptores de la muestra (paso no 7)
Una vez definidas las categorías
es decir, se determinan las Frecuencias de Clase.
Por tanto, uno de los métodos más utilizados para presentar datos es mediante gráficas
Con estos datos se construye la Tabla de datos agrupados

3.3. Tabla de distribución de frecuencias

El proceso anteriormente descrito, conlleva a la construcción de la tabla de datos agrupados, que es una estructura matricial de seis columnas y, preferentemente, entre 5 y 20 filas que permite la presentación de los n datos de la muestra organizados por grupos. Los pasos para su construcción se presentarán, con detalles, a través del siguiente ejemplo:

- Se desea conocer el nivel de aprovechamiento escolar de los alumnos de estadística de cierta universidad; para tal efecto se diseña una muestra cuyo número de estudiantes (unidades) es de 80 individuos, y sus promedios son los siguientes:

8,8	8,0	8,8	6,6	6,0	7,7	6,1	6,1	6,2	7,8	6,2	7,5	6,3	7,9
6,5	7,3	7,6	5,9	6,7	6,7	6,8	6,8	6,8	6,9	7,1	7,1	7,1	7,2
7,2	6,5	7,3	7,3	7,3	7,4	7,4	7,8	7,5	9,3	7,5	6,2	7,5	7,5
7,5	7,6	7,6	6,5	7,6	6,0	9,5	7,4	7,8	6,2	8,8	7,8	7,9	6,3
7,9	5,4	8,1	8,2	9,9	8,3	8,4	8,5	9,6	8,5	8,6	8,7	9,0	5,1
7,8	8,9	5,7	9,3	8,2	9,4	7,7	9,5	8,5	7,5				

1. Ordenar los datos en forma ascendente, esto es: ubicarlos en orden con respecto a su valor relativo, donde el primero en la lista será el de menor valor y el último será el "mayor" de todos los datos, tal y como se encuentran en la siguiente tabla:
   Datos ordenados ascendentemente

5,1	5,4	5,7	5,9	6,0	6,0	6,1	6,1	6,2	6,2	6,2	6,2	6,3	6,3
6,5	6,5	6,5	6,6	6,7	6,7	6,8	6,8	6,8	6,9	7,1	7,1	7,1	7,2
7,2	7,3	7,3	7,3	7,3	7,4	7,4	7,4	7,5	7,5	7,5	7,5	7,5	7,5
7,5	7,6	7,6	7,6	7,6	7,7	7,7	7,8	7,8	7,8	7,8	7,8	7,9	7,9
7,9	8,0	8,1	8,2	8,2	8,3	8,4	8,5	8,5	8,5	8,6	8,7	8,8	8,8
8,8	8,9	9,0	9,3	9,3	9,4	9,5	9,5	9,6	9,9				

2. Cálculo de rango: Se realiza obteniendo la diferencia del valor máximo y el valor mínimo.
   - Valor máximo 9,9
   - Valor mínimo 5,1
   - $9{,}9 - 5{,}1 = 4{,}8$
   - Por tanto: $rango = 4{,}8$

3. Determinar el número k de categorías (intervalos de calase). En la práctica se recomienda la construcción de entre 5 y 20 intervalos de clase. Sin embargo, para determinar el número k de intervalos de clase en los que se organizarán los n datos de una muestra, puede emplearse, si se desea, una de las siguientes fórmulas:

   a) $k \leq 5 \cdot ln(n)$

   b) $k = 1 + 3{,}322 \cdot ln(n)$

   c) $k = \sqrt{n}$

   En este caso, lo más práctico para un fácil manejo de los datos es organizarlos en 10 intervalos de clase, puesto que los n datos de la muestra son 80. De esta manera no será necesario utilizar ninguna de las fórmulas anteriores.

4. Determinar el *ancho* (longitud) h de cada uno de los k intervalos de clase, así como los valores (límites inferior y superior) que los definen. Para ello se utiliza la fórmula:

$$\boxed{h = \frac{rango}{k}}$$

En otras palabras, *la longitud de las clases se calcula dividiendo el rango entre el número de clases* previamente determinadas. Para el presente ejemplo, se desean 10 clases ($k = 10$) y si el $rango = 4{,}8$ el ancho de cada intervalo de clase se calcula como sigue:

$$h = \tfrac{4{,}8}{10} = 0{,}48$$

Entonces, el ancho de cada uno de los 10 intervalos de clase es de 0,5 dado que $0{,}48 \approx 0{,}5$

a) Para construir el primer intervalo de clase, se toma como punto de partida el menor de los datos de la muestra (en este caso 5,1); dicho valor indica el punto en donde comienza el intervalo de clase.

b) Para determinar el punto en donde termina el intervalo, se suma, al punto inicial (5,1), el ancho del intervalo (0,5), con lo que se obtiene el valor 5,6

Ambos valores, conforman los *límites* del primer intervalo de clase. Es decir, el primer intervalo de clase (o subconjunto propuesto) estará determinado por los valores comprendidos entre:

$$5{,}1 \text{ ——— } 5{,}6$$

Donde el valor que se encuentran ubicado a la izquierda se conoce como *límite inferior de clase*, mientras que el situado en el extremo opuesto (derecho) es el *límite superior de clase*. Entonces, es claro que cada intervalo de clase tiene su límite inferior de clase y su límite superior de clase, es decir, cuenta con *límites de clase*.

Para el segundo intervalo de clase, se toma el límite superior de clase del primer intervalo (5,6) como límite inferior. Asimismo, a éste se le suma el ancho del intervalo, esto es: $5{,}6 + 0{,}5 = 6{,}1$ y este valor es el límite superior de clase del segundo intervalo, el cual queda expresado por los valores:

$$5{,}6 \text{ ——— } 6{,}1$$

Para el tercer intervalo, el procedimiento es igual. Se toma el límite superior del segundo intervalo (6,1) como límite inferior de clase y a éste, se le suma el ancho del intervalo ($6{,}1 + 0{,}5 = 6{,}6$) para determinar el límite superior de clase. El proceso se repite hasta completar las 10 categorías o intervalos de clase deseadas. Al final del proceso, se obtuvieron los siguientes intervalos de clase:

a) 5,1 ——— 5,6
b) 5,6 ——— 6,1
c) 6,1 ——— 6,6

d) 6,6 ——— 7,1
e) 7,1 ——— 7,6
f) 7,6 ——— 8,1

g) 8,1 ——— 8,6
h) 8,6 ——— 9,1
i) 9,1 ——— 9,6
j) 9,6 ———10,1

5. A continuación se determina la frecuencia, que es el número de observaciones de la muestra que caen en cada intervalo. Recuérdese que cada observación de la muestra puede ser ubicada en *uno y sólo un intervalo de clase.*

   Antes de proceder a la determinación de las frecuencias de clase, es necesario expresar los intervalos de clase (obtenidos en el paso anterior) en forma de *intervalos semiabiertos*, a saber, si x es una observación o dato de una muestra, y los valores a y b: son el límite inferior y superior de una clase o categoría, respectivamente, se dice que x pertenece a dicha clase si y sólo si x es mayor o igual que el número a y menor que el número b. Esto se expresa en símbolos de la siguiente manera:

$$\boxed{x \in [a, b) \Leftrightarrow a \leq x \text{ ó } x < b}$$

   En la fórmula, anterior, la notación $[a, b)$ significa que se trata de un intervalo que comprende a los valores iguales o mayores que a pero menores que b; a este intervalo se le conoce como intervalo semiabierto (cerrado por la izquierda y abierto por la derecha).

   Entonces, para la determinación de las frecuencias de clase, los intervalos de clase se expresan como intervalos semiabiertos, en donde el límite inferior de cada clase corresponde a la parte *cerrada* del intervalo, y el límite superior corresponde a la parte *abierta* del mismo. Para este caso, se tienen los siguientes intervalos:

   a) [5,1 – 5,6)

   b) [5,6 – 6,1)

   c) [6,1 – 6,6)

d) [6,6 – 7,1)

e) [7,1 – 7,6)

f) [7,6 – 8,1)

g) [8,1 – 8,6)

h) [8,6 – 9,1)

i) [9,1 – 9,6)

j) [9,6 – 10,1)

La conformación de los intervalos anteriores, garantiza que *ningún dato de la muestra pertenezca a más de un intervalo*, e igualmente, que *ningún dato de la muestra quede fuera de alguno de los intervalos obtenidos.*

Una vez definidos los intervalos de clase, se procede a determinar las frecuencias de clase, es decir, contar cuántos elementos de la muestra caen en cada intervalo. Como se verá en este ejemplo, es preciso revisar cada elemento de la muestra y comparar su valor con el de los límites de clase.

Cada elemento pertenecerá al intervalo en cuestión, *únicamente si su valor es igual o mayor que el expresado en el límite inferior de clase y menor que el del límite superior del intervalo.*

En lo que corresponde al primer intervalo de clase, únicamente los elementos que cumplan con la siguiente condición, pertenecerán a éste:

$$x \in [5{,}1 – 5{,}6) \Leftrightarrow 5{,}1 \leq x \text{ ó } x < 5{,}6$$

Ahora bien: se sabe que el primer dato de la muestra es 5,1; así que para determinar su pertenencia o exclusión al primer intervalo, habrá que sustituir la variable (x) por el dato en cuestión (5,1):

$$5{,}1 \in [5{,}1 - 5{,}6) \Leftrightarrow 5{,}1 \leq 5{,}1 \text{ ó } 5{,}1 < 5{,}6$$

Debido a que 5,1 es mayor o igual a 5,1 ($5{,}1 \leq 5{,}1$) y menor que 5,6 ($5{,}1 > 5{,}6$) dicho valor pertenece al primer intervalo de clase ($5{,}1 \in [5{,}1 - 5{,}6)$); en este momento, la frecuencia del intervalo es 1 (un valor que "ha caído" en este intervalo).

Para determinar la pertenencia del segundo elemento de la muestra, que es (5,4), el procedimiento será el mismo:

$$5{,}4 \in [5{,}1 - 5{,}6) \Leftrightarrow 5{,}1 \leq 5{,}4 \text{ ó } 5{,}4 < 5{,}6$$

Debido a que este valor es mayor o igual que 5,1 y menor que 5,6 también cae dentro del primer intervalo, por lo que su frecuencia es ahora 2 (dos valores han caído en este intervalo).

Ahora se revisará el tercer elemento de la muestra (5,7). Se observa que dicho valor cae en el segundo intervalo de clase ($[5{,}6 - 6{,}1)$), ya que 5,7 es mayor o igual que 5,6 y menor que 6,1, es decir, satisface la siguiente condición:

$$5{,}7 \in [5{,}6 - 6{,}1) \Leftrightarrow 5{,}6 \leq 5{,}7 \text{ ó } 5{,}7 < 6{,}1$$

Entonces la frecuencia del segundo intervalo de clase, en este momento, es 1. Este proceso se continúa hasta terminar con todos los elementos de la muestra. Enseguida se muestra la tabla con todas las frecuencias de clase:

i	Clase	f_i
1	[5,1 − 5,6)	2
2	[5,6 − 6,1)	4
3	[6,1 − 6,6)	11
4	[6,6 − 7,1)	7
5	[7,1 − 7,6)	19
6	[7,6 − 8,1)	15
7	[8,1 − 8,6)	8
8	[8,6 − 9,1)	7
9	[9,1 − 9,6)	5
10	[9,6 − 10,1)	2
		$\sum_{i=1}^{k} f_i = n$ $\sum_{i=1}^{10} f_i = 80$

En la tabla anterior, i es el número de intervalo, *clase* es el *i-ésimo intervalo* y f_i es la frecuencia o número de casos que caen en el intervalo número i. Se observa en la tabla que:

$f_1 = 2$

$f_2 = 4$

$f_3 = 11$

$f_4 = 7$

$f_5 = 19$

$f_6 = 15$

$f_7 = 8$

$f_8 = 7$

$f_9 = 5$

$f_{10} = 2$

De tal manera que, como se muestra en la parte inferior de la tabla:

$$\sum_{i=1}^{10} f_i = 80$$

O sea, la suma de las frecuencias f_i debe ser igual al número n de elementos que conforman la muestra, en este ejemplo 80.

El siguiente paso en la construcción de la tabla de datos agrupados, es calcular las *marcas de clase*, las cuales se representan con el símbolo x_i. Una marca de clase es *el punto medio de cada intervalo de clase*. Para determinar las marcas de clase, se suman el límite inferior de clase con el respectivo límite superior y el resultado obtenido se divide entre dos, esto es:

$$\boxed{x_i = \frac{lim_{inf}+lim_{sup}}{2}}$$

En donde x_i es la marca de clase o punto medo del i-ésimo intervalo de clase; lim_{inf} es límite inferior de clase del i-ésimo intervalo de clase y lim_{sup} es el límite superior de clase del i-ésimo intervalo de clase. Para este caso, la primera marca de clase viene dada por:

$x_1 = \frac{5{,}1+5{,}6}{2}$
$= \frac{10{,}7}{2}$
$= 5{,}35$

En este orden de ideas, la segunda marca de clase se calcula en forma similar, esto es:

$x_2 = \frac{5{,}6+6{,}1}{2}$
$= \frac{11{,}7}{2}$
$= 5{,}85$

Este proceso se repite con todos los intervalos, tal como se muestra en la siguiente tabla:

i	Clase	f_i	x_i
1	[5,1 – 5,6)	2	5,35
2	[5,6 – 6,1)	4	5,85
3	[6,1 – 6,6)	11	6,35
4	[6,6 – 7,1)	7	6,85
5	[7,1 – 7,6)	19	7,35
6	[7,6 – 8,1)	15	7,85
7	[8,1 – 8,6)	8	8,35
8	[8,6 – 9,1)	7	8,85
9	[9,1 – 9,6)	5	9,35
10	[9,6 – 10,1)	2	9,85

Una vez hecho lo anterior, se continúa con la siguiente etapa en la construcción de la tabla de datos agrupados. Dicha etapa consiste en calcular la *frecuencia relativa* fr_i, la cual se obtiene al dividir cada una de las frecuencias f_i entre el número total de sujetos, que en este caso son ochenta. Para cerciorarse de que las frecuencias relativas sean correctas, la suma de todas ellas debe dar como resultado 1, que corresponde al 100 %.

Así, para este ejemplo, la primera frecuencia relativa se obtiene por:

$$fr_1 = \frac{2}{80} = 0{,}025$$

Consecuentemente, la segunda frecuencia relativa viene dada por:

$$fr_2 = \frac{4}{80} = 0{,}05$$

Este proceso continúa para las restantes ocho frecuencias relativas, que se muestran en la siguiente tabla:

i	Clase	f_i	x_i	fr_i
1	$[5{,}1 - 5{,}6)$	2	5,35	0,025
2	$[5{,}6 - 6{,}1)$	4	5,85	0,05
3	$[6{,}1 - 6{,}6)$	11	6,35	0,1375
4	$[6{,}6 - 7{,}1)$	7	6,85	0,0875
5	$[7{,}1 - 7{,}6)$	19	7,35	0,2375
6	$[7{,}6 - 8{,}1)$	15	7,85	0,1875
7	$[8{,}1 - 8{,}6)$	8	8,35	0,1
8	$[8{,}6 - 9{,}1)$	7	8,85	0,0875
9	$[9{,}1 - 9{,}6)$	5	9,35	0,0625
10	$[9{,}6 - 10{,}1)$	2	9,85	0,025
				$\sum_{i=1}^{k} fr_i = 1$ $\sum_{i=1}^{10} fr_i = 1$

El último elemento a determinar para la construcción de una tabla de datos agrupados, lo constituye el cálculo de las *frecuencias relativas acumuladas* fra_i. La *i-ésima frecuencia relativa acumulada*, o frecuencia relativa acumulada del *i-ésimo intervalo de clase*, se calcula sumando las frecuencias relativas de todos los intervalos de clase anteriores y la del intervalo de clase actual.

En el presente ejemplo, la primera frecuencia relativa acumulada se calcula por:

$fra_1 = fr_1$
$fra_1 = 0{,}025$

ya que no hay intervalos de clase antes del primero. La segunda frecuencia relativa acumulada se calcula por:

$fra_2 = fr_1 + fr_2$
$= 0{,}025 + 0{,}05$
$= 0{,}075$

La tercera frecuencia relativa acumulada se calcula por:

$fra_3 = fr_1 + fr_2 + fr_3$
$= 0{,}025 + 0{,}05 + 0{,}1375$
$= 0{,}2125$

Este proceso se continúa para todos los intervalos de clase restantes. Es importante observar que la última frecuencia relativa acumulada siempre es 1. Las frecuencias relativas acumuladas se muestran en la siguiente tabla:

i	Clase	f_i	x_i	fr_i	fra_i
1	[5,1 – 5,6)	2	5,35	0,025	0,025
2	[5,6 – 6,1)	4	5,85	0,05	0,075
3	[6,1 – 6,6)	11	6,35	0,1375	0,2125
4	[6,6 – 7,1)	7	6,85	0,0875	0,300
5	[7,1 – 7,6)	19	7,35	0,2375	0,5375
6	[7,6 – 8,1)	15	7,85	0,1875	0,725
7	[8,1 – 8,6)	8	8,35	0,1	0,825
8	[8,6 – 9,1)	7	8,85	0,0875	0,9125
9	[9,1 – 9,6)	5	9,35	0,0625	0,9750
10	[9,6 – 10,1)	2	9,85	0,025	1,0000

A continuación se presenta la tabla completa de datos agrupados, obtenida en los pasos anteriores:

i	Clase	f_i	x_i	fr_i	fra_i
1	[5,1 – 5,6)	2	5,35	0,025	0,025
2	[5,6 – 6,1)	4	5,85	0,05	0,075
3	[6,1 – 6,6)	11	6,35	0,1375	0,2125
4	[6,6 – 7,1)	7	6,85	0,0875	0,300
5	[7,1 – 7,6)	19	7,35	0,2375	0,5375
6	[7,6 – 8,1)	15	7,85	0,1875	0,725
7	[8,1 – 8,6)	8	8,35	0,1	0,825
8	[8,6 – 9,1)	7	8,85	0,0875	0,9125
9	[9,1 – 9,6)	5	9,35	0,0625	0,9750
10	[9,6 –10,1)	2	9,85	0,025	1,0000
		$\sum_{i=1}^{k} f_i = n$ $\sum_{i=1}^{10} f_i = 80$		$\sum_{i=1}^{k} fr_i = 1$ $\sum_{i=1}^{10} fr_i = 1$	

La tabla anterior de datos agrupados también se conoce como *tabla de distribución de frecuencias.*

Esta tabla es muy importante, ya que a partir de ella, se construirán las gráficas para describir la muestra y, como se verá en las siguientes secciones, con sus datos se calcularán los estadígrafos de la muestra o descriptores (medidas de tendencia central y medidas de dispersión).

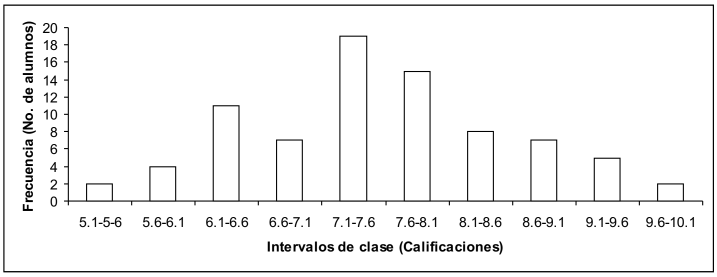
3.4. Gráficas para describir la muestra

6. Construcción de las gráficas para describir las muestras (Histograma, Polígono de Frecuencias, Ojiva Porcentual).

Un *histograma* es una gráfica de barras. Se conforma a partir de los intervalos de clase versus su respectiva frecuencia. Estos valores se toman de la tabla de datos agrupados. En el eje de las variables independientes (eje X) se colocan los intervalos de clase, y en el eje de las variables dependientes se ubican las respectivas frecuencias de los intervalos de clase. Tomando estos de la anterior tabla de distribución de frecuencias:

Clase	f_i
[5,1 − 5,6)	2
[5,6 − 6,1)	4
[6,1 − 6,6)	11
[6,6 − 7,1)	7
[7,1 − 7,6)	19
[7,6 − 8,1)	15
[8,1 − 8,6)	8
[8,6 − 9,1)	7
[9,1 − 9,6)	5
[9,6 − 10,1)	2

Se procede a expresarlos en un histograma:



Cabe destacar que por definición: "un *histograma* o *histograma de frecuencias* consiste en una serie de rectángulos que tienen

a) Sus bases sobre un eje horizontal (el eje de X) con centros en las marcas de clase y longitud igual al intervalo de la marca de clase.

b) Superficies proporcionales a las frecuencias de clase.

"Si los intervalos de clase tienen todos igual tamaño -como en este caso-, las alturas de los rectángulos son proporcionales a las frecuencias de clase y se acostumbra en tal caso a tomar las alturas numéricamente iguales a las frecuencias de clase. "[3]

Cualquier variación de estos aspectos, puede interferir en la correcta interpretación de las gráficas.

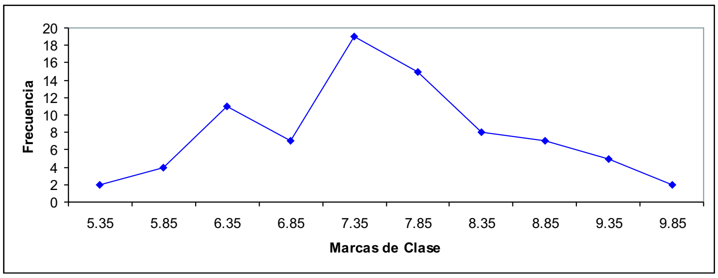
El *Polígono de Frecuencias*, es una gráfica de puntos unidos por segmentos de recta y se genera a partir de las *marcas de clase contra sus frecuencias*. En el eje de las variables independientes se ubican las marcas de clase y en el eje de las variables dependientes, se colocan las f_i.

Al igual que en el caso del histograma, sus valores se toman de la tabla de datos agrupados:

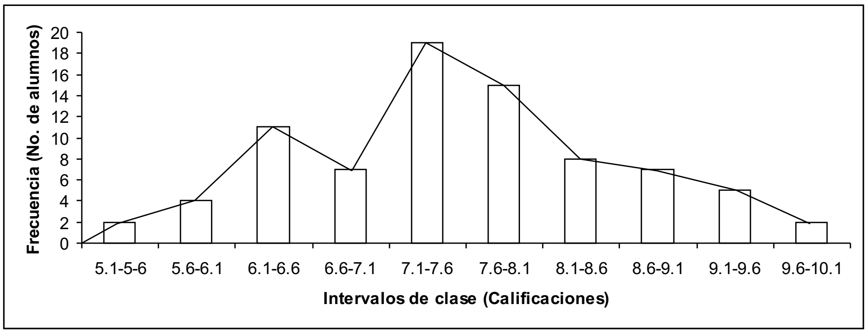
[3]op. cit. SPIEGEL, Murray R.; México; 1970; pp. 29

x_i	f_i
5,35	2
5,85	4
6,35	11
6,85	7
7,35	19
7,85	15
8,35	8
8,85	7
9,35	5
9,85	2

Con los datos anteriores se obtiene el siguiente polígono de frecuencias:

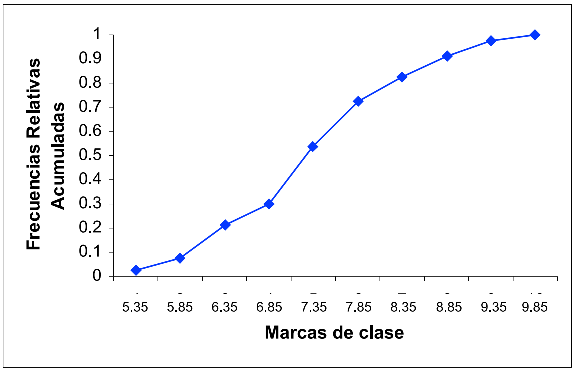


Es común combinar en una sola gráfica el histograma y el polígono de frecuencias; para ello, se acostumbra prolongar el polígono con segmentos de recta que van de la primera y última marca de clase al eje de las abscisas, agregando las marcas de clase que corresponderían a la frecuencia cero (en el presente ejemplo, serían 4,85 y 10,85 para la primera y última marca de clase respectivamente).



"En tal caso, la suma de las áreas de los rectángulos del histograma es igual al área total limitada por el polígono de frecuencias y el eje X"[4]

El *polígono de frecuencia acumulada* u *ojiva porcentual* se genera a partir de las marcas de clase contra las frecuencias relativas acumuladas. Esta también es una gráfica de puntos unidos por segmentos de rectas, que se construye al tomar esos datos de la tabla de distribución de frecuencias:



7. El cálculo de los descriptores de la muestra se verá, como se dijo con anterioridad, en los capítulos 5 y 6.

[4]op. cit. SPIEGEL, Murray R.; México; 1970; pp. 29

Ejercicios 2

1. ¿Qué representa la literal k en el proceso de organización de los n datos de una muestra?

   ____

2. Escribe una de las tres fórmulas (la que sea) con que se puede calcular el número k de categorías en que se pueden ordenar los n datos de una muestra.

   ____

3. Anota la fórmula para determinar el ancho h de los intervalos de clase:

   ____

4. En síntesis ¿Cómo se determinan los intervalos de clase?

   ____

   ____

   ____

   ____

   ____

   ____

   ____

   ____

5. ¿Qué condición formal se debe cumplir para dar paso a la correcta determinación de las frecuencias de clase?

   ____

   ____

   ____

6. ¿Qué indica la expresión $x \in [a, b) \Leftrightarrow a \leq x$ ó $x < b$ al denotar los intervalos de clase como intervalos semiabiertos?

    ____

    ____

    ____

    ____

    ____

    ____

    ____

    ____

    ____

    ____

7. ¿De qué otra forma se le denomina al intervalo de clase (o categorías) de la muestra en la tabla de datos agrupados?

    ____

8. Una vez organizados en las k categorías los n datos de la muestra, ¿cuál deberá ser el resultado de la suma de las frecuencias? Anótalo en lenguaje matemático:

    ____

9. Anota el significado de cada elemento que interviene en la siguiente fórmula: $x_i = \frac{lim_{inf} + lim_{sup}}{2}$

    ____

    ____

    ____

    ____

10. Anota la fórmula para calcular fr_i

    ____

11. Nombre de la gráfica de barras que se conforma a partir de los intervalos de clase y sus respectivas frecuencias.

    a) Histograma.

    b) Polígono de frecuencias.

    c) Ojiva Porcentual.

12. ¿En qué eje del plano cartesiano se ubican los valores correspondientes a las variables independientes?

    ----

13. ¿ Cuales son los valores que se toman de la tabla de datos agrupados para construir el polígono de frecuencias?

Las respuestas a los ejercicios que acabas de resolver se encuentran en la página 273, compáralos y en caso de cualquier error, revisa el procedimiento, los cálculos y corrige tu respuesta.

3.5. Autoevaluación

- Instrucciones: Con la información que se te proporcionará, realiza los procedimientos y cálculos necesarios para ordenar y presentar los n datos de la muestra que se propone; posterirmente revisa que tus respuestas sean las correctas (sección **3.6.1. Respuestas de la autoevaluación**, página 276):

Se desea estimar el gasto promedio semanal, por concepto de pasajes y comida, de los estudiantes preparatorianos en una colonia popular del Distrito Federal. El tamaño N de la población es de 857 individuos que cursan el nivel medio superior; mediante un proceso aleatorio, se obtuvo una tamaño

n de muestra igual a 40 individuos, y los datos aportados por éstos, son los siguientes:

$119, 135, 138, 144, 146, 150, 156, 164, 132, 138, 142, 146, 149, 154, 163, 176$
$125, 135, 140, 144, 147, 150, 157, 165, 128, 136, 142, 145, 148, 153, 161, 175$
$126, 135, 140, 145, 147, 152, 158, 168,$

Con esta información, construye una tabla de distribución de frecuencias y elabora las gráficas correspondientes:

.

.

.

3.6. Apéndice del capítulo

3.6.1. Información remedial de la evaluación diagnóstica

Términos de enlace (preguntas 1-3)

Las *palabras* que sirven para formar *proposiciones compuestas* (moleculares) mediante la unión de *proposiciones simples* (atómicas) o a partir de su simple adición (como es el *caso específico de la negación*), se llaman términos de enlace o conectivos lógicos. Gran parte de lo que se trata en el estudio de la lógica matemática, se refiere a la manera cuidadosa de cómo se han de usar estas palabras. Su nombre correcto es *términos de enlace de proposiciones*, e indica cuál es papel que desempeñan: enlazan proposiciones.

Con frecuencia, dichos términos, se conocen por nombres diferentes en otras disciplinas. Pero basta con verlos, para entender el sentido de su definición; las palabras "*y*", "*o*", "*si... entonces...*" se aplican en la unión de proposiciones atómicas para formar proposiciones moleculares. Un peculiar término de enlace es la palabra "no"; ésta se caracteriza por su efecto en las proposiciones atómicas, pues con su "simple" adición, las transforma en una proporción molecular.[5] Ejemplos:

Considera las proposiciones atómicas que se presentan.

- Llueve de nuevo
- El viento cambia de dirección
- La visibilidad es nula

[5] SUPPES, Patrick; et. al; **Introducción a la Lógica Matemática**; México; Editorial Reverté; 1997; pp. 106

Si se usan los términos de enlace anteriormente señalados, se pueden conformar las subsiguientes proposiciones moleculares:

- Llueve de nuevo *y* la visibilidad es nula
- *Si* el viento cambia de dirección, *entonces* llueve de nuevo
- La visibilidad *no* es nula

De lo anterior se puede observar que el término de enlace "*no*" es el único que al actuar sobre una proposición atómica forma una molecular y el resto de ellos, lo hacen sobre dos. También, es posible notar que la condicional establece, como su nombre lo sugiere, una condición.

En la proposición molecular "*Si* el viento cambia de dirección, *entonces* llueve de nuevo", a la primera proposición atómica se le denomina *antecedente* y a segunda *consecuente*; luego entonces, la condición para que ocurra que llueva de nuevo, es que el viento cambie de dirección.

Cada término de enlace puede ser denotado mediante un símbolo y de acuerdo con el término que sea empleado, al formar una proposición molecular, también se dará un nombre a ésta.

Término de Enlace	Símbolo	Nombre de la Porposición
y	$\wedge$	Conjunción
o	$\vee$	Disjunción
si...entonces...	$\Rightarrow$	Condicional
no	$\neg$	Negación

Hay otro término de enlace, éste es *Si y sólo si.* Las proposiciones que lo utilizan se denominan proposiciones bicondicionales y el símbolo que lo denota es $\Leftrightarrow$. "Una proposición bicondicional se parece extraordinariamente a dos proposiciones condicionales, para ilustrar esto considere el siguiente ejemplo en lenguaje habitual:

"Estos campos se inundan *si y sólo si* el agua alcanza esta altura.

"En forma simbólica la proposición sería:

$$\mathrm{P}\Leftrightarrow\mathrm{Q}$$

"donde **P** es el símbolo de la primera proposición y **Q** es el símbolo de la última proposición. Se puede leer esta proposición: **P** si y sólo si **Q**.

"La proposición bicondicional P$\Leftrightarrow$Q tiene la misma fuerza que dos proposiciones condicionales; primera P $\Rightarrow$ Q y segunda, Q $\Rightarrow$ P. La proposición en castellano significa que si el agua alcanza cierta altura, entonces el campo se inunda. También significa que si el campo se inunda, entonces el agua alcanza cierta altura."[6]

En tal caso, ambas proposiciones cumplen una doble función en la proposición molecular. Son, al mismo tiempo, antecedente y consecuente; de tal manera que para determinar su valor de Verdad, la regla es la siguiente: Para que la proposición bicondicional sea verdadera, el valor de verdad en las proposiciones atómicas debe ser igual (V y V o F y F), en caso contrario, su valor de Verdad será falso.

[6]Ibidem SUPPES, Patrick; et. al; México; 1997; pp. 106

Relación de orden $\leq$, $\geq$ (pregunta 4)

Precisión y/o exactitud, dos palabras con las que se pueden calificar-describir el potencial y la utilidad de la Matemática como ciencia en sí misma y como herramienta científica, esto es: una ciencia al servicio de otras ciencias y disciplinas. Dichas características se deben, entre otras razones, al desarrollo del lenguaje matemático -de forma lógica- que permite expresar con claridad y sin las **ambigüedades** propias del lenguaje natural lo que se desea comunicar.

El lenguaje puro de la matemática, está constituido por símbolos que permiten enunciar definiciones y propiedades con toda exactitud y un mínimo de símbolos. Éste, excluye el uso de las letras en un sentido convencional y las adscribe a él, sin la intencionalidad de formar palabras, de tal manera que un solo símbolo (sea letra o no), **connota** una idea. Buena parte de estos símbolos se encuentran contenidos en la tabla de la página 102.

Uno de estos símbolos es el que expresa la idea de relación entre dos números, donde uno de ellos puede ser *mayor o igual que* el otro, este es $\geq$. Dicho símbolo apararece principalmente en proposiciones abiertas, es decir, en ciertas expresiones que contienen una variable que al ser sustituida por un valor determinado, la expresión puede adquirir valores de verdad, es decir, ésta puede ser una proposición Verdadera o Falsa; ejemplo:

Dados los números q, C relacionados así: $q \geq C$, el valor relativo de q debe ser mayor o igual que el valor relativo de C para que la proposición sea verdadera; dicho lo anterior, verificar el valor de verdad de las ulteriores proposición al sustituir las literales; donde C es una Constante cuyo valor es 15 y q es la variable que para este ejemplo adquiere los siguientes valores: 14, 15, 16, 17; luego entonces:

- $q \geq C$ es Falsa cuando $14 \geq 15$
- $q \geq C$ es Verdadera cuando $15 \geq 15$
- $q \geq C$ es Verdadera cuando $16 \geq 15$
- $q \geq C$ es Verdadera cuando $17 \geq 15$

Ahora bien, si se invierte el sentido del signo de la relación ($q \leq C$) y se conservan las mismas cantidades, el valor de verdad es el siguiente para cada proposición:

- $q \leq C$ es Verdadera cuando $14 \leq 15$
- $q \leq C$ es Verdadera cuando $15 \leq 15$
- $q \leq C$ es Falsa cuando $16 \leq 15$
- $q \leq C$ es Falsa cuando $17 \leq 15$

3.6.2. Respuestas a los ejercicios del capítulo

Ejercicios 1

- Instrucciones. Selecciona con una X el inciso que contenga la respuesta correcta.

1. ¿Cuál es el nombre que recibe el método estadístico que sigue la estrategia de seleccionar una parte representativa del universo a estudiar?
   b) Muestreo.
2. ¿Cómo se llama al conjunto total de elementos (individuos u objetos) que se desea conocer por medio de un tratamiento cuantitativo?
   a) N Población Objetivo.

3. ¿Cuál es el nombre que recibe la parte representativa del conjunto total que se conocer por medio de un tratamiento cuantitativo?
   b) n Muestra Probabilística.

4. ¿Qué son los Estadígrafos o Estimadores?
   d) Datos característicos obtenidos del estudio de una muestra.

5. ¿Qué son los parámetros?
   c) Características que describen a la población cuantitativamente.

6. Anota los primeros tres pasos que se realizan en el proceso de organización de los datos de una muestra, esto es, luego de la obtención de una muestra aleatoria.

   1.- Ordenar los datos en forma ascendente.

   2.- Determinar el *rango* de los datos. Para ello, hay que restar el valor del dato menor al del dato mayor de la muestra.

   3.- Determinar el número k de *categorías* en que se organizarán los n datos de la muestra, esto es proponer x número de subconjuntos en los que se puedan distribuir los n elementos de la muestra.

7. Anota en cada línea, la palabra o palabras que completen de forma correcta la descripción del cuarto paso para organizar los n datos de una muestra probabilística.

   4.- Definir el **ANCHO O LONGITUD** de las categorías, de tal manera que **TODOS** los datos de la muestra se puedan concentrar en ellas;

se debe cuidar que cada uno de los datos, se ubiquen en una categoría y **SÓLO EN UNA**, en otras palabras: establecer el número de elementos que conformarán a cada subconjunto, de forma que **NINGÚN DATO** (elemento) pertenezca a más de una categoría o **QUEDE FUERA** de éstas. En este sentido, se dice que *las categorías deben ser* **MUTUAMENTE EXCLUYENTES**. A las categorías se les denomina, más apropiadamente, *intervalos de clase.*

8. Lee con atención la siguiente red conceptual y con la información allí presentada, llena los espacios de la posterior estructura.
   La respuesta a este reactivo, se puede verificar con la comparación de las dos redes conceptuales.

Ejercicios 2

1. ¿Qué representa la literal k en el proceso de organización de los n datos de una muestra?
   Las Categorías en que se organizarán los n datos de la muestra

2. Escribe una de las tres fórmulas (la que sea) con que se puede calcular el número k de categorías en que se pueden ordenar los n datos de una muestra.

   $k \leq 5 \cdot ln(n)$
   $k = 1 + 3{,}322 \cdot ln(n)$
   $k = \sqrt{n}$

3. Anota la fórmula para determinar el ancho h de los intervalos de clase:

   $h = \frac{rango}{k}$

4. En síntesis ¿Cómo se determinan los intervalos de clase?

   Se toma como límite inferior de clase al dato cuyo valor sea el menor y se le suma el ancho h del intervalo de clase; el resultado indicará dónde termina el intervalo de clase (el límite superior de clase). El siguiente intervalo se calcula tomando como límite inferior de clase al límite superior anteriormente determinado y a este valor se le suma el ancho h de intervalo para determinar el límite superior de clase; este proceso se repite hasta completar el número k de clases establecido.

5. ¿Qué condición formal se debe cumplir para dar paso a la correcta determinación de las frecuencias de clase?
   Es preciso expresar los intervalos de clase en forma de intervalos semiabiertos, es decir, abiertos por la derecha y cerrados por la izquierda.

6. ¿Qué indica la expresión $x \in [a, b) \Leftrightarrow a \leq x$ ó $x < b$ al denotar los intervalos de clase como intervalos semiabiertos?
   Qué *únicamente* los datos cuyos valores sean mayores o iguales que a (límite inferior del intervalo) o menores que b (límite superior de clase mayor que a) pertenecerán al intervalo en cuestión; es decir, si se diera el caso de que $a = 35{,}5$ y $b = 36{,}5$ ([35,5 – 36,5)) y el dato a ordenar fuere 35,49 éste no pertenece al presente intervalo de clase, puesto que no es igual, ni es mayor que 35,5, de igual forma, si el dato a ordenar fuese 36,5 no pertenecería a dicho intervalo puesto que el límite superior es cerrado (no admite valores iguales a sí, a diferencia de el límite inferior que es abierto).

7. ¿De qué otra forma se le denomina al intervalo de clase (o categorías) de la muestra en la tabla de datos agrupados?
   i-ésimo intervalo

8. Una vez organizados en las k categorías los n datos de la muestra, ¿cuál deberá ser el resultado de la suma de las frecuencias? Anótalo en lenguaje matemático:
   $\sum_{i=1}^{k} f_i = n$

9. Anota el significado de cada elemento que interviene en la siguiente fórmula $x_i = \frac{lim_{inf}+lim_{sup}}{2}$
   x_i Marca de clase del i-ésimo intervalo de clase (cociente de la fórmula).
   lim_{inf} Límite inferior del i-ésimo intervalo de clase.
   lim_{sup} Límite superior del i-ésimo intervalo de clase.
   2 Denominador de la fórmula; número entre el cual será dividido el resultado de la suma de los límites (numerador).

10. Anota la fórmula para calcular fr_i
    $fr_i = \frac{f_i}{n}$

11. Nombre de la gráfica de barras que se conforma a partir de los intervalos de clase y sus respectivas frecuencias.

    a) Histograma.

12. ¿En qué eje del plano cartesiano se ubican los valores correspondientes a las variables independientes?
    En el eje de las abscisas, es decir en X

13. ¿Cuáles son los valores que se toman de la tabla de datos agrupados para construir el polígono de frecuencias?
    x_i o marcas de clases versus las f_i que son las frecuencias de clase.

3.6.3. Respuestas de la autoevaluación

Se desea estimar el peso promedio de los alumnos de la carrera de comunicación; de un total de 857 estudiantes, se obtiene aleatoriamente una muestra de 40 alumnos.

Construir una tabla de datos agrupados y el histograma junto con los respectivos polígonos de frecuencias.

1. Datos ordenados en forma ascendente.

119	125	126	128	132	135	135	135	136	138
138	140	140	142	142	144	144	145	145	146
146	147	147	148	149	150	150	152	153	154
156	157	158	161	163	164	165	168	175	176

2. Cálculo de rango:
   Valor máximo: 176
   Valor mínimo: 119
   $176 - 119 = 57$
   Por tanto: $rango = 57$

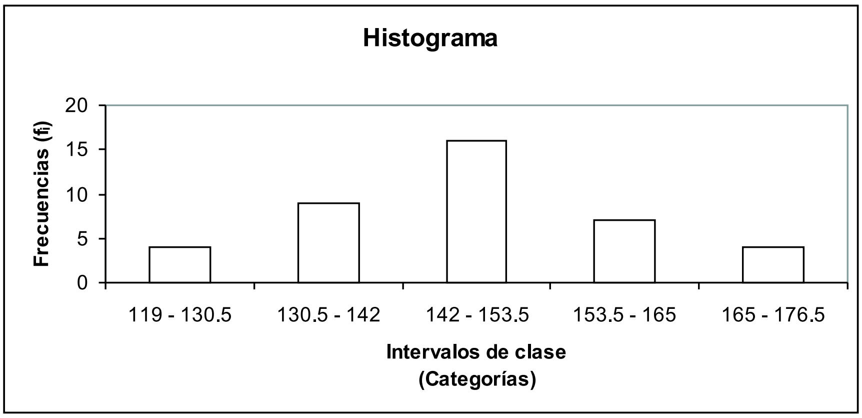
3. Número k de categorías o clases deseadas: Para este caso se optó por el número 5, puesto que es submúltiplo de 40; por tanto:
   $k = 5$

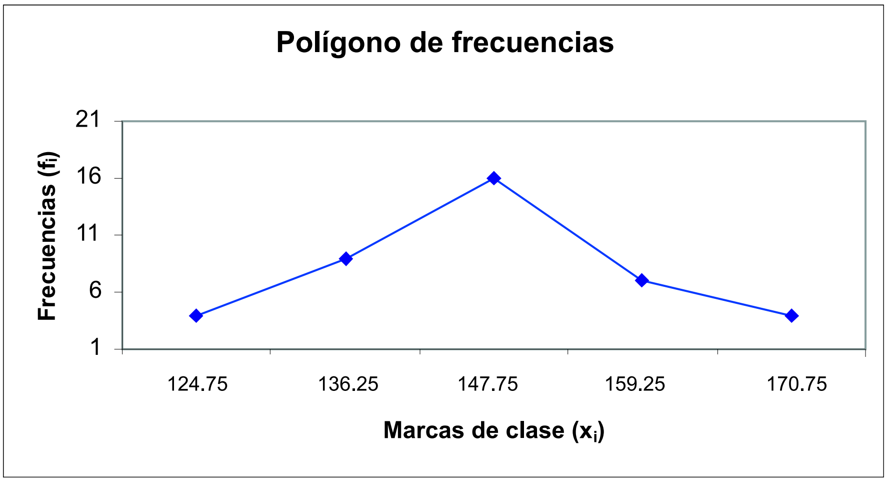
4. Ancho de clase: $h = \frac{rango}{k}$
   Sustitución: $\frac{57}{5} = 11{,}4$
   $11{,}4 \approx 11{,}5$
   Por tanto: $h = 11{,}5$

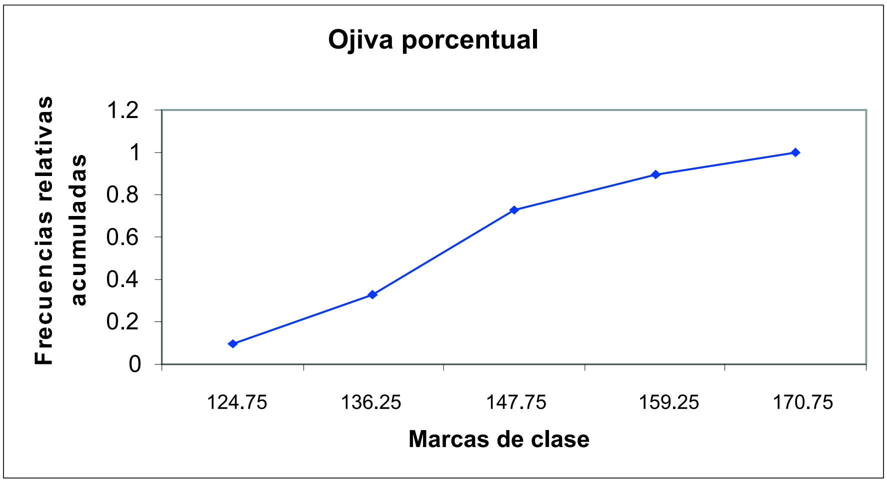
5. Cálculo de frecuencias y construcción de la tabla de datos agrupados.

i	Clase	f_i	x_i	fr_i	fra_i	n
1	$[119 - 130{,}5)$	4	124,75	0,1	0,1	40
2	$[130{,}5 - 142)$	9	136,25	0,225	0,325	
3	$[142 - 153{,}5)$	16	147,75	0,4	0,725	
4	$[153{,}5 - 165)$	7	159,25	0,175	0,9	
5	$[165 - 176{,}5)$	4	170,75	0,1	1,0	
$k = 5$	$h = 11{,}5$	$\sum_{i=1}^{k} f_i = n$ $\sum_{i=1}^{5} f_i = 40$		$\sum_{i=1}^{k} fr_i = 1$ $\sum_{i=1}^{5} fr_i = 1$		

6. Presentación de los datos mediante las gráficas correspondientes:




Polígono de frecuencias
21
16
11
6
1
Frecuencias (fi)
124.75
136.25
147.75
159.25
170.75
Marcas de clase (xi)


Ojiva porcentual
1.2
1
0.8
0.6
0.4
0.2
0
Frecuencias relativas acumuladas
124.75
136.25
147.75
159.25
170.75
Marcas de clase

3.6.4. Glosario

Ambigüedad(es): Término que designa la falta de presición en las expresiones verbales, como producto de la polisemia propia del lenguaje, a causa de los contextos y posibilidades de interpretación.

Connotar: "(De con- y notar). tr. Ling. Dicho de una palabra: Conllevar, además de su significado propio o específico, otro de tipo expresivo o apelativo."[7]

Deducción: "(lat. *deduco*, sacar de) género de razonamiento que pasa por necesidad de principios generales a las cosas particulares"[8] Es el proceso del pensamiento que, mediante el uso de las reglas de inferencia lógica aplicadas a las premisas que conforman el razonamiento, permite obtener una conclusión. "El paso lógico de las premisas a la conclusión es una deducción"[9]

Disciplinas: Para el del presente trabajo, el término *disciplina* se utiliza como sinónimo de *Ciencia*

Intervalos: Dado un conjunto de valores X tal que sus elementos pueden ser representados mediante el uso de los números reales ($X \in \Re \rightarrow x \in \Re$), se denominará intervalo a cualquier subconjunto de X que no sea $\emptyset$ y que contenga más de un elemento.

Semejante: En geometría, se dice que una figura es *semejante* a otra, debido a que tienen la misma forma y diferente tamaño. Análogamente, es

[7]Biblioteca de Consulta Microsoft Encarta 2005. © 1993-2004 Microsoft Corporation.

[8]op. cit. XIRAU, Ramón; México; Universidad Nacional Autónoma de México; 1995; pp. 457

[9]Op. cit. SUPPES, Patrick; et. al; México; 1997; pp. 44

que se utiliza esta palabra para designar el parecido *ideal* de una muestra (probabilística) con respecto a la población de que fue extraída, sin omitir el hecho de que el diseño de tal *modelo* (la muestra) se basa en la teoría de la probabilidad y por ende se encontrarán algunas diferencias (errores) previamente calculadas.

Símbolo: "Se llama símbolo a un signo sin semejanza ni contigüidad, sino solamente con un vínculo convencional entre su significante y su denotado."[10]

[10]op. cit. SEBEOK, Thomas A.; México; Editorial Paidós; 1996; pp. 49

Capítulo 4

Medidas de tendencia central

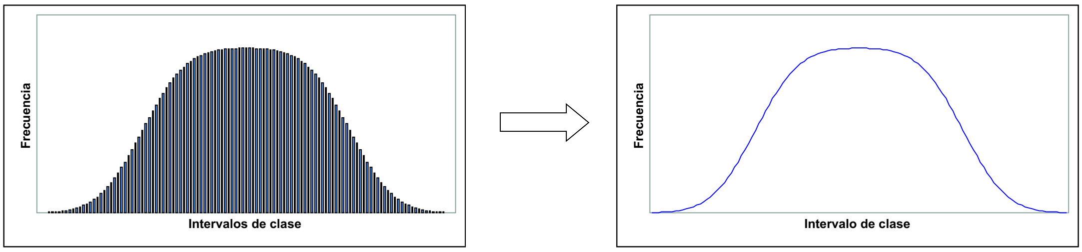
Objetivos específicos

- El estudiante será capaz de calcular las principales *medidas de tendencia central* ($\bar{x}$, x_{med} y x_{mod}) para datos agrupados y no agrupados, mediante la aplicación de sus respectivas fórmulas, así como de la regla empírica, para el análisis de un conjunto de datos.

- El estudiante será capaz de calcular los *cuantiles* (Q_n, D_n *y* P_n) mediante *interpolación lineal*, para la división del conjunto de datos y facilitación de su análisis.

4.1. Preliminar

Al emplear tamaños de muestra cada vez mayores, un polígono de frecuencia tiende a suavizar sus contornos de manera que toma la forma de una curva lisa. Es decir, cuando $n \to N$ (el tamaño n de la muestra tiende al tamaño N de la población) un polígono de frecuencia se transforma en una curva, tal como muestra la siguiente figura:



Como se revisó en el capítulo anterior, las frecuencias de un conjunto de datos muestrales, pueden organizarse en una tabla denominada tabla de datos agrupados. Ésta, puede expresarse visualmente a través de un histograma, un polígono de frecuencias y una ojiva. Los elementos anteriores muestran el comportamiento o distribución de las frecuencias de los datos. Adicionalmente a estos recursos tabulares y gráficos, se han desarrollado ciertas medidas para analizar a un conjunto de datos muestrales.

De estas medidas, destacan aquellas que *cuantifican la concentración de los datos respecto a cierto valor*. A ellas, se conocen con el nombre de *medidas de tendencia central*. Entre las principales medidas de tendencia central, se encuentran *la media* ($\bar{x}$), *la moda* (x_{mod}) *y la mediana* (x_{med}). Éstas, pueden calcularse tanto para conjuntos de datos agrupados como para conjuntos de datos no agrupados.

4.2. Medidas de tendencia central I

4.2.1. La media aritmética ($\bar{x}$) para datos no agrupados

La media ($\bar{x}$) o promedio, es una medida de tendencia central que permite encontrar el punto de equilibrio de un conjunto de datos. Mediante su estimación, se busca obtener un valor común que describa el comportamiento de la muestra en cuestión, ejemplos:

- El promedio de ingresos mensual de un conjunto de personas.
- La inflación mensual promedio en el país.
- La temperatura promedio diaria en la capital del país, por citar algunos casos.

La *media* para un conjunto de *datos no agrupados*, se calcula con la siguiente expresión:

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} xi}{n} = \frac{x_1+x_2+x_3... \; ...x_n}{n}}$$

Donde: n es el número de elementos del conjunto muestral, x_i es el valor del i-ésimo dato muestral.

Dicha medida no sólo se puede calcular para una muestra, sino que también se puede estimar para una población; en tal caso, las únicas variaciones corresponderán a las literales n, que para el caso de las poblaciones será N, así como para la $\bar{x}$, que será $\bar{X}$; esto mismo aplica a las posteriores medidas de tendencia central.

No obstante la frecuencia con que se usa, la media puede ser una medida engañosa, puesto que es sensible a los datos extremos (valores muy pequeños o muy grandes en una muestra), lo que suele ocasionar que la media no ofrezca la verdadera descripción de los datos muestrales, con los consecuentes errores en su interpretación; ejemplos:

- Si se tuviera un trabajador que gana \$500,00 y otro que gana \$5,00 el salario promedio se calcula por:

$$\bar{x} = \frac{\sum_{i=1}^{2} xi}{2} = \frac{x_1+x_2}{2}$$

  - Sustituyendo valores: $\bar{x} = \frac{500{,}00+5{,}00}{2}$
    $\bar{x} = 252{,}50$

  Aunque el promedio es de \$252,50 esta afirmación es engañosa, puesto que al decir "en promedio los trabajadores ganan \$252,50" se tiene una incongruencia tomando en cuenta el salario del trabajador que gana \$5,00

- En una escuela particular, se desea conocer el aprovechamiento de dos grupos escolares; para ello, se decidió comparar las medias de calificación final entre dichos grupos en la asignatura de *Matemáticas I*:

  - $A = \{6, 6, 6, 6, 6, 8, 8, 8, 8, 8\}$
  - $B = \{4, 4, 5, 6, 7, 7, 8, 9, 10, 10\}$

  Empleando la fórmula de la media para datos no agrupados se tiene que:

  - Para el grupo A, $\bar{x} = \sum_{i=1}^{10} xi = \frac{6+6+6+6+6+8+8+8+8+8}{10} = 7{,}0$
  - Para B, $\bar{x} = \sum_{i=1}^{1} 0\ xi = \frac{4+4+5+6+7+7+8+9+10+10}{10} = 7{,}0$

Al respecto, cabe destacar que no obstante que en ambos grupos se tiene como promedio de calificación 7, en el grupo B se encuentran alumnos que reprobaron la asignatura, mientras que en el grupo A, ningún alumno reprobó. Esto significa que, pese a que el promedio de calificaciones es el mismo en ambos grupos, el nivel de aprovechamiento escolar es claramente diferente.

4.2.2. La mediana (x_{med}) para datos no agrupados

La mediana (x_{med}) es el dato que se encuentra exactamente a la mitad del conjunto de observaciones; para datos no agrupados, x_{med} se calcula como sigue:

a) Para un número impar de observaciones, se toma el dato que se encuentra a la mitad, es decir: el valor tal que el número de observaciones por debajo de él es el mismo que el número de observaciones por arriba de él.

b) Para un número par de observaciones, se toman los dos valores centrales y se calcula su media aritmética.

Ejemplos:

- Para el caso *a*): Sea $A = \{1, 5, 7, 9, 10\}$. Se tiene que $n = 5$, es impar, por lo que se toma el valor central $x_{med} = 7$, pues por debajo de él se encuentran dos valores (1 y 5) y asimismo, por arriba de él también existen dos elementos, (9 y 10).
- Para el caso *b*): $B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. En este caso n es par ($n = 10$), por lo tanto se toman los dos valores centrales, el 5 y 6 y se procede a calcular su media aritmética, esto es: $x_{med} = \frac{5+6}{2} = \frac{11}{2}$. Por lo tanto: $x_{med} = 5{,}5$

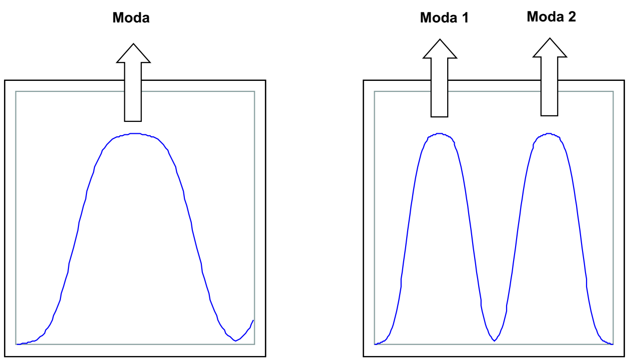
Es importante señalar, que *la mediana no es sensible a los valores extremos.* Por ello, en ocasiones se emplea en lugar de la media, como punto de referencia del término medio de un conjunto de datos muestrales.

4.2.3. La moda (x_{mod}) para datos no agrupados

La moda (x_{mod}) para datos no agrupados, es el valor (o valores) que se repiten con mayor frecuencia en una muestra; de hecho, en un conjunto de observaciones puede no existir la moda o contar con más de una, por ejemplo:

- $A = \{2, 4, 6, 8, 10, 12, 14, 16, 18\}$. En este conjunto de datos no existe la moda.
- $B = \{2, 3, 5, 7, 8, 8, 9, 11, 20\}$. $x_{mod} = 8$. Existe una sólo una moda, por lo que se dice que la distribución de los datos es *unimodal.*
- $C = \{1, 1, 1, 5, 7, 7, 9, 11, 23\}$. Existen dos modas $x_{mod} = 1$ y $x_{mod} = 7$. En este caso se dice que la distribución de los datos es *bimodal.*

La representación gráfica de este tipo de distribución sería la siguiente:



Distribución unimodal

Distribución bimodal

A la distribución de frecuencias que tiene más de dos modas se le denomina *Distribución de frecuencias multimodal.* Al igual que la mediana, la moda es una medida de tendencia central que no es sensible a los valores extremos.

4.3. Medidas de tendencia central II

4.3.1. Media para datos agrupados ($\bar{x}$)

En la práctica, un conjunto de datos muestrales suele organizarse en una tabla de datos agrupados. La construcción de la misma se revisó en el capítulo anterior. Al considerar una tabla de datos agrupados, las fórmulas para el cálculo de la media, mediana y moda, cambian radicalmente.

El primer caso a considerar, es la fórmula para calcular la Media para Datos Agrupados ($\bar{x}$) y su fórmula es la siguiente:

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n} = \frac{f_1 x_1 + f_2 x_2 + \ldots f_n x_n}{n}}$$

Donde:
k = Número de clases.
f_i = Frecuencia de clase.
x_i = Marca de clase.
n = Tamaño de la muestra (suma de las frecuencias).

Como se observa, los elementos de la fórmula, se toman directamente de la tabla de datos agrupados. Si se retoma el ejemplo del capítulo anterior, sobre las calificaciones de los alumnos en la asignatura de estadística en una universidad (página 243), para calcular su media, se retoman las tres primeras columnas de la tabla de datos agrupados (pág. 254) correspondientes al ejemplo citado:

i	Clase	f_i	x_i
1	$[5{,}1 - 5{,}6)$	2	5,35
2	$[5{,}6 - 6{,}1)$	4	5,85
3	$[6{,}1 - 6{,}6)$	11	6,35
4	$[6{,}6 - 7{,}1)$	7	6,85
5	$[7{,}1 - 7{,}6)$	19	7,35
6	$[7{,}6 - 8{,}1)$	15	7,85
7	$[8{,}1 - 8{,}6)$	8	8,35
8	$[8{,}6 - 9{,}1)$	7	8,85
9	$[9{,}1 - 9{,}6)$	5	9,35
10	$[9{,}6 - 10{,}1)$	2	9,85
		$n = \sum_{i=1}^{10} f_i = 80$	

Al aplicar la fórmula $\bar{x} = \frac{\sum_{i=1}^{10} f_i x_i}{80}$, se tiene entonces que:

$\bar{x} = \frac{f_1x_1+f_2x_2+...f_{10}x_{10}}{80}$
$\bar{x} = \frac{2(5{,}35)+4(5{,}85)+11(6{,}35)+7(6{,}85)+19(7{,}35)+15(7{,}85)+8(8{,}35)+7(8{,}85)+5(9{,}35)+2(9{,}85)}{80}$
$\bar{x} = \frac{10{,}7+23{,}4+69{,}85+47{,}95+139{,}65+117{,}75+66{,}8+61{,}95+46{,}75+19{,}7}{80}$
$\bar{x} = \frac{604{,}5}{80}$
$\bar{x} = 7{,}55625$
Redondeando a dos decimales se tiene que: $\bar{x} = 7{,}56$

4.3.2. La mediana para datos agrupados (x_{med})

"Para datos agrupados, la mediana se obtiene mediante **interpolación**"[1] y su estimación viene dada por:

$$\boxed{x_{med} = L_1 + \left(\frac{\frac{n}{2}-(\sum f)_1}{f_{mediana}}\right)c}$$

[1]op. cit. SPIEGEL, Murray R.; México; 1970; pp. 47 Negritas agregadas

donde:

L_1 = Límite inferior de la clase en donde se encuentra mediana (para determinar en qué clase cae la mediana, se divide el número n total de datos entre *dos*. Posteriormente se va sumando, comenzando desde la primera, la frecuencia de cada intervalo. Cuando dicha suma es igual o mayor al número n total de datos entre *dos*, se considera que el último intervalo que compone a dicha suma es el que contiene a la mediana).

n = Tamaño de la muestra.

$(\sum f)_1$ = Suma de las frecuencias por debajo de la clase que contiene a la mediana.

$f_{mediana}$ = Frecuencia de la clase donde se encuentra la mediana.

c = Ancho de la clase.

Determinación de los valores

Respecto a los datos agrupados del ejemplo de las calificaciones de la asignatura de estadística (página 243), se tiene que la quinta clase, es decir $[7{,}1 - - - 7{,}6)$, contiene la mayor frecuencia (19), por lo que ésta es la clase en donde se encuentra la mediana; consecuentemente $L_i = (7{,}1)$*, dado que este valor es el límite inferior de la clase que contiene a la mediana.*

Asimismo, $n = 80$

Se tienen cuatro clases $([5{,}1 - - 5{,}6), [5{,}6 - - 6{,}1), [6{,}1 - - 6{,}6), [6{,}6 - - 7{,}1))$ por debajo de la clase que contiene a la mediana. La suma de sus frecuencias es: $2 + 4 + 11 + 7 = 24$, por lo tanto $\sum f_1 = 24$.

De igual manera, la frecuencia de la clase en donde se encuentra la mediana es 19, por lo que $f_{mediana} = 19$.

Finalmente, el ancho de la clase se calcula por: $7{,}6 - 7{,}1 = 0{,}5$. Por tanto, $c = 0{,}5$. Recuérdese que tal como se explicó en el capítulo anterior, el ancho de la clase es el mismo para cada clase.

Sustitución y cálculo

$x_{med} = 7{,}1 + (\frac{\frac{80}{2}-24}{19})0{,}5$
$x_{med} = 7{,}1 + (\frac{40-24}{19})0{,}5$
$x_{med} = 7{,}1 + (\frac{16}{19})0{,}5$
$x_{med} = 7{,}1 + (0{,}842105)0{,}5$
$x_{med} = 7{,}1 + 0{,}421052631$
$x_{med} = 7{,}521052631$
Redondeando a dos decimales, se tiene que $x_{med} = 7{,}52$

4.3.3. La moda para datos agrupados (x_{mod})

La moda para datos agrupados (x_{mod}) se estima a partir de la siguiente expresión:

$$\boxed{x_{mod} = L_1 + (\frac{\Delta_1}{\Delta_1 + \Delta_2})c}$$

donde:

L_1 = Límite inferior de clase en donde se encuentra la moda (En la tabla de datos agrupados, es la clase con la mayor frecuencia).

Δ_1 = Frecuencia de la clase modal (L_1) menos la frecuencia de la clase contigua inferior.

Δ_2 = Frecuencia de la clase modal menos la frecuencia de la clase contigua superior.

c = Ancho de clase.

Determinación de los valores

Para la tabla de datos agrupados que se ha manejado, se tiene que la clase con mayor frecuencia es la quinta: $[7{,}1 - -7{,}6)$, con una frecuencia de 19, por lo que L_1 = límite inferior de la clase modal, es decir 7,1.

La clase contigua inferior a la clase modal es la cuarta: $[6{,}6 - -7{,}1)$, con una frecuencia de 7. En consecuencia, $\Delta_1 = 7$.

La clase contigua superior a la clase modal es la sexta: $[7{,}6 - -8{,}1)$, con una frecuencia de 15. Esto implica que $\Delta_2 = 15$.

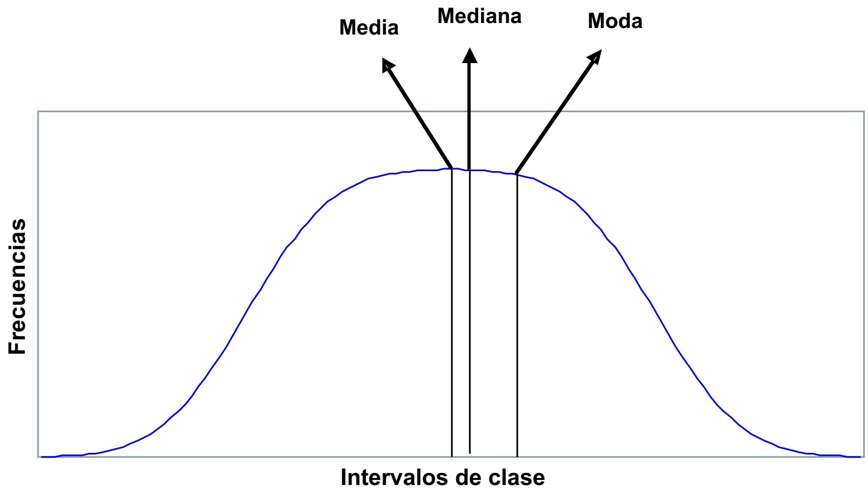
Por lo explicado respecto al cálculo de la moda, se tiene que ancho c del intervalo es 0,5.

Sustitución y cálculo

$x_{mod} = 7{,}1 + (\frac{7}{7+15})0{,}5$
$x_{mod} = 7{,}1 + (\frac{7}{22})0{,}5$
$x_{mod} = 7{,}1 + 0{,}15909$
$x_{mod} = 7{,}25909$
Redondeando a dos decimales se tiene que $x_{mod} = 7{,}26$

Cuando se estudia una muestra, es inevitable –debido a la incertidumbre que implica el uso de datos estadísticos–, que ésta tenga un sesgo, es decir, una cierta tendencia o distorsión. Toda muestra tiene sesgo en mayor o menor grado, dependiendo de la forma en que se obtuvo. En el respectivo capítulo de *Muestreo*, se revisará la metodología para el diseño de muestras y los mecanismos para reducir su sesgo y error.

En muestras cuya distribución de frecuencias se presenta un sesgo moderado (poca asimetría), los valores de la media ($\bar{x}$), la moda (x_{mod}) y la mediana (x_{med}) son cercanos entre sí, y tienden a situarse en la parte central del histograma (en su representación gráfica). Por esta razón, se les denomina *medidas de tendencia central*.



Cuando se estudia una muestra y las medidas de tendencia central guardan una gran diferencia entre sí (una diferencia considerable), se puede pensar que el sesgo es muy grande y en consecuencia que la muestra está mal diseñada o que el cálculo de los estadígrafos está mal realizado.

Al analizar muestras *no es recomendable utilizar sólo una medida de tendencia central para obtener conclusiones*, se sugiere considerar a todas, pues sus propiedades son complementarias. A continuación, se presentan algunas características de las medidas de tendencia central para su consideración al elaborar estudios de base cuantitativa.

- *La media*, es la medida de mayor uso en los estudios estadísticos, sin embargo, es una medida sensible a valores extremos.
- *La mediana*, no depende (su cálculo) de los valores extremos, pero sí del número de individuos (datos) que se está estudiando.
- *La moda*, no es sensible a valores extremos, puede no existir y si existe puede ser única o múltiple (Unimodal, bimodal, multimodal).

4.4. Regla empírica

La *regla empírica* para calcular las medidas de tendencia central, se utiliza al estudiar muestras cuya distribución de frecuencias (la forma como se distribuyen las frecuencias en las distintas categorías o intervalos de clase) es regular, es decir: *sólo es válida para muestras donde el sesgo es pequeño.* La regla empírica es una expresión que relaciona de manera proporcional a la media, la moda y la mediana, a saber:

1. La media menos la moda es aproximadamente igual a tres veces la diferencia entre la media y la mediana:

$$\boxed{\bar{x} - x_{mod} \approx 3(\bar{x} - x_{med})}$$

2. Si se conoce la media y la mediana, la moda puede despejarse de la fórmula de la expresión anterior:

   $\bar{x} - x_{mod} \approx 3(\bar{x} - x_{med})$
   $-x_{mod} \approx 3(\bar{x} - x_{med}) - \bar{x}$
   $x_{mod} \approx -3(\bar{x} - x_{med}) + \bar{x}$
   $x_{mod} \approx -3\bar{x} + 3x_{med} + \bar{x}$
   $x_{mod} \approx -2\bar{x} + 3x_{med}$

   Reordenando:

   $$\boxed{x_{mod} \approx 3x_{med} - 2\bar{x}}$$

3. Por otra parte, si se conoce la moda y la mediana, la media puede despejarse de la fórmula anterior:

   $x_{mod} \approx 3x_{med} - 2\bar{x}$
   $2\bar{x} \approx 3x_{med} - x_{mod}$

   $$\boxed{\bar{x} \approx \tfrac{3x_{med} - x_{mod}}{2}}$$

4. Asimismo, si se conoce la moda y la media, puede despejarse de la expresión del número 1, la mediana:

   $$\bar{x} - x_{mod} \approx 3(\bar{x} - x_{med})$$

   Cambiando términos, se tiene que:
   $3(\bar{x} - x_{med}) \approx \bar{x} - x_{mod}$
   $(\bar{x} - x_{med}) \approx \tfrac{\bar{x} - x_{mod}}{3}$
   $-x_{med} \approx \tfrac{\bar{x} - x_{mod}}{3} - \bar{x}$

   $$\boxed{x_{med} \approx \bar{x} - \tfrac{\bar{x} - x_{mod}}{3}}$$

- Ejemplo: Para el ejercicio sobre la calificación de alumnos en la asignatura de estadística (página 243), se tiene que $\bar{x} = 7{,}56$; $x_{med} = 7{,}52$ y $x_{mod} = 7{,}26$. Empleando la regla empírica realiza lo siguiente:

  - *a*) La aproximación de x_{mod}. Sustituyendo $\bar{x}$ y x_{med} en la fórmula $x_{mod} \approx \bar{x} - 3(\bar{x} - x_{med})$, se obtiene:

    $x_{mod} \approx 3(7{,}52) - 2(7{,}56)$
    $x_{mod} \approx 22{,}56 - 15{,}12$
    $x_{mod} \approx 7{,}44$

    Se observa que 7,44 $\approx$**7.56** (verdadero valor).

  - *b*) La aproximación de $\bar{x}$. Sustituyendo x_{mod} y x_{med} en la fórmula $\bar{x} \approx \frac{3x_{med} - x_{mod}}{2}$, se tiene:

    $\bar{x} \approx \frac{3(7{,}52) - 7{,}26}{2}$
    $\bar{x} \approx \frac{22{,}56 - 7{,}26}{2}$
    $\bar{x} \approx \frac{15{,}3}{2}$
    $\bar{x} \approx 7{,}65$

    Se observa que 7,65 $\approx$**7.56** (verdadero valor).

  - *c*) La aproximación de x_{med}. Sustituyendo $\bar{x}$ y x_{mod} en la fórmula $x_{med} \approx \bar{x} - \frac{\bar{x} - x_{mod}}{3}$, se tiene que:

    $x_{med} \approx 7{,}56 - \frac{7{,}56 - 7{,}26}{3}$
    $x_{med} \approx 7{,}56 - \frac{0{,}3}{3}$
    $x_{med} \approx 7{,}56 - 0{,}1$
    $x_{med} \approx 7{,}46$

    Se observa que 7,46 $\approx$7,52 (verdadero valor).

Ejercicios 1

- *Instrucciones*: Responde breve y suficientemente las preguntas que se te realizan.

- Al finalizar, coteja tus respuestas con las que se te presentan en la página 312

1. ¿Qué significa la expresión $n \to N$ y qué implica para la curva de un polígono de frecuencia?

   ____

   ____

   ____

   ____

   ____

2. ¿Qué son las medidas de tendencia central?

   ____

   ____

   ____

   ____

3. Define lo que es la media aritmética.

   ____

   ____

   ____

   ____

4. Define lo que es la mediana

   ____

   ____

5. ¿Cómo se le denomina al dato que más veces aparece en un conjunto de observaciones?

   ----

6. Representa gráficamente una distribución de frecuencias multimodal.

7. ¿Qué ocurre, con respecto de las medidas de tendencia central, cuando en una muestra la distribución de frecuencias posee un sesgo moderado?

   ----

8. ¿Qué desventaja podría representar el análisis de una muestra a partir de $\bar{x}$?

   ----

9. ¿A qué característica de la muestra presenta sensibilidad el cálculo de x_{med}?

   ----

10. En $\frac{1}{8}$ de cartulina, elabora una tabla que contenga las fórmulas para el cálculo de las medidas de tendencia central para datos agrupados, datos no agrupados y las correspondientes a la regla empírica, así como el significado de las literales que conforman cada expresión (omite aquellas literales que te sean totalmente familiares o las más comunes).

4.5. Cuantiles

Los *cuantiles* son magnitudes estadísticas que *dividen en partes iguales a un conjunto de datos ordenados* (ascendentemente o descendentemente); por ejemplo: *La mediana es un tipo de cuantil*, pues divide en dos partes iguales a un conjunto de datos u observaciones; considérese lo siguiente:

$$A = \{0, 1, 1, 2, 3, 4, 5, 6, 7\}$$

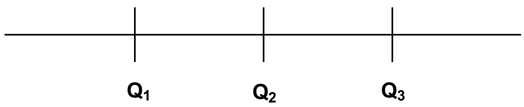
La mediana es = 3 y divide a la muestra en dos partes iguales, pues tanto a la derecha del tres como a la izquierda quedan 4 elementos de la muestra. De acuerdo con el número de segmentos en que se divide al conjunto de datos u observaciones, tenemos que los cuantiles se clasifican en: *cuartiles*, *deciles* y *percentiles*.

La importancia de los cuantiles, radica en la posibilidad que ofrecen –al investigador– de realizar observaciones respecto a cómo se encuentra constituida la muestra y, hasta cierto punto, de la forma en que se comporta de acuerdo a la composición dada por sus unidades muestrales; esto, una vez que se han ordenado los datos en forma ascendente y/o en una tabla de datos agrupados.

4.5.1. Cuartiles

Se denomina con el nombre de *cuartiles* a la serie de datos o magnitudes que *sirven como puntos de referencia* en la división de un conjunto de datos u observaciones en cuatro partes idénticas; estos se denotan por: Q_1, Q_2 y Q_3; siendo el primer cuartil Q_1, el segundo Q_2 y el tercer cuartil Q_3.

Debe quedar claro que no se trata de dividir un segmento en cuatro partes iguales, esto sería un problema de geometría.



Más bien, los cuartiles son valores que dividen a un conjunto de datos en cuatro partes iguales, de tal manera, que cada parte contiene el mismo número de casos u observaciones. Por ejemplo, para el siguiente conjunto de datos:

$$B = \{2, 3, 3, 5, 11, 15, 23, 28, 29, 31, 31, 40, 45, 49, 55\}$$

los cuartiles son $Q_1 = 5$, $Q_2 = 28$ y $Q_3 = 40$, ya que dividen al conjunto en cuatro categorías, con el mismo número de observaciones cada una (3 observaciones), como podrá notarse:

```
(2,3,3),  5,  (11,15,23),  28,  (29,31,31),   40,   (45,49,55)
```

En forma similar los deciles y percentiles pueden dividir en diez y cien partes iguales, respectivamente, a un conjunto de datos u observaciones.

4.5.2. Cálculo de cuartiles mediante interpolación lineal

Cuando para un conjunto de datos agrupados es necesario calcular sus cuartiles (los cuantiles en general), el método para hacerlo es mediante la interpolación lineal; puesto que los datos se encuentran ordenados por grupos (intervalos) y por tanto se pueden conocer aproximadamente los valores entre los que se encuentra el cuartil.

Recuérdese que la *interpolación*, es la acción de estimar un valor desconocido que se encuentra entre dos valores conocidos. A través de este mecanismo, pueden obtenerse los cuartiles de un conjunto de datos. Para explicar y ejemplificar el cálculo de cuartiles, consideremos nuevamente el ejemplo sobre las calificaciones de alumnos en la asignatura de Estadística:

i	Clase	f_i
1	[5,1 − 5,6)	2
2	[5,6 − 6,1)	4
3	[6,1 − 6,6)	11
4	[6,6 − 7,1)	7
5	[7,1 − 7,6)	19
6	[7,6 − 8,1)	15
7	[8,1 − 8,6)	8
8	[8,6 − 9,1)	7
9	[9,1 − 9,6)	5
10	[9,6 − 10,1)	2
		$n = \sum_{i=1}^{10} f_i = 80$

Para el primer cuartil, se deben considerar $\frac{n}{4} = \frac{80}{4} = 20$ de los casos u observaciones. Como las tres primeras clases contienen entre las tres: 17 casos u observaciones $(2 + 4 + 11)$, faltan 3 para completar los 20 a considerar para el primer cuartil. Por tal motivo, se deben tomar 3 observaciones o casos de los 7 que comprende la cuarta clase. Considerando que el ancho c de clase es 0,5 y tomando el límite inferior de la cuarta clase, el primer cuartil se calcula como sigue:

$Q_1 = 6{,}6 + \frac{3}{7}(0{,}5)$
$Q_1 = 6{,}6 + (0{,}4286)(0{,}5)$
$Q_1 = 6{,}6 + (0{,}2143)$
$Q_1 = 6{,}8143$
Redondeando a dos decimales, $Q_1 = 6{,}81$

Para el segundo cuartil, se deben considerar $\frac{2n}{4} = \frac{n}{2} = \frac{80}{2} = 40$ de los casos u observaciones. Como las cuatro primeras clases contienen entre todas: 24 casos u observaciones $(2 + 4 + 11 + 7)$, faltan 16 para completar los 40 a considerar para el segundo cuartil. Por tal motivo, se deben tomar 16 observaciones o casos de los 19 que comprende la quinta clase. Considerando que el ancho c de clase es 0,5 y tomando el límite inferior de la quinta clase, el segundo cuartil se calcula como sigue:

$Q_2 = 7{,}1 + \frac{16}{19}(0{,}5)$
$Q_2 = 7{,}1 + (0{,}8421)(0{,}5)$
$Q_2 = 7{,}1 + (0{,}42105)$
$Q_2 = 7{,}52105$
Redondeando a dos decimales, $Q_2 = 7{,}52$

‡ *Obsérvese que el valor del segundo cuartil corresponde al valor de la mediana, es decir*:

$$\boxed{Q_2 = x_{med}}$$

Para el tercer cuartil, se deben considerar $\frac{3n}{4} = \frac{3(80)}{4} = \frac{240}{4} = 60$

$Q_3 = 8{,}1 + \frac{2}{8}(0{,}5)$
$Q_3 = 8{,}1 + (0{,}25)(0{,}5)$
$Q_3 = 8{,}1 + (0{,}125)$
$Q_3 = 8{,}225$
Redondeando a dos decimales, $Q_3 = 8{,}23$

De lo anterior se desprende que:

- El 25 % de los alumnos tiene una calificación de 6,81 o menor
- El 50 % de los alumnos tiene una calificación de 7,52 o menor
- El 75 % de los alumnos tiene una calificación de 8,23 o menor

4.5.3. Cálculo de deciles mediante interpolación

Se designa con el nombre de *Deciles* a la serie de datos o magnitudes que sirven como puntos de referencia en la división de un conjunto de datos u observaciones en diez partes idénticas; estos se denotan por: D_1, D_2, D_3,$D_4 \dots D_9$ siendo el Primer decil D_1, el segundo D_2 el tercero D_3 y así hasta el Decil D_9.

Para ejemplificar y explicar el cálculo de los deciles, consideraremos el anterior conjunto de observaciones referentes a las calificaciones de alumnos en la asignatura de estadística.

Para el primer decil, D_1, se consideran $\frac{n}{10} = \frac{80}{10} = 8$. Del ejemplo anterior se tiene que $c = 0{,}5$ por lo tanto, se tiene que:
$D_1 = 6{,}1 + \frac{2}{11}(0{,}5) = 6{,}19.$

Para el segundo decil, D_2, se consideran $\frac{2n}{10} = \frac{160}{10} = 16$ observaciones o casos. Por lo tanto, se tiene que:

$D_2 = 6{,}1 + \frac{10}{11}(0{,}5) = 6{,}55.$

Para el tercer decil, D_3, se consideran $\frac{3n}{10} = \frac{240}{10} = 24$ observaciones o casos. Por lo tanto, se tiene que:
$D_3 = 7{,}1 + \frac{0}{19}(0{,}5) = 7{,}1.$

Para el cuarto decil, D_4, se consideran $\frac{4n}{10} = \frac{320}{10} = 326$ observaciones o casos. Por lo tanto, se tiene que:
$D_4 = 7{,}1 + \frac{8}{19}(0{,}5) = 7{,}31.$

Para el quinto decil, D_5, se consideran $\frac{5n}{10} = \frac{400}{10} = 40$ observaciones o casos. Por lo tanto, se tiene que:
$D_5 = 7{,}1 + \frac{16}{19}(0{,}5) = 7{,}52.$

Para el sexto decil, D_6, se consideran $\frac{6n}{10} = \frac{480}{10} = 48$ observaciones o casos. Por lo tanto, se tiene que:
$D_6 = 7{,}6 + \frac{5}{15}(0{,}5) = 7{,}77.$

Para el séptimo decil, D_7, se consideran $\frac{7n}{10} = \frac{560}{10} = 56$ observaciones o casos. Por lo tanto, se tiene que:
$D_7 = 7{,}6 + \frac{13}{15}(0{,}5) = 8{,}03.$

Para el octavo decil, D_8, se consideran $\frac{8n}{10} = \frac{640}{10} = 64$ observaciones o casos. Por lo tanto, se tiene que:
$D_8 = 8{,}1 + \frac{6}{8}(0{,}5) = 8{,}48.$

Para el noveno decil, D_2, se consideran $\frac{9n}{10} = \frac{720}{10} = 72$ observaciones o casos. Por lo tanto, se tiene que:
$D_9 = 8{,}6 + \frac{6}{7}(0{,}5) = 9{,}03.$

Los anteriores cálculos implican que:

- 10 % de los alumnos obtuvieron una calificación de 6,19 o menor
- 20 % de los alumnos obtuvieron una calificación de 6,55 o menor
- 30 % de los alumnos obtuvieron una calificación de 7,1 o menor
- 40 % de los alumnos obtuvieron una calificación de 7,31 o menor
- 50 % de los alumnos obtuvieron una calificación de 7,52 o menor
- 60 % de los alumnos obtuvieron una calificación de 7,77 o menor
- 70 % de los alumnos obtuvieron una calificación de 8,03 o menor
- 80 % de los alumnos obtuvieron una calificación de 8,48 o menor
- 90 % de los alumnos obtuvieron una calificación de 9,03 o menor

‡ *Puede observarse que la mediana, el segundo cuartil y el quinto decil son iguales a 7.52, es decir:*

$$\boxed{D_5 = Q_2 = x_{med}}$$

4.5.4. Cálculo de percentiles mediante interpolación

Se denomina con el nombre de *Percentiles* a la serie de datos o magnitudes que sirven como puntos de referencia en la división de un conjunto de datos u observaciones en cien partes idénticas; estos se denotan por: P_1, P_2, P_3,... siendo el Primer centil P_1, el segundo P_2,...

- Ejemplo: Calcular el percentil treinta y cinco y el percentil cincuenta de la tabla de datos con la calificación de alumnos en la asignatura de estadística.

  - Para el percentil treinta y cinco, se tienen que considerar $\frac{35(n)}{100} = \frac{2800}{100} = 28$ casos u observaciones, por lo que:
    $P_{35} = 7{,}1 + \frac{4}{19}(0{,}5) = 7{,}21$

  - Para el percentil cincuenta, se tienen que considerar $\frac{50(n)}{100} = \frac{4000}{100} = 40$ casos u observaciones, por lo que:
    $P_{50} = 7{,}1 + \frac{16}{19}(0{,}5) = 7{,}52$

De lo anterior se desprende que:

- 35 % de los alumnos obtuvieron una calificación de 7,21 o menor
- 50 % de los alumnos obtuvieron una calificación de 7,52 o menor

‡ *Puede observarse que la mediana, el segundo cuartil, el quinto decil y el percentil cincuenta son iguales a* 7,52 *es decir:*

$$\boxed{P_{50} = D_5 = Q_2 = x_{med}}$$

Ejercicios 2

- *Instrucciones*: Responde breve y suficientemente las preguntas que se te realizan.

- Al finalizar, coteja tus respuestas con las que se te presentan en la página 315

1. ¿Qué son los cuantiles?

2. ¿Cuántas clases de cuantiles hay? Descríbelos brevemente.

3. Elabora un ejemplo de los deciles para datos no agrupados.

4. ¿Por qué para datos agrupados los cuantiles se calculan mediante interpolación lineal?

____

____

____

____

5. En el reverso de la tabla que elaboraste como parte de la anterior serie de ejercicios, anota las correspondencias que hay entre los diferentes tipos de cuantiles y las medidas de tendencia central.

6. Diseña un esquema donde se vean los pasos del procedimiento para calcular los percentiles, así como las partes que los conforman; para ello, puedes retomar alguno de los ejemplos presentados en este material.

7. Finalmente repasa la forma en que se calculan las medidas de tendencia central y los cuantiles para resolver los ejercicios de Autoevaluación.

4.6. Autoevaluación

- *Instrucciones*: Realiza las operaciones necesarias para el cálculo de lo que se te pide. Para ello, te puedes apoyar en las tablas de fórmulas que ya construiste en los ejercicios del capítulo.

- Finalmente, coteja tus respuestas con las que se te proporcionan en la página 316

1. Considera los siguientes conjuntos de *datos no agrupados* para calcular sus medidas de tendencia central.

   - $A = \{1, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27\}$

     a) $\bar{x} =$

     b) $x_{mod} =$

     c) $x_{med} =$

   - $B = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24\}$

     a) $\bar{x} =$

     b) $x_{mod} =$

     c) $x_{med} =$

   - $C = \{1, 1, 2, 2, 3, 4, 5, 6, 6, 6, 7, 8, 9, 10, 11, 12, 13\}$

     a) $\bar{x} =$

     b) $x_{mod} =$

     c) $x_{med} =$

2. Considera la siguiente tabla de distribución de frecuencias:

i	$Clase$	f_i	x_i
1	[300 – 400)	14	
2	[400 – 500)	46	
3	[500 – 600)	58	
4	[600 – 700)	76	
5	[700 – 800)	68	
6	[800 – 900)	62	
7	[900 – 1000)	48	
8	[1000 – 1100)	22	
9	[1100 – 1200)	6	
		$n = \sum_i = 1^9\ f_i = 400$	

Empleando las fórmulas para datos agrupados, calcula la *media, moda* y *mediana*, no olvides calcular las *marcas de clase*.

a) $\bar{x} =$

b) $x_{mod} =$

c) $x_{med} =$

3. Para la anterior tabla de distribución de frecuencias, calcula los *cuartiles*, sus *deciles* y los *percentiles* 30, 50, 70 y 90.

a) $Q_1 =$

b) $Q_2 =$

c) $Q_3 =$

d) $D_1 =$

e) $D_2 =$

f) $D_3 =$

g) $D_4 =$

h) $D_5 =$

i) $D_6 =$

j) $D_7 =$

k) $D_8 =$

l) $D_9 =$

m) $P_{30} =$

n) $P_{50} =$

ñ) $P_{70} =$

o) $P_{90} =$

4. Considera los valores que se te presentan para calcular con la regla empírica, los valores de las medidas de tendencia central.

a) Sea que $\bar{x} = 23{,}5$ y $x_{mod} = 24$, mediante la aplicación de la regla empírica, calcula x_{med}

- $x_{med} =$

b) Supóngase que $\bar{x} = 12$ y $x_{med} = 11$, mediante la aplicación de la regla empírica, calcula x_{mod}

- $x_{mod} =$

c) Si $x_{med} = 3{,}5$ y $x_{mod} = 3{,}7$ mediante la aplicación de la regla empírica, calcula $\bar{x}$

- $\bar{x} =$

4.7. Apéndice del capítulo

4.7.1. Respuestas a los ejercicios del capítulo

Ejercicios 1

- *Instrucciones*: En la presente sección, se te ofrecen las respuestas a las preguntas que anotaste en la sección de *Ejercicios 1* (pág 297). Recuerda que para las preguntas cuya respuesta es abierta, lo importante es el *sentido* de tus respuestas, es decir, que éstas sean congruentes con las aquí proporcionadas. En el caso de las tablas y esquemas, cuida que contengan todos los elementos propuestos.

1. ¿Qué significa la expresión $n \to N$ y qué implica para la curva de un polígono de frecuencia?
   Significa que el tamaño n de la muestra tiende al tamaño N de la población, esto provoca que el contorno de la curva de un polígono de frecuencias se suavice, de manera que toma la forma de una curva lisa dado que el tamaño n de la muestra se acerca al tamaño N de la población.

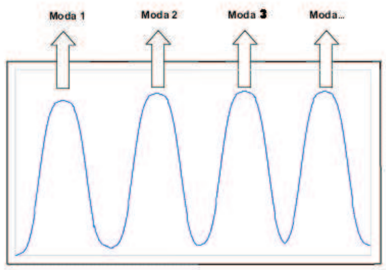
2. ¿Qué son las medidas de tendencia central?
   Son medidas que sirven para analizar a un conjunto de datos muestrales. Éstas cuantifican la concentración de los datos respecto de un cierto valor, tal es el caso de la media $\bar{x}$, la Mediana x_{med} y la Moda x_{mod}.

3. Define lo que es la media aritmética.
   La media $\bar{x}$ es una medida de tendencia central con que se puede calcular el punto de equilibrio en un conjunto de datos. Mediante su estimación, se busca obtener un valor común que describa el comportamiento de la muestra en cuestión

4. Define lo que es la mediana
   La mediana x_{med} es el dato que se encuentra justamente a la mitad del conjunto de observaciones.

5. ¿Cómo se le denomina al dato que más veces aparece en un conjunto de observaciones?
   Moda x_{mod}

6. Representa gráficamente una distribución de frecuencias multimodal.



Distribución Multimodal

7. ¿Qué ocurre, con respecto de las medidas de tendencia central, cuando en una muestra la distribución de frecuencias posee un sesgo moderado?
   Que los valores de $\bar{x}$, x_{med} y x_{mod} son muy cercanos entre sí y por ello, el grado de asimetría en la curva del histograma se reduce, es decir: la ubicación de dichas medidas se aproxima a la parte central de la curva; en caso contrario, se puede pensar que se ha realizado mal el cálculo de los estadígrafos o que la muestra está mal diseñada.

8. ¿Qué desventaja podría representar el análisis de una muestra a partir de $\bar{x}$?
   Que la $\bar{x}$ es una medida de tendencia central sensible a valores extremos, lo podría afectar la obtención de conclusiones.

9. ¿A qué característica de la muestra presenta sensibilidad el cálculo de x_{med}? Al número de individuos o datos, que se están estudiando

10. Formato propuesto para la tabla de fórmulas

Fórmulas			
Modalidad	**Atributo**	**Fórmula**	**Literales**
Datos no agrupados	Media	$\bar{x} = \frac{\sum_{i=1}^{n} xi}{n} = \frac{x_1+x_2+x_3\ldots\ \ldots x_n}{n}$	***x_i*** **=** Valor del i-ésimo dato muestral
	Mediana (x_{med})	***a) Si*** *n* es **impar**, se toma el dato del centro, el que se encuentra en medio del conjunto de observaciones. ***b) Si*** *n* es **par**, se toman los dos valores centrales y se calcula su media aritmética. x_{med}=v+w/2	***v*** = Primer valor central ***w*** = Segundo valor central **2** = Numero de datos del cual se calculará la media aritmética.
	Moda (x_{mod})	a) Unimodal = Sólo un dato se repite b) Bimodal = Dos datos predominana c) Multimodal = Más de dos datos predominan	
Datos agrupados	Media	$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n} = \frac{f_1x_1+f_2x_2+\ldots f_nx_n}{n}$	***K*** **=** Número de clases ***f_i*** **=** Frecuencia de clases ***x_i*** **=** Marca de clase
	Mediana (xmed)	$x_{med} = L_1 + (\frac{\frac{n}{2}-(\sum f)_1}{f_{mediana}})c$	***L_1*** **=** límite inferior de clase donde se encuentra la mediana **(*f*)1 =** *Suma de las frecuencias por debajo de la clase que contiene a la mediana* ***$f_{mediana}$***= Frecuencia de clase donde se encuentra la mediana. ***c*** = Ancho de clase
	Moda (xmod)	$x_{mod} = L_1 + (\frac{\Delta_1}{\Delta_1+\Delta_2})c$	Frecuencia de la clase modal *(L_1)* Menos la frecuencia de la clase contigua inferior = Frecuencia de la clase modal menos la frecuencia de la clase contigua superior.
Regla empírica	Media	$\bar{x} = \frac{3x_{med}-x_{mod}}{2}$	***$3x_{med}$*** = Tres veces la mediana
	Moda (xmod)	$x_{mod} = 3x_{med} - 2\bar{x}$	$-2\bar{x}$ Menos dos veces ma media
	Mediana (xmed)	$x_{med} = \bar{x} - \frac{\bar{x}-x_{mod}}{3}$	

Ejercicios 2

1. ¿Qué son los cuantiles?
   Son magnitudes estadísticas que dividen en partes iguales a un conjunto de datos, de tal manera que cada parte contiene el mismo número de casos u observaciones; en otras palabras los cuantiles, sirven como punto de referencia en dicha división.

2. ¿Cuántas clases de cuantiles hay? Descríbelos brevemente.
   La media (x_{med}), los cuartiles (Q_n), deciles (D_n) y percentiles (P_n), que sirven como puntos de referencia al dividir a un conjunto de datos (que pueden estar o no agrupados) en dos, cuatro, diez y cien partes exactamente iguales, respectivamente.

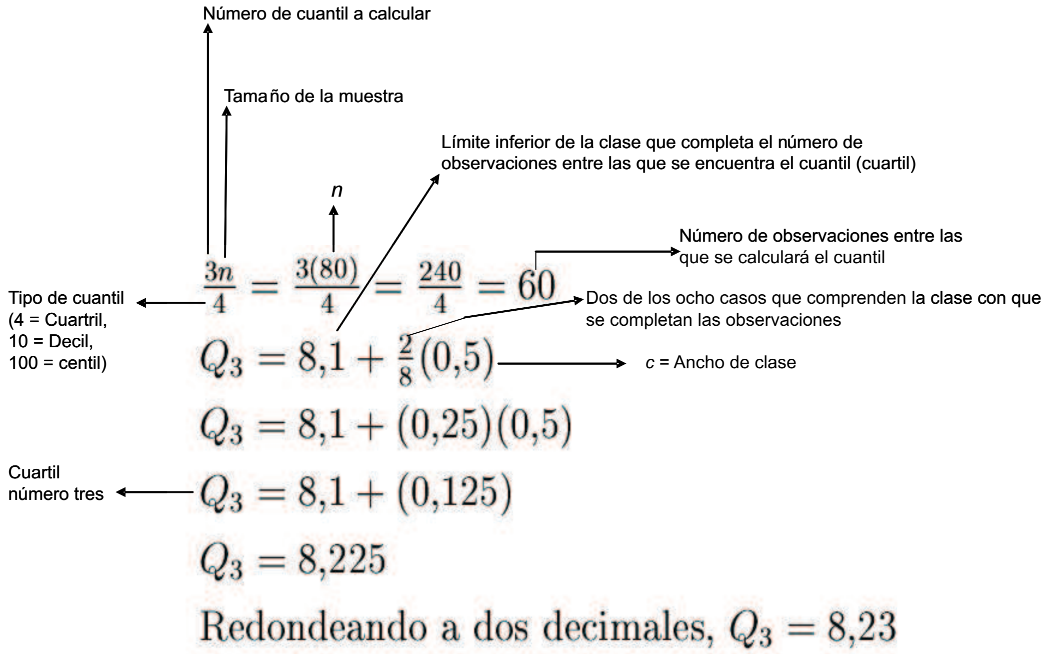
3. Elabora un ejemplo de los deciles para datos no agrupados.
   Éste debe ser semejante a los presentados en las páginas 271 (media) y/o 272 (Conjunto B referente a los cuartiles).

4. ¿Por qué para datos agrupados los cuantiles se calculan mediante interpolación lineal?
   Porque los datos se encuentran ya organizados en intervalos, lo que implica la ubicación del cuantil, valor desconocido, entre dos valores conocidos.

5. En el reverso de la tabla que elaboraste para la anterior serie de ejercicios, anota las correspondencias que hay entre los diferentes tipos de cuantiles y las medidas de tendencia central.

Equivalencia de los cuantiles
$Q_2 = x_{med}$
$D_5 = Q_2 = x_{med}$
$P_{50} = D_5 = Q_2 = x_{med}$

6. Diseña un esquema donde se vean los pasos del procedimiento para calcular los percentiles, así como las partes que los conforman; para ello, puedes retomar alguno de los ejemplos presentados en este material.



4.7.2. Respuestas de la autoevaluación

- *Instrucciones*: Realiza las operaciones necesarias para el cálculo de lo que se te pide. Para ello, te puedes apoya en las tablas de fórmulas que ya construiste en los ejercicios de capítulo.

1. Considera los siguientes conjuntos de *datos no agrupados* para calcular sus medidas de tendencia central.

   - $A = \{1, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27\}$

a) $\bar{x} = 13{,}13$

b) $x_{mod} = 1; Unimodal$

c) $x_{med} = 13$

- $B = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24\}$

a) $\bar{x} = 13$

b) $x_{mod} =$ *No tiene moda*

c) $x_{med} = 13$

- $C = \{1, 1, 2, 2, 3, 4, 5, 6, 6, 6, 7, 8, 9, 10, 11, 12, 13\}$

a) $\bar{x} = 6{,}2353$

b) $x_{mod} = 1, 2, 6$; *Multimodal*

c) $x_{med} = 13$

2. Considera la siguiente tabla de distribución de frecuencias:

i	*Clase*	f_i	x_i
1	$[300 - 400)$	14	350
2	$[400 - 500)$	46	450
3	$[500 - 600)$	58	550
4	$[600 - 700)$	76	650
5	$[700 - 800)$	68	750
6	$[800 - 900)$	62	850
7	$[900 - 1000)$	48	950
8	$[1000 - 1100)$	22	1050
9	$[1100 - 1200)$	6	1150
		$n = \sum_i = 1^9 \ f_i = 400$	

Empleando las fórmulas para datos agrupados, calcula la media, moda y mediana, no olvides calcular las marcas de clase.

a) $\bar{x} = 715{,}5$

b) $x_{mod} = 669{,}2307692$

c) $x_{med} = 707{,}8947368$

3. Para la anterior tabla de distribución de frecuencias, calcula los cuartiles, sus deciles y los percentiles 30, 50, 70 y 90.

   a) $Q_1 = 568{,}9655172$

   b) $Q_2 = 708{,}8235294$

   c) $Q_3 = 861{,}2903226$

   d) $D_1 = 456{,}3043$

   e) $D_2 = 534{,}13$

   f) $D_3 = 602{,}2368$

   g) $D_4 = 654{,}73$

   h) $D_5 = 708{,}08$

   i) $D_6 = 766{,}764$

   j) $D_7 = 828{,}36$

   k) $D_8 = 893{,}77$

   l) $D_9 = 975{,}20$

   m) $P_{30} = 602{,}2368$

   n) $P_{50} = 708{,}08$

   ñ) $P_{70} = 828{,}36$

   o) $P_{90} = 975{,}2$

4. Considera los valores que se te presentan para calcular con la regla empírica, los valores de las medidas de tendencia central.

   a) Sea que $\bar{x} = 23{,}5$ y $x_{mod} = 24$, mediante la aplicación de la regla empírica, calcula x_{med}

- $x_{med} = 23{,}667$

b) Supóngase que $\bar{x} = 12$ y $x_{med} = 11$, mediante la aplicación de la regla empírica, calcula x_{mod}

- $x_{mod} = 9$

c) Si $x_{med} = 3{,}5$ y $x_{mod} = 3{,}7$ mediante la aplicación de la regla empírica, calcula $\bar{x}$

- $\bar{x} = 3{,}4$

4.7.3. Glosario

Interpolación: "Calcular el valor aproximado de una magnitud en un intervalo cuando se conocen algunos de los valores que toma a uno y otro lado de dicho intervalo."[2]

[2]Biblioteca de Consulta Microsoft Encarta 2005. © 1993-2004 Microsoft Corporation. Reservados todos los derechos.

Capítulo 5

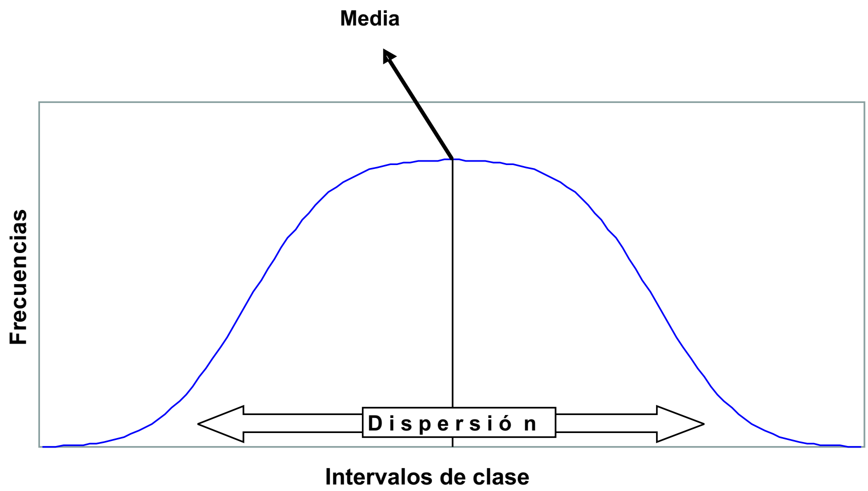
Medidas de dispersión

Objetivos específicos

- El estudiante empleará las principales *medidas de dispersión* (rango, desviación estándar y varianza) como un complemento de las *medidas de tendencia central* en el estudio de diferentes *conjuntos de datos muestrales*; para lo cual, aprenderá cómo se calculan e interpretan mediante la resolución de ejercicios.

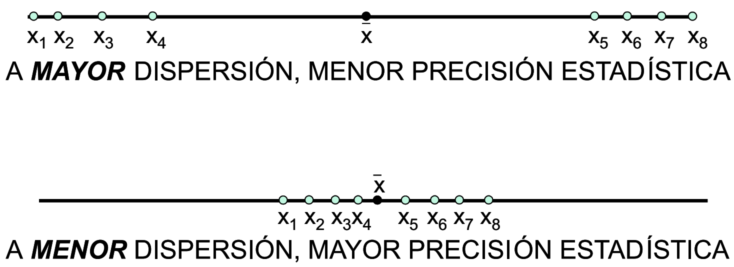
5.1. Preliminar

En el capítulo anterior se explicó cómo las *medidas de tendencia central*, describen y miden la concentración de las observaciones o datos respecto a determinados valores. Se dice que son medidas de tendencia central, ya que tienden a ubicarse en la parte media de la gráfica (histograma y polígono de frecuencias). Un tipo de medidas complementarias a las de tendencia central, lo constituyen las *medidas de dispersión* o variabilidad. Éstas describen y miden el grado de *variación* –la distancia en la distribución– de las observaciones o datos muestrales respecto a la media, es decir, qué tan separados están de $\bar{x}$.



Medir la variación de un conjunto de datos muestrales, permite conocer, por ejemplo, qué tan bien se encuentra representada una población (los datos típicos de ésta) por la muestra. Las medidas de dispersión más comunes y representativas son: *el rango*, *la desviación estándar* (también conocida como *desviación típica* o *promedio de dispersión*) y la *varianza*.

Las medidas de dispersión brindan un referente sobre la precisión estadística de las observaciones o datos muestrales. Cuando la dispersión de un conjunto de datos es muy alta, las predicciones que se realizan en función de éstas, serán sumamente imprecisas; más cuando ocurre lo contrario, la incertidumbre disminuye considerablemente. Luego entonces, mayor dispersión implica menor precisión estadística y, por complemento, a menor dispersión en los datos u observaciones, mayor precisión estadística.



A continuación se presentan las principales medidas de dispersión. Como se podrá notar, éstas son de gran valía para el análisis estadístico de los datos pertinentes a una investigación; si sólo se tomasen en cuenta las medidas de tendencia central para elaborar conclusiones y/o hacer predicciones, el análisis de los datos sería parcial y poco confiable.

5.2. Rango para datos agrupados

Como se analizó en el capítulo 3, **Organización de los datos**, el rango (véase rango en la pág. 237) de un conjunto de datos se calcula con la diferencia o resta entre el mayor de los datos y el menor. El rango es la más sencilla de las medidas de dispersión, es "una medida elemental de dispersión *basada en la ordenación* de las observaciones"[1]. Si los datos han sido ordenados ascendentemente, el rango se calcula mediante la sustitución de los valores correspondientes a la siguiente expresión:

$$\boxed{r = x_n - x_1}$$

donde x_n es el dato mayor (es decir el último de los datos ordenados ascendentemente) y x_1 es el menor dato (es decir, el primero de los datos ordenados ascendentemente). Retomando el ejemplo sobre la calificación de alumnos en la asignatura de estadística (página 243), debe recordarse que el conjunto de datos u observaciones ordenados ascendentemente es:

5,1	5,4	5,7	5,9	6,0	6,0	6,1	6,1	6,2	6,2	6,2	6,2	6,3	6,3
6,5	6,5	6,5	6,6	6,7	6,7	6,8	6,8	6,8	6,9	7,1	7,1	7,1	7,2
7,2	7,3	7,3	7,3	7,3	7,4	7,4	7,4	7,5	7,5	7,5	7,5	7,5	7,5
7,5	7,6	7,6	7,6	7,6	7,7	7,7	7,8	7,8	7,8	7,8	7,8	7,9	7,9
7,9	8,0	8,1	8,2	8,2	8,3	8,4	8,5	8,5	8,5	8,6	8,7	8,8	8,8
8,8	8,9	9,0	9,3	9,3	9,4	9,5	9,5	9,6	9,9				

De aquí se observa que $x_1 = 5{,}1$ y $x_n = 9{,}9$ por lo que el rango es igual a $9{,}9 - 5{,}1$ o sea, $r = 4{,}8$; asimismo, se puede notar que el rango da cuenta de la distancia numérica que existe entre el primer dato (el menor) y el último (o mayor).

En función de ello, es que se puede caracterizar a esta medida de dispersión como la más sencilla de obtener, pero también como la más inestable; el segundo punto, se debe a que el valor del rango depende, únicamente, de

[1]Peña, Daniel; et. al. **Introducción a la Estadística para Ciencias Sociales**; Madrid; McGraw-Hill/Interamericana, S. A. V.; 1999; pp. 68 Cursivas agregadas.

los valores extremos, y "Cualquier cambio en alguna de las observaciones, sea el dato más alto o el más bajo, afectará al rango aunque todas las otras observaciones puedan permanecer sin alteración."[2]

- Ejemplo: Calcular el rango de los subsiguientes conjuntos de datos:
  - $A = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$
    - $r = x_n - x_1$
      - $x_n = 20$
      - $x_1 = 2$
    - $r = 20 - 2$
    - $r = 18$
  - $B = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 200\}$
    - $r = x_n - x_1$
      - $x_n = 200$
      - $x_1 = 2$
    - $r = 200 - 2$
    - $r = 198$

Pese a que en ambos casos el número de elementos es el mismo, el rango para ambos es ampliamente diferente. Esto se debe a que en el conjunto B se dio una variación importante con respecto a x_n, cuestión que permite ilustrar 3 cosas:

1. Que hay un alto grado de dispersión en el conjunto B; si éste fuera una muestra, las predicciones realizadas a partir de ella, serían de poca fiabilidad.

[2]KIESS, Harold O.; **Statistical Concepts for the Vehavioral Sciencies**; Allyn and Cacon Inc.; 1989; pp. 100

2. La relativa inestabilidad del rango como medida de dispersión, debida a la sensibilidad que presenta ante los valores elevados de los límites del conjunto ordenado.

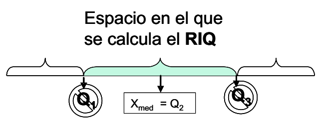
3. Los valores intermedios permanecen (en comparación con el conjunto A) sin cambios o alteraciones, no obstante la variación del rango. Por ello, la interpretación del rango obtenido a partir de un conjunto grande, puede ser difícil de realizar adecuadamente.

Una medida de dispersión, basada en la ordenación, que proporciona mayor nivel de confianza es el *Rango Intercuartil* (RIQ). Éste, permite calcular la distancia que hay entre Q_1 y Q_2, una vez que se han agrupado los datos de una muestra; su expresión formal es la siguiente:

$$\boxed{RIQ = Q_3 - Q_1}$$

donde Q_3 es el tercer cuartil y Q_1 el primero.

"El rango intercuartil proporciona información útil sobre la variabilidad de las observaciones en una distribución, para ello revela el rango de los datos típicos para el 50 por ciento de las individualidades agrupadas al rededor de la mediana. Debido a que el rango intercuartilíco excluye tanto al 25 por ciento inferior como al 25 por ciento superior de las observaciones en una distribución, este valor no es afectado tan fácilmente por observaciones extremas. Obviamente, el rango intercuartil será más corto que el rango para el grupo de datos."[3]



[3]ibidem. KIESS, Harold O.; 1989; pp. 100

Para el caso de las calificaciones de los alumnos de estadística, el RIQ se calcula como sigue:

▷ Datos: $Q_3 = 8{,}23$ y $Q_1 = 6{,}81$

▷ Fórmula: $RIQ = Q_3 - Q_1$

▷ Sustitución: $RIQ = 8{,}23 - 6{,}81$

▷ Resultado: $RIQ = 1{,}42$

Si se divide al rango intercuartil entre dos, se obtiene otra medida de dispersión denominada *Rango Semi-intercuartil* ($RSIQ$), esto es:

$$\boxed{RSIQ = \frac{RIQ}{2} = \frac{Q_3 - Q_1}{2}}$$

"El $RSIQ$ describe el promedio de dispersión que hay en el 25 por ciento de los datos que están por arriba y de bajo de la mediana. Únicamente cuando la distribución es perfectamente simétrica, sin embargo, el $RSIQ$ representará exactamente el rango del 25 por ciento de los datos anteriores a la mediana y el 25 por ciento de los datos que están por debajo de la mediana."[4]

Si se aplica dicho cálculo al ejemplo de las calificaciones, el rango seimi-intercuartil para el 25 por ciento de las calificaciones será $RSIQ = 0{,}71$; esto indica que la mitad del 50 por ciento de los datos, no se extienden más de 0,71 puntos en derredor de la media, tanto por arriba como por debajo.

[4]ibidem. KIESS, Harold O.; 1989; pp. 101

Ejercicios 1

- Instrucciones: Contesta completamente y con brevedad, a las preguntas que a continuación se te formulan. Asimismo, realiza los cálculos necesarios para obtener los resultados que se te solicitan.

- Al finalizar, coteja tus respuestas con las proporcionadas en la página 344

1. ¿Cuál es la relación que guardan las medidas de tendencia central y las medidas de dispersión?

   ____

   ____

   ____

   ____

2. ¿Qué miden o describen las medidas de dispersión?

   ____

   ____

   ____

   ____

   ____

3. En general ¿Cómo se pueden interpretar los resultados numéricos de las medidas de dispersión?

   ____

   ____

   ____

   ____

   ____

4. ¿Cuáles son las medidas de dispersión más comunes y representativas?

   ____

   ____

5. ¿Qué es el rango?

   ____

   ____

   ____

   ____

   ____

   ____

6. Menciona 2 características del rango:

   ____

   ____

   ____

   ____

   ____

   ____

7. Anota las tres expresiones matemáticas mediante las cuales se puede calcular el rango como medida de dispersión.

   ____

   ____

   ____

8. Considera la siguiente tabla de distribución de frecuencias

i	$Clase$	f_i	x_i
1	$[300-400)$	14	350
2	$[400-500)$	46	450
3	$[500-600)$	58	550
4	$[600-700)$	76	650
5	$[700-800)$	68	750
6	$[800-900)$	62	850
7	$[900-1000)$	48	950
8	$[1000-1100)$	22	1050
9	$[1100-1200)$	6	1150
		$n=\sum_{i=1}^{9} f_i = 400$	

y calcula RIQ y $RSIQ$ para estos datos, recuerda que sus cuantiles ya fueron calculados para la sección 4.6 del capítulo anterior.

1) $RIQ =$

2) $RSIQ =$

5.3. Desviación estándar para datos agrupados

La *desviación estándar* es una de las medidas de dispersión más comunes en la estadística. También conocida como *promedio de desviación*, se denota con la letra s. Esta medida brinda el promedio de alejamiento, dispersión o espaciamiento de los datos u observaciones muestrales respecto a la media de la muestra. La desviación estándar para datos agrupados se calcula con la siguiente expresión cuando la muestra es chica ($n < 30$):

$$\boxed{s = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i-\bar{x})^2}{n-1}}}$$

Donde:
k = Número de clases.
f_i = Frecuencia de clase.
x_i = Marca de clase.
$\bar{x}$ = Media para datos agrupados.
n = Número de elementos de la muestra.

Como se observa, los elementos de la fórmula se toman directamente de la tabla de datos agrupados.

- Ejemplo: Considérese la tabla de datos agrupados del ejercicio sobre calificaciones de alumnos de la asignatura de estadística:

i	Clase	f_i	x_i
1	$[5{,}1 - 5{,}6)$	2	5,35
2	$[5{,}6 - 6{,}1)$	4	5,85
3	$[6{,}1 - 6{,}6)$	11	6,35
4	$[6{,}6 - 7{,}1)$	7	6,85
5	$[7{,}1 - 7{,}6)$	19	7,35
6	$[7{,}6 - 8{,}1)$	15	7,85
7	$[8{,}1 - 8{,}6)$	8	8,35
8	$[8{,}6 - 9{,}1)$	7	8,85
9	$[9{,}1 - 9{,}6)$	5	9,35
10	$[9{,}6 - 10{,}1)$	2	9,85
		$n = \sum_{i=1}^{10} f_i = 80$	

  - Calcular la desviación estándar.

Determinación de los valores

$k = 10$
f_i = Frecuencias de clase

x_i = Marcas de clase

$\bar{x} = 7{,}56$ dato que se obtuvo en la sección anterior.

$n = 80$

Sustitución y cálculo

$$s = \sqrt{\frac{\sum_{i=1}^{10} f_i(x_i-\bar{x})^2}{80-1}}$$

sustitución de valores en la suma:

$\sum_{i=1}^{10} f_i(x_i-\bar{x})^2 = 2(5{,}35-7{,}56)^2+4(5{,}85-7{,}56)^2+11(6{,}35-7{,}56)^2+7(6{,}85-7{,}56)^2+19(7{,}35-7{,}56)^2+15(7{,}85-7{,}56)^2+8(8{,}35-7{,}56)^2+7(8{,}85-7{,}56)^2+5(9{,}35-7{,}56)^2+2(9{,}85-7{,}56)^2 = 86{,}348$

Ahora, sustituyendo este resultado en el resto de la fórmula:

$s = \sqrt{\frac{86{,}348}{79}}$

$s = \sqrt{1{,}093012658}$

$s = 1{,}045472457$

Aunque la fórmula empleada para el ejemplo anterior se utiliza debido a sus propiedades matemáticas y estadísticas, también se puede usar –con buenos resultados– la siguiente expresión para el cálculo de la desviación estándar, esta fórmula sólo varía en el denominador cuando la muestra es grande ($n > 30$):

$$\boxed{s = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i-\bar{x})^2}{n}}}$$

Aplicando ésta en el ejercicio anterior, se obtiene:

$s = \sqrt{\frac{86{,}348}{80}}$
$s = \sqrt{1{,}07935}$
$s = 1{,}038917706 \approx 1{,}04$

Como puede observarse, los resultados obtenidos con las diferentes fórmulas varían poco entre sí.

Una vez que se han calculado la media y la desviación estándar, se procede al reconocimiento de la dispersión que guardan los datos con respecto de la media. La expresión formal que permite realizar la inferencia en torno a la precisión estadística es la siguiente:

$$\boxed{(\bar{x} - s) < \bar{x} < (\bar{x} + s)}$$

Donde para sustituir los valores en la expresión, se debe realizar la siguiente operación entre los factores en cuestión ($\bar{x} = 7{,}56$ y $s = 1{,}04$):
$7{,}56 \pm 1{,}04$
$= 7{,}56 + 1{,}04 = 8{,}6$
$= 7{,}56 - 1{,}04 = 6{,}52$
resultados que en la expresión original se denotan así:

$$6{,}52 < \bar{x} < 8{,}6$$

y se interpretan de la siguiente manera:

Se tiene que el valor de la desviación típica correspondiente a los datos de las calificaciones en cuestión, indica que los estudiantes de estadística obtuvieron calificaciones que se encuentran entre 6,52 y 8,6; por lo que se puede afirmar que la muestra tiene un alto grado de precisión en cuanto a las observaciones que se recogieron. Si el resultado de los anteriores cálculos indicara que los estudiantes obtuvieron calificaciones entre 0 y 15,12 con toda seguridad se podría afirmar que (si los cálculos se realizaron correctamente) la muestra está mal diseñad, es decir, carece por completo de precisión.

En el primer caso, la desviación estándar de 1,04 indica que la muestra es precisa, pues los datos tienen una baja dispersión con respecto de la media. En el supuesto de que $s = 7{,}56$ se ha establecido que no hay precisión alguna, pues los estudiantes de la muestra, habrían obtenido calificaciones que van de 0 a de 15,12 y eso no tiene sentido –aún cuando fueran de 0 a 10, eso no diría nada, pues cualquiera lo podría afirmar sin necesidad de cálculo alguno. Finalmente, cabe destacar que si la desviación estándar fuese igual a cero ($s = 0$), nos encontraríamos ante un fenómeno determinista, pues el resultado sería una constante.

5.4. La varianza para datos agrupados

La *varianza* es una medida de dispersión íntimamente relacionada con la desviación estándar, dado que es el cuadrado de ésta. La fórmula de la varianza se expresa por:

$$\boxed{s^2 = \frac{\sum_{i=1}^{k} f_i(x_i - \bar{x})^2}{n-1}}$$

Donde:

k = Número de clases.

f_i = Frecuencia de clase.

x_i = Marca de clase.

$\bar{x}$ = Media para datos agrupados.

n = Número de elementos de la muestra.

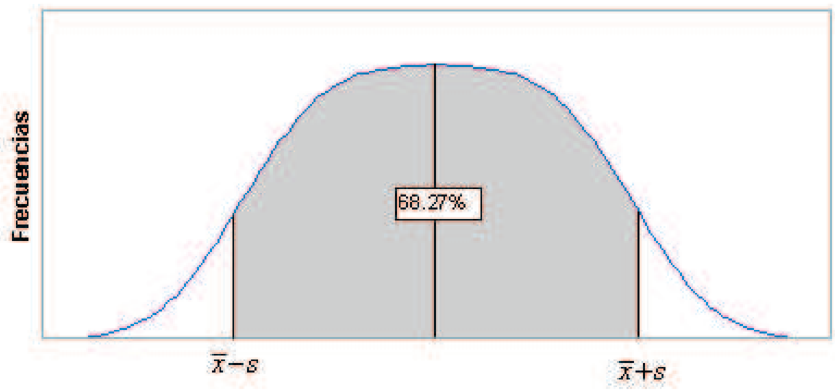
Como se observa, los elementos de la fórmula, se toman directamente de la tabla de datos agrupados. Sin embargo, se suele calcular primero la desviación estándar y posteriormente se eleva ésta al cuadrado para obtener la varianza directamente de la desviación estándar.

- Ejemplo: Retomando el caso de las calificaciones de alumnos en la asignatura de estadística, se obtuvo una desviación estándar de 1,045472457; por lo tanto la varianza es:

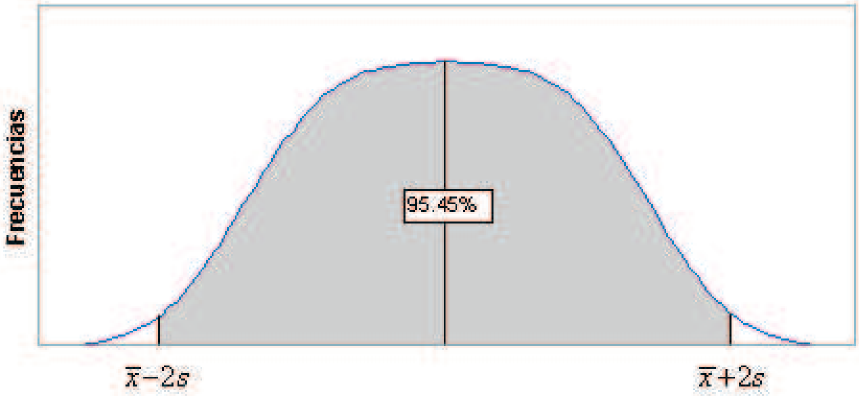
$s^2 = 1{,}045472457^2$
$s^2 = 1{,}093012658$

5.5. Teorema de Chebyshev

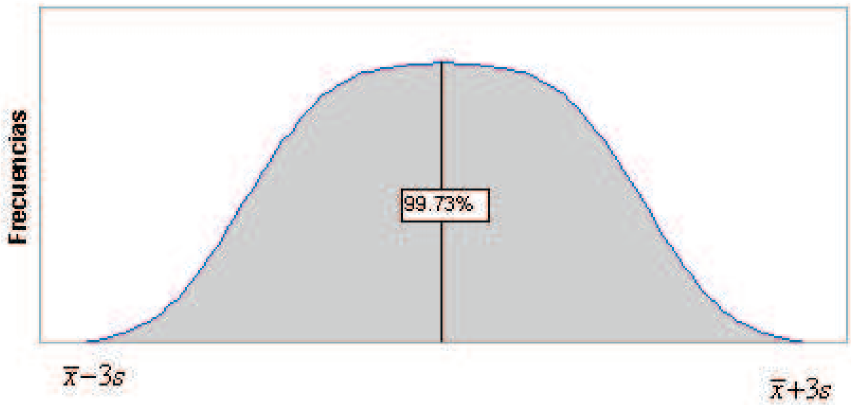
Directamente relacionado con algunas propiedades matemáticas de la desviación estándar, se encuentra el teorema del Matemático ruso Pafnuti Lvovich Chebyshev; éste indica que: En distribuciones de datos u observaciones muestrales con *asimetría baja o moderada*, se puede establecer una relación entre la media y la desviación estándar, tal que:
a) El 68.27 % de las observaciones se encuentran entre $\bar{x} - s$ y $\bar{x} + s$



b) El 95.45 % de las observaciones se encuentran entre $\bar{x} - 2s$ y $\bar{x} + 2s$



c) El 99,73 % de las observaciones se encuentran entre $\bar{x} - 3s$ y $\bar{x} + 3s$



En otras palabras, para el primer caso ($\bar{x} - s$ y $\bar{x} + s$), si al valor que se obtuvo al calcular la media por una parte se le suma y por otra se le resta el de la desviación estándar ($\bar{x} \pm s$), el 68,27 % de los datos que conforman la muestra, se encontrarán entre los valores obtenidos. Esta misma lógica aplica para la comprensión de los casos *b*) y *c*), con la diferencia de que al valor de la media se le sumarán 2 y 3 veces, respectivamente, el de la desviación estándar, y el porcentaje de datos que se encuentran entre dichos valores cambiará de acuerdo a como se expresó en la relación.

- Ejemplo: Como se recordará, para el caso de las calificaciones de los estudiantes que cursaron la asignatura de estadística, los valores de la media y la desviación estándar fueron: $s = 1{,}038917706 \approx 1{,}04$ y $\bar{x} = 7{,}56$. Al sumar y restar una vez el valor de la desviación estándar a la media, resultado fue: $6{,}52 < \bar{x} < 8{,}6$ y sirvió para determinar la precisión de las observaciones contenidas en la muestra; si se aplica el teorema de Chebyshev, se puede afirmar que entre dichos valores (6,52 y 8,6), se encuentran el 68,27 % de los datos que conforman la muestra y ello se ilustra con la gráfica del caso a)

Ejercicios 2

- *Instrucciones*: Resuelve los ejercicios propuestos y compara tus respuestas con las de la sección correspondiente (página 346)

1. Une los signos y fórmulas de la columna izquierda con los términos de la derecha.

$s^2 = \frac{\sum_{i=1}^{k} f_i(x_i-\bar{x})^2}{n-1}$	Media para datos agrupados
$k =$	Marca de clase
$f_i =$	Número de elementos de la muestra.
$x_i =$	Varianza.
$\bar{x} =$	Frecuencia de clase.
$s = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i-\bar{x})^2}{n}}$	Número de clases.
n	Desviación estándar.

2. Calcula la *varianza* y la *desviación estándar* para el siguiente conjunto de datos agrupados:

i	$Clase$	f_i	x_i
1	$[300 - 400)$	14	350
2	$[400 - 500)$	46	450
3	$[500 - 600)$	58	550
4	$[600 - 700)$	76	650
5	$[700 - 800)$	68	750
6	$[800 - 900)$	62	850
7	$[900 - 1000)$	48	950
8	$[1000 - 1100)$	22	1050
9	$[1100 - 1200)$	6	1150
		$n = \sum_{i=1}^{9} f_i = 400$	

a) $s^2 =$

b) $s =$

c) Interpretar el resultado de la desviación estándar.

____

____

____

____

____

____

3. Explica con tus propias palabras, y mediante la gráfica correspondiente, la relación que se verifica para el conjunto de datos conformado por las calificaciones de los estudiantes de estadística cuando $\bar{x} - 2s$ y $\bar{x} + 2s$

   ____

   ____

   ____

   ____

   ____

   ____

5.6. Autoevaluación

- *Instrucciones*: Calcula lo que se te pide y es necesario; para ello, considera el siguiente caso:

  El análisis de la comunicación en las organizaciones (instituciones, empresas, organizaciones no gubernamentales y más), puede ser una herramienta de gran relevancia para conseguir el éxito en los objetivos que éstas persiguen. A continuación, se presenta un ejemplo de la clase de investigaciones que en dicho tenor se pueden realizar.

  - Se desea conocer la opinión pública del personal de *Mexalite Industrial S. A. de C. V.*, Planta Santa Clara, sobre el sistema de calidad ISO 9000 implantado en el año 2000, para sentar un precedente comunicativo y constituirlo como el principio de una constante participación activa de todo el personal, acerca del mejoramiento de sistema de calidad implantado, de tal forma que se llegue a la calidad total.[5]

  Para conseguir lo anterior, se diseñó una herramienta (cuestionario) que se aplicó a una muestra de 149 individuos. Entre las variables sociodemográficas que la investigación maneja, se encuentran las edades de los trabajadores (empleados y sindicalizados) que en la planta se desempeñan; éstas, fluctúan entre los 18 y 61 años, según reporta la investigadora en la sección *Levantamiento de datos*[6]

[5]FRAGOSO, Almaraz Cruz Elvira **Opinión Pública del personal de *Mexalite Industrial S. A. de C. V.* Planta Santa Clara, Sobre el Sistema de Calidad**, Seminario - Taller extracurricular; México; UNAM-ENEP Acatlán; 2001; pp. II, III

[6]Ibidem. FRAGOSO, Almaraz Cruz Elvira; México; UNAM-ENEP Acatlán; 2001; pp. 142, 152

A partir de los datos reportados en la investigación, se sabe que la composición de la muestra (a propósito de las edades de quienes la integraron) es la siguiente:

1. los individuos cuyas edades fluctúan entre los 18 y 28 años de edad, son 20
2. aquellos cuya edad está entre los 29 y los 39, son 74
3. el grupo de personas que posee una edad que oscila entre los 40 y 50 es igual a 47
4. y entre los 51 y 61 años de edad, sólo hay 8 personas en la muestra.

Ahora bien: Supóngase que es necesario describir el comportamiento de la muestra en función de las edades que los individuos tienen y que éstas, fueron las siguientes:

33	34	35	38	39	37	36	29	32	30
31	31	37	31	31	37	36	35	37	31
35	36	34	35	39	37	30	39	31	38
33	39	39	34	39	37	36	35	32	29
35	39	35	31	38	30	37	39	38	39
36	37	33	31	34	38	38	34	37	34
34	36	33	36	35	34	35	32	34	34
39	34	29	29	41	46	45	45	47	47
50	50	50	41	46	43	48	44	49	40
47	49	40	49	50	50	41	43	50	45
49	42	40	40	47	45	44	43	45	46
40	49	47	46	49	44	44	49	47	46
50	28	28	24	25	25	20	21	26	23
22	20	18	22	24	22	22	24	22	27
51	51	58	59	57	56	58	51	18	

1. Efectúa el tratamiento estadístico que permita calcular las diferentes medidas de dispersión que en éste capítulo se revisaron.

   a) Ordenar los datos

   b) Diseñar tabla de datos agrupados

   c) Calcular medidas de tendencia central necesarias.

2. Calcular los siguientes estadígrafos, pertinentes a la edad de los trabajadores, para el conjunto de datos que se te presentan.

   a) Rango semi-intercuartil.

   b) Desviación típica para datos agrupados.

   c) Varianza para datos agrupados.

   d) Con base en los resultados de s y $\bar{x}$, anota tus observaciones respecto a la muestra.

   ____

   ____

   ____

   ____

   ____

   ____

   ____

3. Aplica la regla de Chebyshev y explica, con tus propias palabras, la relación que se verifica para el conjunto de datos conformado por las edades de los trabajadores cuando $\bar{x} - 3s$ y $\bar{x} + 3s$

   ____

   ____

   ____

   ____

Coteja tus respuestas, con las de la página 348

5.7. Apéndice del capítulo

5.7.1. Respuestas a los ejercicios del capítulo

Ejercicios 1

- *Instrucciones*: Contesta completamente y con brevedad, a las preguntas que a continuación se te formulan. Asimismo, realiza los cálculos necesarios para obtener los resultados que se te solicitan.

1. ¿Cuál es la relación que guardan las medidas de tendencia central y las medidas de dispersión?
   Estas medidas son complementarias al momento de evaluar las características de una muestra, es decir, permiten saber si la muestra está bien diseñada (conocer su sesgo y la precisión estadística de las observaciones).

2. ¿Qué miden o describen las medidas de dispersión?
   El grado de *variación* –la distancia en la distribución– de las observaciones o datos muestrales respecto a la media, es decir, qué tan separados están de $\bar{x}$, lo que permite conocer qué tan bien representados están los datos típicos en el conjunto de datos que se está analizando.

3. En general, ¿Cómo se pueden interpretar los resultados numéricos de las medidas de dispersión?
   A mayor dispersión, menor precisión estadística y a menor dispersión, mayor precisión, es decir, que si existe un alto grado de dispersión en las observaciones, esto es señal de que los datos típicos del conjunto en cuestión, no están bien representados.

4. ¿Cuáles son las medidas de dispersión más comunes y representativas? Rango; Desviación estándar (desviación típica o promedio de dispersión) y la varianza.

5. ¿Qué es el rango?
Es la medida de dispersión más sencilla de calcular ya que se basa en la ordenación de las observaciones. El rango mide la distancia numérica que hay entre las observaciones o datos de menor y mayor valor. Éste se puede calcular para datos no agrupados y agrupados, en el segundo caso existen por lo menos dos modalidades: 1) El rango intercuartil y 2) El rango semi-intercuartil.

6. Menciona 2 características del rango:

   1) Es la medida de dispersión más sencilla de calcular.
   2) Es una medida de dispersión relativamente inestable, debido a que sólo considera los valores extremos y cualquier variación en alguno de ellos, afectará el resultado de su cálculo (aunque las observaciones intermedias permanezcan sin cambios).

7. Anota las tres expresiones matemáticas mediante las cuales se puede calcular el rango como medida de dispersión.

   1) $r = x_n - x_1$
   2) $RIQ = Q_3 - Q_2$
   3) $RSI = \frac{Q_3 - Q_2}{2}$

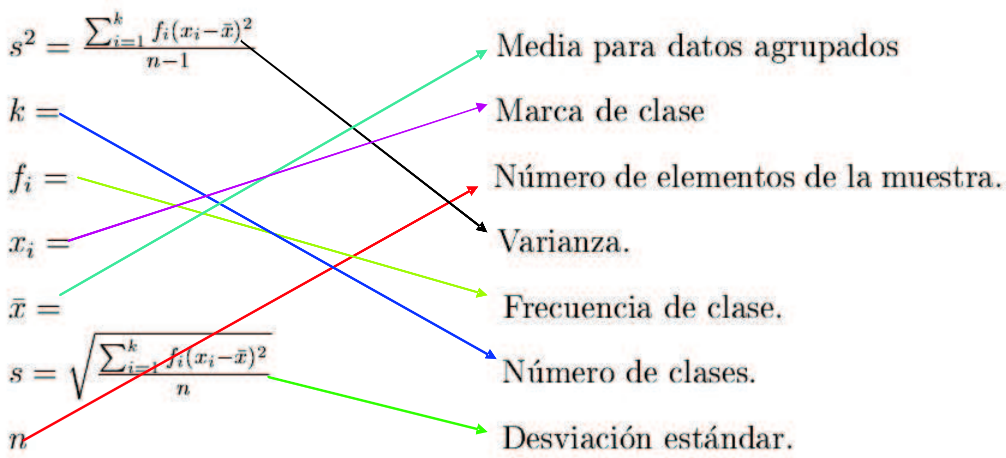
8. Considera la siguiente tabla de distribución de frecuencias...y calcula RIQ y $RSIQ$ para éstos datos, recuerda que sus cuantiles ya fueron calculados para la sección 4.6 del capítulo anterior.

   1) $RIQ = 292{,}32$
   2) $RSIQ = 146{,}16$

Ejercicios 2

- *Instrucciones*: Resuelve los ejercicios propuestos y compara tus respuestas con las de la sección correspondiente.

1. Une los signos y fórmulas de la columna izquierda con los términos de la derecha.



2. Calcula la *varianza* y la *desviación estándar* para el siguiente conjunto de datos agrupados...

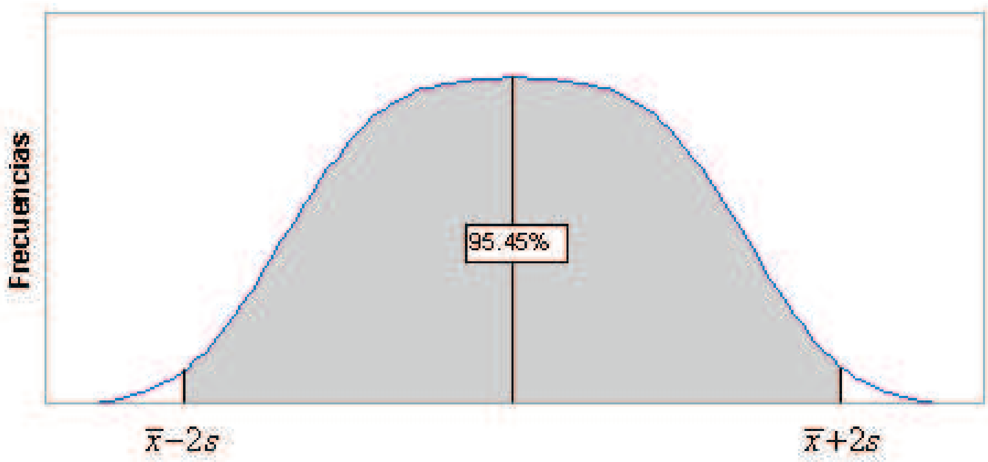
   a) $s^2 = 7655359{,}218$

   b) $s = 2766{,}83$

   c) Interpretar el resultado de la desviación estándar.
   Puesto que $\bar{x} = 715{,}5$ y $s = 2766{,}83$ es posible de determinar que la dispersión de los datos con respecto de la media es muy alta, como se puede ver en la esta expresión $-2051{,}33 < 715{,}5 < 3482{,}33$; si se tratase de una muestra, ésta sería sumamente imprecisa.

3. Explica con tus propias palabras, y mediante la gráfica correspondiente, la relación que se verifica para el conjunto de datos conformado por las calificaciones de los estudiantes de estadística cuando $\bar{x} - 2s$ y $\bar{x} + 2s$

   Si a la desviación estándar de la muestra (que presenta una baja asimetría) se le resta 2 veces el valor de la desviación estándar ($\bar{x} - 2s \equiv 7{,}56 - 2(1{,}04) = 5{,}48$) y posteriormente se le suma, al valor original de la media, dos veces el de la desviación típica ($\bar{x} + 2s \equiv 7{,}56 + 2(1{,}04) = 9{,}64$), es posible afirmar que el 95,45 % de las observaciones se encuentran entre dichos valores. En otras palabras, de los 80 estudiantes que conforman la muestra, 76 (95,45 %) obtuvieron calificaciones que van de 5,48 a 9,64.



5.7.2. Respuestas de la autoevaluación

- *Instrucciones*: Calcula lo que se te pide y es necesario; para ello, considera el siguiente caso...

1. Efectúa el tratamiento estadístico que permita calcular las diferentes medidas de dispersión que en éste capítulo se revisaron.

   a) Ordenar los datos

18	18	20	20	21	22	22	22	22	22
23	24	24	24	25	25	26	27	28	28
29	29	29	29	30	30	30	31	31	31
31	31	31	31	31	32	32	32	33	23
33	33	34	34	34	34	34	34	34	34
34	34	34	35	35	35	35	35	35	35
35	35	36	36	36	36	36	36	36	37
37	37	37	37	37	37	37	37	38	38
38	38	38	38	39	39	39	39	39	39
39	39	39	39	40	40	40	40	40	41
41	41	43	43	43	43	44	44	44	44
45	45	45	45	45	46	46	46	46	46
47	47	47	47	47	47	48	49	49	49
49	49	49	49	50	50	50	50	50	50
50	51	51	51	56	57	58	58	59	

b) Diseñar tabla de datos agrupados.

i	clase	f_i	x_i
1	$18 - 21{,}6$	5	19,8
2	$21{,}6 - 25{,}2$	11	23,4
3	$25{,}2 - 28{,}8$	4	27
4	$28{,}8 - 32{,}4$	18	30,6
5	$32{,}4 - 36$	24	34,2
6	$36 - 39{,}6$	32	37,8
7	$39{,}6 - 43{,}2$	12	41,4
8	$43{,}2 - 46{,}8$	14	45
9	$46{,}8 - 50{,}4$	21	48,6
10	$50{,}4 - 54$	3	52,2
11	$54 - 57{,}6$	2	55,8
12	$57{,}6 - 61{,}2$	3	59,4

c) Calcular medidas de tendencia central necesarias.

$\bar{x} = 37{,}85$
$Q_1 = 32{,}25$
$Q_3 = 45{,}32$

2. Calcular los siguientes estadígrafos, pertinentes a la edad de los trabajadores, para el conjunto de datos que se te presentan.

   a) Rango semi-intercuartil $RSIQ = 6{,}535 \approx 6{,}54$

   b) Desviación típica para datos agrupados $s = 8{,}827 \approx 8{,}83$

   c) Varianza para datos agrupados $s^2 = 2{,}971 \approx 2{,}97$

   d) Con base en los resultados de s y $\bar{x}$, anota tus observaciones respecto a la muestra.

Si $s = 8{,}83$ y recordamos que $\bar{x} = 37{,}85$ entonces podemos afirmar que la dispersión de los datos con respecto a la media es elevada $(29{,}02 < \bar{x} < 46{,}68)$, pues el $66{,}27\,\%$ de los trabajadores se encuentran entre los 29 y los 47 años, cuando la mínima y la máxima son 18 y 59, respectivamente.

3. Aplica la regla de Chebyshev y explica, con tus propias palabras, la relación que se verifica para el conjunto de datos conformado por las edades de los trabajadores cuando $\bar{x} - 3s$ y $\bar{x} + 3s$

   Cuando $\bar{x} - 3s \equiv 37{,}85 - 3(8{,}83)$ y $\bar{x} - 3s \equiv 37{,}85 + 3(8{,}83)$, el $99{,}73\,\%$ de los datos (las edades de los trabajadores), se encuentra entre 11 y 64 años, es decir, que de los 149 trabajadores, $148{,}5977 \approx 149$ tienen edades entre 11 y 64 años. Afirmación que es totalmente verdadera –aunque de poca o nula utilidad– ya que se sabe que la muestra está integrada por personas cuyas edades están entre los 18 y 61 años.

Capítulo 6

Momentos

Objetivos específicos

- El estudiante calculará los *momentos de orden* r y los *momentos no centrales*, mediante la aplicación de sus respectivas fórmulas, para comprobar la relación de la media y la varianza con dichas magnitudes.

- El estudiante reconocerá elementos generales de las *medidas de asimetría* y forma, tales como *sesgo* y *curtosis*, mediante la lectura del texto y aplicación de la teoría en la resolución de los ejercicios correspondientes al presente capítulo.

6.1. Preliminar

En el contexto de la estadística, un *momento* es una magnitud específica que se calcula sobre un conjunto de datos u observaciones. Puede establecerse una clara relación entre algunos momentos y ciertos estadígrafos muestrales, específicamente, la media y la varianza.

En el presente capítulo, se conocerá y practicará el cálculo momentos para datos agrupados y no agrupados, así como su relación con las medidas anteriormente mencionadas; además, se revisarán elementos generales de las medidas de asimetría y forma como el sesgo y la curtosis.

6.2. Momentos de orden r para datos no agrupados

Sean $x_1, x_2, \ldots x_n$ los n elementos u observaciones de una muestra.

- Se define al momento de orden r, de la siguiente manera:

$$\boxed{\overline{x^r} = \frac{\sum_{i=1}^{n} x_i^r}{n} = \frac{x_1^r + x_2^r + x_3^r \ldots x_n^r}{n}}$$

Por ejemplo, sea $X = \{-1, 0, 1, 3, 4, 5, 7, 7, 8\}$. Determina los momentos de orden: $1, 2$ y 3.

Solución. Se tienen nueve elementos muestrales u observaciones, por lo que $n = 9$. Entonces, el momento de orden 1 implica que $r = 1$, por tanto:

$$\overline{x^1} = \frac{\sum_{i=1}^{9} x_i^1}{9}$$

$$\bar{x} = \frac{\sum_{i=1}^{9} x_i}{9}$$

Sustituyendo los valores en la fórmula:
$\bar{x} = \frac{-1+0+1+3+4+5+7+7+8}{9}$
$\bar{x} = \frac{34}{9}$
$\bar{x} = 3{,}7778$, redondeado a 4 decimales.

‡ *Obsérvese que el momento de orden 1, es equivalente a la media para datos no agrupados.*

Para el caso del momento de orden dos, se tiene que $r = 2$, por lo que:

$$\overline{x^2} = \frac{\sum_{i=1}^{9} \ x_i^2}{9}$$

Desarrollando la expresión, se tiene:
$\overline{x^2} = \frac{-1^2+0^2+1^2+3^2+4^2+5^2+7^2+7^2+8^2}{9}$
$\overline{x^2} = \frac{1+1+9+16+25+49+49+64}{9}$
$\overline{x^2} = \frac{214}{9}$
$\overline{x^2} = 23{,}7778$

Finalmente, para calcular el momento de orden tres, se tiene que $r = 3$, por tanto:

$$\overline{x^3} = \frac{\sum_{i=1}^{9} \ x_i^3}{9}$$

Desarrollada la expresión:
$\overline{x^3} = \frac{-1^3+0^3+1^3+3^3+4^3+5^3+7^3+7^3+8^3}{9}$
$\overline{x^3} = \frac{1+1+27+64+125+343+343+512}{9}$
$\overline{x^3} = \frac{1414}{9}$
$\overline{x^3} = 157{,}11$

6.3. Momento de orden r respecto a la media ($\bar{x}$)

Definición:

Sean $x_1, x_2, \ldots x_n$, las n observaciones o elementos de una muestra. El momento de orden r con respecto a la media ($\bar{x}$) o *momento r no central*, se define por la siguiente expresión:

$$\boxed{m_r = \frac{\sum_{i=1}^{n} (xi-\bar{x})^r}{n} = (x_1 - \bar{x})^r + (x_2 - \bar{x})^r + \ldots + (x_n - \bar{x})^r}$$

Donde $\bar{x}$ corresponde a la media para datos no agrupados. Por ejemplo: Calcular el primer y segundo momentos de orden $r = 1$ y $r = 2$ no centrales para los datos del ejemplo anterior.

Solución. Como puede comprobarse, $\bar{x} = 3{,}78$. Entonces:

$$m_1 = \frac{\sum_{i=1}^{9} (x_i-\bar{x})^1}{9}$$

Desarrollando la expresión anterior:

$$m_1 = \frac{(-1-3{,}78)+(0-3{,}78)+(1-3{,}78)+(3-3{,}78)+(4-3{,}78)+(5-3{,}78)+(7-3{,}78)+(7-3{,}78)+(8-3{,}78)}{9}$$
$$m_1 = \frac{-4{,}78+(-3{,}78)+(-2{,}78)+(-0{,}78)+0{,}22+1{,}22+3{,}22+3{,}22+4{,}22}{9}$$
$$m_1 = \frac{-0{,}02}{9}$$
$$m_1 = -0{,}002222$$

Por lo que respecta al segundo momento de orden no central de orden dos para datos no agrupados, se tiene que:

$$m_2 = \frac{\sum_{i=1}^{9} (x_i-\bar{x})^2}{9}$$

Desarrollando la expresión anterior:

$m_2 = \frac{(-1-3{,}78)^2+(0-3{,}78)^2+(1-3{,}78)^2+(3-3{,}78)^2+(4-3{,}78)^2+(5-3{,}78)^2+(7-3{,}78)^2+(7-3{,}78)^2+(8-3{,}78)^2}{9}$
$m_2 = \frac{(-4{,}72)^2+(-3{,}78)^2+(-2{,}78)^2+(-0{,}78)^2+(0{,}22)^2+(1{,}22)^2+(3{,}22)^2+(3{,}22)^2+(4{,}22)^2}{9}$
$m_2 = \frac{22{,}848+14{,}288+7{,}728+0{,}608+0{,}048+1{,}488+10{,}368+10{,}368+17{,}808}{9}$
$m_2 = \frac{85{,}552}{9}$
$m_2 = 9{,}506$

‡ *Obsérvese que cuando $r = 2$, es decir al calcular m_2, equivale a determinar la varianza s^2, para datos no agrupados. En consecuencia, a la varianza también se le conoce como segundo momento no central.*

Ejercicios 1

- *Instrucciones*: Para el siguiente conjunto de datos: $X = \{1, 1, 2, 3, 4, 5\}$ Calcular los momentos que se te solicitan:

  1. 2^o Momento.
  2. 3^{er} Momento.
  3. 1^{er} Momento no central.
  4. 2^o Momento no central.

- Para asegurarte que los resultados obtenidos son correctos, comparalos con los que se te proporcionan en el apartado **6.7.1.** (página 370)

6.4. Momentos para datos agrupados

Si $x_1, x_2, \ldots x_k$, son k marcas de clase, con las respectivas frecuencias: $f_1, f_2 \ldots f_k$, se calculan los momentos y los momentos no centrales para datos no agrupados mediante las siguientes expresiones:

Momento de orden r para datos agrupados:

$$\boxed{\overline{X^r} = \frac{\sum_{i=1}^{k}}{n} = \frac{f_1 x_1^r + f_2 x_2^r \ldots + f_k x_k^r}{n}}$$

Donde $n =$ número de elementos u observaciones de la muestra.

‡ *Nótese que si $r = 1$, x^r equivale al cálculo de la media para datos agrupados.*

Para *momentos con respecto a la media* o no centrales para datos agrupados, éstos se calculan mediante la siguiente expresión:

$$\boxed{m_r = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^r}{n} = \frac{f_1 (x_i - \bar{x})^r + f_2 (x_2 + \bar{x})^r + \ldots + f_k (x_k - \bar{x})^r}{n}}$$

Donde $n =$ número de elementos de la muestra.

- Ejemplo: Retomando los datos del ejemplo de las calificaciones de alumnos en la asignatura de Estadística:

i	Clase	f_i	x_i
1	[5,1 − 5,6)	2	5,35
2	[5,6 − 6,1)	4	5,85
3	[6,1 − 6,6)	11	6,35
4	[6,6 − 7,1)	7	6,85
5	[7,1 − 7,6)	19	7,35
6	[7,6 − 8,1)	15	7,85
7	[8,1 − 8,6)	8	8,35
8	[8,6 − 9,1)	7	8,85
9	[9,1 − 9,6)	5	9,35
10	[9,6 − 10,1)	2	9,85
		$n = \sum_{i=1}^{10} f_i = 80$	

Calcular:

1. Primer momento.
2. Segundo momento.
3. Primer momento no central.
4. Segundo momento no central.

Tomando los datos de la tabla anterior:
$\overline{X^1}$:
$\overline{X^1} = \frac{2(5,35)+4(5,85)+11(6,35)+7(6,85)+19(7,35)+15(7,85)+8(8,35)+7(8,85)+5(9,35)+2(9,85)}{80}$
$\overline{X^1} = \frac{10,7+23,4+69,85+47,95+139,65+117,75+66,8+61,95+46,75+19,7}{80}$
$\overline{X^1} = \frac{604,5}{80}$
$\overline{X^1} = 7,56$, redondeado a dos decimales.

‡ *Como se puede ver, cuando* $r = 1$, $\overline{X^1}$ *equivale a la media* ($\bar{x}$) *para datos agrupados.*

Ahora, para el segundo momento:

$\overline{X^2}$:

$\overline{X^2} = \frac{2(5{,}35)^2+4(5{,}85)^2+11(6{,}35)^2+7(6{,}85)^2+19(7{,}35)^2+15(7{,}85)^2+8(8{,}35)^2+7(8{,}85)^2+5(9{,}35)^2+2(9{,}85)^2}{80}$

$\overline{X^2} = \frac{57{,}25+136{,}89+443{,}55+328{,}46+1026{,}43+924{,}34+557{,}78+548{,}26+437{,}11+194{,}05}{80}$

$\overline{X^2} = \frac{4654{,}1}{80}$

$\overline{X^2} = 58{,}18$

Para el primer momento no central para datos agrupados se tiene:

m_1:

$m_1 = \frac{2(5{,}35-7{,}56)+4(5{,}85-7{,}56)+11(6{,}35-7{,}56)+7(6{,}85-7{,}56)+19(7{,}35-7{,}56)+15(7{,}85-7{,}56)+...}{80}$

$m_1 = \frac{-4{,}42+(-6{,}84)+(-13{,}31)+(-4{,}97)+(-3{,}99)+4{,}35+6{,}32+9{,}03+8{,}95+4{,}58}{80}$

$m_1 = \frac{-0{,}3}{80}$

$m_1 = -0{,}00375$

Finalmente para el segundo momento no central para datos agrupados:

m_2:

$m_2 = \frac{2(5{,}35-7{,}56)^2+4(5{,}85-7{,}56)^2+11(6{,}35-7{,}56)^2+7(6{,}85-7{,}56)^2+19(7{,}35-7{,}56)^2+...}{80}$

$m_2 = \frac{86{,}348}{80}$

$m_2 = 1{,}07935$

‡ *Se observa que cuando* $r = 2$, m_2 *equivale a la varianza* s^2

Ejercicios 2

- *Instrucciones*: Calcular para los datos del ejemplo anterior:

  1. 3^{er} Momento.
  2. 4° Momento.
  3. 3^{er} Momento no central.
  4. 4° Momento no central.

- Ahora confronta tus respuestas con las que se te brindan en la página 370

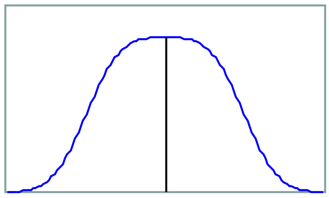
6.5. Medidas de asimetría y forma

Así como las medidas de tendencia central cuantifican el grado de concentración de los datos respecto a un valor, y las medidas de variación, por su parte, determinan el grado de dispersión de los datos respecto a un valor, ambas medidas pueden emplearse para cuantificar la asimetría (falta de simetría) y la forma de la gráfica de un conjunto de datos muestrales. Las medidas de asimetría miden el *sesgo* de una distribución de datos muestrales, mientras que la *curtosis* determina qué tan puntiaguda es la gráfica de una distribución de datos muestrales.

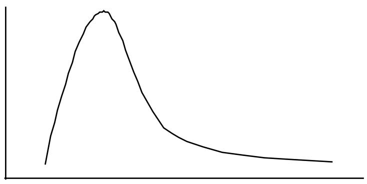
Las medidas de asimetría y forma, son complementarias a las de tendencia central y dispersión. Mediante el cálculo, observación y estudio de las medidas de asimetría y forma (conjuntamente a las anteriormente enunciadas), se pueden conocer con mayor precisión las características de una muestra y, sobre dicha base, determinar si ésta fue bien diseñada y por tanto si será representativa de la población o no.

6.5.1. Sesgo

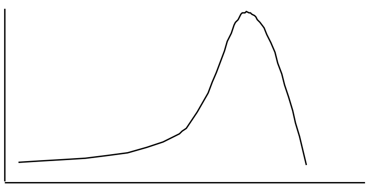
El sesgo es el grado de asimetría de una distribución de datos u observaciones muestrales. Una distribución sin sesgo, presenta una simetría notable.



Por otra parte, si la curva de frecuencias (polígono suavizado) tiene una *cola* más larga a la derecha del máximo central que a la izquierda, implica que la distribución está *sesgada a la derecha*, es decir, tiene *sesgo positivo*:



En caso contrario, si su cola izquierda es más larga que su cola derecha, se dice que tiene sesgo negativo o a la izquierda:



Coeficiente de sesgo intercuartílico

El *coeficiente de sesgo intercuartílico* es una medida para cuantificar el sesgo de una distribución de datos u observaciones muestrales y se calcula con la siguiente expresión:

$$\boxed{\text{Coeficiente de sesgo intercuartílico} = \frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1} = \frac{Q_3-2Q_2+Q_1}{Q_3-Q_2}}$$

Donde Q_1, Q_2 y Q_3 son el primero, segundo y tercer cuartil, respectivamente.

Cabe destacar que el cálculo del coeficiente

Coeficiente de sesgo percentílico

Adicionalmente a la anterior, se tiene otra medida de sesgo, a saber, el *coeficiente de sesgo percentlico*:

$$\boxed{\text{Coeficiente de sesgo percentílico} = \frac{(P_{90}-P_{50})-(P_{50}-P_{10})}{P_{90}-P_{10}} = \frac{P_{90}-2P_{50}+P_{10}}{P_{90}-P_{10}}}$$

Donde P_{10}, P_{50} y P_{90}, son los percentiles $10, 50$ y 90 respectivamente.

- Ejemplos: Para los datos del ejemplo de las calificaciones de alumnos en la asignatura de estadística, calcular:

  1. El coeficiente de sesgo intercuartílico
  2. El coeficiente de sesgo percentílico

  Solución:

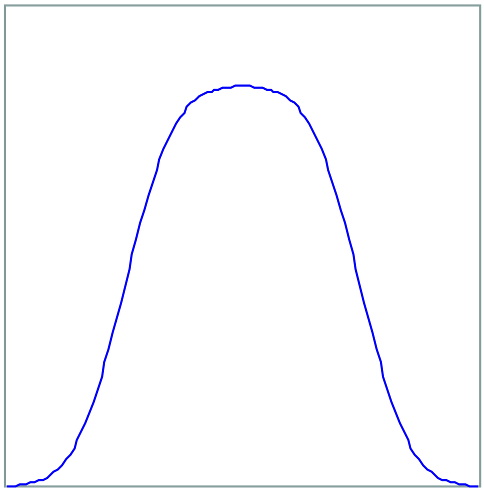
  1. De la sección correspondiente a medidas de tendencia central, en la sección de cuartiles, se obtuvieron los valores:
     $Q_1 = 6{,}81$
     $Q_2 = 7{,}52$
     $Q_3 = 8{,}23$

Los cuales son sustituidos en la fórmula respectiva:
Coeficiente de sesgo intercuartílico=$\frac{8{,}23-2(7{,}52)+6{,}81}{8{,}23-7{,}52}$
$= \frac{0}{0{,}71}$
$= 0$
Un coeficiente cero o cercano a cero, significa que la distribución de los datos presenta un sesgo muy bajo.

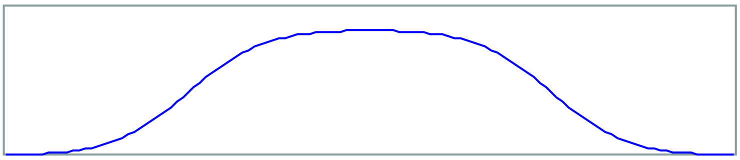
2. Aplicando las fórmulas para el cálculo de percentiles, revisadas en el capítulo de medidas de tendencia central y considerando que el percentil 90 es igual a la mediana y al segundo cuartil, se tiene que:
$P_{10} = 6{,}19$
$P_{50} = 7{,}52$
$P_{90} = 9{,}03$
Los cuales son sustituidos en la fórmula correspondiente:
Coeficiente de sesgo percentílico=$\frac{9{,}03-2(7{,}52)+6{,}19}{9{,}03-6{,}19}$
$= \frac{0{,}18}{2{,}84}$
$= 0{,}06$
Que como puede observarse, es un valor cercano a cero, lo que indica un sesgo poco pronunciado o muy pequeño.

6.5.2. Curtosis

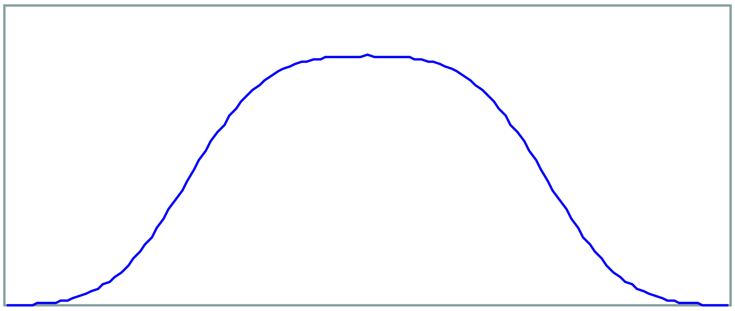
La *curtosis* es una medida asociada a la forma de una distribución de datos u observaciones muestrales. Dicha medida cuantifica qué tan *puntiaguda* es el polígono de frecuencias suavizado de una distribución de datos, la cual puede asumir una de las siguientes formas generales: Leptocúrtica, Platicúrtica y Mesocúrtica.



Leptocúrtica



Platicúrtica



Mesocúrtica

Una de las principales medidas de curtosis es el *coeficiente de momento de curtosis*, el cual se denota con el símbolo a_4 y se calcula mediante la determinación del segundo y cuarto momentos no centrales para datos agrupados:

$$\boxed{a_4 = \frac{m_4}{m_2^2}}$$

Ejemplo: Determine el coeficiente de momento de curtosis para los datos del ejemplo de calificaciones de alumnos de la asignatura de estadística:
Solución. Como ya se expresó, el segundo momento no central, es equivalente a la varianza de la distribución de datos, calculada con la fórmula:

$$\boxed{s^2 = \frac{\sum_{i=1}^{k} f_i(x_i-\bar{x})^2}{n}}$$

En el capítulo anterior, se calculó la varianza para las calificaciones de los alumnos de la asignatura de estadística, y se obtuvo que: $s^2 = 1{,}07935$. Por otra parte, como ya se explicó en este capítulo, el cuarto momento no central se obtiene mediante la fórmula:

$$m_4 = \frac{f_1(x_i-\bar{x})^4+f_2(x_2+\bar{x})^4+...+f_k(x_k-\bar{x})^4}{n}$$

Como resultado de la fórmula anterior se tiene que $m_4 = 2{,}9531$.

Sustituyendo los valores de s^2 y m_4 en la fórmula correspondiente:

$$\begin{aligned} a_4 &= \frac{2{,}9531}{(1{,}07935)^2} \\ &= 2{,}535 \end{aligned}$$

Si el resultado es negativo, la distribución de los datos es de tipo leptocúrtica. Si es positivo, la distribución de los datos es de tipo platicúrtica. si el resultado es cero, la distribución de los datos corresponde a una *distribución normal*, concepto que se abordará en su respectivo capítulo.

Ejercicios 3

- *Instrucciones*: Responde brevemente, a lo que se te pregunta

1. Describe lo que es el sesgo positivo.

   ____

   ____

   ____

   ____

   ____

   ____

   ____

2. Dibuja una curva de frecuencias cuyo sesgo sea negativo.

3. Anota las expresiones que permiten calcular el coeficiente de sesgo intercuartílico y percentílico.

   ____

   ____

4. ¿Qué es la curtosis?

   ____

   ____

5. Describe con tus palabras cuántas y cuáles son las formas generales que puede asumir la curva de un polígono de frecuencias suavizado.

    ____
    ____
    ____
    ____
    ____
    ____
    ____

6. Anota la expresión matemática que permite calcular el coeficiente de momento de curtosis, así como los nombres de todos los símbolos que le conforman.

    ____
    ____
    ____
    ____

7. ¿Qué indica el coeficiente de momento de curtosis igual a cero?

    ____
    ____
    ____

- Coteja tus respuestas con las que se te proporcionan en la sección **6.7.1** (página 370) y repasa el contenido del capítulo antes de resolver la autoevaluación. No olvides hacer énfasis en los puntos que se te dificultaron y/o aquellos en los que fallaste al resolver los ejercicios.

6.6. Autoevaluación

- *Instrucciones*: Para la siguiente serie de ejercicios, evita regresar al contenido de la unidad para retomar la información y fórmulas que te permitan responder, pues ya se resolvieron ejercicios y se realizó un repaso; si lo haces, deberás tomar como errónea dicha respuesta.

1. Para el siguiente conjunto de datos: $X = \{1, 1, 2, 2, 2, 3, 4, 5, 6\}$ Calcular los momentos que se te solicitan:

   a) 1^{er} Momento.

   b) 2^{o} Momento no central.

   - Considera el siguiente conjunto de datos agrupados para calcular las medidas asimetría y forma que se te solicitan.

i	clase	f_i	x_i
1	18 – 21,6	5	19,8
2	21,6 – 25,2	11	23,4
3	25,2 – 28,8	4	27
4	28,8 – 32,4	18	30,6
5	32,4 – 36	24	34,2
6	36 – 39,6	32	37,8
7	39,6 – 43,2	12	41,4
8	43,2 – 46,8	14	45
9	46,8 – 50,4	21	48,6
10	50,4 – 54	3	52,2
11	54 – 57,6	2	55,8
12	57,6 – 61,2	3	59,4

Recuerda que para este conjunto de datos agrupados se han calculado:

- $\bar{x} = 37{,}85$
- $Q_1 = 32{,}25$
- $Q_2 = 37{,}41$
- $Q_3 = 45{,}32$
- Rango semi-intercuartil $RSIQ = 6{,}535 \approx 6{,}54$
- Desviación típica para datos agrupados $s = 8{,}827 \approx 8{,}83$
- Varianza para datos agrupados $s^2 = 2{,}971 \approx 2{,}97$

2. Calcular o deducir los siguientes momentos para la anterior tabla de datos agrupados.

   a) Primer momento.

   b) Segundo momento.

   c) Primer momento no central.

   d) Segundo momento no central.

3. Coeficiente de sesgo intercuartílico

4. Coeficiente de momento de curtosis.

- Verifica que tus respuestas sean las correctas, para ello, te puedes apoyar en la página 372

6.7. Apéndice del capítulo

6.7.1. Respuestas a los ejercicios del capítulo

Ejercicios 1

- *Instrucciones*: Para el siguiente conjunto de datos: $X = \{1, 1, 2, 3, 4, 5\}$ Calcular los momentos que se te solicitan:

1. 2^o Momento. Respuesta: ($\overline{x^2} = 9{,}333$)
2. 3^{er} Momento. Respuesta ($\overline{x^3} = 37{,}6667$)
3. 1^{er} Momento no central. Respuesta ($m_1 = -0{,}02$ puesto que $\bar{x} = 2{,}67$)
4. 2^o Momento no central. Respuesta ($m_2 = 13{,}34$)

Ejercicios 2

- *Instrucciones*: Calcular para los datos del ejemplo anterior:

1. 3^{er} momento no central. $m_3 = 0{,}0309455$
2. 4° momento no central. $m_4 = 0{,}946713601$.
3. 3^{er} Momento. $\overline{X^3} = 456{,}0269531$
4. 4° Momento. $\overline{X^4} = 3209{,}784004$

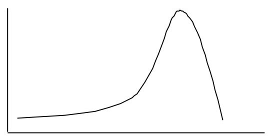
Ejercicios 3

- *Instrucciones*: Responde brevemente, a lo que se te pregunta

1. Describe lo que es el sesgo positivo
   Se dice que la distribución de datos presenta un sesgo positivo, cuando en su representación gráfica (polígono suavizado) se presenta una cola

más larga hacia la derecha del máximo central, es decir, la mayor parte de los datos de la distribución, tienden a alejarse del origen y la media de las observaciones es muy alta.

2. Dibuja una curva de frecuencias cuyo sesgo sea negativo.



3. Anota las expresiones que permiten calcular el coeficiente de sesgo intercuartílico y percentílico.
   Coeficiente de sesgo intercuartílico $= \frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1} = \frac{Q_3-2Q_2+Q_1}{Q_3-Q_2}$
   Coeficiente de sesgo percentílico $= \frac{(P_{90}-P_{50})-(P_{50}-P_{10})}{P_{90}-P_{10}} = \frac{P_{90}-2P_{50}+P_{10}}{P_{90}-P_{10}}$

4. ¿Qué es la curtosis?
   La curtosis es el grado de achatamiento o apuntalamiento que presenta la curva de un polígono de frecuencias suavizado.

5. Describe con tus palabras cuántas y cuáles son las formas generales que puede asumir la curva de un polígono de frecuencias suavizado.
   En general, la curva de un polígono de frecuencias puede adquirir tres formas y éstas son: Leptocúrtica, Platicúrtica y Mesocúrtica; la primera (leptocúrtica), se caracteriza por contar con una “joroba” puntiaguda, que indica una media sumamente elevada y la mayor parte de los datos concentrados en torno a ésta; la curva Mesocúrtica, es aquella cuya “joroba” es baja y ancha, lo que indica una media no elevada y, hasta cierto punto, baja concentración de los datos en torno a ésta medida; finalmente, la curva mesocúrtica es la más equilibrada de las tres, su media no es muy alta o baja y es la más regular de las curvas.

6. Anota la expresión matemática que permite calcular el coeficiente de momento de curtosis, así como los nombres de todos los símbolos que le conforman.

   El *coeficiente de momento de curtosis* se denota con el símbolo a_4 y se calcula mediante la determinación del segundo y cuarto momentos no centrales para datos agrupados: $a_4 = \frac{m_4}{m_2^2}$, donde el segundo momento (que equivale a la varianza de la distribución de datos) se encuentra elevado al cuadrado.

7. ¿Qué indica el coeficiente de momento de curtosis igual a cero?
   Que la distribución de los datos corresponde a una distribución normal.

6.7.2. Respuestas de la autoevaluación

- *Instrucciones*: Para la siguiente serie de ejercicios, evita regresar al contenido de la unidad para retomar la información y fórmulas que te permitan responder, pues ya se resolvieron ejercicios y se realizó un repaso; si lo haces, deberás tomar como errónea dicha respuesta.

1. Para el siguiente conjunto de datos: $X = \{1, 1, 2, 2, 2, 3, 4, 5, 6\}$ Calcular los momentos que se te solicitan:

   a) 1^{er} Momento. $\overline{X^1} = 2{,}8888888889 \approx 2{,}9$

   b) 2^o Momento no central. $m_2 = 2{,}765433 \approx 3{,}765$

2. Calcular o deducir los siguientes momentos para la anterior tabla de datos agrupados.

   a) Primer momento. $\overline{X^1} = \bar{x}$, y como para este caso $\bar{x} = 37{,}85$ entonces $X = 37{,}85$

b) Segundo momento. $\overline{X^2} = 1510{,}427114 \approx 1510{,}43$

c) Primer momento no central. $m_1 = -0{,}001677852$

d) Segundo momento no central. $m_2 = s^2$, y como para este caso $s^2 = 2{,}971 \approx 2{,}97$ entonces $m_2 = 2{,}97$

3. Coeficiente de sesgo intercuartílico = 0,339263024

4. Coeficiente de momento de curtosis $a_4 = 1811{,}236873628609782090902014975 \approx 1811{,}2369$

- ¡Felicidades por tu esfuerzo! Los aciertos y errores que tuviste son producto de tu trabajo, dedicación y atención. Si todas tus respuestas fueron correctas continúa así, de lo contrario, analiza el porqué de tus resultados.

  Para las respuestas que fueron incorrectas debido a que olvidaste alguna fórmula o relación entre medidas, diseña una estrategia que te permita recordar ese elemento –o elementos– de "difícil comprensión", nadie mejor que tú sabe de qué manera se te facilita el aprendizaje.

  Si el resultado de algún cálculo está mal, revisa el procedimiento –que es muy importante– y verifica las operaciones que realizaste; ante cualquier error en el desarrollo, repite el ejercicio y para errores en las operaciones, se te recomienda que pongas más atención mientras las resuelves (apaga el radio, televisión, si tienes hambre come, en síntesis: elimina todos los estímulos que te distraigan al estudiar; procúrate un espacio adecuado y las herramientas que ocuparás mientras lo haces).

Capítulo 7

Regresión lineal y correlación

Objetivos específicos

- El estudiante practicará la *técnica de regresión lineal*, y sus elementos básicos, al estudio de conjuntos de datos donde existe una relación entre variables independientes y dependientes, mediante la lectura del texto y resolución de ejercicios.

- El estudiante determinará el tipo de *correlación* que guardan los conjuntos de datos propuestos, mediante el cálculo su *coeficiente de correlación*, para el estudio del comportamiento de las variables en cuestión.

7.1. Preliminar

En diversos fenómenos puede establecerse una relación entre dos o más variables; un ejemplo de esto, es el peso de una persona adulta que depende, entre otros factores, de su edad; otro ejemplo lo constituye la cantidad de accidentes viales en una ciudad, cuya ocurrencia depende del día de la semana. De aquí se desprende que, en este tipo de fenómenos, existe una relación causal (relación causa-efecto) entre dos variables, a saber, la *variable independiente* y la *variable dependiente*. En los dos ejemplos anteriores, las variables independientes corresponden a la edad y día de la semana y, por otro lado, las variables dependientes, corresponden al peso y al número de accidentes. Generalmente, este tipo de fenómenos se describen en tablas, por ejemplo, para los casos anteriores, se tienen las siguientes tablas:

Relación Edad-Peso

Edad de la persona (X)	Peso en Kg. (Y)
12	51
15	59
18	62
25	65
30	71

Relación Día-Accidentes

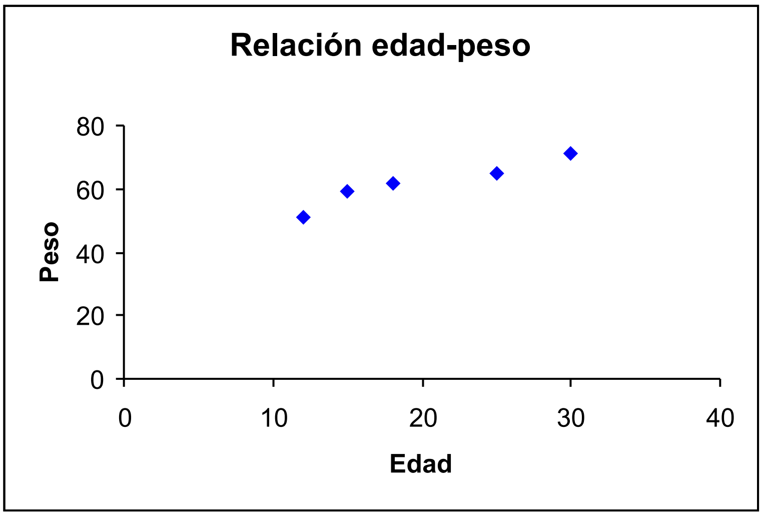
Día de la semana (X)	No. de Accidentes (Y)
Lunes	3
Martes	2
Miércoles	5
Jueves	7
Viernes	15
Sábado	13
Domingo	2

7.2. Diagrama de dispersión

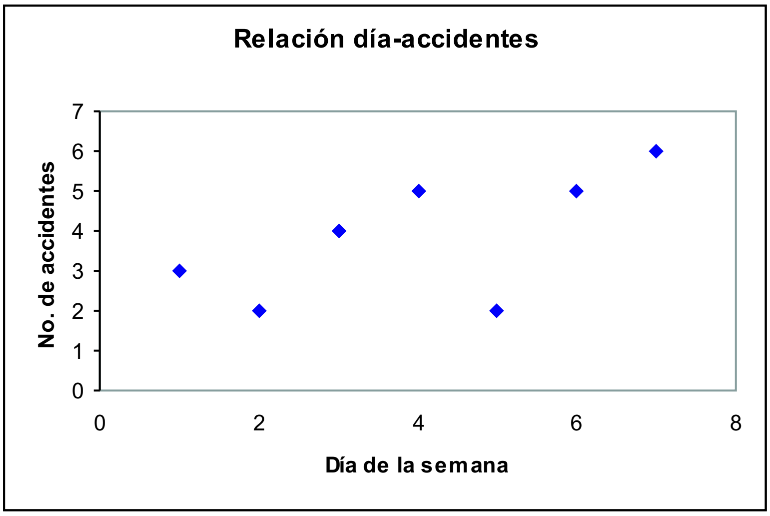
Las tablas anteriores presentan cierta utilidad para el análisis del fenómeno, no obstante, la representación gráfica de las mismas, proporciona elementos adicionales –para el estudio de su comportamiento– como pueden ser la identificación de patrones (repeticiones) y la identificación de la tendencia (creciente, decreciente o constante) en los datos.

La representación gráfica de un fenómeno que involucra la relación de dos variables, se conoce como *diagrama de dispersión*. Es un gráfico que consiste en una colección de puntos definidos por coordenadas x, y, donde la primera corresponde a la variable independiente y la segunda a la variable dependiente. A continuación se muestran los respectivos diagramas de dispersión de los dos ejemplos iniciales del capítulo:

Para el caso del peso de las personas:



Para el caso del número de accidentes:



Nótese que para el ejemplo del número de accidentes, el nombre del día se ha sustituido, siendo el lunes el día uno y el domingo el día siete. En ambos ejemplos, x es variable explicativa y por su parte, y es la variable de respuesta. Es decir, se presume que x explica el comportamiento de y (y depende de x) bajo una regla matemática, la cual es de gran importancia determinar, pues ésta permitirá obtener conclusiones sobre el fenómeno y realizar proyecciones o estimaciones sobre el comportamiento del mismo.

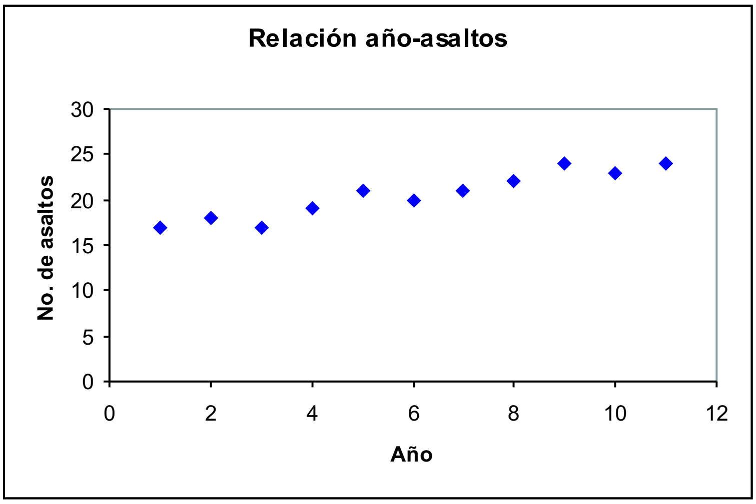
7.3. Regresión

Una de las técnicas que permite determinar la regla matemática que describe un fenómeno con relación causal, lo constituye la *regresión.* Si existe una relación entre dos elementos. Para poder determinar la regla matemática o ecuación, que describe la relación entre dos variables, un primer paso consiste en la recolección de datos, generalmente organizados en forma tabular.

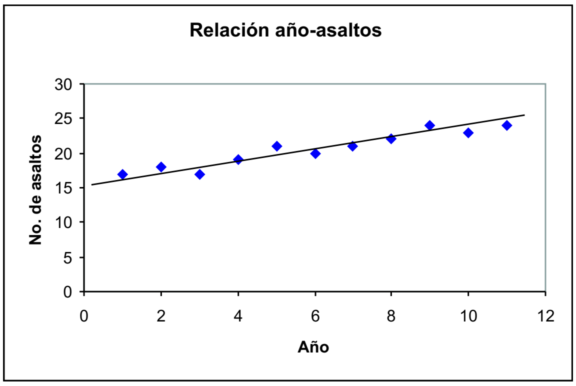
- Ejemplo. Considérese el número de asaltos a instituciones bancarias por año:

Año	No. de Año (X)	No. de asaltos (Y)
1994	1	17
1995	2	18
1996	3	17
1997	4	19
1998	5	21
1999	6	20
2000	7	21
2001	8	22
2002	9	24
2003	10	23
2004	11	24

Posteriormente, se realiza el diagrama de dispersión de los datos:



Una vez hecho lo anterior, se determina la tendencia o forma de los mismos, observando qué tipo de figura (recta o curva) describe la tendencia en el comportamiento de los datos sujetos a estudio:



Se observa en la correspondiente figura, que una recta puede ajustarse al comportamiento de los datos del ejemplo. El proceso matemático que consiste en ajustar una recta a una curva o a un conjunto de puntos, para obtener la mejor representación del comportamiento de los datos se denomina *regresión*. Es importante mencionar que al ajustar a través de la regresión, una recta o curva al conjunto de los datos, no es necesario *tocar* al total, de puntos. La recta o curva a ajustar, sólo debe reflejar la tendencia en el comportamiento de los mismos, como se observa en la figura, en donde la recta ajustada no toca a todos los puntos, sólo muestra la tendencia en el comportamiento de los mismos.

7.3.1. Regresión lineal

El presente capítulo se enfocará a datos cuya tendencia en comportamiento es lineal, es decir, que puede ser descrito por una línea recta. El proceso de ajustar una línea recta a un conjunto de datos se conoce como *regresión lineal*.

Toda recta puede representarse por medio de la siguiente ecuación:

$$y = \beta_0 + \beta_1 x$$

Donde β_0 se conoce como el intercepto de la recta (determina el punto donde la recta *corta* al eje y) y $beta_1$ se conoce como la *pendiente* de la recta (define la inclinación de la recta); ejemplos:

a) $y = 5 + x \quad \beta_0 = 5, \quad \beta_1 = 1$

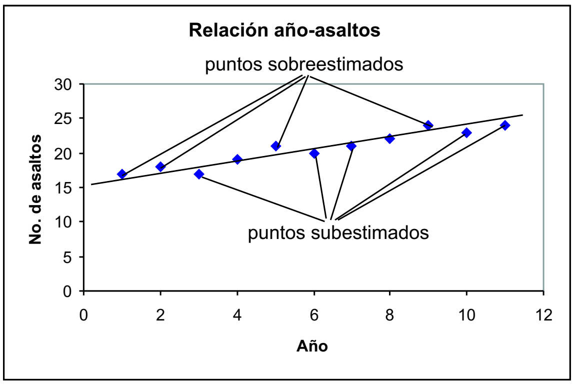
b) $y = -x \quad \beta_0 = 0, \quad \beta_1 = -1$

c) $y = -2 + 3x \quad \beta_0 = -2, \quad \beta_1 = 3$

A través de la regresión lineal, se busca obtener la recta que mejor se ajuste o mejor describa el comportamiento de una colección de datos. Dado que en un plano cartesiano pueden definirse un número infinito de rectas, se debe emplear un método matemático que garantice que de toda la gama de posibles rectas, se ha determinado la que mejor se ajusta al conjunto de datos. El método a emplear se conoce como *método de los mínimos cuadrados.*

Método de los mínimos cuadrados.

Dado un conjunto de datos a los cuales se les puede ajustar una recta:



Se observa que entre cada uno de los puntos y la recta, existen diversas distancias. Estas distancias son los *errores* de ajuste de la recta. Cuando la recta pasa por arriba de un punto, se dice que se está *sobreestimando* un valor, por otra parte, cuando la recta pasa por debajo de un punto, se está *subestimando* su valor. El error del punto i respecto a la recta, se denota con el símbolo e_i.

Para determinar una recta de ajuste, se deben encontrar los valores β_0 y β_1 que corresponden a la recta que garantiza que pasa lo más cerca posible de todos y cada uno de los puntos, es decir, una recta que *minimiza* los errores respecto a cada punto. Por razones de carácter matemático, uno de los criterios para seleccionar dicha recta será el de *elegir* a la recta cuya suma de los errores e_i, elevados al cuadrado, sea la menor de todas las posibles rectas. Es decir, se obtiene un valor β_0 y β_1, que definan una recta tal que $min\{\sum e_i^2\}$ Esta recta se define con el nombre de *recta de mínimos cuadrados*. Para encontrar dicha recta, se emplean métodos de cálculo diferencial, que no serán descritos en el presente texto por no ser el objetivo del mismo, llegando a las siguientes expresiones que permiten obtener los valores que definen la mejor recta de ajuste:

$$\boxed{\beta_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$
$$\boxed{\beta_0 = \bar{y} - \beta_1\bar{x}}$$

Donde:
n = Número de observaciones
x_i = El i-ésimo valor de la variable independiente o variable explicativa
y_i = El i-ésimo valor de la variable dependiente o variable de respuesta
$\bar{x}$ = Media del conjunto de observaciones X, es decir, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
$\bar{y}$ = Media del conjunto de observaciones Y, es decir, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

De las anteriores expresiones, puede construirse la siguiente ecuación fundamental:

$$\boxed{y = \beta_0 + \beta_1 x + \epsilon}$$

Donde ϵ (épsilon) representa el *error* del modelo ajustado, dado que, como ya se mencionó, la recta es sólo una aproximación a los datos verdaderos, no obstante, por la estrategia del método matemático de los mínimos cuadrados para la regresión lineal, el error se reduce a la menor magnitud posible.

Esta expresión, es el modelo matemático que describe el comportamiento de un fenómeno que presenta relación lineal entre dos variables. Se trata de la ecuación de una recta definida por los valores β_0 y β_1. Dicha ecuación, permitirá realizar estimaciones y proyecciones sobre el comportamiento de los datos objeto de estudio.

- Ejemplo: retomando el caso de asaltos a instituciones bancarias, se tienen los siguientes datos:

Año	No. de Año (X)	No. de asaltos (Y)
1994	1	17
1995	2	18
1996	3	17
1997	4	19
1998	5	21
1999	6	20
2000	7	21
2001	8	22
2002	9	24
2003	10	23
2004	11	24

Empleando el método de los mínimos cuadrados, ajustar la recta que describe el comportamiento de los datos y realizar una estimación sobre el número de asaltos que ocurrirán en el año 2005 y en el año 2006.

- **Solución**.

Se recomienda la construcción de una tabla como la siguiente para realizar de forma ordenada el proceso de cálculo:

Año (X)	No. de asaltos (Y)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	17	-5	$-3{,}55$	17,75	25
2	18	-4	$-2{,}55$	10,2	16
3	17	-3	$-3{,}55$	10,65	9
4	19	-2	$-1{,}55$	3,1	4
5	21	-1	0,45	$-0{,}45$	1
6	20	0	$-0{,}55$	0	0
7	21	1	0,45	0,45	1
8	22	2	1,45	2,9	4
9	24	3	3,45	10,35	9
10	23	4	2,45	9,8	16
11	24	5	3,45	17.25	25
				$\sum(x_i - \bar{x})(y_i - \bar{y})$ =82	$\sum(x_i - \bar{x})^2$ =110

Donde: $n = 11$, (11 observaciones)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$
$$= \frac{1+2+3+4+5+6+7+8+9+10+11}{11}$$
$$= \frac{66}{11} = 6$$

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$
$= \frac{17+18+17+19+21+20+21+22+24+23+24}{11}$
$= \frac{226}{11}$
$= 20{,}55$, redondeado a dos decimales.

Ambos valores permiten el cálculo de la tercera y cuarta columna de la tabla respectivamente.

- La tercera columna se obtiene al restar a cada valor de X, (Año), el valor de $\bar{x} = 6$.

- La cuarta columna se obtiene al restar a cada valor de Y, (No. de asaltos), el valor de $\bar{y} = 20{,}55$.

- Por otra parte, la quinta columna de la tabla se calcula al multiplicar respectivamente los valores de la tercera y cuarta columna.

- En este orden de ideas, la sexta y última columna, se obtiene al elevar al cuadrado a cada elemento de la tercera columna.

Una vez hecho lo anterior, se obtiene la suma total, tanto de la quinta como de la sexta columna. Como se observa, la expresión:

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Se reduce a tomar de la tabla la suma total de la quinta columna, así como de la sexta y realizar su cociente:

$\beta_1 = \frac{82}{110}$
$\beta_1 = 0{,}745454545$ (0,75 redondeado a dos decimales)

Con este valor, es posible calcular el valor de β_0. Sustituyendo β_1, $\bar{x}$ y $\bar{y}$ en la expresión revisada anteriormente:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Se tiene entonces que:
$\beta_0 = 20{,}55 - 0{,}75(6)$
$\beta_0 = 16{,}05$

Con los valores β_0 y β_1, obtenidos en los pasos anteriores, se procede a construir la recta de regresión de mínimos cuadrados que describe el comportamiento del número y de asaltos bancarios:

$$y = 16{,}05 + 0{,}75x$$

Ahora bien, esta ecuación puede emplearse para realizar una estimación o proyección sobre el número de asaltos a instituciones bancarias para el año 2005 y 2006.

El año 2005 corresponde al año doce y el año 2006 corresponde al año trece (recuérdese que se tienen datos de once años previos). Entonces el año doce ($x = 12$), se sustituye en la ecuación de la recta y se tiene:

$y = 16{,}05 + 0{,}75(12)$
$y = 16{,}05 + 9$
$y = 25{,}05$, redondeando: $y = 25$

Esto significa que se estima ocurrirán 25 asaltos bancarios en el año 2005.

Por otra parte, proyectando el número de asaltos a instituciones bancarias para el año trece (2006), se tiene entonces que *x=13*.

Sustituyendo este valor en la ecuación de la recta obtenida, se tiene que:

$y = 16{,}05 + 0{,}75(13)$
$y = 16{,}05 + 9{,}75$
$y = 25{,}8$, redondeando: $y = 26$

Esto significa que se estima ocurrirán 26 asaltos bancarios en el año 2006.

Ejercicios 1

- *Instrucciones*: Responde brevemente a las preguntas que se te realizan y efectúa los cálculos necesarios para obtener los resultados que se requieren.

1. ¿Qué es una variable independiente?

____

____

____

____

2. ¿Qué es una variable dependiente?

____

____

____

____

3. ¿Qué es un diagrama de dispersión?

____

____

____

____

4. ¿Qué es la regresión?

____

____

____

____

5. Anota la ecuación matemática mediante la cual, en la regresión lineal, es posible representar a cualquier recta.

____

6. ¿Qué son los errores de ajuste de la recta?

____

____

7. ¿En qué casos se dice que un punto está subestimado?

____

____

8. ¿Cuál es la mejor recta para un conjunto de observaciones?

____

____

____

____

- Considera el siguiente caso y efectúa el tratamiento necesario para obtener lo que se te pide:

♡ Nancy Lucia Badillo Centeno (Lucy) nació el 14 de agosto de 2001. A partir de entonces, la pediatra lleva su registro de edad y peso. A continuación se presentan los datos registrados desde los cuatro meses.

Edad (meses)	Peso (Kg.)
4	5,900
6	6,400
7	6,720
8	7,020
9	7,360
11	7,560
12	7,800
14	8,320

9. Elabora la representación gráfica de la dispersión de los puntos.

10. Mediante el método de los mínimos cuadrados, ajusta le recta que describe el comportamiento de los datos.

    a) Ecuación:____

b) Representación gráfica.

11. Realiza la estimación para el peso de Lucy en los meses que se te indica a continuación:

    a) 17

    b) 20

    c) 24

Para verificar que tus respuestas sean las correctas, compáralas con las que se te ofrecen en la página 399

7.4. Correlación

Se define a la *correlación*, como el grado de relación lineal entre dos variables. Se emplea para determinar en qué medida una ecuación lineal (una recta), describe o explica la relación entre dos variables.

7.4.1. Coeficiente de correlación

El *coeficiente de correlación* cuantifica el nivel de relación lineal entre dos variables. La siguiente expresión para el cálculo del coeficiente de correlación se denomina *fórmula producto-momento* o *coeficiente de correlación de Pearson* y se denota con la literal r:

$$\boxed{r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^{n} (x_i - \bar{x})^2][\sum_{i=1}^{n} (y_i - \bar{y})^2]}}}$$

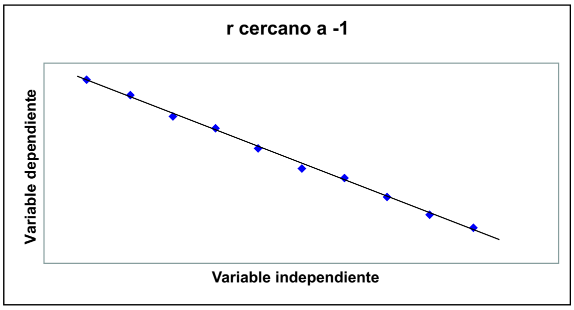
El coeficiente de correlación r, siempre cumple con la desigualdad:

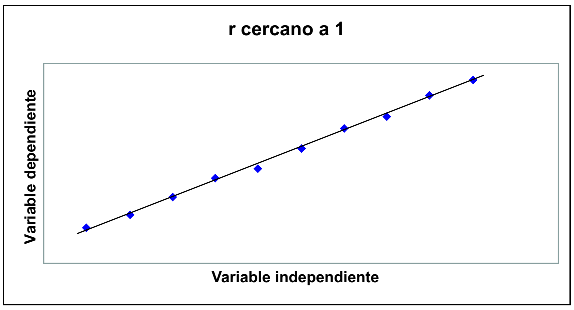
$$r \in [-1, 1]$$
$$-1 \leq r \leq 1$$

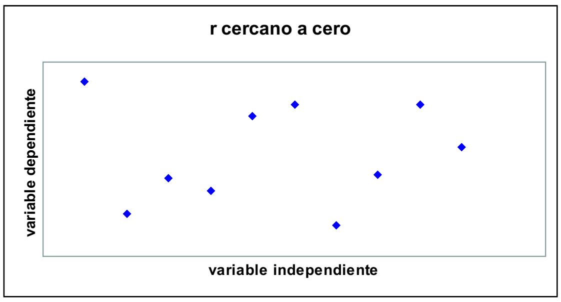
Es decir, el valor de r siempre se encuentra entre -1 y 1.

Un coeficiente de correlación cercano a -1, significa *correlación negativa* (cuando aumenta el valor de la variable independiente, disminuye el valor de la variable dependiente). Y por otro lado, un coeficiente de correlación cercano a 1, implica *correlación positiva* (al aumentar el valor de la variable dependiente, aumenta el valor de la variable idependiente).

Finalmente, si el valor del coeficiente de correlación es cero, o un valor cercano a cero, implica que no existe correlación (no existe relación lineal entre las variables del fenómeno, por lo que una recta no puede describir su comportamiento). Véanse los siguientes diagramas de dispersión:


r cercano a -1
Variable dependiente
Variable independiente


r cercano a 1
Variable dependiente
Variable independiente


r cercano a cero
variable dependiente
variable independiente

Retomando el ejemplo sobre asaltos a instituciones bancarias, puede calcularse su coeficiente de correlación ampliando la tabla que se construyó para determinar la recta de regresión lineal de la siguiente manera:

1	2	3	4	5	6	7
1	17	-5	$-3{,}55$	17,75	25	12,6025
2	18	-4	$-2{,}55$	10,2	16	6,5025
3	17	-3	$-3{,}55$	10,65	9	12,6025
4	19	-2	$-1{,}55$	3,1	4	2,4025
5	21	-1	0,45	$-0{,}45$	1	0,2025
6	20	0	$-0{,}55$	0	0	0,3025
7	21	1	0,45	0,45	1	0,2025
8	22	2	1,45	2,9	4	2,1025
9	24	3	3,45	10,35	9	11,9025
10	23	4	2,45	9,8	16	6,0025
11	24	5	3,45	17.25	25	11,9025
				$\sum(x_i-\bar{x})(y_i-\bar{y})$ =82	$\sum(x_i-\bar{x})^2$ =110	$\sum(y_i-\bar{y})^2$ = 66,7275

Donde:
Columna 1 = Año (X)
Columna 2 = No. de asaltos (Y)
Columna 3 = $x-\bar{x}$
Columna 4 = $y-\bar{y}$
Columna 5 = $(x-\bar{x})(y-\bar{y})$
Columna 6 = $(x-\bar{x})^2$
Columna 7 = $(y-\bar{y})^2$

La columna siete de la tabla se obtiene al elevar al cuadrado cada valor de la columna cuatro.

De la tabla se tiene que:
$\sum_{i=1}^{n}\ (x_i - \bar{x})(y_i - \bar{y}) = 82$
$\sum_{i=1}^{n}\ (x_i - \bar{x})^2 = 110$
$\sum_{i=1}^{n}\ (y_i - \bar{y})^2 = 66{,}7275$

Sustituyendo estos valores en la expresión:

$$r = \frac{\sum_{i=1}^{n}\ (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^{n}\ (x_i - \bar{x})^2][\sum_{i=1}^{n}\ (y_i - \bar{y})^2]}}$$

se tiene entonces que:

$$r = \frac{82}{\sqrt{110(66{,}7275)}}$$

$$r = 0{,}957117116$$

Como se observa, el valor de r es cercano a 1, por lo que se tiene presente una correlación positiva.

Ejercicios 2

1. ¿Qué es la correlación lineal?

   ____

   ____

2. ¿Qué implica un coeficiente de correlación cercano a -1?

   ____

   ____

3. Elige la situación que se pueda clasificar como una correlación positiva.

   a) A mayor tamaño de una pantalla, distinto número de puntos provocados por la interferencia electromagnética.

   b) A mayor tiempo de lectura, menor número de faltas de ortografía.

   c) A mayor tiempo de lectura, mayor número de páginas recorridas.

4. Observa la fórmula *Producto-momento* y anota en su lugar los elementos faltantes:

$$r = \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=}(x \quad - \bar{x})^2][\sum_{i=1} \quad (y_i - y) \quad]}}$$

5. ¿Qué implica un valor cercano a cero para r?

____

____

____

____

6. Calcular el coeficiente de correlación r para las variables: Edad y peso del caso Lucy (Ejercicios 1, página 390).

____

____

- Las respuestas para esta serie de ejercicios, se encuentran en la página 402, asegúrate de que las respuestas y resultados que obtuviste, sean los correctos. Si tienes dudas en algún tema, revisa la parte correspondiente del capítulo antes de iniciar la resolución de la autoevaluación.

7.5. Autoevaluación

- *Instrucciones*: Analiza el siguiente caso:
  - Supóngase que luego de asistir a la conferencia *La Investigación Sobre Infraestructura Física Educativa*[1], realizada en el salón 901 de la FES Acatlán; los estudiantes de comunicación inscritos en la preespecialidad de *investigación y docencia*, realizarán una investigación con el propósito de elaborar una propuesta para rediseñar el espacio físico educativo en que se imparte la preespecialida.
  - Para conseguir su objetivo, deben –entre otros factores– hacer una proyección de la *demanda* que dicha preespecialidad podría tener a corto, mediano y largo plazo. En función de ello, el grupo de investigadores se ha formulado las siguientes preguntas y actividades.
- Responde las preguntas:

1. ¿Existe en el aspecto de la *demanda*, una relación con otra variable al hacer la proyección a futuro?
   a) Sí
   b) No
   c) Justifica tu respuesta

   ____
   ____
   ____
   ____

[1] Conferencia dictada por la Maestra en Educación, María del Rocío Ávila (Octubre de 2005), quien afirmó que "el espacio físico educativo determina los procesos de enseñanza aprendizaje"

2. ¿Cuál de las siguientes técnicas, permitirá al grupo determinar la demanda futura que podría tener la preespecialidad?

   a) Estudio de correlación.

   b) Técnica de regresión.

   c) Diagrama de dispersión.

3. ¿Cuál es el paso que sigue a la organización tabular y representación gráfica de la dispersión de los datos?

   ----

4. ¿Cuál es el propósito de la regresión lineal?

- Los datos recolectados por los estudiantes fueron los siguientes:

Año	No. de Año (X)	No. Estudiantes por semestre (Y)
1995	1	10
1996	2	09
1997	3	12
1998	4	12
1999	5	14
2000	6	07
2001	7	16
2002	8	18
2003	9	19
2004	10	20
2005	11	23

- Calcula lo que se te pide empleando el método de los mínimos cuadrados:

5. Ajusta la recta que describa el comportamiento de los datos.

    a) Ecuación: ____

    b) Representación Gráfica:

6. Realiza una estimación sobre el número de estudiantes que podrían inscribirse a la preespecialidad en los siguientes años:

    a) En 2010

    b) En 2015

    c) En 2025

7.6. Apéndice del capítulo

7.6.1. Respuestas a los ejercicios del capítulo

Ejercicios 1

- *Instrucciones*: Coteja tus respuestas con las de esta sección. Recuerda que el sentido de éstas es muy importante, pero la precisión es substancial para la investigación formal; por tal motivo, te debes seguir esforzando en conseguir que tus respuestas sean lo más apegadas que sea posible al texto en que ahora trabajas.

1. ¿Qué es una variable independiente?
   Una característica propia del objeto que se encuentra libre de la influencia de otra variable. En otras palabras, sus propiedades o magnitudes no dependen de otra cosa que su misma naturaleza.

2. ¿Qué es una variable dependiente?
   En una relación de causa-efecto entre dos propiedades pertinentes a x fenómeno, se denomina variable dependiente a la característica del fenómeno que cambia en función de las variaciones que le infringe la propiedad con que interactúa en la relación.

3. ¿Qué es un diagrama de dispersión?
   La representación gráfica de un fenómeno que involucra la relación de dos variables; ésta consiste en una colección de puntos definidos por coordenadas x, y, que corresponden a la variable dependiente e independiente, respectivamente.

4. ¿Qué es la regresión?
   Es un proceso matemático que consiste en ajustar una recta a una curva o a un conjunto de puntos, para obtener la mejor representación del comportamiento (tendencia) de los datos.

5. Anota la ecuación matemática mediante la cual, en la regresión lineal, es posible representar a cualquier recta.
   $y = \beta_0 + \beta_1 x$

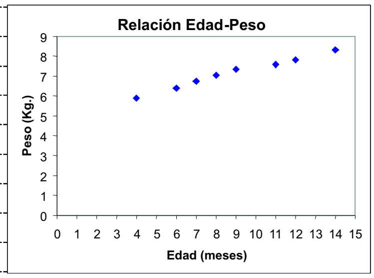
6. ¿Qué son los errores de ajuste de la recta?
   Se llama así a la distancia que existe entre un punto y la mejor recta posible determinada.

7. ¿En qué casos se dice que un punto está subestimado?
   Cuando la mejor recta posible pasa por debajo de un punto.

8. ¿Cuál es la mejor recta para un conjunto de observaciones?
Aquella que garantiza la máxima cercanía posible de todos los puntos en cuestión, es decir, la recta que minimiza los errores e_i al máximo.

9. Elabora la representación gráfica de la dispersión de los puntos.

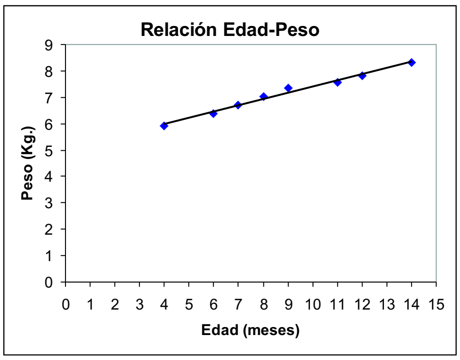
Meses	Peso (Kg.)
4	5,900
6	6,400
7	6,720
8	7,020
9	7,360
11	7,560
12	7,800
14	8,320



10. Mediante el método de los mínimos cuadrados, ajusta le recta que describe el comportamiento de los datos.

   a) Ecuación: $y = 05{,}049 + 0{,}235x$

   b) Representación gráfica



11. Realiza la estimación para el peso de Lucy en los meses que se te indica a continuación:

    a) 17 es a 9,040Kg.

    b) 20 es a 9,750Kg.

    c) 24 es a 10,689Kg.

    Para este caso, cabe destacar que los valores reales en la cartilla de *"Control de Crecimiento, Desarrollo e Inmunizaciones"* son muy próximos a los resultados obtenidos mediante el método de los mínimos cuadrados.

    Registros del pediatra:

    a) 1 año 5 meses 9,180Kg.

    b) 1 año 8 meses 10,060Kg.

    c) 2 años 10,650Kg.

Ejercicios 2

1. ¿Qué es la correlación lineal?
   El grado de relación lineal que guardan dos variables y ésta puede ser positiva o negativa.

2. ¿Qué implica un coeficiente de correlación cercano a -1?
   Que la correlación es negativa, es decir, que cuando el valor de la variable independiente aumenta, el de la dependiente disminuye.

3. Elige la situación que se pueda clasificar como una correlación positiva.
   c) A mayor tiempo de lectura, mayor número de páginas recorridas. La variable independiente (número de páginas) aumenta conforme aumenta el tiempo de lectura.

4. Observa la fórmula *Producto-momento* y anota en su lugar los elementos faltantes:

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^{n} (x_i - \bar{x})^2][\sum_{i=1}^{n} (y_i - \bar{y})^2]}}$$

5. ¿Qué implica un valor cercano a cero para r?
Que no existe una correlación lineal entre las dos variables del fenómeno, es decir, que el comportamiento de los datos no puede ser descrito por una recta en su respectivo diagrama de dispersión.

6. Calcular el coeficiente de correlación r para las variables: Edad y peso del caso Lucy (Ejercicios 1, página 370).
$r = 1{,}041178062 \approx 1{,}0$ Por lo que se puede afirmar que la correlación entre las variables edad y peso de Lucy, es una correlación lineal positiva.

7.6.2. Respuestas de la autoevaluación

- *Instrucciones*: Analiza el siguiente caso...
- Responde las preguntas:

1. ¿Existe en el aspecto de la *demanda*, una relación con otra variable al hacer la proyección a futuro?

   a) Sí

   b) ...

   c) Justifica tu respuesta

   La otra variable en cuestión es el tiempo, pues al incrementarse este último, el número de estudiantes que se inscriban a la preespecialidad cambiará; por lo que el tiempo es la variable independiente y la demanda será la variable dependiente.

2. ¿Cuál de las siguientes técnicas, permitirá al grupo determinar la demanda futura que podría tener la preespecialidad?
   b)Técnica de regresión.

3. ¿Cuál es el paso que sigue a la organización tabular y representación gráfica de la dispersión de los datos?
   Determinar la tendencia o forma de éstos, observando qué tipo de figura (recta o curva) describe la tendencia en su comportamiento.

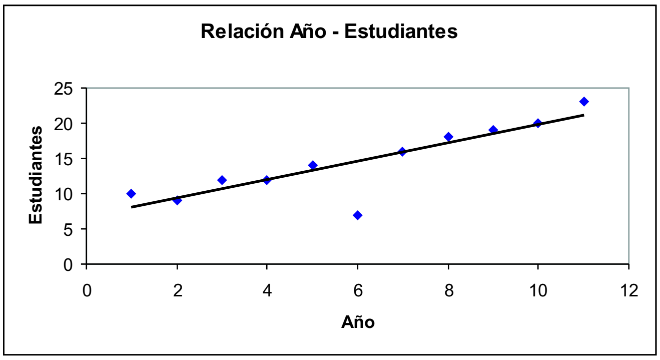
4. ¿Cuál es el propósito de la regresión lineal?
   Determinar la recta que mejor se ajuste o describa el comportamiento de los datos.

   - Los datos recolectados por los estudiantes fueron...
   - Calcula lo que se te pide empleando el método de los mínimos cuadrados:

5. Ajusta la recta que describa el comportamiento de los datos.

   a) Ecuación: $y = 6{,}719 + 1{,}306x$

   b) Representación gráfica



6. Realiza una estimación sobre el número de estudiantes que podrían inscribirse a la preespecialidad en los siguientes años años:

   a) En 2010 $y = 27{,}615 \approx 28$, es decir que habrá 28 estudiantes.

   b) En 2015 $y = 34{,}145 \approx 34$, es decir que habrá 45 estudiantes.

   c) En 2025 $y = 47{,}205 \approx 47$, es decir que habrá 47 estudiantes.

Capítulo 8

Introducción a la probabilidad

Objetivos específicos

- El estudiante reconocerá elementos básicos de la *teoría de la probabilidad*, mediante la lectura del texto y resolución de preguntas, para el posterior abordaje del *muestreo probabilístico*.

- El estudiante calculará la *probabilidad* de un evento y el valor de una *variable aleatoria*, mediante el seguimiento de los pasos requeridos para conocer la posibilidad de ocurrencia de éste.

8.1. Preliminar

Como ya se indicó en el primer capítulo, el "boom" del pensamiento racionalista que se generó con el advenimiento del proceso histórico denominado *la ilustración*, derivó en una nueva forma de ver y comprender al universo. La cosmovisión teológica con que se explicaba la realidad, fue desplazada –en cierta medida– por la positivista caracterizada por su tendencia al **mecanicismo materialista ateo**; ésta, implica un mundo real en que los fenómenos acaecidos obedecen a una *dinámica lineal*, donde las causas y los efectos pueden ser determinadas con toda certeza y son, sin lugar a dudas, independientes a caprichos divinos.

Sin embargo, el progreso en el estudio del universo (la ciencia en sus diferentes ámbitos), la progresiva liberación del pensamiento, diversos avances tecnológicos, así como cierta "incapacidad" para solucionar problemas, evidenciaron la insuficiencia del paradigma positivista, y algunas de sus vertientes, al tratar de explicar, predecir y controlar las características y consecuencias de *fenómenos-objetos de estudio (constituidos como sistemas para su abordaje) que no obedecen a la pretendida dinámica lineal imperante* en el universo.

Lo anterior, puso en jaque la pertinencia del método único para el estudio de los sistemas que interesan a las ciencias y abrió paso al estudio formal de la *probabilidad* por sí misma y como una herramienta formal para el estudio de los fenómenos no deterministas. En este capítulo, se tratará el tema desde una perspectiva histórica - epistemológica que permita comprender lo que es la probabilidad, para distinguir con claridad los fenómenos aleatorios de los deterministas; su definición clásica en lenguaje formal, lo que es la distribución de probabilidad y se dará un acercamiento a la aplicación de la probabilidad en el campo del muestreo.

8.2. Probabilidad, hombre, conocimiento

Una muy concreta necesidad del ser humano es la seguridad; de ahí su tenacidad en la investigación, en la búsqueda, sistematización y aplicación del conocimiento referente al mundo que le rodea. Señalar, marcar, nombrar y medir, son algunos medios por los cuales, como se mencionó en la primera unidad, el hombre se apropia del mundo y, dicho sea de paso, se procura la ansiada seguridad. La investigación científica es producto de la conjugación de sus aspiraciones y habilidades propiamente humanas; ella, le ha permitido generar diversos constructos ideológicos para explicar la realidad y, gradualmente, acceder a la **comprensión**, control y/o predicción de fenómenos que aprovecha o, de otro modo, le amenazan como especie.

En el pasado más remoto, la naturaleza se presentaba como algo incomprensible al ser humano. Debido a la manera en que se percibían los fenómenos naturales, estos fueron caracterizados como entes divinos, volubles y caprichosos que difícilmente o con nada se podían explicar, predecir o controlar; seguramente el ser humano se sintió, al respecto, abandonado a su suerte. Es así como el concepto de *caos* o *desorden* y el de *azar* o *aleatoriedad*, han acompañado al ser humano desde sus orígenes. "En *La teogonía*, Hesiodo dice: 'El Caos fue lo primero, y luego la Tierra'. Caos es una palabra de origen griego, equivalente a abismo, o también a una entidad sin forma; cosmos, por su parte, designa el orden y, por extensión, el Universo."[1]

El crecimiento **cultural**, producto de la paulatina evolución integral, permitió al ser humano identificar ciertas regularidades que lo alentaron en la búsqueda y construcción del conocimiento.

[1] JOSÉ, Sametband Moisés; **Entre el Orden y el Caos. La Complejidad**; México; FCE, SEP, CONACYT, 2^a edición; 1999; pp. 13

La transición de las estaciones, fases de la luna, emigraciones animales, erupción de volcanes, sismos, huracanes, la conducta humana y toda clase de fenómenos y eventos que podían o no ser "regulares" y por tanto conocidos o no, ocuparon –y aún lo hacen– el ingenio característico de nuestra especie. Al tomar como indicadores estos objetos de referencia en la construcción del conocimiento, se puede ver el desarrollo histórico del conocimiento científico.[2]

La primera etapa de su desarrollo se caracteriza por la intensa búsqueda por el conocimiento del Universo (**filosofía especulativa o científica**), mientras que en la segunda se hace lo propio respecto del ser humano (**filosofía práctica**), pero como producto de toda la actividad intelectual que en torno al Universo se generaba, la filosofía avanzó de tal manera que un nuevo objeto de sus razonamientos "surgió": el interés acerca del conocimiento en sí, o lo que hoy se denomina teoría del conocimiento.

"He dicho que la filosofía, al desarrollarse, descubre una tercera vertiente. Es ésta la *filosofía crítica*, que incluye la lógica y la epistemología o teoría del conocimiento. Sólo cuando alcanza una fase relativamente avanzada del pensamiento, empieza el hombre a preguntarse a sí mismo cuál es la eficacia de los instrumentos con que le dotó la naturaleza para entrar en contacto con el mundo exterior."[3] por lo que preguntas en torno al origen, la esencia y posibilidad del conocimiento derivan en nuevas corrientes del pensamiento filosófico-científico.

[2]Considérese lo visto en la primera unidad del texto en cuanto jonios y pitagóricos, así como la reacción hacia el humanismo cuyo punto de referencia más contundente se puede encontrar en Sócrates

[3]op. cit. GUTHRIE, William, K. C; México; Fondo de Cultura Económica; 1995; pp. 26

Uno de los aspectos que le son de interés a la filosofía crítica es el que obliga al hombre a cuestionarse por *la posibilidad del conocimiento*, dada la intensa actividad de los sistemas que se constituyen como objetos de estudio para las nacientes disciplinas. En la Grecia antigua, filósofos como Anaximandro, Heráclito, Parménides, Platón y Aristóteles, entre otros, concibieron elaborados razonamientos en los cuales se encuentran cuestiones respecto a la posibilidad del conocimiento insertas en contextos **metafísicos** (intentos por responder a preguntas como ¿De qué están hechas las cosas? o ¿Cuál es el origen del universo?); ejemplo de ello es la filosofía platónica.

“El mundo en que vivimos está hecho de cambio o, como dice Platón, de generación y corrupción. Todo cuanto nos rodea, y también nosotros mismos, está de tránsito. [...] Todo lo que existe *deviene* y el devenir cesa de ser lo que era. [...] Por eso podemos decir que – las cosas – son y no son, que su modo de existir es un modo de existir a medias entre el no-ser de lo que fueron y el no-ser de lo que todavía no alcanzan a ser. Ahora bien, por hermoso que sea el mundo sensible que nos rodea, por hermosas que sean la planta, el mar, la montaña o la persona, *no son explicables* por sí mismas.”[4]

El pensamiento platónico constituye un notable esfuerzo por combatir la tendencia intelectual de su época, que negaba la posibilidad del conocimiento fundándose en que no había ninguna realidad permanente que conocer[5], mediante la postulación de la existencia de un orden permanente (el mundo de las ideas) tras el *aparente caos y confusión* que se presentan a la percepción humana; en este sentido, se pueden notar los principios de dos posturas epistemológicas como son el *dogmatismo* y el *escepticismo*.

[4]op. cit. XIRAU, Ramón; México; 1995; pp. 54 Cursivas agregadas

[5]op. cit. GUTHRIE, William, K. C; México; Fondo de Cultura Económica; 1995; pp. 27

El dogmatismo, es aquella postura epistemológica en que *la posibilidad del conocimiento no se cuestiona*, pues sin lugar a dudas el individuo cognoscente es capaz de aprehender lo que desee. En la teoría del conocimiento el conflicto, o foco de atención, se encuentra centrado en la relación sujeto - objeto, cuestión que en el dogmatismo es irrelevante dado que el conocimiento es producto del raciocinio, de la actividad intelectual que permite discernir el orden existente en el universo –tal y como lo profesaban los jonios y Platón (entre otros) en la historia griega, o los racionalistas del siglo XVIII en adelante.

Por otra parte, en oposición al dogmatismo, el escepticismo se caracteriza por la negación de la posibilidad del conocimiento. Ésta postura, sí considera la relación antes mencionada y cuestiona severamente la preeminencia e importancia que se le concede al sujeto con respecto al objeto en el proceso de aprehensión –es decir, en al construcción del conocimiento. Primordialmente existen tres posiciones generales de esta corriente: El escepticismo antiguo, el escepticismo medio y el escepticismo posterior; de ellas, la que interesa es la segunda, puesto que niega la posibilidad del conocimiento en términos de certeza absoluta, pero la conciente en términos de probabilidad.

"El escepticismo medio o *académico*, cuyos principales exponentes son *Arcesilao* (242 d. de J. C.) y *Carneades* (129 d. de J. C.), no es tan radical como el escepticismo antiguo o pirrónico. El escepticismo académico afirma que no es posible el conocimiento exacto. Jamás podremos tener la certeza de que nuestros juicios concuerdan con la realidad. Por lo tanto jamás podremos afirmar que tal o cual proposición es verdadera; pero sí podemos afirmar que parece verdadera, que es probable. En consecuencia, *no existe la certeza absoluta, únicamente la probabilidad.*"[6]

[6]JESSEN, Johan; **Teoría del Conocimiento**; México; Editores Unidos Mexicanos; 1999; pp. 39

Es esta la actitud epistemológica que, en cierta medida, corresponde con el pensamiento-acción de la metodología estadística (en particular la del muestreo probabilístico), pues no hay verdades absolutas ya que a toda investigación le es inherente cierto error (llamado sesgo) y a las conclusiones de todo proceso formal de estudio se les concede validez no por ser verdaderas para situaciones particulares (evento) de un fenómeno, sino por ser posibles en cierta medida para los elementos de un conjunto de observaciones, motivo por el cual las conclusiones se asumen y expresan en términos de probabilidad, es decir de posibilidad distribuida o "suerte equitativa".

Como se ha mencionado en su momento, esta forma de pensar no es la propia de los racionalistas del siglo XVIII, pero es la que probablemente inspiró a esa comunidad que, bajo la consideración de la complejidad que se presenta a los sentidos, comenzó el estudio de los fenómenos irregulares, de todos aquellos cuyos resultados parecían depender de la suerte. De hecho, es la suerte-azar el motor primero de los estudios formales en cuanto a los fenómenos probabilísticos. "A Girolamo Cardano (1501 - 1576), físico, astrólogo y matemático, se le atribuye la primera discusión sobre probabilidad en su manual para jugadores **Liber De Ludo Aleae** (Manual para tirar los dados)."[7]

Quizá la incorporación de estos términos (suerte, azar, aleatoriedad) al ser involucradas con cuestiones científicas pueda causar cierto recelo. La visión tradicional de la ciencia remite a las nociones de orden, linealidad y seguridad, mientras que lo relacionado con el azar tiene una connotación negativa, diametralmente opuesta cuando no se le conoce bien, dado que la irregulari-

[7]op. cit. WILLOUGHBY, Stephen S; **Probabilidad y Estadística**; México; Publicaciones Cultural S. A.; 1977; pp. VIII

dad sólo es aparente, es decir, obedece a un tipo de orden diferente, cuestión que choca con el sentido común (que sigue interfiriendo con la verdadera comprensión científica de los sistemas objeto de estudio de las ciencias). Sin embargo, el estudio de la probabilidad ha evolucionado y en la actualidad va más allá de la incertidumbre intrínseca en los juegos de azar.

"En 1654, Antoine Chevalier de Méré, quién tenía un particular interés por los juegos de azar, planteó a Blaise Pascal y a Pierre de Ferrmat, un dilema sobre un juego que consistía en lanzar un par de dados 24 veces. El problema consistía en decidir si ó (sic) no apostar a que durante los 24 lanzamientos ocurrirían al menos un par de seises. [...] De las discusiones entre Pierre de Fermat y Blaise Pascal surgieron los principios fundamentales de Teoría de Probabilidad planteados por primera vez, no obstante que en los siglos XV y XVI algunos italianos y alemanes habían discutido la cuestión en algunos problemas de juegos."[8]

A partir de entonces, comenzó el desarrollo formal de la teoría de la probabilidad. Ésta, como se podrá notar, encarnó la actitud escéptica académica en tormo al conocimiento del universo aunque no como teoría del conocimiento, sino como una herramienta teórico-metodológica con la cual se puede cuantificar la posibilidad de que ocurra cierto evento o resultado. Las novedades teóricas en el campo de la probabilidad y la estadística, desencadenaron profundas discusiones en torno a la naturaleza del universo y por ende del conocimiento. El término *probabilidad* fue sintetizado en lenguaje formal y aplicado al estudio de sistemas cuyos resultados no era posible predecir con exactitud, aún cuando se parte de las mismas condiciones iniciales.

[8]Tomado de http://www.mor.itesm.mx/~logica/logyprob/page3.html Jueves 29 de septiembre de 2005; 14:40 PM

Para los científicos del renacimiento, las propuestas filosóficas griegas cobraron una gran importancia como paradigma. La teoría de las ideas de Platón se retomó para el estudio del universo. Isaac Newton, por ejemplo, estaba convencido de la naturaleza determinista del Universo, de la existencia de un orden que permite predecir el resultado de cualquier fenómeno. Así, convencido de ello, y de que este orden puede ser expresado mediante el lenguaje matemático, sintetizó sus leyes en ecuaciones lineales ($f = m \cdot a$, por ejemplo) aplicables a sistemas particulares.

"Newton aplicó además un concepto que ha sido esencial para el método científico: el de aislar idealmente el sistema dinámico que se desea examinar del resto del universo del que forma parte. Esto permite que su comportamiento se pueda comprender, ya que no es necesario considerar todas sus infinitas relaciones con el universo, lo que sólo sería posible para un ser infinito. Basta entonces con que sólo se consideren las características del sistema que son relevantes al fenómeno que se desea estudiar."[9]

Aislar al sistema que se desea estudiar, implica delimitación, certeza y control del número y naturaleza de la o las variables a considerar en un experimento o investigación científica. Conocer y/o determinar las condiciones iniciales de un fenómeno, permite la observación de los resultados y las variantes (en caso de haberlas) por cada evento; esto permite, en primer lugar, la clasificación del fenómenos en función de sus resultados: Si el resultado es constante, entonces el sistema-fenómeno será determinista, de lo contrario será aleatorio; adicionalmente, se puede pasar a la realización de inferencias en función de las observaciones pertinentes al resultado del conjunto de eventos.

[9] JOSÉ, Sametband Moisés; **Entre el Orden y el Caos. La Complejidad**; México; FCE, SEP, CONACYT, 2^a edición; 1999; pp.17

La resistencia a aceptar que el conocimiento del universo es imposible dada su complejidad, se exacerbó hacia el siglo de la ilustración. En esta época Pierre Simon Laplace (1749-1827) en su texto *Teoría Analítica de las Probabilidades*, afirmaba que el Universo es determinista y decía: "Debemos considerar el estado presente del Universo como efecto de su estado anterior y como causa de su estado futuro." Mas la aparente linealidad con que esta expresión describe al cosmos, no es tan sencilla como parece, este personaje sí veía al Universo como una máquina perfectamente predecible –como un reloj–, pero estaba conciente de la complejidad que posee y debido a ello, manejaba la idea de que sólo una inteligencia superior –el llamado demonio matemático de Laplace– que, por un instante, conociese *todas* las fuerzas de que está animada la naturaleza y la situación respectiva de los seres que la componen, sería capaz de conocer mediante un análisis matemático el estado pasado y futuro de todo cuanto compone el universo.[10]

Es, quizá, en este sentido que el filósofo y matemático francés Henri Poincaré (1854 -1912) caracterizara al término *azar* como "la medida de la ignorancia." En su libro *Ciencia y Método*, afirmaba que los pequeños detalles que se descuidan –o simplemente no se pueden medir ni controlar– al efectuar un experimento, pueden causar importantes variaciones en sus resultados; por lo que, al observar comportamientos imprevistos (debidos a esos detalles) en eventos de fenómenos que se suponen controlados inicialmente, se dice que son producto de la suerte, del azar. Así, dicho personaje se adelantó en su época al estudio de un antiguo concepto: el *caos*, dado que esta observación, es uno de los indicadores que coadyuvan a efectuar la distinción entre lo aleatorio y lo caótico: "el caos es una medida para extraer y mostrar la aleatoriedad inherentes en las condiciones iniciales"[11] de un fenómeno.

[10] LAPLACE en JOSÉ, Sametband Moisés; ibidem; México; FCE; 1999; pp. 22

[11] STEWART, Ian; **¿Juega Dios a los Dados? La Nueva Matemática del Caos**;

A la luz de estas consideraciones (las características, causas y efectos del azar), la concepción de la ciencia cambia. Inicia la búsqueda de nuevos enfoques para abordar fenómenos cuya complejidad de afectaciones, debido a su interacción con el medio y la sensibilidad a las variaciones en las condiciones iniciales, inciden significativamente en los resultados de la dinámica del sistema; es aquí donde la mancuerna Probabilidad - Estadística comienza a cobrar importancia; asimismo, la metodología y técnica, ampliaron su horizonte hermenéutico ante la complejidad de las relaciones que los componentes de ciertos sistemas anteponen a su estudio; ejemplo de ello (en las Ciencias Sociales) es Quètelet y su interés estadístico por los sistemas sociales, donde las **variables ocultas**[12] obligan a su estudio a partir de la teoría de las probabilidades. En adelante, los avances en la teoría de la probabilidad, serán aplicados en la estadística inferencial, ciencias sociales, la física (mecánica cuántica), psicología y más.

8.3. Fenómenos

Ante tal panorama y antes de revisar algunos conceptos fundamentales sobre la *teoría de la probabilidad*, los cuales se emplean en combinación con los métodos estadísticos en el estudio de problemas concernientes a las Ciencias Sociales, se realizará una breve puntualización de los tipos de fenómenos objeto de estudio de la ciencia.

8.3.1. Fenómenos deterministas

Los fenómenos deterministas son aquellos que al repetirse en igualdad de condiciones, (o lo más cercano a la igualdad de condiciones) producirán los

Barcelona; Crítica ; 2001; pp. 317

[12]Ibidem; STEWART, Ian; Barcelona; Crítica; 2001; pp. 316

mismos resultados. Por su carácter, una notable cantidad de fenómenos deterministas permiten realizar *predicciones sobre su comportamiento.* Ejemplos de fenómenos deterministas:

- La temperatura a la que hierve el agua al nivel del mar
- Velocidad de la caída de un cuerpo
- Ocurrencia de eclipses
- Ciclos de la luna y por lo tanto de las mareas

No obstante, es importante señalar que no todos los fenómenos deterministas son *predecibles.* El determinismo es una condición verificable a corto plazo; se puede ejecutar un experimento en igualdad de condiciones (o lo más cercano a la igualdad de condiciones) e iguales resultados (bis), pero la estabilidad de los resultados dependerá, entre otras cosas, de pequeñas variaciones en las condiciones iniciales (como se verá en lo referente a fenómenos caóticos), de la posibilidad de medir con exactitud y precisión los resultados y de la intervención de factores que están fuera de nuestro conocimiento o control.

8.3.2. Fenómenos probabilísticos

Son aquellos en los que su repetición en igualdad de condiciones (o lo más cercano a la igualdad de condiciones) no necesariamente produce los mismos resultados. Por ejemplo:

- Número de la lotería que saldrá premiado.
- La caída de un rayo.
- Ocurrencia de terremotos, huracanes y otros desastres naturales.
- Número de accidentes en una cierta avenida.

Los fenómenos probabilísticos también son denominados fenómenos *aleatorios* o *estocásticos*. Por su naturaleza, los fenómenos probabilísticos no son propiamente *predecibles a corto plazo* y por lo tanto, se han desarrollado técnicas para la *estimación* de su comportamiento.

Los fenómenos aleatorios, *ofrecen resultados múltiples* (pese a la igualdad de las condiciones iniciales bajo las cuales se desarrollan) a los experimentos observados, debido a la complejidad de las relaciones que el sistema guarda con su medio ambiente (el macrosistema).

La aleatoriedad es producto de la *falta de información* respecto a las afectaciones que elementos no considerados infringen en los sistemas observados. Dichas afectaciones se hacen manifiestas luego de largos periodos de tiempo, por lo que las regularidades propias de los sistemas estocásticos sólo se pueden notar tras "varias" repeticiones; luego entonces, aleatorio no es sinónimo de desordenado o caótico.

8.3.3. Fenómenos caóticos

Es importante mencionar que es común confundir los *fenómenos caóticos* con los probabilísticos. Sin embargo, ambos son de naturaleza muy diferente. Los fenómenos caóticos son, en esencia, fenómenos deterministas con *un alto grado de sensibilidad a las condiciones iniciales*, es decir, un pequeño cambio en las condiciones iniciales de un fenómeno, no producirá un pequeño cambio en sus resultados, sino que ocasionará un dramático cambio en sus resultados. Ejemplos de fenómenos caóticos:

- Las condiciones atmosféricas y climáticas
- El comportamiento de los mercados bursátiles

- El comportamiento sobre reproducción y mortalidad de ciertas especies de peces
- Ciertos padecimientos asociados al ritmo del corazón humano

Los fenómenos caóticos *no son predecibles*, sin embargo, debe aclararse que caótico no significa *desordenado*, ya que como han demostrado científicos como Edward Lorenz, Benoit Mandelbrot, Richard Feigenbaum y otros, las técnicas para el estudio de fenómenos caóticos producen modelos con patrones ordenados. De estas reflexiones surge la actual *Teoría del Caos* y el estudio de su representación geométrica: los **fractales**. Una vez revisados los tipos de fenómenos anteriores, puede puntualizarse lo siguiente:

- Únicamente son predecibles los fenómenos deterministas que tienen un comportamiento periódico (Eclipses, fases de la luna, mareas…)
- Los fenómenos deterministas que no son periódicos, se dice que son caóticos y no son predecibles por su alta sensibilidad a las condiciones iniciales (movimientos bursátiles, fenómenos climáticos, y otros)
- Los fenómenos probabilísticos siguen un orden diferente a los fenómenos deterministas. Debido a su complejidad por la gran cantidad de “minúsculas” variables que intervienen en su desarrollo, estos fenómenos tienden a mostrar su patrón de comportamiento a largo plazo, es decir, después de un gran número de ocurrencias. Es importante señalar que un fenómeno probabilístico no es **fortuito**.

Ejercicios 1

- *Instrucciones*: Responde breve y suficientemente a las preguntas que se te presentan.
- Coteja tus respuestas con las de la página 437

1. ¿Cuál es la principal característica de la primera etapa que experimentó el desarrollo histórico del conocimiento científico?

   ____

   ____

   ____

   ____

2. ¿De qué otra forma se le conoce a la tercera etapa de la filosofía que incluye al conocimiento como objeto de estudio?

   ____

   ____

3. ¿Cuáles son los tres problemas fundamentales de la teoría del conocimiento?

   ____

4. Explica por qué (intrínsecas) en la filosofía platónica es posible vislumbrar corrientes epistemológicas del pensamiento tales como el dogmatismo y el escepticismo.

   ____

   ____

   ____

   ____

   ____

   ____

5. ¿Cuál es la postura epistemológica que corresponde con el pensamiento-acción de la metodología estadística como actitud y por qué?

6. ¿Cuál es el papel que desempeña la suerte o azar en la conformación de la teoría de la probabilidad?

7. ¿Cuál fue una de las más importantes aportaciones de Isaac Newton a la *metodología* científica en general?

8. ¿De qué forma, el llamado demonio matemático de Laplace, sería capaz de conocer el estado pasado y futuro de todo cuanto compone al Universo?

9. Explica por qué el matemático francés Henrry Poincaré, caracterizó al término azar como la medida de la ignorancia.

10. ¿Qué tipo de fenómenos deterministas permiten realizar predicciones en torno a sus resultados?

11. Enuncia dos factores que dan pie a la clasificación de algunos fenómenos como aleatorios:

12. ¿Cuál es la principal característica de los fenómenos caóticos?

8.4. Conceptos de probabilidad

La teoría de la probabilidad no puede garantizar o negar la ocurrencia de un event, sólo puede asignar un valor numérico indicando, en forma de porcentaje, qué tan factible es la ocurrencia o no de un tal o cual resultado.

- Por ejemplo, al lanzar una moneda, la teoría de la probabilidad no puede garantizar que aparezca un *"sol"*, únicamente puede afirmar que esta cara de la moneda tiene una *probabilidad* del 50 % de aparecer.

El porcentaje o probabilidad de ocurrencia de un evento no se determina ***a priori***, sino que es el resultado de un proceso de realización de un gran número de ensayos del experimento y registro de resultados. Para el caso de la moneda, si la ésta se lanza 5 veces, puede que todos los resultados sean *sol*, o que aparezcan cuatro *águilas* y un *sol* u otros resultados más. Es evidente que cinco lanzamientos de monedas no son suficientes para poder apreciar un patrón de comportamiento en el fenómeno. Asimismo, si se lanza la moneda diez veces en lugar de cinco, es posible que todavía aparezca una cantidad de *águilas* mayor a la de los *soles*, por tanto, diez lanzamiento todavía no puede considerarse una cantidad adecuada para identificar un patrón de comportamiento en el experimento de lanzar una moneda.

Si se aumenta el número de ensayos a cien, mil, diez mil, un millón y así sucesivamente, el número de veces que aparece *águila* y el número de veces que aparece *sol*, se estabiliza y resulta ser igual para ambos casos, es decir, la mitad de las veces aparece *águila* y la otra mitad aparece *sol*. Esto significa, que en experimentos que tienen un conjunto posible de resultados, su repetición un elevado número de veces, arrojará la proporción de ocurrencia de cada uno de los posibles resultados. A esta proporción se le denomina *probabilidad de ocurrencia de un evento*. Una vez determinada la probabilidad de ocurrencia de un evento, pueden determinarse *a priori*, características sobre el comportamiento del mismo. Este proceso de estabilización y aparición de un patrón de comportamiento de un evento, cuando es repetido una gran cantidad de veces, que permite definir la probabilidad de ocurrencia de un evento, se conoce con la denominación de *ley de los grandes números*.

Debe señalarse que el término *repetir una gran cantidad de veces un experimento*, es poco explícito, ya que el concepto *una gran cantidad*, no hace referencia a un número en particular y más bien se trata de un concepto subjetivo. Asimismo, la determinación de la probabilidad de ocurrencia de un evento no requiere que en realidad se realice muchas veces un experimento, sino que a través del estudio matemático de un fenómeno, pueden establecerse las probabilidades de ocurrencia de sus posibles resultados. Finalmente, como un dato curioso e ilustrativo, considérense lo siguiente:

- Ejemplos del término *muchas veces* y su relatividad. Resultados empíricos del lanzamiento de una moneda.
    - "El naturalista francés Buffon (1707 - 1788) lanzó una moneda 4040 veces. Resultado: 2048 caras, una proporción de $2048\backslash 4040 = 0{,}5069$ de caras.
    - Al rededor de 1900, el estadístico inglés Karl Pearson en un acto

sin precedentes lanzó una moneda 24000 veces. Resultado: 12012 caras, una proporción de 0,5005

- El matemático inglés Johon Kerrich, mientras fue prisionero de los alemanes en la segunda guerra mundial, lanzó una moneda 10000 veces. Resultado: 5067 caras, una proporción de 0,5067"[13]

8.5. Definición clásica de probabilidad

"La palabra *probabilidad* se usa para indicar la posibilidad de que ocurra un evento o resultado. La definición clásica de probabilidad y, en cierto modo, la más simple, se usa cuando un experimento puede tener solamente ciertos resultados definidos, cada uno de los cuales es igualmente probable: Si hay N elementos en el conjunto de resultados posibles, la probabilidad para cualquiera de ellos es $\frac{1}{N}$."[14] Para lograr una mejor comprensión y manejo de éste término, es necesario revisar las siguientes definiciones.

Definiciones:

Experimento.- Es un fenómeno o acción que tiene un conjunto de posibles resultados, los cuales son de carácter aleatorio. Se denotan con la literal $\boldsymbol{F}$.

Evento.- Es un resultado en particular de un experimento. Se denota con letra $\boldsymbol{E}$.

[13] MOORE, David S.; en ARTHUR, Steen Lynn (editor); **La Enseñanza Agradable de las Matemáticas**; México; Limusa Noriega Editores, Colección Textos Politécnicos, Serie Matemáticas; 1998; pp. 106

[14] op. cit. WILLOUGHBY, Stephen S; México; 1977; pp. VIII

Espacio muestral.- Es el conjunto de todos los posibles resultados de un experimento. Se denota con el símbolo Ω.

Probabilidad de ocurrencia de un evento E.- Se determina al dividir el número total de casos favorables del espacio muestral entre el número total de elementos del mismo. La probabilidad de ocurrencia de un evento E, se denota por ***P(E)***.

- Ejemplo 1: Supón que se lanza una moneda legal. Determina F, Ω, y la probabilidad de ocurrencia del evento *águila*.
  - Solución:
    - $F =$ Lanzamiento de una moneda.
    - $\Omega = \{A, S\}$. Se observan dos posibles resultados del experimento: $A =$*águila* y $S =$*sol*.

    Si $E =$ *Cae águila*, entonces, número de casos favorables (número de águilas en el espacio muestral) $= 1$, número de elementos del espacio muestral (número de posibles resultados) $=2$.

    - Por lo tanto, $P(E) = \frac{1}{2} = 0{,}5$ es decir, existe un $50\,\%$ de probabilidad de que al lanzar una moneda aparezca *águila*.
- Ejemplo 2: Se lanzan dos monedas. Determina F, Ω, y la probabilidad de que no aparezca ninguna águila.
  - Solución:
    - $F =$ Lanzamiento de dos monedas.
    - $\Omega = \{AA, AS, SA, SS\}$. Se observan cuatro posibles resultados del experimento: *águila-águila*, *águila-sol*, *sol-águila*, *sol-sol*. Donde $A =$*águila* y $S =$*sol*.

Si E = *No aparece águila*, entonces, número de casos favorables (número parejas de caras de la moneda en las que no aparece *águila*en el espacio muestral) = 1, número de elementos del espacio muestral (número de posibles resultados) =4.

- Por lo tanto, $P(E) = \frac{1}{4} = 0{,}25$, es decir, existe un 25 % de probabilidad de que al lanzar dos monedas, no aparezca *águila*.

- Ejemplo 3: Supón que una urna se depositan nueve canicas de colores, de la siguiente forma: dos canicas blancas (denotadas por B), cuatro canicas negras (denotadas por N) y tres canicas rojas (denotadas por R). Se agita la urna y se selecciona al azar una canica. Determinar F, Ω y la probabilidad del evento E_1 = sale canica blanca, E_2 = sale canica negra y E_3 = sale canica roja.

  - Solución:

    - F =Urna con nueve canicas de colores.
    - $\Omega = \{B, B, N, N, N, N, R, R, R\}$. Se observa un total de nueve resultados.
    - $P(E_1)$ = Número de casos favorables (canicas blancas en la urna)=2, entre el número de posibles resultados del espacio muestral (nueve canicas). Es decir, $P(E_1) = \frac{2}{9}$.
    - $P(E_2)$ = Número de casos favorables (canicas negras en la urna)=4, entre el número de posibles resultados del espacio muestral (nueve canicas). Es decir, $P(E_2) = \frac{4}{9}$.
    - $P(E_3)$ = Número de casos favorables (canicas rojas en la urna)=3, entre el número de posibles resultados del espacio muestral (nueve canicas). Es decir, $P(E_3) = \frac{3}{9}$.

Obsérvese que la suma de las probabilidades de todos los eventos del espacio muestral es igual a la unidad: $\frac{2}{9} + \frac{4}{9} + \frac{3}{9} = 1$.

8.6. Variables aleatorias

Sea un experimento F y su espacio muestral Ω. Una *función* X que asocia a cada uno de los elementos de $\omega \in \Omega$ un número real $X(\omega)$, se denomina *variable aleatoria* (donde ω es la letra minúscula del alfabeto griego: omega y denota a los elementos del conjunto Universo Ω). Por ejemplo, retomando el ejemplo de la sección anterior sobre el experimento que consiste en lanzar dos monedas, puede definirse la variable aleatoria: $X =$Número de *soles* que aparecen. Se tiene que $\Omega = \{AA, AS, SA, SS\}$, por tanto, la variable aleatoria X asume los siguientes valores:

- $X(AA) = 0$ (Cero *soles*)
- $X(AS) = 1$ (Un *sol*)
- $X(SA) = 1$ (Un *sol*)
- $X(SS) = 2$ (Dos *soles*)

Si se calcula la probabilidad de cada uno de los eventos anteriores, se tiene que:

- $P(AA) = \frac{1}{4}$
- $P(AS) = \frac{1}{4}$
- $P(SA) = \frac{1}{4}$
- $P(SS) = \frac{1}{4}$

Las probabilidades anteriores se conocen como la *distribución de probabilidad* de la variable aleatoria X.

Ejercicios 2

- *Instrucciones*: Responde brevemente y obtén los resultados requeridos sin omitir paso alguno para su cálculo. Al finalizar, compara tus respuestas con las de la página 440

1. Sintetiza la ley de los grandes números.

   ____

   ____

   ____

   ____

2. ¿Cómo se determina la probabilidad de ocurrencia de un evento?

   ____

   ____

   ____

   ____

3. Se lanza un dado de seis caras. Determina F, Ω y la probabilidad del evento: *La cara que cae es menor que tres.*

4. Se lanzan tres monedas. Determina F, Ω y la probabilidad del evento: *aparecen dos soles.*

5. Una urna contiene diez canicas, 3 blancas, 6 negras y una roja. Determina F, Ω y la probabilidad de que salga una canica blanca, la probabilidad de salga una canica negra y la probabilidad de que salga una canica roja.

6. Retoma el reactivo 4 (se lanzan tres monedas) y define la variable aleatoria $X =$ número de *águilas* que aparecen.

8.7. Autoevaluación

- *Instrucciones*: Repasa los temas que se te hayan dificultado o aquellos en los que te equivocaste al resolver los ejercicios de la unidad.

- Una vez hecho lo anterior, desarrolla las actividades que a continuación se te proponen y verifica que tus respuestas sean correctas comparándolas con las de la página 443

1. Enuncia tres elementos socioculturales que señalaron la insuficiencia del paradigma positivista para explicar el universo.

   ____

   ____

   ____

   ____

2. ¿La puesta en jaque del método único para el estudio del Cosmos desplazó al paradigma positivista con el surgimiento de la teoría de la probabilidad?

3. ¿Cómo se le denomina al error inherente a toda investigación?

4. ¿En qué sentido se le confiere validez a los estudios realizados bajo la metodología estadística (en particular con el muestreo probabilístico), que se encuentra relacionada con la postura epistemológica denominada escepticismo medio o académico?

5. ¿Cómo caracterizó el filósofo y matemático Henri Poincaré al término azar?

6. ¿Qué son los fenómenos caóticos?

7. ¿Cuál es la función de la probabilidad?

8. Sintetiza la ley de los grandes números.

9. Se toman todas las cartas correspondientes a las figuras: corazones y diamantes, de una baraja; es decir: As, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, sota (11), reina (12) y rey (13), de cada figura y *se saca una carta*. Determina F, Ω, y las probabilidades de ocurrencia:

   a) la carta que sale es mayor que diez *b*) la carta que sale es menor que 8 y mayor que 2

c) la carta que sale es un as.

10. Supóngase que se debe hacer un análisis a 2 películas de diferente nacionalidad, para ello se tiene que elegir, al azar, a la pareja de entre 4 filmes con las siguientes nacionalidades: Francesa (denotada por F), Búlgara (denotada por B), Mexicana (denotada por M) y Norteamericana (denotada por N); determinar Ω, F, la variable aleatoria $X =$ número de mexicanas que aparecen y la distribución de probabilidad.

8.8. Apéndice del capítulo

8.8.1. Respuestas a los ejercicios del capítulo

Ejercicios 1

1. ¿Cuál es la principal característica de la primera etapa que experimentó el desarrollo histórico del conocimiento científico?
   Es la intensa búsqueda de por el conocimiento del universo en su aspecto físico, a esta etapa se le conoce como Filosofía Especulativa o Científica.

2. ¿De qué otra forma se le conoce a la tercera etapa de la filosofía que incluye al conocimiento como objeto de estudio?
   Filosofía Crítica.

3. ¿Cuáles son los tres problemas fundamentales de la teoría del conocimiento?
   El origen, la esencia y posibilidad del conocimiento.

4. Explica por qué (intrínsecas) en la filosofía platónica es posible vislumbrar corrientes epistemológicas del pensamiento tales como el dogmatismo y el escepticismo.
   Porque Platón decía que todo cuanto nos rodea, igual que nosotros, está de tránsito; que el mundo es devenir, cambio; que las cosas son y no son (esepticismo en cuanto a la posibilidad de explicar los fenómenos por sí mismos). Sin embargo, tras el aparente caos y confusión, existe un orden permanente (el mundo de las ideas) discernible por la mente huma (dogmatismo).

5. ¿Cuál es la postura epistemológica que corresponde con el pensamiento-acción de la metodología estadística como actitud y porqué?
   El escepticismo medio o académico, que si bien niega la posibilidad del conocimiento, no es tan exacerbado como el escepticismo pirrónico, pues afirma que no es posible el *conocimiento exacto* de la realidad, es decir, no existe la certeza absoluta, únicamente la *probabilidad* de que nuestras afirmaciones, en torno un fenómeno-objeto en cuestión, sean verdaderas.

6. ¿Cuál es el papel que desempeña la suerte o azar en la conformación de la teoría de la probabilidad?
   Es fundamental (que funda) en cuanto a que la observación de fenómenos en los que parece intervenir lo que se denomina suerte, desata una serie

de especulaciones en cuanto a la complejidad que se presenta a los sentidos para comprender dichos fenómenos, lo que deriva en la construcción de teorías y técnicas destinadas al estudio de sistemas-fenómenos estocásticos (en principio destinados al abordaje de los juegos de azar) y que evolucionó hasta llevarlos a la aplicación en las ciencias.

7. ¿Cuál fue una de las más importantes aportaciones de Isaac Newton a la *metodología* científica en general?
   La idea de aislar idealmente al sistema que se estudiará; lo que facilita su estudio, puesto que debido a su delimitación, no es necesario considerar las infinitas relaciones que guarda con el medio.

8. ¿De qué forma, el llamado demonio matemático de Laplace, sería capaz de conocer el estado pasado y futuro de todo cuanto compone al Universo?
   Mediante un análisis matemático basado en el conocimiento de todas las fuerzas de que está animada la naturaleza y la situación respectiva de los seres que la componen.

9. Explica por qué el matemático francés Henrry Poincaré, caracterizó al término azar como la medida de la ignorancia.
   Él afirmaba que los *pequeños detalles* que se descuidan –o simplemente no se pueden medir ni controlar– al efectuar un experimento, pueden causar importantes variaciones en sus resultados; por lo que, al observar comportamientos imprevistos (debidos a esos detalles) en eventos de fenómenos que se suponen controlados inicialmente, se dice que son producto de la suerte, del azar, cuando en realidad son producto de variables ocultas al investigador.

10. ¿Qué tipo de fenómenos deterministas permiten realizar predicciones en torno a sus resultados?

No todos los fenómenos deterministas permiten la predicción de sus resultados, sólo aquellos que presentan comportamiento cíclico permiten realizar predicciones sobre su comportamiento.

11. Enuncia dos factores que dan pie a la denominación de algunos fenómenos como aleatorios:
    1) Multiplicidad en los resultados pese a la cercanía de igualdad de condiciones en que se da su repetición; 2) Sus resultados no son predecibles a corto plazo; 3) La falta de información (conocimiento) que prevalece en torno a las condiciones bajo las cuales se desarrollan y que les afectan.

12. ¿Cuál es la principal característica de los fenómenos caóticos?
    Que son deterministas y no predecibles debido a la gran sensibilidad que presentan a las condiciones iniciales en que se realizan, esto es; una pequeña variación a las condiciones iniciales, produce dramáticos cambios en su resultado.

8.8.2. Ejercicios 2

1. Sintetiza la ley de los grandes números.
   Cuando la frecuencia de un experimento con determinado número de posibles resultados (eventos definidos) tiende a infinito, la ocurrencia de los eventos tiende a estabilizarse con respecto al valor denominado probabilidad.

2. ¿Cómo se determina la probabilidad de ocurrencia de un evento?
   Se determina al dividir el número total de casos favorables del espacio muestral entre el número total de elementos del mismo. La probabilidad de ocurrencia de un evento E, se denota por $\boldsymbol{P(E)}$.

3. Se lanza un dado de seis caras. Determina F, Ω y la probabilidad del evento: *La cara que cae es menor que tres.*
   ◦ F =Se lanza un dado
   ◦ $\Omega = \{1, 2, 3, 4, 5, 6\}$ Se observan seis posibles resultados o eventos; donde 1 representa la cara del dado con un punto, 2 a la cara con dos puntos y así hasta el seis.
   ◦ Si E = La cara que cae es menor que tres, entonces el número de casos favorables = 2 y el número de elementos que componen el espacio muestral = 6.
   ◦ Por lo tanto, $P(E) = \frac{2}{6} = 0{,}33$ es decir, 33 % de probabilidad de que al lanzar un dado de 6 caras, aparezca una cara con un número menor que tres.

4. Se lanzan tres monedas. Determina F, Ω y la probabilidad del evento: *aparecen dos soles.*
   ◦ F = Se lanzan tres monedas.
   ◦ $\Omega = \{AAA, AAS, ASS, SSS, SAA, SSA\}$ se observan seis posibles resultados del experimento A = *Águila* y S = *Sol*.
   ◦ Si E = Aparecen 2 *soles*, entonces en número de casos favorables = 2 y el número de elementos del espacio muestral = 6
   ◦ Por lo tanto $P(E) = \frac{2}{6} = {,}333\ldots$ es decir, que existe un 33 % de probabilidades de que al lanzar tres monedas aparezcan dos *soles*.

5. Una urna contiene diez canicas, 3 blancas, 6 negras y una roja. Determina F, Ω y la probabilidad de que salga una canica blanca, la probabilidad de salga una canica negra y la probabilidad de que salga una canica roja.
   ◦ F = Una urna con diez canicas, 3 blancas (denotadas por B), 6 negras (denotadas por N) y una roja (denotada por R)
   ◦ $\Omega = B, B, B, N, N, N, N, N, N, R\}$

∘ Si $E =$ sale una canica blanca, entonces el número de casos favorables $= 3$ y el número de elementos en el espacio muestral $= 10$.
∘ Por lo tanto, $P(E) = \frac{3}{10} = 0{,}3$ por lo que se puede concluir que hay un 30 % de probabilidades de sacar una canica blanca.
∘ Si $E =$ sale una canica negra, entonces el número de casos favorables $= 6$ y el número de elementos en el espacio muestral $= 10$.
∘ Por lo tanto, $P(E) = \frac{6}{10} = 0{,}6$ por lo tanto, hay un 60 % de probabilidades de sacar una canica color negro.
∘ Si $E =$ sale una canica roja, entonces el número de casos favorables $= 1$ y el número de elementos en el espacio muestral $= 10$.
∘ Por lo tanto, $P(E) = \frac{1}{10} = 0{,}1$ es decir, hay un 10 % de probabilidades de que al sacar una canica, su color sea el rojo.

6. Retoma el reactivo 4 (se lanzan tres monedas) y define la variable aleatoria $X =$ número de *águilas* que aparecen.
∘ $\Omega = \{AAA, AAS, ASS, SSS, SAA, SSA\}$
∘ Valores de la variable aleatoria X
⋄ $X(AAA) = 3$ (Tres *Águilas*)
⋄ $X(AAS) = 2$ (Dos *Águilas*)
⋄ $X(ASS) = 1$ (Una *Águila*)
⋄ $X(SSS) = 0$ (Cero *Águilas*)
⋄ $X(SAA) = 2$ (Dos *Águilas*)
⋄ $X(SSA) = 1$ (Una *Águilas*)
∘ Si se calcula la probabilidad de cada uno de los eventos anteriores, se tiene que:
⋄ $P(AAA) = \frac{1}{3}$
⋄ $P(AAS) = \frac{1}{3}$
⋄ $P(ASS) = \frac{1}{3}$
⋄ $P(SSS) = \frac{1}{3}$
⋄ $P(SAA) = \frac{1}{3}$

$\diamond\ P(SSA) = \frac{1}{3}$

8.8.3. Respuestas de la autoevaluación

1. Enuncia tres elementos socioculturales que señalaron la insuficiencia del paradigma positivista para explicar el universo.

   El progreso en el estudio del universo (la ciencia en sus diferentes ámbitos), la progresiva liberación del pensamiento, diversos avances tecnológicos, así como cierta "incapacidad" para solucionar problemas, evidenciaron la insuficiencia del paradigma positivista, y algunas de sus vertientes, al tratar de explicar, predecir y controlar las características y consecuencias de *fenómenos-objetos de estudio (constituidos como sistemas para su abordaje) que no obedecen a la pretendida dinámica lineal imperante* en el universo.

2. ¿La puesta en jaque del método único para el estudio del Cosmos desplazó al paradigma positivista con el surgimiento de la teoría de la probabilidad?

   No, ésta se amplío con las herramientas brindadas por la introducción de la teoría de la probabilidad. Aunque la probabilidad encarna (de cierta manera) una actitud escéptica entorno a la posibilidad del conocimiento, no está en contra del positivismo, más bien, amplia sus posibilidades en cuanto a la posibilidad de explicar fenómenos no deterministas, no lineales, de manera formal.

3. ¿Cómo se le denomina al error inherente a toda investigación?
   Sesgo.

4. ¿En qué sentido se le confiere validez a los estudios realizados bajo la metodología estadística (en particular con el muestreo probabilístico), que se encuentra relacionada con la postura epistemológica denominada escepticismo medio o académico?
   La metodología estadística no postula un conocimiento absoluto, lineal, ni determinista a ultranza para los fenómenos y sus eventos, sino que a partir de teorías y técnicas formales, se producen conclusiones posibles en cierta medida para elementos de un conjunto de observaciones, las cuales son aplicables al conjunto de observaciones y no de manera particular a cada elemento del conjunto.

5. ¿Cómo caracterizó el filósofo y matemático Henri Poincaré al término azar?
   Como la medida de la ignorancia, pues él afirmaba que eso a lo que se denomina suerte, no es otra cosa que falta de información (al realizar un experimento u observar un fenómeno) en torno a "pequeños" detalles que se descuidan o no se pueden medir ni controlar, pero que sí afectan significativamente los resultados de la situación en cuestión.

6. ¿Qué son los fenómenos caóticos?
   Los fenómenos caóticos son, en esencia, fenómenos deterministas con *un alto grado de sensibilidad a las condiciones iniciales*, es decir, un pequeño cambio en las condiciones iniciales de un fenómeno no producirá un pequeño cambio en sus resultados, sino que ocasionará un *dramático* cambio en sus resultados.

7. ¿Cuál es la función de la probabilidad?
   Determinar cuantitativamente la posibilidad de ocurrencia de un evento (resultado), dado cierto experimento con determinado número de posibles resultados (eventos), no la de garantizar que éste ocurra.

8. Sintetiza la ley de los grandes números.
   Cuando la frecuencia de un experimento con determinado número de posibles resultados (eventos definidos) tiende a infinito, la ocurrencia de los eventos tiende a estabilizarse con respecto al valor denominado probabilidad.

9. Se toman todas las cartas correspondientes a las figuras: corazones y diamantes, de una baraja; es decir: As, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, sota (11), reina (12) y rey (13), de cada figura y *se saca una carta*. Determina F, Ω y las probabilidades de ocurrencia:

   a) la carta que sale es mayor que diez.
   ◦ $F =$ Se saca una carta
   ◦ $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,$
   $1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$
   ◦ Si $E =$ Sale una carta mayor que 10, entonces el número de casos favorables = 6 y el número de elementos en el espacio muestral =26.
   ◦ Por lo tanto $P(E) = \frac{6}{26} = 0{,}23076923 \approx 0{,}23$ es decir que hay un 23 % de probabilidades de que salga una carta mayor que 10

   b) la carta que sale es menor que 8 y mayor que 2
   ◦ $F =$ Se saca una carta
   ◦ $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,$
   $1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$
   ◦ Si $E =$ Sale una carta menor que 8 y mayor que 2, entonces el número de casos favorables = 10 y el número de elementos en el espacio muestral = 26.
   ◦ Por lo tanto $P(E) = \frac{10}{26} = 0{,}384615384 \approx 0{,}385$ es decir que hay un 38,5 % de probabilidades de que salga una carta menor que 8 y mayor que 2

c) la carta que sale es un as.

◦ F = Se saca una carta

◦ $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,$
$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$

◦ Si E = La carta que sale es un as, entonces el número de casos favorables = 2 y el número de elementos en el espacio muestral =26.

◦ Por lo tanto $P(E) = \frac{2}{26} = 0{,}076923076 \approx 0{,}077$ es decir que hay un 7,7 % de probabilidades de que salga un as.

10. Supóngase que se debe hacer un análisis a 2 películas diferente nacionalidad, para ello se tiene que elegir, al azar, a la pareja de entre 4 filmes con las siguientes nacionalidades: Francesa (denotada por F), Búlgara (denotada por B), Mexicana (denotada por M) y Norteamericana (denotada por N); determinar F,Ω y la variable aleatoria: X = número de mexicanas que aparecen y la distribución de probabilidad.

◦ F = Elegir un par de películas.

◦ $\Omega = \{FB, FM, FN, BM, BN, MN\}$

◦ Variable aleatoria: X = Número de mexicanas que aparecen.

$X(F - B) = 0$ (Cero mexicnas)
$X(F - M) = 1$ (Una mexicana)
$X(F - N) = 0$ (Cero mexicnas)
$X(B - M) = 1$ (Una mexicana)
$X(B - N) = 0$ (Cero mexicnas)
$X(M - N) = 1$ (Una mexicana)

∘ distribución de probabilidad:
$P(F - B) = \frac{1}{6}$
$P(F - M) = \frac{1}{6}$
$P(F - N) = \frac{1}{6}$
$P(B - M) = \frac{1}{6}$
$P(B - N) = \frac{1}{6}$
$P(M - N) = \frac{1}{6}$

8.8.4. Glosario

A priori: "(lat. antes de) ideas que provienen de la experiencia pero que no dependen de ella (*cf.* Kant)."[15]

Comprensión: Actividad intelectual que se manifiesta mediante la posesión de un concepto, su explicación y ejemplificación.

Cultural: Término que puede asumir varios significados, pero que en su sentido antropológico –que es el apropiado para el presente texto– hace referencia a todo producto humano emanado de su espíritu (Razón, sentimiento y voluntad).

Fortuito: Hecho o evento acaecido

Fractal(es): "Fractal, en matemáticas, figura geométrica con una estructura compleja y pormenorizada a cualquier escala. Normalmente los fractales son autosemejantes, es decir, tienen la propiedad de que una pequeña sección de un fractal puede ser vista como una réplica a menor escala de todo el fractal. Un ejemplo de fractal es el "copo de nieve", curva que se obtiene tomando

[15]op. cit. XIRAU, Ramón; México; 1995; pp. 454

un triángulo equilátero y colocando sucesivos triángulos, cada vez de menor tamaño, en el tercio medio de los lados cada vez más pequeños. En teoría, el resultado es una figura de superficie finita pero con un perímetro de longitud infinita, y con un número infinito de vértices. En el lenguaje matemático del cálculo, dicha curva no se puede diferenciar. Se pueden construir muchas de estas figuras repetitivas aunque desde su aparición en el siglo XIX se habían considerado como un concepto extravagante."[16]

Mecanicismo materialista ateo: Tendencia del pensamiento caracterizada por una visión mecánica de los fenómenos como producto de una serie de relaciones psicofísicas, que nada tiene que ver con entes, divinos y/o ultraterrenos.

Metafísica: "parte fundamental de la filosofía que se ocupa de los primeros principios y las primeras causas de todas las cosas (Aristóteles). Teoría del ser en cuanto ser. Las principales ramas de la metafísica de la Cosmología, la Psicología Racional y la Teología Natural."[17]

Variables ocultas: "La aleatoriedad es el resultado de una *falta de información* acerca de algún sistema más amplio que incluye al que pensamos que estamos considerando. Si supiéramos qué estaban haciendo estas «variables ocultas», entonces dejaríamos de imaginar que el sistema es aleatorio."[18]

[16]Biblioteca de Consulta Microsoft Encarta 2005. © 1993-2004 Microsoft Corporation. Reservados todos los derechos.

[17]op. cit. XIRAU, Ramón; México; 1995; pp. 454

[18]op. cit. STEWART, Ian; Barcelona; Crítica ; 2001; pp. 316

Capítulo 9

Distribución normal

Objetivos específicos

- El estudiante empleará los diferentes conceptos que designan y denotan la *distribución normal*, así como sus expresiones matemáticas aplicadas a conjuntos de datos agrupados, para el cálculo de las probabilidades que tienen los datos de ocurrir en la realidad.

9.1. Fenómenos con distribución normal

En el presente capítulo, se revisarán fenómenos que tienen un comportamiento denominado *normal*, en el sentido de que, dada una característica, el número de personas u bjetos que presentan un *valor extremo* (muy alto o muy bajo) respecto a dicha característica es menor, en comparación con el número de personas o cosas que presentan un *valor promedio* para dicha característica.

- Ejemplos de comportamiento normal:
    - La estatura de las alumnas de preparatoria. La mayoría de ellas tiene una estatura promedio, pocas de ellas tienen una estatura muy baja. Asimismo, pocas de ellas tienen una estatura muy alta.
    - El peso de personas de sexo masculino de una cierta edad. La mayor parte de ellos tiene un peso promedio. Pocos de ellos tienen un peso muy bajo y, por otra parte, muy pocos de ellos tienen un peso muy alto.
    - El coeficiente intelectual de alumnos de licenciatura. La mayoría tiene un coeficiente intelectual promedio. Muy pocos tienen un coeficiente intelectual muy bajo y muy pocos tienen un coeficiente intelectual muy alto.

Los fenómenos que presentan la característica de que la mayoría de las observaciones se concentran respecto a un valor promedio y que adicionalmente, pocos casos se alejan por arriba o por abajo de dicha promedio, se denominan fenómenos con *distribución normal.*

La distribución normal está asociada a una gráfica con forma de *campana* (también conocida como *curva normal*), en la cual se observa que los datos tienden a concentrase en el centro y los valores extremos tienden a disminuir.

Se dice que es *normal* debido a que muchos fenómenos de la naturaleza se comportan de esta forma, entre ellos se encuentran los de interés para el comunicólogo.

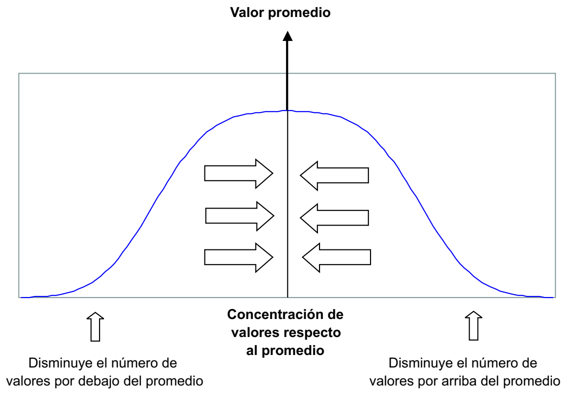


Figura 9.1: Distribución normal

Existen diversos fenómenos que, por su naturaleza, pueden expresarse a través de funciones. Por ejemplo: fenómenos físicos, químicos, fisiológicos, biológicos y sociales entre otros. Como se revisó en el capítulo *Antecedentes Matemáticos*, las funciones pueden representarse gráficamente en un plano cartesiano, en donde el eje de las x representa a la *variable independiente* y el eje de las y a la variable dependiente.

La gráfica de una función es de gran utilidad para el análisis visual de diversos fenómenos que pueden representarse a través de una función. Los *fenómenos con distribución normal* pueden estudiarse desde el concepto de función y su gráfica. La regla de correspondencia que define a una distribución normal se denota con la siguiente expresión:

$$\boxed{y = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2}}$$

y su representación gráfica:

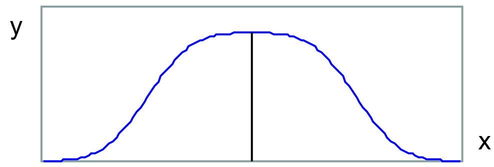


Figura 9.2: Gráfica de la distribución normal

9.2. Distribución de probabilidad normal

En el capítulo anterior se definió el concepto de distribución de probabilidad y de variable aleatoria. Emplearemos dichos conceptos para estudiar las características de diversos fenómenos probabilísticos y, en este capítulo concretamente, para fenómenos con una distribución normal. Sea X una variable aleatoria que se distribuye (comporta) con una distribución de probabilidad normal, con media μ y desviación estándar σ. Esto se denota de la siguiente manera:

$$\boxed{X \sim N(\mu, \sigma)}$$

Donde:
X = Variable aleatoria (peso, altura, coeficiente intelectual...)
μ = Media poblacional (letra del alfabeto griego llamada *mu*)
σ = Desviación estándar poblacional (sigma minúscula)

Una vez que se sabe que un fenómeno tiene distribución normal, si se conoce la media y la desviación estándar de la población, pueden calcularse las probabilidades de ocurrencia de ciertos eventos, como puede ser la probabilidad de que al seleccionar a una alumna de una preparatoria, su estatura sea mayor que un cierto valor, o menor que el mismo, o que se encuentre entre dos valores específicos. Lo mismo puede aplicarse para el peso de personas de sexo masculino o para el coeficiente intelectual de alumnos de licenciatura, entre otros.

9.2.1. Área bajo una curva normal

Antes de explicar cómo se calculan las probabilidades de ocurrencia de fenómenos con una distribución normal, se explicará el concepto de *cálculo de área bajo la curva*, ya que de él se desprenden las técnicas para el cálculo de probabilidades.

Dado un fenómeno que tiene una distribución normal, éste puede representarse con la regla de correspondencia:

$$y = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2}$$

y su respectiva gráfica se muestra a continuación:

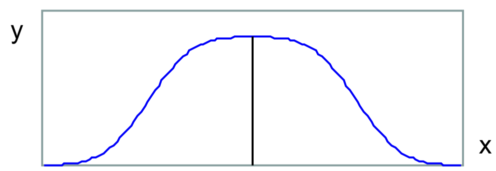


Figura 9.3: Representación gráfica de la distribución normal

Se trata de una curva simétrica respecto a la media, sin embargo, no corresponde a un polígono regular con una fórmula determinada para calcular su área, como es el caso de un triángulo, cuadrado o pentágono entre otros. Entonces, cuando se desea calcular el área de figuras cuya regla de correspondencia genera formas que no corresponden a figuras con una fórmula general, se emplean herramientas de *Cálculo Diferencial e Integral.*

Pese a que estas herramientas no corresponden a un curso de Estadística para la licenciatura en Comunicación, es importante describir de manera general, los principios sobre los que puede calcularse el área de figuras como lo son la distribución normal. El concepto que permite calcular las áreas mencionadas se denomina *integral de una función.* Por ejemplo, dada la siguiente curva normal, su área puede aproximarse, trazando segmentos rectangulares dentro de la curva:

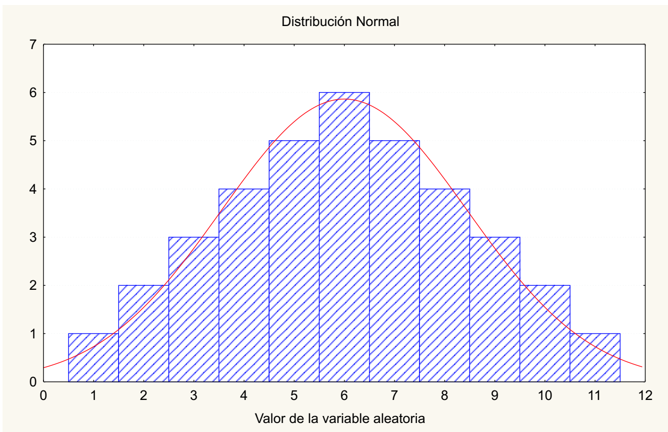


Figura 9.4: Aproximación al área bajo la curva

De tal manera que la suma de las áreas de cada rectángulo se aproxima al área bajo la curva:

Área total $= \sum_{i=1}^{n} A_i$, donde A_i corresponde al área del *i-ésimo rectángulo*

Puede apreciarse que entre más pequeña sea la base de los rectángulos, aumenta el número de éstos y, en conjunto, se ajustan más a la curva, por lo que la suma de sus respectivas áreas se aproxima más al área verdadera. Bajo supuestos y técnicas de Cálculo Diferencial e Integral, que no se detallarán pues no es el objetivo del presente texto, se hace que la base de cada rectángulo tienda a cero (se acerque a cero lo más posible sin llegar a dicho valor) con lo que el número de rectángulos tiende a infinito (el número de rectángulos se acerca a infinito). Esto significa que la suma de todas las áreas de este *número infinito* de rectángulos arroja el área exacta bajo la curva. Esta suma infinita se expresa de la siguiente manera:

$$\boxed{A = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2} dx}$$

Sin entrar a detalles técnicos ajenos a la asignatura de Estadística para la Licenciatura en Comunicación, puede comentarse que la expresión anterior simboliza el proceso de llevar a cabo una *suma* con un número *infinito* de sumandos. Se observa que el símbolo de suma común (Σ), se sustituye con el símbolo de *suma infinita* ($\int$), denominado *integral*.

Debe aclararse que en realidad no se lleva a cabo la suma de un número infinito de sumandos, ya que a través de conceptos propios del Cálculo Diferencial e Integral, dichas sumas infinitas pueden evaluarse sin necesidad de desarrollar un número infinito de sumandos.

Regresando al contexto de distribución normal de probabilidad, se mencionó que dicho concepto implica el cálculo de la probabilidad de ocurrencia de eventos asociados a la distribución normal, como lo es calcular la probabilidad de que en un fenómeno con distribución normal, el resultado de un experimento (estatura, peso, coeficiente intelectual, entre otros), sea mayor o menor que un determinado valor, o que se encuentre entre dos valores determinados. Pues bien, dichas probabilidades se obtienen al calcular el área bajo la curva normal, delimitada por los valores de interés. Y, como se acaba de explicar, las áreas bajo la curva de una función, se determinan a través del cálculo de una integral.

Dada la complejidad de la función que define una distribución normal, el área bajo la curva normal, no puede calcularse a través de técnicas matemáticas convencionales. Esto no debe preocupar al estudiante de Comunicación, ya que se han desarrollado tablas, denominadas *tablas de distribución normal* (al final del presente texto se incluye una en la página 486), que contienen el valor del área bajo la curva normal, por lo que no es necesario para casos prácticos del ámbito de la Comunicación, manejar conceptos técnicos de Cálculo Diferencial e Integral; ejemplo:

- Ejemplo, supóngase que en cierta universidad, se tiene registrado el coeficiente intelectual de todos los alumnos y se han calculado la media μ y desviación estándar σ poblacionales del mencionado coeficiente. Puede ser de interés, calcular la probabilidad de que al seleccionar, al azar, un alumno X de esta universidad:
    - Su coeficiente intelectual sea menor que un valor x_1 denotado por $P(X \leq x_1)$
    - Su coeficiente intelectual sea mayor que un valor x_2 denotado por $P(X \geq x_2)$
    - Su coeficiente intelectual se encuentre entre los valores x_1 y x_2, denotado por $P(x_1 \leq X \leq x_2)$

- En términos formales, esto *implicaría* calcular las siguientes integrales:
  - $P(X \leq x_1) = \displaystyle\int_{-\infty}^{x_1} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2} dx$
  - $P(X \geq x_2) = \displaystyle\int_{x_2}^{+\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2} dx$
  - $P(x_1 \leq X \leq x_2) = \displaystyle\int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2} dx$

  Como ya se explicó, el cálculo de las áreas asociadas a estas integrales, se obtiene directamente de las tablas de distribución normal (pág. 486).

Ejercicios 1

- Responde las siguientes preguntas y al finalizar, asegúrate de que las respuestas sean correctas comparándolas con las de la página 476

1. ¿A qué se le denomina comportamiento normal en un conjunto de observaciones?

   ____

   ____

   ____

2. ¿Qué ocurre con la distribución de los datos, bajo la curva de una gráfica, cuando presentan una distribución normal?

   ____

   ____

3. Observa la siguiente regla de correspondencia y dibuja su representación gráfica: $y = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2}$

4. Explica para qué se utiliza la siguiente expresión: $X \sim N(\mu, \sigma)$

   ____

   ____

   ____

5. ¿Bajo qué condiciones puede calcularse la probabilidad de ocurrencia de ciertos eventos en la realidad?

   ____

   ____

   ____

6. Dada una curva normal ¿cómo se puede obtener una aproximación del área bajo ésta?

   ____

   ____

   ____

7. ¿Qué ocurre si se reduce el número de rectángulos trazados bajo el área de una curva normal para estimar aproximadamente el área bajo la curva?

   ____

   ____

8. ¿Qué expresa la siguiente integral para el cálculo del área bajo la curva normal?

$$A = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2} dx$$

----

9. ¿Qué implica el cálculo de la distribución normal de probabilidad?

----

10. Expresa formalmente las tres implicaciones del anterior reactivo, tomando como punto de comparación x_1 y x_2 respecto de la probabilidad de que ocurra un evento $P(X)$

    a)

    b)

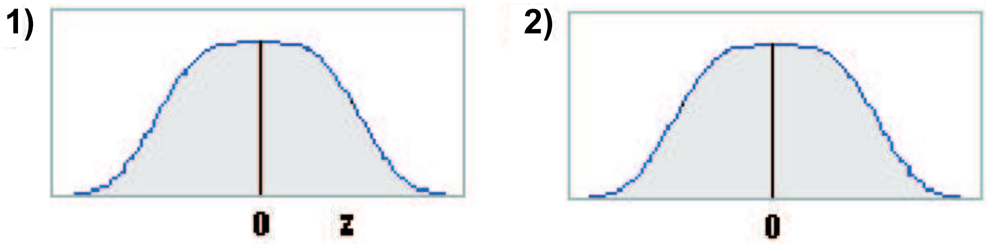
    c)

9.3. Normalización y cálculo de probabilidad

Para calcular la probabilidad de ocurrencia de eventos asociados a un fenómeno con distribución normal, es necesario considerar, por lo menos, dos propiedades y/o características de la curva normal, que son de gran utilidad para dicho objetivo:

- 1) El área total debajo de la curva normal es igual a 1 ó 100 %

- 2) La curva es simétrica respecto a la media μ, por lo que cada mitad corresponde a un área de 0,5 ó 50 %



Ahora bien, el primer paso para realizar el cálculo de probabilidad mencionado, es: *trasladar* los datos originales (variable aleatoria X, como puede ser estatura, peso, o coeficiente intelectual, entre otros) del fenómeno a una escala *adecuada*, denotada por la literal z. Este proceso se denomina *normalización* o *estandarización*, y se lleva a cabo con la siguiente expresión:

$$\boxed{z = \frac{x-\mu}{\sigma}}$$

El valor de z obtenido en el proceso anterior, se distribuye como una normal con media $\mu = 0$ y desviación estándar $\sigma = 1$, y se emplea (como se mostrará en breve) en las tablas de distribución normal para calcular la probabilidad de ocurrencia de eventos, la cual puede manifestarse de las siguientes formas:

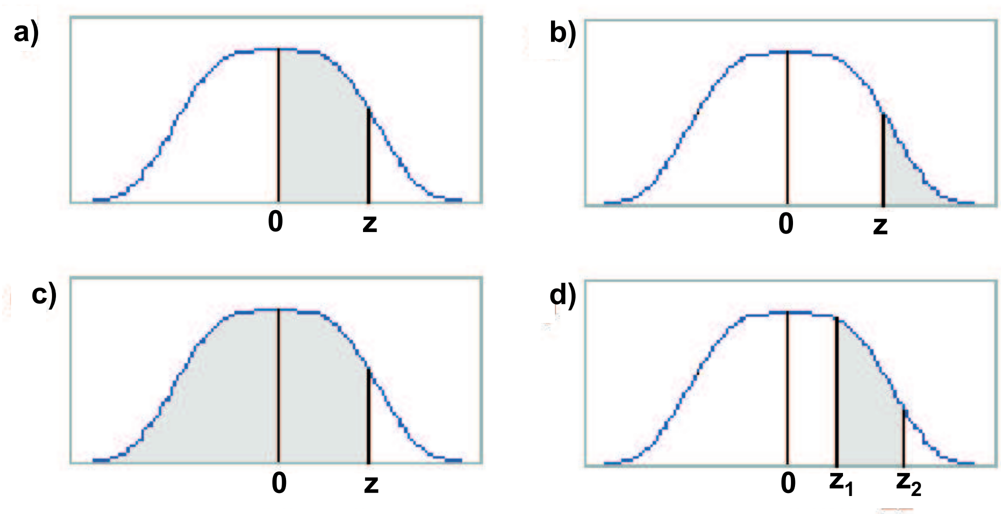


Figura 9.5: Casos de distribución normal

que corresponden a estas expresiones:

- a) $P(0 \leq a \leq z)$ (Probabilidad de que un valor a, se encuentre entre 0 y z)

- b) $P(z \leq a) = 0{,}5 - P(0 \leq a \leq z)$ (Probabilidad de que un valor a, sea mayor o igual que z)

- c) $P(a \leq z) = 0{,}5 + P(0 \leq a \leq z)$ (Probabilidad de que un valor a, sea menor o igual que z)

- d) $P(z_1 \leq a \leq z_2) = P(0 \leq a \leq z_2) - P(0 \leq a \leq z_1)$ (Probabilidad de que un valor a, se encuentre entre z_1 y z_2)

Donde a, z, z_1 y z_2, son variables estandarizadas. Asimismo, por simetría, se tienen las siguientes equivalencias:

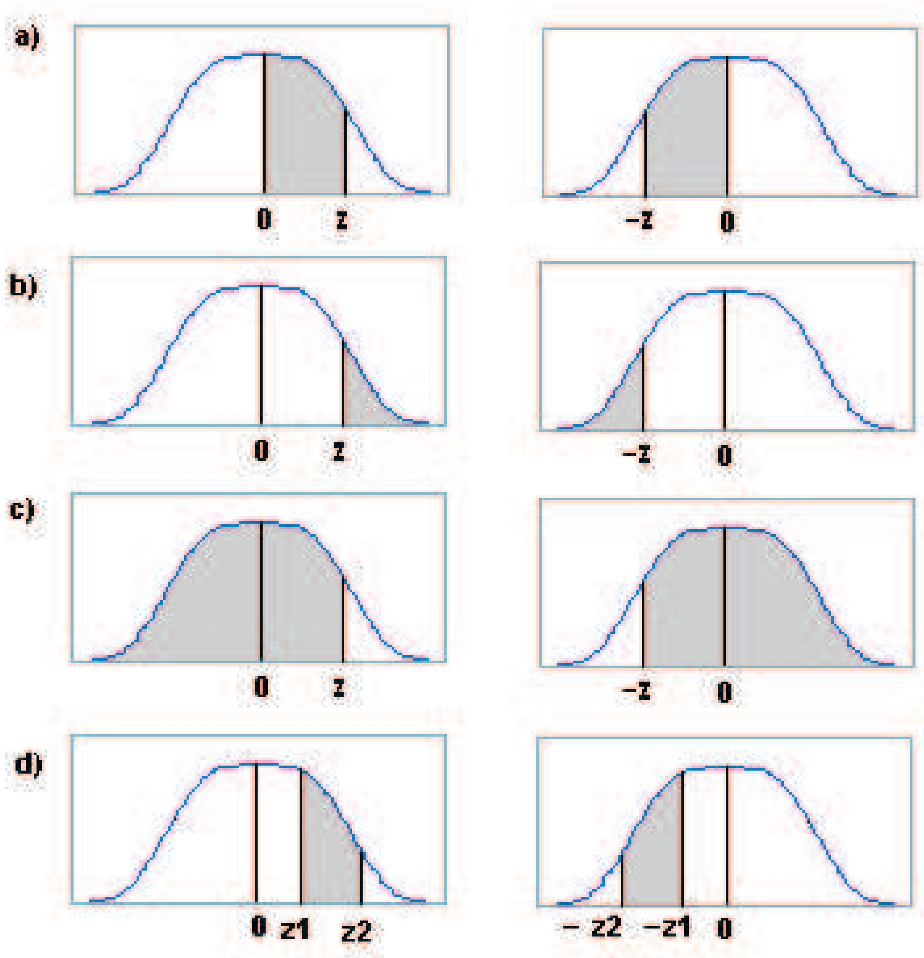


Figura 9.6: Simetría de la distribución normal

que corresponde a las siguientes expresiones:

- a) $P(0 \leq a \leq z) = P(-z \leq a \leq 0)$
- b) $P(z \leq a) = P(a \leq -z)$
- c) $P(a \leq z) = P(-z \leq a)$
- d) $P(z_1 \leq a \leq z_2) = P(-z_2 \leq a \leq -z_1)$

Donde a, z, z_1 y z_2, son variables estandarizadas.

Las tablas de distribución normal, generalmente proporcionan la probabilidad de que un valor normalizado a, se encuentre entre 0 y z, es decir $P(0 \leq a \leq z)$. En las tablas correspondientes se observa que se tienen valores para z, desde 0 hasta 3,99.

- Ejemplo:

  El coeficiente intelectual es un número que resulta de la realización de una prueba diseñada para medir las habilidades cognitivas de una persona (*inteligencia*) en relación con su grupo de edad. Se expresa en una escala estándar para que el valor promedio en un grupo de edad sea 100, es decir, una persona con un coeficiente igual a 115 está por encima de la media entre las personas de su edad. Asimismo, una persona con un coeficiente intelectual de 67, está por debajo del promedio. Como ya se ha establecido, puede considerarse que el coeficiente intelectual sigue una distribución normal. A partir de este punto pueden calcularse algunas probabilidades sobre el coeficiente intelectual.

Entonces, X es la variable aleatoria asociada al coeficiente intelectual de los alumnos de licenciatura de una universidad. Si X sigue una distribución normal con media $\mu = 100$ y desviación estándar $\sigma = 16$ es decir, $X \sim N(100, 16)$.

Calcula mediante el uso de tablas de distribución normal (pág. 486), las siguientes probabilidades:

1. Probabilidad de que el coeficiente intelectual de un alumno sea mayor que 80, es decir $P(80 \leq x)$
2. Probabilidad de que el coeficiente intelectual de un alumno sea mayor que 105, es decir $P(105 \leq x)$
3. Probabilidad de que el coeficiente intelectual de un alumno sea menor que 80, es decir $P(x \leq 80)$
4. Probabilidad de que el coeficiente intelectual de un alumno sea menor que 105, es decir $P(x \leq 105)$
5. Probabilidad de que el coeficiente intelectual de un alumno se encuentre entre 105 y 110, es decir $P(105 \leq x \leq 110)$
6. Probabilidad de que el coeficiente intelectual de un alumno se encuentre entre 70 y 80, es decir $P(70 \leq x \leq 80)$

Solución:
1) De acuerdo a los datos del problema, se tiene que $\mu = 100$, $\sigma = 16$ y $x = 80$. Primero se normalizan los datos:
$z = \frac{x-\mu}{\sigma}$
$= \frac{80-100}{16}$
$= -1{,}25$

Este valor transforma la expresión $P(80 \leq x)$, a su equivalente normalizada: $P(z \leq a)$, donde $z = -1{,}25$ es decir, se debe calcular $P(-1{,}25 \leq a)$. Véase en la gráfica 9.6 que por simetría corresponde al caso c).

Entonces: $P(-1{,}25 \leq a) = 0{,}5 + P(0 \leq a \leq 1{,}25)$. Consultando en la tabla correspondiente, se tiene que $P(0 \leq a \leq 1{,}25) = 0{,}3944$ por lo que $P(-1{,}25 \leq a) = 0{,}5 + P(0 \leq a \leq 1{,}25) = 0{,}5 + 0{,}3944 = 0{,}8944$.

O sea, la probabilidad de que un alumno de la universidad sujeta a estudio tenga un coeficiente intelectual de 80 puntos o superior, es del 89,44 %.

2) Normalizando datos, se tiene:
$z = \frac{x-\mu}{\sigma}$
$= \frac{105-100}{16}$
$= 0{,}3125$ que redondeado a dos decimales es igual a 0,31

Este valor transforma la expresión $P(105 \leq x)$, a su equivalente normalizada: $P(z \leq a)$, donde $z = 0{,}31$ es decir, se debe calcular $P(0{,}31 \leq a)$. Véase en la gráfica 9.5 que corresponde al caso b).

Entonces: $P(0{,}31 \leq a) = 0{,}5 - P(0 \leq a \leq 0{,}31)$. De las tablas se observa que $P(0 \leq a \leq 0{,}31) = 0{,}1217$ por lo que $P(0{,}31 \leq a) = 0{,}5 - 0{,}1217 = 0{,}3783$ es decir, la probabilidad de que un alumno de la universidad sujeta a estudio tenga un coeficiente intelectual de 105 puntos o superior es del 37,83 %.

3) Normalizando datos, se tiene:
$z = \frac{x-\mu}{\sigma}$
$= \frac{80-100}{16}$
$= -1{,}25$

Este valor transforma la expresión $P(x \leq 80)$, a su equivalente normalizada: $P(a \leq z)$, donde $z = -1{,}25$ es decir, se debe calcular $P(a \leq -1{,}25)$. Véase en la gráfica 9.6 qué corresponde al caso b).

Entonces: $P(a \leq -1{,}25) = 0{,}5 - P(0 \leq a \leq 1{,}25)$. De las tablas se observa que $P(0 \leq a \leq 1{,}25) = 0{,}3944$ por lo que $P(a \leq -1{,}25) = 0{,}5 - 0{,}3944 = 0{,}1056$ es decir, la probabilidad de que un alumno de la universidad sujeta a estudio tenga un coeficiente intelectual de 80 puntos o menor es del 10,56 %.

4) Normalizando datos, se tiene:
$z = \frac{x-\mu}{\sigma}$
$= \frac{105-100}{16}$
$= 0{,}3125$ que redondeado a dos decimales es igual a 0,31

Este valor transforma la expresión $P(x \leq 105)$, a su equivalente normalizada: $P(a \leq z)$, donde $z = 0{,}31$ es decir, se debe calcular $P(a \leq 0{,}31)$. Véase en la gráfica 9.5 que corresponde al caso c).

Entonces: $P(a \leq 0{,}31) = 0{,}5 + P(0 \leq a \leq 0{,}31)$. De las tablas se observa que $P(0 \leq a \leq 0{,}31) = 0{,}1217$ por lo que $P(a \leq 0{,}31) = 0{,}5 + 0{,}1217 = 0{,}6217$ es decir, la probabilidad de que un alumno de la universidad sujeta a estudio tenga un coeficiente intelectual de 105 puntos o menor es del 62,17 %.

5) Normalizando datos, se tiene:
$z_1 = \frac{x-\mu}{\sigma}$
$= \frac{105-100}{16}$
$= 0{,}3125$ que redondeado a dos decimales es igual a 0,31

Asimismo:
$z_2 = \frac{x-\mu}{\sigma}$
$= \frac{110-100}{16}$
$= 0{,}625$ que redondeado a dos decimales es igual a 0,63

Estos valores transforman la expresión $P(105 \leq x \leq 110)$ a su equivalente normalizada: $P(z_1 \leq a \leq z_2)$, donde $z_1 = 0{,}31$ y $z_2 = 0{,}63$) es decir, se debe calcular $P(0{,}31 \leq a \leq 0{,}63)$. Véase en la gráfica 9.5 que corresponde al caso d).

Entonces: $P(0{,}31 \leq a \leq 0{,}63) = P(0 \leq a \leq z_2) - P(0 \leq a \leq z_1)$. De las tablas se observa que $P(0 \leq a \leq 0{,}31) = 0{,}1217$ y $P(0 \leq a \leq 0{,}63) = 0{,}2357$ por lo que $P(0{,}31 \leq a \leq 0{,}63) = 0{,}2357 - 0{,}1217 = 0{,}114$ es decir, la probabilidad de que un alumno de la universidad sujeta a estudio tenga un coeficiente intelectual entre 105 y 110 puntos es del 11,4 %.

6) Normalizando datos, se tiene:
$z_1 = \frac{x-\mu}{\sigma}$
$= \frac{80-100}{16}$
$= -1{,}25$

Asimismo:
$z_2 = \frac{x-\mu}{\sigma}$
$= \frac{70-100}{16}$
$= -1{,}875$ que redondeado a dos decimales es igual a $-1{,}88$

Estos valores transforman la expresión $P(70 \leq x \leq 80)$, a su equivalente normalizada: $P(z_2 \leq a \leq z_1)$, donde $z_1 = -1{,}25$ y $z_2 = -1{,}88$) es decir, se debe calcular $P(-1{,}88 \leq a \leq -1{,}25)$. Véase en la gráfica 9.6 que corresponde al caso d).

Entonces: $P(-1{,}88 \leq a \leq -1{,}25) = P(0 \leq a \leq -z_2) - P(0 \leq a \leq -z_1)$, es decir $P(0 \leq a \leq 1{,}88) - P(0 \leq a \leq 1{,}25)$. De las tablas de distribución se obtienen respectivamente los valores: 0,4699 y 0,3944 por lo que $P(-1{,}88 \leq a \leq -1{,}25) = 0{,}4699 - 0{,}3944 = 0{,}0755$ es decir, la probabilidad de que un alumno de la universidad sujeta a estudio tenga un coeficiente intelectual entre 70 y 80 puntos es del 7,55 %.

Ejercicios 2

1. ¿Cuál es el primer paso para calcular la probabilidad de ocurrencia de eventos asociados a un fenómeno con distribución normal?
____
____
____
____

2. Elabora una tabla que contenga todas las expresiones matemáticas que se han revisado hasta el momento y asegúrate de que dicha tabla contenga lo que indican las literales que en ella se encuentran, así como la utilidad de la fórmula (en qué caso se puede utilizar).

3. Para el mismo ejemplo del coeficiente intelectual, calcula las siguientes probabilidades y exprésalas en términos de porcentaje:

   a) $P(x \leq 115)$

   ____

   ____

   b) $P(115 \leq x)$

   ____

   ____

   c) $P(x \leq 90)$

   ____

   ____

   a) $P(90 \leq x)$

   ____

   ____

   b) $P(90 \leq x \leq 110)$

   ____

   ____

   c) $P(80 \leq x \leq 90)$

   ____

   ____

   d) $P(105 \leq x \leq 110)$

   ____

   ____

Consulta las respuestas a estos ejercicios en la página 479

9.4. Autoevaluación

- Supóngase que para aplicar el análisis de contenido, respecto al tratamiento que los diarios de circulación nacional dan a los temas de la figura presidencial, se desea conformar una muestra representativa; para ello, se encontró que el número de notas publicadas quincenalmente sigue una distribución normal con media poblacional de $\mu = 60$ y desviación estándar poblacional $\sigma = 9$ por lo que $X \sim N(60; 9)$. A partir de lo anterior, calcula las siguientes probabilidades y exprésalas en términos de porcentaje:

1. La probabilidad de que el periódico elegido publique, quincenalmente, cuarenta o más notas referentes a la figura presidencial, es decir:

   ____

   Normalizando datos, se tiene que:

   Lo cual permite transformar la expresión $P($________$)$ a su equivalente normalizada:

   ____

   Que gráficamente se expresa:

Por lo que: $P($________$)=$

____

Luego entonces, consultando en la tabla correspondiente:

____

Por lo tanto: $P($________$)=$

____

Es decir: Se tiene el ____________ de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

2. La probabilidad de que el periódico elegido publique, quincenalmente, cuarenta notas o menos, referentes a la figura presidencial, es decir:

____

Normalizando datos, se tiene que:

Lo cual permite transformar la expresión $P($________$)$ a su equivalente normalizada:

____

Que gráficamente se expresa:

Por lo que: $P($________$)=$

____

Luego entonces, consultando en la tabla correspondiente:

____

Por lo tanto: $P(\_\_\_\_\_\_\_)=$

____

Es decir: Se tiene el ____________ de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

3. La probabilidad de que el periódico elegido publique, quincenalmente, entre 40 y 50 notas referentes a la figura presidenciales decir:

____

Normalizando datos, se tiene que:

Lo cual permite transformar la expresión $P\_\_\_\_\_\_\_$ a su equivalente normalizada:

____

Que gráficamente se expresa:

Por lo que: $P(\_\_\_\_\_\_\_)=$

____

Luego entonces, consultando en la tabla correspondiente:

Por lo tanto: $P(\_\_\_\_\_\_\_\_)=$

Es decir: Se tiene el ____________ de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

4. La probabilidad de que el periódico elegido publique, quincenalmente, entre 50 y 70 notas referentes a la figura presidencial.

Normalizando datos, se tiene que:

Lo cual permite transformar la expresión $P\_\_\_\_\_\_\_\_$ a su equivalente normalizada:

Que gráficamente se expresa:

Por lo que: $P(\_\_\_\_\_\_\_\_)=$

Luego entonces, consultando en la tabla correspondiente:

____

Por lo tanto: $P($______$)=$

____

Es decir: Se tiene el __________ de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

5. La probabilidad de que el periódico elegido publique, quincenalmente, 80 notas, o más, referentes a la figura presidencial.

____

Normalizando datos, se tiene que:

Lo cual permite transformar la expresión $P($______$)$ a su equivalente normalizada:

____

Que gráficamente se expresa:

Por lo que: $P($______$)=$

____

Luego entonces, consultando en la tabla correspondiente:

____

Por lo tanto: P(________)=

____

Es decir: Se tiene el ____________ de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

Consulta las respuestas a la autoevalución en la página 480

9.5. Apéndice del capítulo

9.5.1. Respuestas a los ejercicios del capítulo

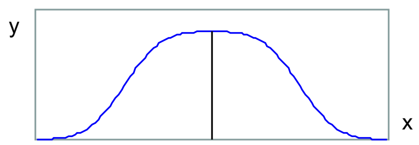
Ejercicios 1

- Responde las siguientes preguntas y al finalizar, asegúrate de que las respuestas sean correctas comparándolas con las de la página 476

1. ¿A qué se le denomina comportamiento normal en un conjunto de observaciones?
   Dada una característica, el número de objetos que presentan un valor extremo (muy alto o muy bajo), es menor en comparación con el de aquellos que presentan un valor promedio.

2. ¿Qué ocurre con la distribución de los datos, bajo la curva de una gráfica, cuando presentan una Distribución Normal?
   La mayoría de los datos tienden a concentrarse en el centro y los datos con valores extremos disminuyen su frecuencia.

3. Observa la siguiente regla de correspondencia y dibuja su representación gráfica: $y = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2}$



4. Explica para qué se utiliza la siguiente expresión: $X \sim N(\mu, \sigma)$
   Para denotar que una variable aleatoria (X) tiene una distribución de probabilidad normal (N) con media y desviación estándar poblacional μ, σ, respectivamente.

5. ¿Bajo qué condiciones puede calcularse la probabilidad de ocurrencia de ciertos eventos en la realidad?
   Cuando se sabe que el fenómeno tiene: distribución normal, se conoce su desviación estándar y media poblacional.

6. Dada una curva normal ¿cómo se puede obtener una aproximación del área bajo ésta?
   Trazando segmentos rectangulares dentro de la curva, ya que las suma de sus áreas (de los segmentos regctangulares) se aproximará al área bajo la curva normal.

7. ¿Qué ocurre si se reduce el número de rectángulos trazados bajo el área de una curva normal para estimar aproximadamente el área bajo la curva?
   El número de éstos aumenta y la precisión en el cálculo del área es mayor (se aproxima más al área verdadera).

8. ¿Qué expresa la siguiente integral para el cálculo del área bajo la curva normal?

$$A = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}[\frac{x-\mu}{\sigma}]^2} dx$$

   Una suma infinita de áreas, que tienden a cero, pertenecientes a rectángulos (sumandos) dispuestos bajo la curva descrita por la gráfica, que dará como resultado el área exacta bajo la curva, donde $\sum$ es sustituida por el signo de la suma infinita $\int$

9. ¿Qué implica el cálculo de la distribución normal de probabilidad?
   Calcular la probabilidad de ocurrencia de eventos asociados a un fenómeno que presenta distribución normal.

10. Expresa formalmente las tres implicaciones del anterior reactivo, tomando como punto de comparación x_1 y x_2 respecto de la probabilidad de que ocurra un evento $P(X)$

    a) $P(X \leq x_1)$

    b) $P(X \geq x_2)$

    c) $P(x_1 \leq X \leq x_2)$

Ejercicios 2

1. ¿Cuál es el primer paso para calcular la probabilidad de ocurrencia de eventos asociados a un fenómeno con distribución normal?
   Normalizar o estandarizar los datos originales de la distribución, esto es, trasladarlos a una escala adecuada que se denota por la literal z

2. Elabora una tabla que contenga todas las expresiones matemáticas que se han revisado hasta el momento y asegúrate de que dicha tabla contenga lo que indican las literales que en ella se encuentran, así como la utilidad de la fórmula (en qué caso se puede utilizar).

3. Para el mismo ejemplo del coeficiente intelectual, calcula las siguientes probabilidades y exprésalas en términos de porcentaje:

   a) $P(x \leq 115) = 0{,}8264$ ó $82{,}64\,\%$ de probabilidades de que el coeficiente intelectual de un alumno sea menor o igual que 115
   b) $P(115 \leq x) = 0{,}1736$ ó $17{,}36\,\%$ de probabilidades de que el coeficiente intelectual de un alumno sea mayor o igual que 115
   c) $P(x \leq 90) = 0{,}2667$ ó $26{,}67\,\%$ de probabilidades de que el coeficiente intelectual de un alumno sea mayor o igual que 90
   a) $P(90 \leq x) = 0{,}7324$ ó $73{,}24\,\%$ de probabilidades de que el coeficiente intelectual de un alumno sea menor o igual que 90
   b) $P(90 \leq x \leq 110) = 0{,}4648$ ó $46{,}48\,\%$ de probabilidades de que el coeficiente intelectual de un alumno esté entre 90 y 110
   c) $P(80 \leq x \leq 90) = 0{,}0531$ ó $5{,}31\,\%$ de probabilidades de que el coeficiente intelectual de un alumno esté entre 80 y 90
   d) $P(105 \leq x \leq 110) = 0{,}1107$ ó $11{,}07\,\%$ de probabilidades de que el coeficiente intelectual de un alumno esté entre 105 y 110

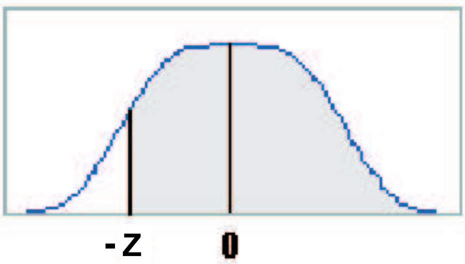
9.5.2. Respuestas de la autoevaluación

- Supóngase que para aplicar el análisis de contenido, respecto al tratamiento que los diarios de circulación nacional dan a los temas de la figura presidencial, se desea conformar una muestra representativa; para ello, se encontró que el número de notas publicadas quincenalmente sigue una distribución normal con media poblacional de $\mu = 60$ y desviación estándar poblacional $\sigma = 9$ por lo que $X \sim N(60; 9)$. A partir de lo anterior, calcular las siguientes probabilidades:

1. La probabilidad de que el periódico elegido publique, quincenalmente, cuarenta o más notas referentes a la figura presidencial, es decir:
$P(40 \leq X)$
Normalizando datos, se tiene que:

$$z = \frac{40-60}{9} = -2{,}2 \text{ es decir } z = -2{,}2$$

Lo cual permite transformar la expresión $P(40 \leq X)$ a su equivalente normalizada:
$P(40 \leq X) \equiv P(-z \leq a) = P(-2{,}2 \leq a)$
Que gráficamente se expresa:



Por lo que: $P(40 \leq X) = P(2{,}2 \leq a) = 0{,}5 + P(0 \leq a \leq 2{,}2)$ dado que por simetría, se tiene que $-2{,}2 = 2{,}2$

Luego entonces, consultando en la tabla correspondiente:

$P(0 \leq a \leq 2{,}2) = 0{,}4861$

Por lo tanto: $P(40 \leq X) \equiv P(2{,}2 \leq a) = 0{,}5 + P(0 \leq a \leq 2{,}2) = 0{,}5 + 0{,}4861 = 0{,}9861$

Es decir: Se tiene el 98,61 % de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

2. La probabilidad de que el periódico elegido publique, quincenalmente, cuarenta notas o menos, referentes a la figura presidencial, es decir: $P(X \leq 40)$

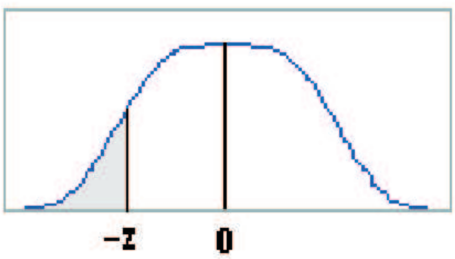
   Normalizando datos, se tiene que:

$$z = \frac{40-60}{9} = -2{,}2 \text{ es decir } z = -2{,}2$$

   Lo cual permite transformar la expresión $P(X \leq 40)$ a su equivalente normalizada:

   $P(X \leq 40) \equiv P(a \leq -z) = P(a \leq -2{,}2)$

   Que gráficamente se expresa:

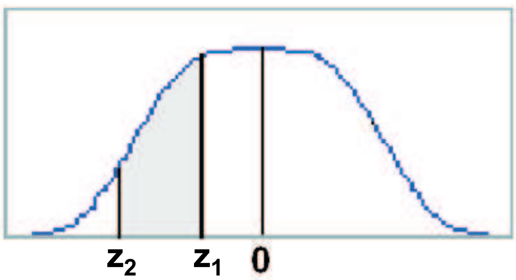


Por lo que: $P(X \leq 40) = P(a \leq -2{,}2) = 0{,}5 - P(0 \leq a \leq 2{,}2)$ dado que por simetría, se tiene que $-2{,}2 = 2{,}2$
Luego entonces, consultando en la tabla correspondiente:
$P(0 \leq a \leq 2{,}2) = 0{,}4861$
Por lo tanto: $P(X \leq 40) = P(a \leq 2{,}2) = 0{,}5 - P(0 \leq a \leq 2{,}2) = 0{,}5 - 0{,}4861 = 0{,}0139$
Es decir: Se tiene el 1,39 % de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

3. La probabilidad de que el periódico elegido publique, quincenalmente, entre 40 y 50 notas referentes a la figura presidenciales decir:
$P(40 \leq X \leq 50)$
Normalizando datos, se tiene que:

$$z_1 = \tfrac{50-60}{9} = -1{,}1 \text{ es decir } z = -1{,}1$$
$$z_2 = \tfrac{40-60}{9} = -2{,}2 \text{ es decir } z = -2{,}2$$

Lo cual permite transformar la expresión $P(40 \leq X \leq 50)$ a su equivalente normalizada:
$P(40 \leq X \leq 50) \equiv P(z_1 \leq a \leq z_2) = P(-1{,}1 \leq a \leq -2{,}2)$
Que gráficamente se expresa:



Por lo que: $P(-1{,}1 \leq a \leq -2{,}2) = P(1{,}1 \leq a \leq 2{,}2)$ dado que por simetría, se tiene que $-2{,}2 = 2{,}2$ y $-1{,}1 = 1{,}1$

$P(1{,}1 \leq a \leq 2{,}2) = P(0 \leq a \leq z_2) - P(0 \leq a \leq z_1)$

Luego entonces, consultando en la tabla correspondiente:

$P(0 \leq a \leq 2{,}2) = 0{,}4861$ y $P(0 \leq a \leq 1{,}1) = 0{,}3643$

Por lo tanto: $P(z_1 \leq a \leq z_2) = 0{,}4861 - 0{,}3643 = 0{,}1281$

Es decir: Se tiene el 12,81 % de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

4. La probabilidad de que el periódico elegido publique, quincenalmente, entre 50 y 70 notas referentes a la figura presidencial.

   $P(50 \leq X \leq 70)$

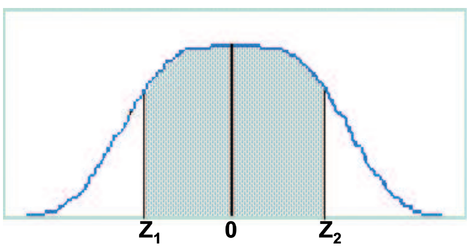
   Normalizando datos, se tiene que:

$$z_1 = \tfrac{50-60}{9} = -1{,}1 \text{ es decir } z = -1{,}1$$
$$z_2 = \tfrac{70-60}{9} = -1{,}1 \text{ es decir } z = -1{,}1$$

   Lo cual permite transformar la expresión $P(50 \leq X \leq 70)$ a su equivalente normalizada:

   $P(50 \leq X \leq 70) \equiv P(z_1 \leq a \leq z_2) = P(-1{,}1 \leq a \leq -1{,}1)$

   Que gráficamente se expresa:



   Por lo que: $P(-1{,}1 \leq a \leq -1{,}1) = P(0 \leq a \leq z_1) + P(0 \leq a \leq z_2)$

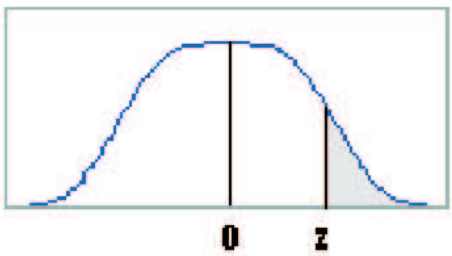
   Luego entonces, consultando en la tabla correspondiente:

$P(0 \leq a \leq 1{,}1) = 0{,}3643$
Por lo tanto: $P(z_1 \leq a \leq z_2) = P(z_1 \leq a \leq z_1) = 0{,}3643 + 0{,}3643 = 0{,}7286$ dado que por simetría, se tiene que $-1{,}1 = 1{,}1$
Es decir: Se tiene el 72,86 % de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

5. La probabilidad de que el periódico elegido publique, quincenalmente, 80 notas, o más, referentes a la figura presidencial.
$P(80 \leq X)$
Normalizando datos, se tiene que:

$$z = \frac{80-60}{9} = 2{,}2 \text{ es decir } z = 2{,}2$$

Lo cual permite transformar la expresión $P(80 \leq X)$ a su equivalente normalizada:
$P(40 \leq X) \equiv P(z \leq a) = P(2{,}2 \leq a)$
Que gráficamente se expresa:



Por lo que: $P(2{,}2 \leq a) = P(2{,}2 \leq a) = 0{,}5 - P(0 \leq a \leq 2{,}2)$
Luego entonces, consultando en la tabla correspondiente:
$P(0 \leq a \leq 2{,}2) = 0{,}4861$
Por lo tanto: $P(2{,}2 \leq a) = P(2{,}2 \leq a) = 0{,}5 - P(0 \leq a \leq 2{,}2) =$

$0{,}5 - 0{,}4861 = 0{,}0139$

Es decir: Se tiene el 1,39 % de probabilidad de elegir a un periódico que, quincenalmente, publique 40 notas o más, referentes a la figura presidencial.

9.5.3. Tabla de áreas bajo la curva normal

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999
3.1	0.99903	0.99906	0.9991	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.9994	0.99942	0.99944	0.99946	0.99948	0.9995
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.9996	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.9997	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.9998	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.9999	0.9999	0.9999	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

Capítulo 10

Técnicas de muestreo

Objetivo específico

- El estudiante identificará las principales características del *muestreo*, mediante la lectura del texto y resolución de cuestionarios, para el reconocimiento de sus funciones e importancia al interior de la investigación científica.

- El estudiante describirá los principales tipos de *muestreo*, a partir de la lectura del texto y resolución de ejercicios, para la comprensión de su aplicación a problemas de carácter social.

- El estudiante aplicará los *fundamentos probabilísticos del muestreo*, mediante el cálculo del tamaño para diferentes tipos de muestras, para la valoración de éstos como una herramienta en el proceso de inferencia estadística.

10.1. Preliminar

En el presente capítulo, se desarrolla el tema de *técnicas de muestreo*. En primer lugar, se revisa el concepto de *investigación*, como una actividad inherente al ser humano, ergo una perspectiva epistemológica, es decir, la investigación inserta en el contexto del método científico y su papel en la generación del conocimiento. Enseguida, se realiza un acercamiento al muestreo y su papel en el proceso de la *investigación científica*. Se efectúa, además, una reflexión sobre la investigación desde las perspectivas *cualitativa* y *cuantitativa*, describiendo la metodología que sigue esta última dada su aplicación en las Ciencias Exactas, Sociales, Humanísticas y la Comunicación.

Posteriormente, y antes de abordar el tema de muestreo, se desarrolla una contextualización sobre la metodología de la estadística y su aplicación en el proceso de investigación científica. Más adelante se exponen los principales métodos de muestreo, llevando a cabo una clasificación del muestreo en probabilístico y no probabilístico y, a su vez, detallando las principales técnicas, características, aplicaciones, ventajas y desventajas del primero, debido a que es éste es el método que más se emplea en la investigación científica.

Sobre el concepto de distribución muestral, que se explica en su respectiva sección, se revisan las propiedades de los estadígrafos, así como de sus distribuciones de probabilidad. Una vez hecho lo anterior, se analiza el teorema del límite central (uno de los teoremas más importantes de la estadística), se explica como una consecuencia del concepto de distribución muestral y su papel en la inferencia estadística a través de un proceso de muestreo. Finalmente se expone el error estándar y se explica su función como grado de precisión deseada para una investigación por muestreo o **cota** máxima de error a tolerar en un estudio.

10.2. Muestreo e investigación

10.2.1. La investigación científica

Desde las primeras etapas de su vida, el ser humano posee una inclinación natural hacia la investigación. En sus primeros meses de vida, los niños, con gran curiosidad, exploran mediante sus sentidos el mundo que les rodea. Una vez que son un poco mayores, continúan la identificación de su entorno auxiliándose con frecuentes preguntas dirigidas a personas de mayor edad. Ya en su juventud y edad adulta, continúan dicha actividad constantemente, aunque no siempre de manera planeada y/o formal, con la finalidad de cubrir sus necesidades en los diversos aspectos de la vida cotidiana (ya sea en lo social, laboral o afectivo entre otros).

En su vida escolar, principalmente en el nivel superior, las exigencias académicas obligan a los estudiantes a realizar investigación, que si bien ya sigue un orden más definido, no puede considerarse investigación formal. Una vez que se han culminado los estudios profesionales, se busca obtener el título de la carrera. Éste se obtiene, generalmente, a través de la elaboración de un trabajo de investigación denominado tesis, el cual, debido a que su estructura incluye elementos como objetivo, hipótesis, delimitación, marco teórico y metodológico, puede considerarse como un proceso de investigación más en forma.

Empero, los procesos de investigación científica (denominada así por aplicar y reconstruir **paradigmas** teóricos y metodológicos) trascienden el ámbito personal ante las complejas necesidades actuales de la sociedad; en ellos, el ser humano debe realizar el proceso de investigación con exacerbada rigurosidad formal, para la obtención de conclusiones que permitan tanto la generación del conocimiento teórico y/o metodológico, como la toma de decisiones y la consecuente resolución de problemas de fondo en diversos sectores.

En las diferentes disciplinas, tanto de las Ciencias Exactas como de las Ciencias Sociales y las Humanidades, se llevan a cabo investigaciones en sus respectivos campos de interés. Dichas investigaciones, además de arrojar resultados que permiten la resolución de problemas específicos, incrementan el acervo de conocimientos de su área específica de estudio, con lo cual se da continuidad al permanente proceso de construcción de la ciencia, ya que ésta es acumulable. Es importante entonces ubicar a la investigación científica, en el status de generadora de nuevos conocimientos, y en consecuencia, como uno de los principales ejes que permiten el desarrollo de una nación.

10.2.2. Enfoques de la investigación científica

Como ya se mencionó anteriormente, la investigación científica sigue una metodología sistemática, que la hace ordenada, rigurosa y contribuye a la validación científica de sus resultados. En este punto (referente al ámbito metodológico), la investigación científica puede clasificarse en dos categorías: *Investigación Cualitativa* e *Investigación Cuantitativa*. Por un lado, la investigación con enfoque cualitativo parte del hecho de que hay una realidad **intersubjetiva** que descubrir. Como premisa sostiene que la realidad es construida por los individuos, quienes dan significado al fenómeno social. El manejo de los datos es a través de lenguaje natural. Su finalidad es entender el contexto y punto de vista del actor social.

"La investigación cualitativa, dicen Denzin y Lincoln, es un campo muy amplio que atraviesa disciplinas, problemas de investigación, métodos y perspectivas epistemológicas. Es un *conjunto de prácticas interpretativas* que no se encuentra ligado con una determinada teoría o paradigma en particular, ni es privativo de una u otra área del conocimiento, ni posee sus propios métodos, sino que se vale de las aproximaciones, los métodos y las técnicas de diversas disciplinas y perspectivas teóricas"[1] para el abordaje de la realidad y construcción del conocimiento.

En contraparte, la investigación con enfoque cuantitativo parte de que hay una realidad que conocer. La realidad del fenómeno social puede *describirse* con la aplicación de teorías, metodología, técnicas y demás elementos científicos. El manejo de los datos se realiza a través de la medición y cuantificación. Su meta es obtener conclusiones del objeto de estudio con el fin de explicar y predecir su comportamiento.

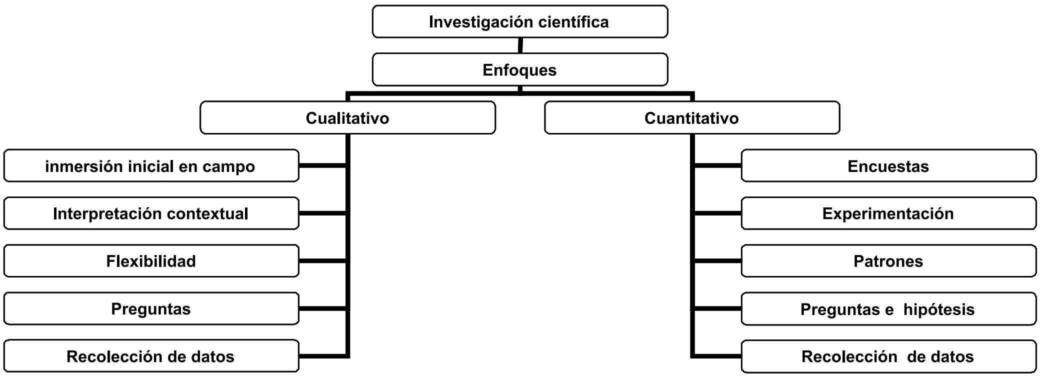
Además, "la perspectiva *cientificista/cuantitativa* defiende la existencia de un único *método* (el de las ciencias naturales y exactas), general a todas las ciencias: al igual que el principio de ***causalidad***, y la formulación de *leyes generales* en el análisis de la realidad social. El énfasis se pone en la '*explicación*, en la *contrastación* empírica y en la medición objetiva de los fenómenos sociales'."[2] De acuerdo a Hernández Sampieri,[3] los enfoques de la investigación pueden representarse con el siguiente esquema:

[1] ZAS, Ivone; et al, **Para Comprender la Subjetividad: Investigación Cualitativa en Salud Reproductiva y Sexualidad**; México; El Colegio de México; 1999; pp. 36 Cursivas agregadas

[2] op. cit. CEA, D' Ancona; María Ángeles; Madrid; 1998; pp. 44 Negritas agregadas

[3] HERNÁNDEZ Sampieri; et. al; **Metodología de la Investigación**; México; McGraw Hill; 2002.

Investigación según Hernández Sampieri



10.2.3. Metodología de la investigación con enfoque cuantitativo

En el enfoque cuantitativo, la investigación científica recurre a la recolección de datos para responder preguntas de investigación y comprobar hipótesis planteadas inicialmente. Emplea técnicas de medición y conteo y se basa en el uso de métodos estadísticos para detectar patrones de comportamiento de una población sometida a estudio.

Por lo anterior, es que el enfoque cuantitativo de la investigación científica, es de gran importancia en las Ciencias Exactas, Sociales y de la Comunicación. Padua, ha desarrollado una sólida metodología para realizar investigación en Ciencias Sociales. La importancia de su investigación radica en el hecho de que ha adaptado sus técnicas a la cultura e idiosincrasia de América Latina. Para la realización de un estudio de investigación, Padua[4] establece las siguientes fases:

[4]PADUA, Jorge. **Técnicas de Investigación Aplicadas a las Ciencias Sociales**; México; 2002; El Colegio de México, Fondo de Cultura Económica

I. Planteamiento del problema y construcción de un cuerpo de hipótesis
II. Construcción, evaluación y manejo del instrumento de recolección de datos (cuestionario) y realización del muestreo
III. Recolección de datos
IV. Procesamiento de datos
V. Análisis de datos
VI. Presentación de datos

10.3. Tipos de muestreo

La importancia del proceso de muestreo es tal, que de su correcta planeación, diseño y ejecución dependen tanto la generación de los datos sujetos a estudio, como las conclusiones que de éste se desprendan. Las técnicas de muestreo proporcionan soporte formal al proceso de investigación científica, ya que debido a sus mecanismos para la selección de los datos, se evita en gran medida el sesgo en la medición de los **atributos** del objeto de estudio. Asimismo, permite **acotar** y reducir el error de dicha medición. De no realizar un proceso de muestreo para la generación de datos, o de hacerlo en forma incorrecta, las conclusiones obtenidas por la investigación serán erróneas.

Cada técnica de muestreo permite diferentes grados de exactitud y precisión al abordar un objeto de estudio. *Existen dos tipos de muestreo: el probabilístico y no probabilístico.* El muestreo probabilístico es el de mayor confiabilidad, puesto que sus técnicas fundamentan la elección de las unidades muestrales en procedimientos lógico-matemáticos donde el nivel de confianza y error pueden ser estimados. El segundo, con sus técnicas, también puede brindar cierto "soporte formal" a la investigación, sin embargo, su nivel de confiabilidad es bajo puesto que la elección de las unidades muestrales es, incluso, arbitraria.

Lo anterior, no indica que las investigaciones de base cualitativa o los estudios no probabilísticos, carezcan por completo de validez. Para diferentes objetos de estudio, distintas formas de tratamiento. Existen fenómenos aún intactos por la investigación científica y un primer acercamiento se puede valer de técnicas muestrales no probabilísticas para obtener cierta idea de sus características. En cambio, hay otros casi agotados por ésta, o cuyos requerimientos de amplitud y profundidad exigen un acercamiento cualitativo.

10.3.1. Muestreo probabilístico

Los métodos estadísticos y las técnicas de muestreo son fundamentales en el desarrollo de un proceso de investigación en Ciencias Exactas, Sociales y de la Comunicación. A través de ellos, se recolectan datos, se organizan, analizan y se presentan, para la formulación de conclusiones y la consecuente toma de decisiones. En general, en un estudio de carácter social, se desea investigar las características que posee una población. Si se estudia a todos los miembros de una población, se dice que se realiza un *censo*. Sin embargo, las más de las veces no se disponen de los recursos, necesarios para ello.

Una alternativa consiste en seleccionar sólo a una parte de la población, es decir, estudiar y analizar una muestra para determinar las características que la describen. Si ésta es de tamaño adecuado y fue seleccionada aleatoriamente, para minimizar el sesgo y las tendencias propias del investigador, entonces las características que la describen pueden generalizarse al total de la población, y, por tanto, se puede afirmar que la muestra es representativa.

La parte de la estadística que se encarga de analizar y describir a una muestra se conoce con el nombre de *Estadística Descriptiva*, mientras que a la parte de la estadística que se ocupa de generalizar al total de la población las características obtenidas al analizar a una muestra se conoce como *Estadística Inferencial.*

Es importante reflexionar en los siguientes puntos:

- Una muestra perfecta sería como una versión a escala de la población y reproduciría cada una de las características de ésta, sin embargo, en la realidad *las muestras perfectas no existen.* "En cambio, una buena muestra reproduce las características de interés que existen en la población de la manera más cercana posible. Esta muestra será representativa, en el sentido de que cada unidad muestreada representará las características de una cantidad conocida de unidades en la población."[5]

[5]LOHR, Sharon; **Muestreo: Diseño y Análisis**; México; Thomson Learning, 2000; pp. 3

- Si las muestras perfectas no existen, entonces los datos arrojados al estudiar los elementos de una muestra, conllevan un determinado grado de incertidumbre; ésta, se debe a factores entre los que se pueden contar diversos errores personales al recabar o registrar datos, posibles errores en los instrumentos de medición y el error inherente a la conformación de una muestra obtenida de forma aleatoria.

- Dado que los datos de una muestra conllevan determinado grado de incertidumbre, *el proceso de generalizar las características que describen a una muestra al total de una población debe expresarse en términos de probabilidad*, de tal manera que aunque nunca se tiene el cien por ciento de seguridad de que las características de un muestra puedan describir exactamente al total de una población, sí se tiene una aceptable probabilidad de ello.

- Un proceso de muestreo estadístico permite medir, establecer y acotar, en términos de probabilidad, tanto la confianza (nivel de confianza) de las inferencias de una muestra generalizadas a una población, como el error que se está dispuesto a tolerar en dicho proceso de generalización (error muestral).

- En un proceso de muestreo, nunca se puede pensar en un nivel de confianza del cien por ciento ni en un error muestral de cero. Es claro que si se desea un alto nivel de confianza y un bajo error muestral, se debe aumentar el tamaño de la muestra (*ley de los grandes números*), lo que conlleva la inversión de mayores recursos materiales y humanos.

10.3.2. Muestreo no probabilístico

A diferencia del muestreo probabilístico, en el no probabilístico la selección de los elementos de la muestra no está en función de la probabilidad o de la ley de los grandes números, sino de otros factores asociados a las características de quien realiza el estudio. Este tipo de muestreo no emplea conceptos ni fórmulas de probabilidad y su empleo depende del objetivo de la investigación a realizar. Se considera que este tipo de muestreo tiene poco valor desde el punto de vista de la estimación, ya que no permite la determinación del error en la estimación de los respectivos parámetros poblacionales.

Sobre este entendido, se sabe que no hay un proceso de muestreo formal; por ende, *la muestra no aparece como representativa del universo*, lo que postula la inexistencia de una posible relación de equivalencia entre la muestra y el universo. Se emplea cuando el investigador desea utilidad en términos de inmersión en el tema de estudio, es decir, con fines de **sondeo**, de exploración y/o de un primer acercamiento al fenómeno-objeto de estudio; el caso del sondeo, es particularmente especial, puesto que de él se abusa en los medios de comunicación masiva, en tanto que para la exploración y/o un primer acercamiento, este tipo de técnicas de muestreo, pueden ser útiles cuando las fronteras del fenómeno-objeto de estudio no son claramente visibles; en general, el muestro no probabilístico se clasifica en tres categorías:

1. Muestreo casual
   El muestreo casual es, comúnmente, el más empleado por los medios periodísticos y algunas empresas dedicadas a la investigación de mercados. Consiste en entrevistar personas de forma casual, por ejemplo, uno de cada diez personas que pasan por un determinado lugar público. El muestreo casual no requiere ni de personal ni técnicas altamente especializados, por lo que su realización es de bajo costo, sin embargo tiene la desventaja de no poder generalizar sus resultados al total de una población.

2. Muestras intencionales

   En este caso, la selección de los elementos de la muestra es llevada a cabo por un experto, el cual escoge casos que, desde su particular criterio son típicos. Al igual que el caso anterior, este tipo de muestreo tiene poco o nulo valor estadístico, por lo que también se considera de carácter exploratorio o de sondeo.

3. Muestreo por cuotas

   En cierto modo es un tipo de muestreo estratificado. Tiene un importante uso en algunas empresas de investigación de mercados. A un conjunto de entrevistadores, con o sin experiencia, se les asigna una cuota o cantidad de individuos a entrevistar, indicándoles previamente las características que los entrevistados deben reunir, tal como ser de un cierto sexo, filiación o edad.

   Cada encuestador selecciona por su cuenta a sus individuos y los entrevista hasta cumplir con su cuota de sujetos. Este tipo de muestreo presenta el principal inconveniente de estar sujeto sesgo introducido por el entrevistador al emplear su criterio para seleccionar a sus individuos.

   Puede afirmarse que el muestreo no probabilístico presenta un importante grado de subjetividad, por lo que se recomienda su aplicación únicamente con fines de sondeo. Por dichas razones, a este tipo de muestreo también se le denomina *muestreo subjetivo*.

Ejercicios 1

- *Instrucciones*: Señala con una X la opción que consideres correcta y posteriormente verifica que tus respuestas sean las correctas con las que se te proporcionan en la página 554

1. ¿Cuáles son las categorías en que se puede dividir a la investigación científica desde un punto de vista metodológico?

   a) Investigación cualitativa e investigación cuantitativa.

   b) Investigación intersubjetiva y paradigmática.

   c) Prácticas interpretativas y epistemológicas.

   d) Investigación de las ciencias naturales e investigación de las ciencias exactas.

2. ¿En cuál de estas perspectivas o categorías de la investigación científica se defiende la existencia de un solo método general a todas las ciencias?

   a) La perspectiva de la causalidad hermenéutica

   b) La perspectiva cientificista cuantitativa.

   c) El principio de la intersubjetividad.

   d) La perspectiva cualitativa.

3. ¿En qué se debe fundamentar la elección de las unidades muestrales de una investigación para poder estimar su nivel de confianza y el error?

   a) En las técnicas de muestreo.

   b) En procedimientos lógico-matemáticos.

   c) En la teoría de la causalidad.

   d) En el muestreo no probabilístico.

4. ¿Por qué las técnicas de muestreo no probabilístico tienen un bajo nivel de confiabilidad?

    a) Porque se encuentran ligadas a una sola teoría o paradigma.

    b) Sólo se basan en un método único y general para toda investigación.

    c) La elección de las unidades muestrales es arbitraria.

    d) Se fundamenta en el azar y la probabilidad.

5. ¿De qué manera(s) se pueden minimizar el sesgo y las tendencias propias del investigador en una investigación?

    a) Sistematizando el proceso de investigación con objetivos, hipótesis y marcos teóricos.

    b) Investigando las características de una población con cuestionarios aleatorios.

    c) Seleccionando a la N población razonadamente y de forma proporcional.

    d) Diseñando una muestra de tamaño adecuado y seleccionando las unidades muestrales en forma aleatoria.

6. Identifica una de las premisas más importantes para las técnicas de muestreo cuantitativo.

    a) Investigación cualitativa es indispensable para la ciencia.

    b) Las muestras perfectas no existen.

    c) Las muestras representativas son copia exacta del universo a estudiar.

    d) La Investigación de las ciencias sociales es sistemática.

7. ¿Cuál es el punto de partida de la investigación cualitativa?

    a) Hay una realidad intersubjetiva qué conocer.

    b) La perspectiva social especulativa de la realidad humana.

    c) El principio de la mesurabilidad social.

    d) La perspectiva cualitativa.

8. ¿Cómo se denomina a la parte de la estadística que se encarga de analizar y describir a una parte de la población conocida como muestra?

    a) Estadística inferencial.

    b) Estadística probabilística.

    c) Estadística descriptiva.

    d) Estadística muestral.

9. ¿Cómo se denomina a la parte de la estadística que se ocupa de generalizar al total de la población las características obtenidas de una muestra?

    a) Estadística inferencial.

    b) Estadística probabilística.

    c) Estadística descriptiva.

    d) Estadística muestral.

10. ¿En el muestreo por cuotas, de qué manera son seleccionados los individuos que forman parte de la muestra?

    a) Sistematizadamente.

    b) Investigando las características de una población con cuestionarios aleatorios.

c) Seleccionando a n porciones del universo de forma proporcional.

d) El encuestador selecciona a los sujetos que entrevistará.

11. ¿Por qué se dice que una muestra es representativa de una población?

a) Es una versión que reproduce (fielmente) a escala las características del universo o población de que fue extraída.

b) Cada unidad muestreada representará las características de una cantidad conocida de unidades en la población.

c) Porque posee el tamaño adecuado para representar a los elementos de que se conforma la población.

d) Porque fue elegida de forma aleatoria y no azarosa para evitar las tendencias del investigador.

12. ¿Cómo se debe expresar el proceso de generalización de las características que describen a una muestra al total de la población?

a) En el lenguaje científico proporcionado por las ciencias formales.

b) Matemáticamente.

c) En términos de probabilidad.

d) Lógica y formalmente.

13. ¿Qué se debe hacer para aumentar el nivel de confianza y disminuir el error muestral?

a) Sondear el objeto-fenómeno de estudio antes de elaborar el cuestionario para recopilar los datos.

b) Incrementar los recursos materiales y humanos.

c) Determinar el tamaño de la población.

d) Aumentar el tamaño de la muestra.

14. Señala una característica de las técnicas de muestreo no probabilístico

    a) Existe una relación de equivalencia entre el tamaño n de la muestra y el de la N población

    b) Incrementar la incertidumbre del estudio que se realiza

    c) La muestra no aparece como representativa del Universo

    d) Aumentar el tamaño de la n muestra para incrementar su nivel de confianza

15. ¿Cuáles son las categorías, generales, en que se clasifica el muestreo no probabilístico.

    a) Muestreo Azaroso, Muestreo Subjetivo, Muestreo Intencional

    b) Muestreo Casual, Muestras Intencionales, Muestreo por Cuotas

    c) Muestreo de Inmersión, Sondeo, Muestreo Exploratorio

    d) Muestreo Simple, Muestreo Representativo, M. Estratificado

10.4. Elementos del muestreo

El proceso de muestreo comienza con la localización de las fuentes adecuadas, tales como listados de población, registros, marcas y otros. Las muestras se pueden extraen de estos marcos. Si el marco es inadecuado debido a que ciertos individuos de la población no deberían aparecer en él, o algunos sujetos no aparecen y deberían figurar en dicho registro, o cuando ciertos elementos aparecen más de una vez (en varias ocasiones), esto implica que se podría obtener una muestra defectuosa a partir de un marco con dichas características. Por tal motivo, lo que procede es la corrección de dichas anomalías. A este proceso se le denomina *depuración de marcos imperfectos*.

En la investigación de base cuantitativa, se cuenta con dos tipos de fuentes de datos: Las *fuentes primarias*, que se refieren a la toma de datos directa, mediante la observación encuestas y cuestionarios (recopilación) y las *fuentes secundarias*; que designan a los datos cuyo origen se encuentra en la consulta a estadísticas emitidas por fuentes oficiales o en su defecto por consultoras privadas (compilación), en otras palabras: a datos retomados de investigaciones realizadas por otros agentes. Los datos obtenidos a partir de las fuentes antes mencionadas se clasifican en:

- Categóricos: que indican características cualitativas de los elementos.

- Numéricos: que se dividen en dos clases: los *discretos*, denotados por números enteros ($x \in N$) y, en segundo lugar, los *continuos*, que se aplica para los datos que pueden adoptar todos los valores de un intervalo, esto es: $x \in R$ (las magnitudes como la estatura o el peso), ejemplo:

Tipo de datos	Pregunta	Respuesta
Categórico	¿Usted fuma?	Sí □ No □
Numérico discreto	¿Cuántos hijos tiene?	Número de hijos: 2
Numérico continuo	¿Cuál es su estatura?	1,82mts.

10.4.1. Características de la muestra

Para poder realizar un adecuado proceso de inferencia estadística, es decir, dados los descriptores o estadígrafos de una muestra generalizarlos al total de una población, la muestra debe ser representativa de la población, como ya se ha expresado anteriormente. Para que una muestra sea representativa, debe cumplir con los siguientes puntos:

- La muestra debe ser aleatoria. Para evitar lo más posible el sesgo en las inferencias, ya sea por error de medición o porque el investigador, aunque de manera involuntaria, pueda "contaminar" el estudio al emplear su criterio personal para seleccionar a los elementos de estudio, la conformación de la muestra debe ser al azar.

Esto da imparcialidad al proceso de investigación y permite establecer, por un lado el nivel de confianza del estudio y por el otro, fijar el error estándar o discrepancia máxima a tolerar entre el verdadero parámetro población y el estimador obtenido a través del análisis de la muestra.

- El tamaño n de la muestra debe ser lo suficientemente grande. En tanto mayor sea el tamaño de una muestra, mayor será su semejanza con el total de la población. No obstante, puede calcularse el mínimo tamaño que debe tener una muestra para que sea representativa de la población. La determinación del tamaño n de una muestra depende del atributo estadístico o parámetro poblacional a medir. Los más frecuentes son:
  - Media, μ
  - Varianza, σ^2
  - Proporción, P
  - Diferencia de, $\sigma^2 - \sigma^2$
  - Diferencias de medias, $\mu_1 - \mu_2$
  - Diferencia de proporciones, $P_1 - P_2$
- De tal manera que para estimar el valor de cada uno de los parámetros anteriores, se usa su fórmula correspondiente.

10.4.2. Conceptos de muestreo probabilístico

Para realizar un proceso de investigación, primeramente se deberá definir detallada e íntegramente la población sujeta a estudio. La *población* es la colección completa de elementos sobre los que se desea realizar alguna inferencia. A esta *población se le denomina población objetivo.*

Por lo general, el muestreo de toda la población objetivo no es posible debido, entre otras causas, a las negativas de algunos individuos a participar, ausencia de los mismos en su domicilio, posible incapacidad física o por motivos de salud del sujeto para responder.

Por tal motivo, el investigador se puede ver limitado en el proceso de obtener información de la población objetivo y sus elementos, dando paso al concepto de *población investigada*, que es en realidad, la población sometida a un proceso de muestreo.

A un elemento de una población se le denomina *unidad elemental de muestreo*, sin embargo en ocasiones es conveniente agrupar a las unidades elementales de muestreo en conjuntos. En este caso se habla de una *unidad de muestreo compuesta*, que también se conoce como *unidad de muestreo primaria.*

Por otra parte, al conjunto de las unidades que conformarán la muestra, se les denomina *unidades muestrales o unidades de muestreo.* Para poder seleccionar a este conjunto, se necesita un listado completo, sean elementales o compuestas, de las unidades que constituyen la población sujeta a estudio.

A este listado, se le conoce con el nombre de *marco muestral.* Un ejemplo de marco muestral, puede ser un directorio telefónico, un listado de alumnos inscritos a una universidad o una lista del padrón electoral.

Por lo general, el marco muestral contiene “impurezas” por falta de actualización, errores y omisiones entre otras causas; tal situación ocasiona que el marco muestral no coincida con los elementos de la población objetivo. En este sentido, se dice que se trata de un *marco muestral imperfecto.*

Un marco muestral imperfecto se debe a situaciones como la repetición de los elementos listados, por ejemplo, que una misma persona aparezca listada dos o más veces en el mismo marco. A este tipo de error se le llama *duplicación.*

Asimismo, algunos elementos que deberían aparecer listados no lo son, regularmente por un error al registrarse. En este caso el error se conoce como *omisión.* En contraparte, puede suceder que aparezca listado, en el marco muestral, un elemento que no debería aparecer. A este elemento se le denomina *unidad vacía.*

Se denomina *unidad extraña*, a un elemento que pertenece al marco pero no puede participar en un proceso de muestreo. Por ejemplo, si se lleva a cabo una encuesta a domicilio en la que el marco muestral es la lista de viviendas de una comunidad, se considera unidad extraña a las casa vacías o deshabitadas.

Si se elimina del marco muestral las unidades erróneamente incluidas en él, ya sea unidades extrañas, vacías o duplicaciones y simultáneamente se incorporan las omisiones, se obtiene la población objetivo. Este proceso se denomina *depuración de marcos imperfectos.*

Una vez que se ha conformado un marco muestral y que se ha depurado, se procede, en función del parámetro poblacional que se desea investigar, a *diseñar la muestra*; esto, en primer lugar, hace necesario determinar el tamaño de la misma. Para ello, el investigador empleará la fórmula correspondiente a dicho parámetro. Aunque las fórmulas son diferentes, todas incluyen para su cálculo, dos elementos fundamentales que son establecidos por el investigador, a saber, el nivel de confianza y el error estándar.

El *nivel de confianza*, se refiere a la confiabilidad que el investigador desea que posean los estimadores de los parámetros poblacionales. Dicho de otra manera, significa el porcentaje de veces que una afirmación o inferencia acerca de la población, será verdadera. Por ejemplo, si se estima la media poblacional con un nivel de confianza del 95 %, significa que de cada 100 muestras, 95 de ellas arrojarán un estimador muy cercano a la verdadera media poblacional.

El término *error estándar*, hace referencia a la precisión buscada en la estimación. Se busca que el error estándar sea bajo, es decir, que la precisión sea alta. El error estándar indica el porcentaje máximo de discrepancia que se está dispuesto a aceptar entre el parámetro poblacional y su respectivo estimador. Por ejemplo, sea θ un parámetro poblacional, y sea $\hat{\theta}$ su estimador. La discrepancia o diferencia entre ellos se denomina error muestral, y se expresa por:

$$e = |\theta - \hat{\theta}|$$

Asimismo, sea e_α el error estándar o margen de error a aceptar entre el parámetro poblacional y su estimador. Entonces significa que dado $e - e_\alpha$ debe satisfacer:

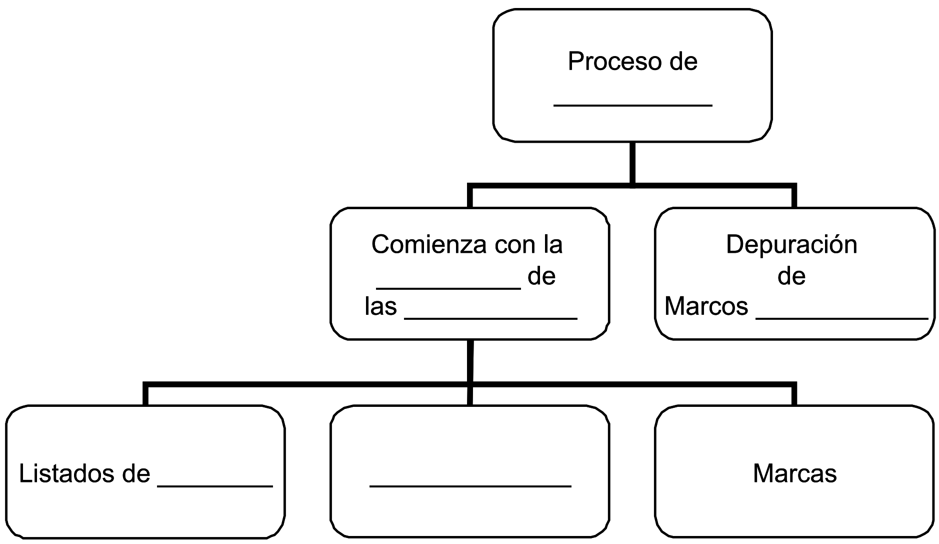
$$e \leq e_\alpha$$

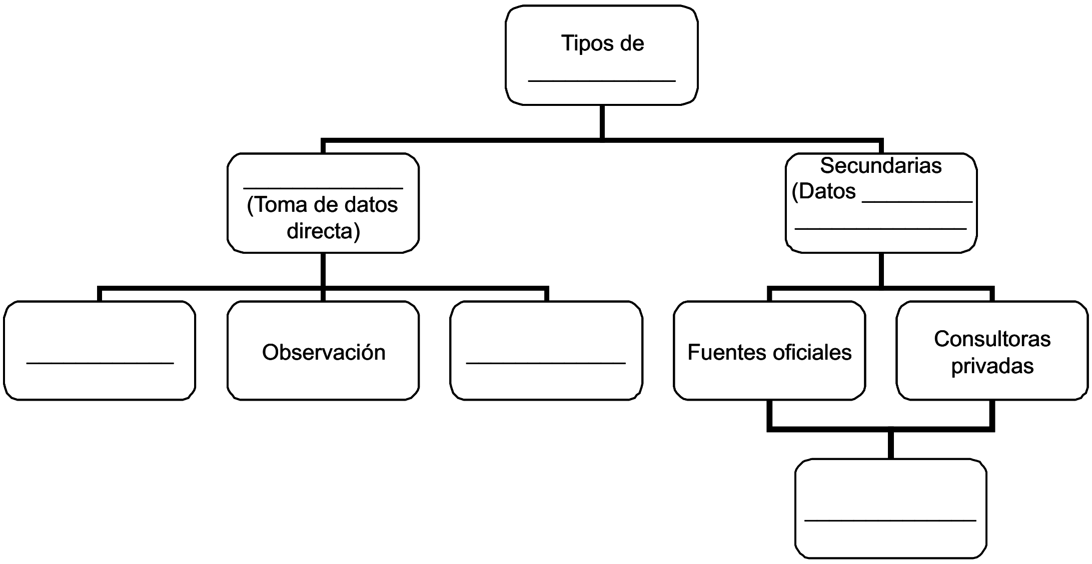
El error estándar, al igual que el nivel de confianza, se expresa en términos de porcentaje. Es razonable pensar que al realizar una investigación por muestreo, se desea un alto nivel de confianza y un bajo error estándar. Sin embargo, como se explicó anteriormente, en la práctica nunca se puede llegar al cien por ciento de confianza ni manejar un error estándar igual a cero; debido a la naturaleza aleatoria del proceso, los resultados conllevan cierto grado de incertidumbre (en los datos involucrados) cuantificado con métodos de la teoría de la probabilidad. De igual manera, si se desea un error estándar bajo y un nivel de confianza alto, implica aumentar el tamaño de la muestra, con los respectivos aumentos en costo y tiempo de realización del muestreo como consecuencia. En la práctica se manejan niveles de confianza del 90; 95; 99 y 99,7 %.

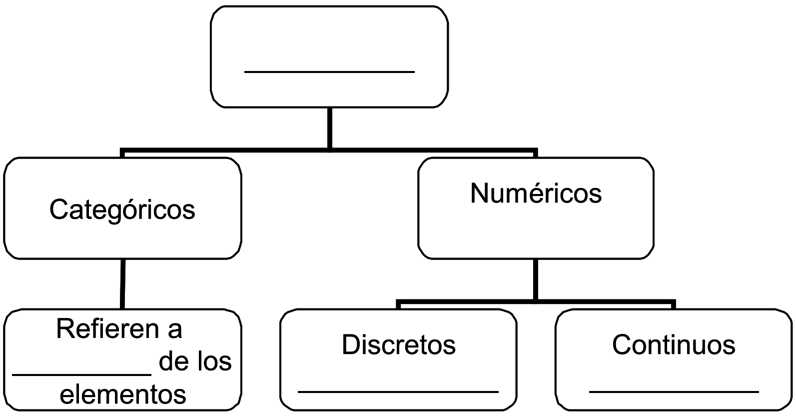
Ejercicios 2

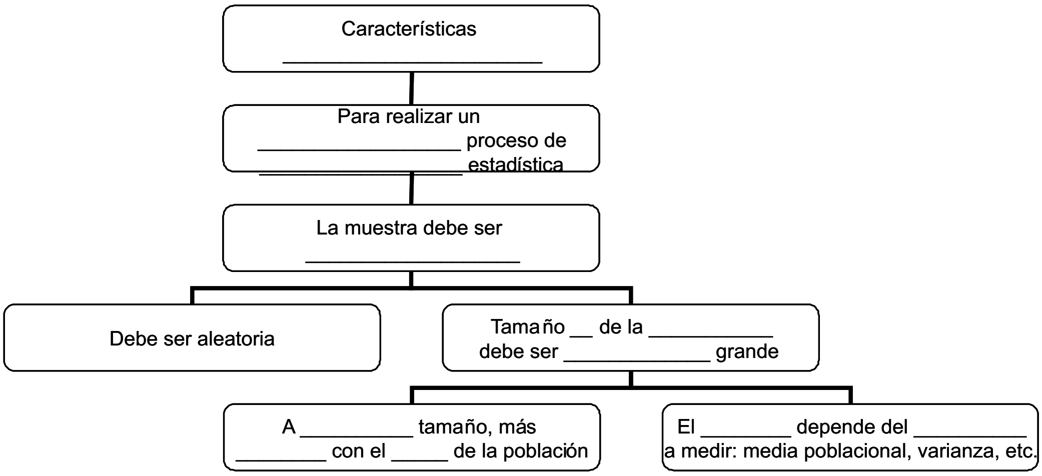
- *Instrucciones*: Antes de resolver los ejercicios de la presente sección, elabora (en tu cuaderno) un resumen que te permita completar las redes conceptuales y conceptos propuestos a continuación.

1. Anota en las líneas de la red conceptual, la palabra que complete correctamente el concepto.

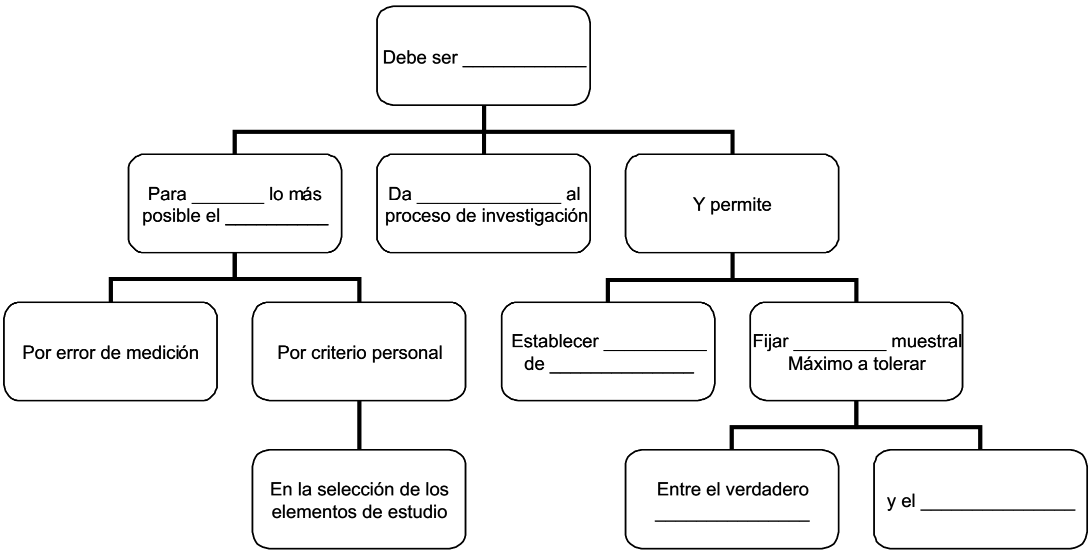



Tipos de
(Toma de datos directa)
Secundarias (Datos
Observación
Fuentes oficiales
Consultoras privadas


Categóricos
Numéricos
Refieren a de los elementos
Discretos
Continuos



Características ____________________
Para realizar un ________________ proceso de ________________ estadística
La muestra debe ser ________________
Debe ser aleatoria
Tamaño __ de la __________ debe ser ____________ grande
A _________ tamaño, más ________ con el _____ de la población
El ________ depende del __________ a medir: media poblacional, varianza, etc.



Debe ser ___________
Para _______ lo más posible el _________
Da _____________ al proceso de investigación
Y permite
Por error de medición
Por criterio personal
Establecer _________ de _____________
Fijar _________ muestral Máximo a tolerar
En la selección de los elementos de estudio
Entre el verdadero ______________
y el ______________

- Completa el concepto que se te presenta, colocando en la línea correspondiente, la palabra que falta.

2. Para realizar un proceso de ____________________, primeramente se deberá definir ____________________ e íntegramente la ____________________sujeta a ____________________. La población es la ____________________completa de elementos sobre los que se desea realizar alguna inferencia. A esta ____________________ *se le denomina población objetivo.*

3. A un elemento de una población se le denomina ____________________ *elemental de* ____________________, sin embargo en ocasiones es conveniente ____________________ a las unidades elementales de muestreo en conjuntos. En este caso se habla de una unidad de muestreo____________________, que también se conoce como *unidad de muestreo primaria.*

4. Para poder seleccionar a las unidades de muestreo, se necesita un __________________ completo, sean elementales o compuestas, de las unidades que constituyen la población sujeta a estudio. A este listado, se le conoce con el nombre de ______________________________.

5. Se denomina ____________________, a un elemento que pertenece al marco pero no puede participar en un proceso de muestreo.

6. Si se ____________________ del marco muestral las unidades erróneamente incluidas en él, ya sea unidades ____________________, vacías o ______________ __________ y simultáneamente se ____________________ las omisiones, se obtiene la población objetivo. Este proceso se denomina ____________________ de marcos ____________________.

7. Una vez que se ha conformado un ____________________ y que se ha depurado, se procede, en función del ____________________ que se desea investigar, a *diseñar la muestra*; esto, en primer lugar, hace necesario determi-

nar el ____________ de la misma. Para ello, el investigador empleará la ________________ correspondiente a dicho parámetro. Aunque las fórmulas son diferentes, todas incluyen para su cálculo, dos elementos fundamentales que son establecidos por el investigador, a saber, el ________________________ y el ________________ estándar.

8. El término ________________, hace referencia a la precisión buscada en la estimación. Se busca que el error estándar sea ________________, es decir, que la precisión sea ________________. El error estándar indica el ________________ máximo de discrepancia que se está dispuesto a aceptar entre el ________________ poblacional y su respectivo estimador.

9. El ________________ estándar, al igual que el nivel de ________________, se expresa en términos de ________________. Es razonable pensar que al realizar una investigación por muestreo, se desea un alto nivel de confianza y un bajo error estándar. Sin embargo, en la práctica ____________ se puede llegar al ________________ de confianza ni manejar un error estándar igual a cero; debido a la naturaleza aleatoria del proceso, los resultados conllevan cierto grado de ________________ (en los datos involucrados) cuantificado con métodos de la teoría de la ______________.

10. De igual manera, si se desea un error estándar bajo y un nivel de confianza alto, implica ________________ el tamaño de la ________________, con los respectivos aumentos en costo y tiempo de realización del muestreo como consecuencia. En la práctica se manejan ________________ de ________________ del 90, 95, 99 y 99,7 %.

- Ahora compara tus **respuestas** con las de la página 556

10.4.3. Metodología del muestreo probabilístico

No existe un estándar único para la realización de una investigación a través de un proceso de muestreo. Pese a ello, el autor del presente trabajo, considera adecuadas las siguientes fases:

- La formulación del planteamiento del problema. "En este apartado se trata de explicar los aspectos, factores o elementos relevantes relacionados con el problema de investigación [. . .] La formulación de un problema debe concretarse a la enunciación que responda de manera clara al qué y para qué de la investigación"[6] Esto permite avanzar en la elección de paradigmas científicos y las respectivas técnicas pertinentes para el estudio del fenómeno, así como en la construcción de objetivos.

- Establecer los objetivos de la investigación. "Los objetivos son el sistema de acciones que indican los *propósitos, las características y los medios* para lograr la resolución del problema planteado en la investigación. La formulación de estos es fundamental en el diseño, ya que *son los referentes que guiarán el desarrollo del estudio.* Además tienen el cometido de establecer las pretensiones de una investigación."[7]

- Delimitar la población objetivo y la población investigada. "Una de las primeras decisiones a tomar en cualquier investigación es la *especificación* y *acotación* de la *población* a analizar. La concreción de ésta, vendrá determinada por cuál sea el *problema* y los *objetivos* principales de la investigación."[8]

[6] op. cit. AGUILAR, Vieyra Griselda; México; 1999; pp. 40
[7] op. cit. AGUILAR, Vieyra Griselda; México; 1999; pp. 40
[8] op. cit. CEA, D' Ancona María Ángeles; Madrid; 1998; pp. 159

Tal delimitación, se debe realizar de manera formal, es decir, mediante una *proposición abierta* que permita reconocer a la población de que se tomarán los elementos que conformarán la muestra. Dicha proposición, deberá contener la(s) característica(s) esencial(es) que permita(n) la observación del fenómeno; dos de ellas (insoslayables), son las pertinentes al espacio y el tiempo en que se estudiará el fenómeno. Luego entonces, será posible localizar o elaborar el marco muestral que contenga el total de los elementos de interés para la investigación.

- Conformación del marco muestral. Una vez trazadas las fronteras de la porción de la realidad a estudiar, se debe contar con el listado de los elementos que conforman el objeto de interés para la investigación. Como ya se ha mencionado, existen dos tipos de fuentes para la conformación de un marco muestral; la elección de la fuente a que se recurrirá para la conformación del marco muestral (sea de forma directa o de un listado elaborado previamente), dependerá del problema a tratar y la acotación que se haya hecho del universo.

- Depuración de marcos imperfectos:

  "*a*) El *marco* ha de ser lo más completo posible, en orden a facilitar la representatividad de la muestra. [...] la *muestra* escogida sólo podrá considerarse 'representativa' de la *población* comprendida en el *marco de muestreo* elegido; es decir, de aquellos que han tenido probabilidad de participar en la *muestra*. Por esta razón, la **comprehensividad** se convierte en una exigencia básica de todo *marco muestral*.

  "*b*) La *comprehensividad del marco muestral* conlleva, necesariamente la exigencia de su actualización. En la medida en que el marco muestral se halle actualizado (en cortos periodos de tiempo, preferentemente) las posibilidades de omisión se restringen. Por el contrario, aumenta la probabilidad de que éste contenga a los miembros reales de la población que 'representa'.

"*c*) Cuando la investigación persigue la *generalización de los datos muestrales* (a la población que conforma el marco muestral), es preciso que cada componente de la población esté igualmente representado en el *marco de muestreo.* Han de evitarse las *duplicidades.* Este es un problema habitual cuando, para una misma muestra, se combinan dos o más listados diferentes. Aquellos que (por cualquier motivo aparezcan más de una vez, tendrán una oportunidad superior de ser elegidos. Lo que favorece su sobrerepresentación en la muestra.

"*d*) El *marco muestral* tampoco ha de incluir *unidades* que no correspondan a la *población* que se analiza. La inclusión de estas *unidades* reduce la probabilidad de elección de las *unidades* que sí pertenecen a la *población.*"[9]

- Identificación del parámetro poblacional a investigar. Los principales parámetros poblacionales a estudiar (por prácticos y comunes), son la media μ y la proporción P; por ejemplo: supóngase que se realiza un estudio de opinión pública, respecto a cierta campaña publicitaria en una colonia, si se deseara determinar a cuántas personas de esa colonia les agrada el mensaje publicitario, entonces estaríamos hablando de una proporción (30 de cada 100 gustan del mensaje), pero si se deseara conocer el número de veces que lo ven y/o escuchan a la semana, se estaría hablando de la media o promedio.

[9]op. cit. CEA, D' Ancona María Ángeles; Madrid; 1998; pp. 161, 162, 163 Negritas agregadas

- Diseño de la muestra
    - o Establecer el nivel de confianza y error estándar deseados; dichos elementos, son establecidos directamente por el investigador, en función de la precisión deseada y los recursos con que se cuenta para realizar la investigación.
    - o Determinación del tamaño de la muestra en función del parámetro poblacional a investigar, así como del nivel de confianza y error establecidos. Hay que tener en cuenta mientras mayor sea la precisión deseada, el tamaño de la muestra aumentará así como sus costos.
    - o Definición de un mecanismo aleatorio para la selección de los elementos que integrarán la muestra, es decir, definir el tipo de muestreo a emplear.
    - Diseñar el instrumento de medida o cuestionario
    - Procesamiento de los datos.
    - Análisis de datos.
    - Presentación de resultados.

En el caso de encuestas de gran dimensión, se sugiere la aplicación de una *encuesta* ***piloto*** para probar el instrumento de medición o cuestionario, así como para la obtención de estadígrafos que permitirán la determinación del tamaño de la muestra. En la sección 10.5. Técnicas de Muestreo Aleatorio, se revisarán los diversos tipos de muestreo empleados para seleccionar los elementos que serán sometidos a estudio, así como las fórmulas para determinar el tamaño de la muestra para que ésta sea representativa; mientras que en la sección 10.6. Error Estándar, se explicará dicho concepto y se revisarán las expresiones para su cálculo.

10.4.4. Consideraciones en el diseño del instrumento de medición en el muestreo probabilístico

Efectivamente, el éxito de una investigación a través de muestreo depende en gran medida del tamaño de la muestra y del método de selección de la misma. Sin embargo, un paso de gran trascendencia para tal proceso lo constituye el diseño del instrumento de medición, el cual, como su nombre lo indica, es el mecanismo que permite extraer y cuantificar información sobre cada uno de los elementos de la muestra.

Si el instrumento está bien diseñado, los resultados desprendidos de él tendrán gran utilidad en el proceso de inferencia. Por el contrario, un mal diseño implica la obtención de información errónea. Uno de los instrumentos de medición de mayor uso en la práctica, lo constituyen los cuestionarios. Algunas normas sugeridas para la correcta elaboración de un cuestionario, se presentan a continuación:

- Evitar la *autoselección*. En determinados estudios, algunos investigadores o agencias envían cuestionarios a diversos domicilios, solicitando que los habitantes lo contesten y devuelvan por correo. Este es un proceso equivocado, ya que se cae en la autoselección, es decir, los involucrados deciden si forman o no parte de una muestra, con lo que se introduce en el estudio sesgo en un alto grado. Otro ejemplo lo constituyen el tipo de sondeo realizado por algunos noticiarios del radio o televisivos, en los que se invita al auditorio a que exprese su opinión, vía telefónica, sobre algún tema de interés. Al igual que el caso anterior, se incurre en el sesgo de la información, ya que el auditorio decide si forma parte de la muestra o no y adicionalmente, una persona puede participar llamando más de una vez, con lo que el sesgo en el estudio es aún mayor.

- Evitar la sugerencia de algún tipo de respuesta. Un cuestionario nunca debe reflejar la opinión del investigador o presentar preguntas que en determinada forma, "obliguen" al entrevistado a dar cierta respuesta, por ejemplo, preguntas con cierta obviedad en la respuesta como: ¿Usted está de acuerdo en que se reduzca la contaminación en la ciudad? ¿Le gustaría que se redujeran los índices delictivos en su ciudad? o preguntas de este tipo.

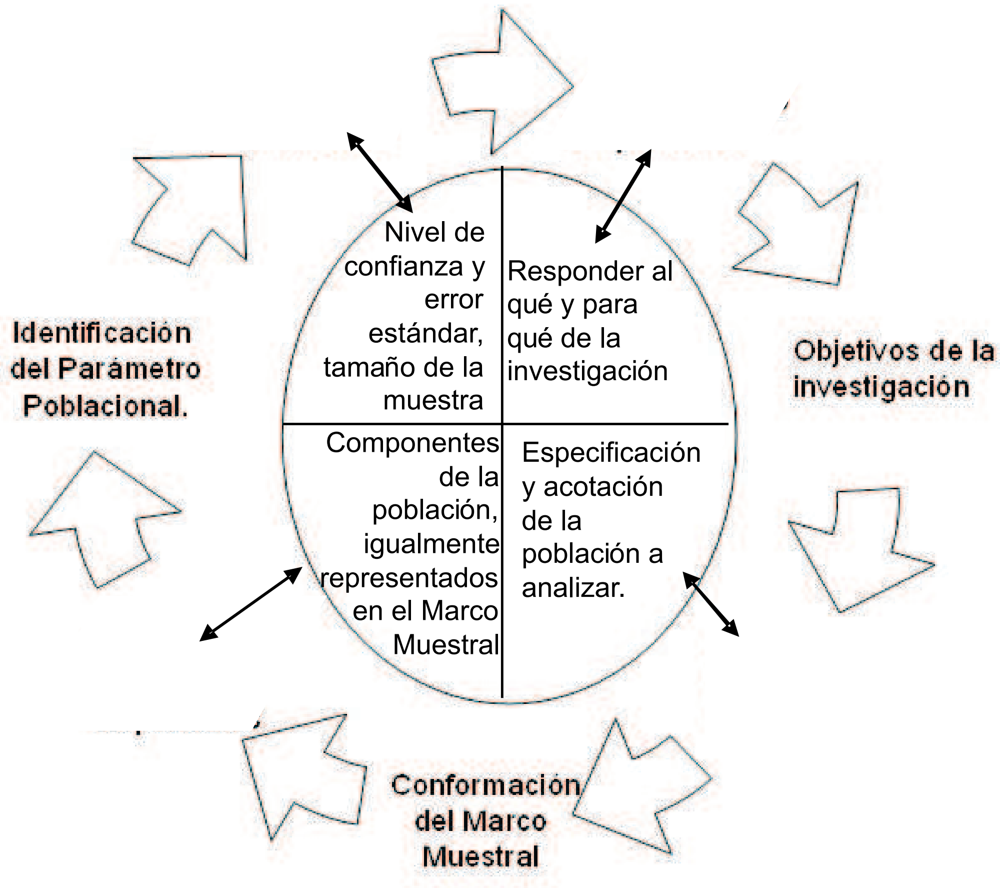
- Claridad en las preguntas. Un cuestionario debe siempre permitir al entrevistado identificar de manera clara, sencilla y concreta el aspecto sobre el que debe responder. Para tal efecto, se debe pensar en las característica socio-culturales de la población objetivo (el público a que se dirigen las preguntas). Evitar los tecnicismos cuando las unidades muestrales no se especializan en algún el tema (estudios de opinión), así como el uso de términos ambiguos.

- Tamaño razonable del cuestionario. Si se desea reducir el grado de error en las respuestas o de falsedad en las mismas por parte del entrevistado, se deben diseñar cuestionarios que no se excedan en el número de preguntas, ya que un cuestionario grande, puede ocasionar errores o falsedad en la respuesta del entrevistado e incluso, indisposición a responder.

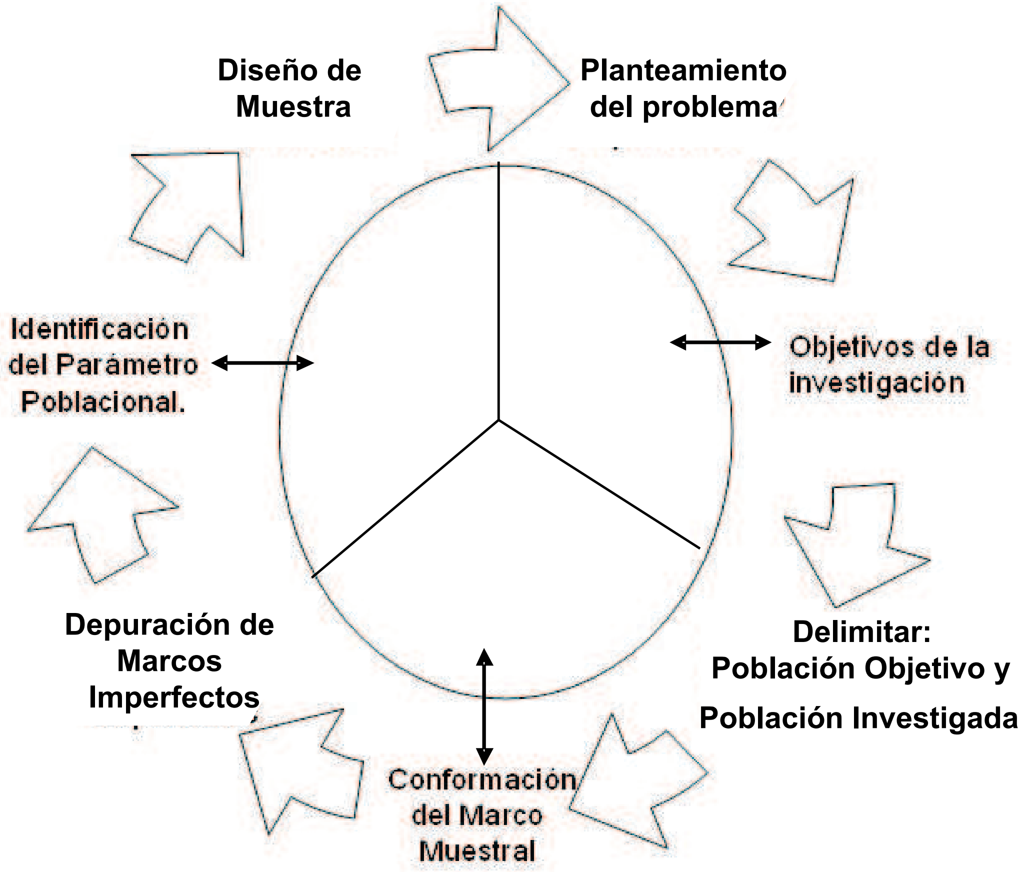
- Preguntas abiertas o cerradas. El investigador debe diseñar cuidadosamente el cuestionario y decidir sobre la inclusión o no de preguntas abiertas, ya que éstas, en determinadas circunstancias pueden ocasionar que el entrevistado divague o se desvíe del tema sobre el que se le cuestiona.

Ejercicios 3

1. Anota los componentes que faltan en el siguiente diagrama del Muestreo Probabilístico. Apóyate en las características que al interior del círculo se presentan y señalan la con el signo $\leftrightarrow$.



2. Ahora, coloca al interior del círculo: a)Qué indican los objetivos, b) Qué es un marco muestral y c) Principales parámetros poblacionales.



3. Investiga y anota, por lo menos, tres mecanismos para generar números aleatorios que pueden auxiliar en la selección de los elementos que integrarán una muestra.

____

____

____

4. ¿Qué es una encuesta piloto?

5. ¿Cuál es el mecanismo que permite extraer y cuantificar información sobre cada uno de los elementos de la muestra?

6. ¿En qué caso(s), se puede afirmar que el investigador incurrió en el uso de un mecanismo de autoselección para extraer información de la muestra?

7. Proporciona dos ejemplos de mecanismos de autoselección.

8. Observa la siguiente pregunta y señala el error en que incurre: ¿Cree usted que el acervo de esta biblioteca se debe actualizar anualmente?

9. ¿Qué se debe considerar para que las preguntas de un cuestionario sean claras?

10. Enuncia una posible desventaja de las preguntas abiertas incluidas en los cuestionarios?

- Compara tus respuestas con las de la página 560

10.5. Técnicas de muestreo aleatorio

Métodos de muestreo probabilístico

Como ya se explicó anteriormente, para evitar el sesgo y las actitudes tendenciosas, un proceso de muestreo debe ser aleatorio o probabilístico. Esto significa que cada elemento de la muestra tiene una probabilidad conocida de ser seleccionado. Existen diversos métodos de muestreo probabilístico para seleccionar los elementos que conformarán una muestra. Entre los principales pueden mencionarse los siguientes:

- Muestreo aleatorio simple
- Muestreo sistemático
- Muestreo estratificado
- Muestreo por conglomerados
    - Muestreo multietápico

10.5.1. Muestreo aleatorio simple

El *muestreo aleatorio simple* puede ser con reposición o sin reposición. A éste último se le conoce como muestreo irrestricto aleatorio, aunque en la práctica, al mencionar el término muestreo aleatorio simple, se da por entendido que se refiere al muestreo sin reposición, que es el que se considerará en el presente trabajo de investigación. El muestro aleatorio simple, *es el mecanismo en el que todos los elementos de una población tienen la misma probabilidad de ser seleccionados para la muestra*, sin importar en orden de colocación de los elementos. Defínase a N, como el tamaño total de la población y a n, como el tamaño de la muestra. En consecuencia, el espacio muestral asociado a tal método tiene:

$$\boxed{C_n^N = \begin{pmatrix} N \\ n \end{pmatrix} = \tfrac{N!}{n!(N-n)!}}$$

posibles muestras de tamaño n, es decir, se calculan las combinaciones de n elementos tomados de un total de N. Nótese que en la ecuación anterior, se usa otra notación para las *combinaciones* (técnica de conteo revisada en el capítulo *Antecedentes Matemáticos*). Es decir, C_k^n se puede escribir de la forma:

$$\begin{pmatrix} N \\ n \end{pmatrix}$$

Asimismo, la probabilidad de ocurrencia de cualquier muestra de tamaño n, viene dada por:

$$\boxed{p(u_1 \ldots u_n) = \tfrac{c_f}{c_t} = \tfrac{1}{C_k^n} = \frac{1}{\begin{pmatrix} N \\ n \end{pmatrix}}}$$

Donde, $u_1 \ldots u_n$, corresponde a las unidades muestrales, c_f al número de casos favorables y c_t al número de casos totales.

La teoría del muestreo aleatorio simple se formaliza al determinar la probabilidad π_i de seleccionar la i-ésima unidad muestral u_i. Ya se sabe cómo calcular el número de muestras posibles de tamaño n. Entonces el número de muestras posibles que se pueden formar con los elementos de la población y que contengan al elemento u_i, será:

$$\boxed{C_{n-1}^{N-1} = \begin{pmatrix} N-1 \\ n-1 \end{pmatrix}}$$

Por lo que para calcular la probabilidad i_π de que el elemento u_i pertenezca a la muestra $(\bar{x})$, se divide el número de casos favorables entre el número de casos totales, es decir, el cociente del número de muestras que contiene la i-ésima unidad u_i, entre el número total de muestras, o sea:

$$\boxed{\pi_i = p(u_i \in (\bar{x})) = \frac{\begin{pmatrix} N-1 \\ n-1 \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} = \frac{\frac{(N-1)!}{(n-1)(N-n)!}}{\frac{N!}{n!(N-n)!}} = \frac{\frac{(n-1)!}{(n-1)!(N-n)!}}{\frac{N(N-1)!}{n(n-1)!(N-n)!}} = \frac{1}{\frac{n}{N}} = \frac{n}{N}}$$

Por otra parte, en la mayor parte de las investigaciones por muestreo, los parámetros poblaciones a investigar son la media y la proporción, estableciendo previamente un nivel de confianza y un error estándar deseados. Por tal motivo, el presente trabajo se enfocará a tales métodos.

10.5.2. Determinación del tamaño de la muestra para la estimación de la media poblacional

Dado un error estándar o margen de error e_α, la varianza poblacional σ^2 y un nivel de confianza $1-\alpha$, el tamaño mínimo n, de la muestra para que ésta sea representativa al estimar la media poblacional viene dado por la siguiente expresión:

$$\boxed{n = \frac{\sigma^2 Z^2}{{e_\alpha}^2}}$$

Donde:
n = Tamaño de la muestra.
Z = Valor que se obtiene de las tablas de distribución normal, tomando las coordenadas que corresponden al valor del nivel de confianza dividido entre dos, es decir: $\frac{1-\alpha}{2}$.
e_α = Error estándar o precisión.
σ^2 = Varianza de la población. En la mayor parte de los casos, este valor se desconoce, por lo que se extrae una muestra aleatoria piloto a la que se calcula su desviación estándar s, el cual se emplea en lugar de σ.

Ejemplo: Se desea conocer la cantidad promedio de cigarrillos que consume una persona fumadora al día en una cierta ciudad. Se observa entonces que el atributo estadístico a estimar es la media μ de la población. Para ello se realizará un estudio a través de muestreo aleatorio, por lo que es importante determinar el tamaño de la muestra (número de fumadores que se entrevistarán). Antes de ello, se realiza al azar un muestreo piloto a 60 fumadores, y al analizar esta muestra piloto, se obtienen los siguientes estadísticos:

- Media: $\bar{x} = 5$ (cigarrillos en promedio al día)

- Desviación estándar: $s = 2{,}1$ (cigarrillos)

Es decir, una persona fumadora de la muestra piloto, consume 5 $\pm$ 2,1 cigarrillos al día, es decir, el número x de cigarrillos se encuentra en el intervalo: $2{,}9 < x < 7{,}1$, que en términos más prácticos significa que una persona de la muestra piloto, fuma entre tres y siete cigarrillos al día.

Debe recordarse que esta aseveración sólo es respecto a la muestra piloto y no puede generalizarse al total de la población. La muestra piloto se emplea únicamente para calcular estadígrafos (media y desviación estándar) que serán empleados en la respectiva fórmula que permitirá determinar el tamaño n de la muestra aleatoria (número de personas fumadoras a entrevistar) sobre el cual se estimará el valor de la media muestral (promedio de cigarrillos consumidos al día por las personas fumadoras). Dicha media, sí se puede generalizar al total de la población.

De acuerdo a la fórmula, el tamaño n de la muestra que permitirá estimar la media poblacional (número de cigarrillos en promedio que consumen al día los fumadores), *estará en función de dos valores establecidos por el investigador*: el nivel de confianza y el error estándar. Como ya se ha comentado, el primero indica la probabilidad que se desea de que el valor de la media poblacional estimada sea correcto. El segundo indica en términos porcentuales, el grado máximo que se está dispuesto a aceptar de discrepancia o diferencia entre el verdadero valor de la media poblacional y el valor estimado en la muestra.

Si se desea en nuestro ejemplo que la muestra tenga un nivel de confianza del 95 % y un error o precisión del 40 %, se debe determinar el tamaño n de la muestra que satisfaga las condiciones planteadas.

Como no se conoce el valor de la desviación estándar poblacional σ, en lugar de emplear la expresión:

$$n = \frac{\sigma^2 Z^2}{e_\alpha^2}$$

se utilizará la fórmula:

$$n = \frac{s^2 Z^2}{e_\alpha^2}$$

De acuerdo a las características del ejemplo, los valores:

- $1-\alpha$ (nivel de confianza) = 95 % (el cual puede expresarse como 0,95). Esto implica que z es igual a las coordenadas en la tabla normal correspondiente al valor del área $\frac{1-\alpha}{2}$, es decir a las coordenadas correspondientes al valor $\frac{0{,}95}{2} = 0{,}475$. Las coordenadas de dicho valor son: 1,96, es decir, $z = 1{,}96$.
- $s = 2{,}1$ (Desviación estándar de la muestra piloto)
- e_α (Error estándar)=0.4 (es decir 40 %)

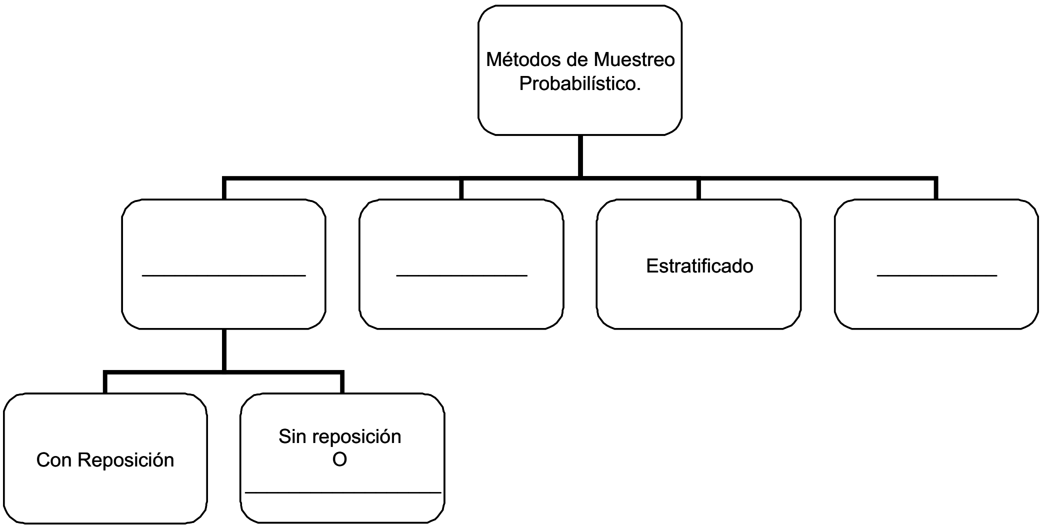
Se sustituyen en la respectiva fórmula:

$n = \frac{(2{,}1)^2(1{,}96)^2}{(0{,}4)^2}$
$n = \frac{4{,}41(3{,}8416)}{0{,}16}$
$n = 105{,}88 \approx 106$
$n = 106$

Esto significa que con una muestra de tamaño 106, se tiene una confianza del 95 % respecto al valor estimado de la media (número promedio de cigarrillos diarios que consume un fumador) y el máximo error a cometer es del 40 % respecto al valor verdadero.

Ejercicios 4

1. Anota los métodos de muestreo probabilístico que se mencionan en el texto



2. ¿Qué permite calcular la siguiente fórmula? $C_n^N = \begin{pmatrix} N \\ n \end{pmatrix} = \frac{N!}{n!(N-n)!}$

3. ¿Qué indican el nivel de confianza y el error estándar?

4. Coloca el significado de las literales de la fórmula para que el tamaño mínimo n de la muestra sea representativo de la media poblacional.

$$n = \frac{\sigma^2 Z^2}{{e_\alpha}^2}$$

5. Retoma el ejemplo de la cantidad promedio de cigarrillos (página 527) con los datos necesarios y calcula n para un error del 20 % y un nivel de confianza del 99 %. Compara el resultado de este ejercicio con el valor del anterior y coméntalo.

Comentarios:

- Compara tus respuestas con las de la página 563

10.5.3. Determinación del tamaño de la muestra para la estimación de la proporción poblacional

También se le denomina *muestreo de estimación por atributos*. Mide qué *proporción* del universo posee (o no) un atributo o característica. Se llama $p\,\%$ a la proporción de la población que cumple con algún atributo de interés y, por complemento, el resto de la población que no cumple dicho atributo corresponde al $(100 - p)\,\%$. En otras palabras, se desea estimar el valor $p\,\%$ de la población que posee un atributo, a través de una muestra aleatoria de n unidades.

Estimación del tamaño de la muestra para universos infinitos

Al realizar estudios sobre poblaciones infinitas o de gran tamaño, la fórmula para calcular el tamaño de la muestra con que se determina la proporción $p\,\%$ de elementos que cumplen o poseen algún atributo de interés, viene dada por la siguiente expresión:

$$\boxed{n = z^2 \tfrac{p(1-p)}{e_\alpha^2}}$$

Donde:

El estimador de p se obtiene a través de una muestra piloto aleatoria de tamaño n_p, donde dicho valor debe ser mayor a 30. Entonces, el estimador de p viene dado por $p = \frac{x}{n_p} 100\,\%$. Aquí x es el número de elementos de la muestra piloto que cumplen con el atributo de interés.

Z = Valor que se obtiene de las tablas de distribución normal, tomando las coordenadas que corresponden al valor del nivel de confianza dividido entre dos, es decir: $\frac{1-\alpha}{2}$.

e_α = Error estándar o precisión.

Estimación del tamaño de la muestra para universos finitos

Cuando se trabaja con universos finitos, la fórmula anterior se emplea como una aproximación, la cual será denominada ahora n_0, es decir:

$$\boxed{n_0 = z^2 \frac{p(1-p)}{e_\alpha^2}}$$

El cual se emplea en la siguiente fórmula para calcular el tamaño de la muestra que permitirá la estimación de la proporción p de elementos de un universo finito, que cumplen algún atributo de interés:

$$\boxed{n = \frac{n_0}{1+\frac{n_0-1}{N}}}$$

Donde:

n_0 se obtuvo de la muestra piloto, como se explicó líneas atrás y N, es el tamaño total del universo de estudio.

- Ejemplo:

Supóngase que una agencia de publicidad está interesada en determinar la preferencia de los habitantes de un municipio, que cuenta con 750000 habitantes, en cuanto a su consumo de medios impresos. Se desea, en particular, investigar qué porcentaje de estas personas, lee una determinada revista de publicación periódica. Aquí se puede observar que el atributo de interés, es sí una persona lee o no lee la revista sujeta a estudio (es decir, los elementos del universo -lectores-, leen la revista o no).

Para ello, se debe realizar un estudio a través de una muestra aleatoria que permita determinar la proporción p de personas que leen la revista. Como se observa, se trata de un universo de 750000 personas. Supóngase además, que no se dispone de tiempo para realizar una encuesta piloto que permitiría realizar una estimación inicial de $p\,\%$. Es válido en estos casos, dar a p el valor inicial de $50\,\%$ y, en consecuencia, $(1-p)=50\,\%$. Si se desea un nivel de confianza $(1-\alpha)$ del $95{,}45\,\%$ (que al dividir entre dos y tomar de las tablas de distribución normal las coordenadas correspondientes a este valor, se obtiene $z=2$) y una precisión o error estándar (e_α del $4\,\%$; estos valores se sustituyen primeramente en la fórmula para universos infinitos:

$$n = 2^2 \tfrac{50(100-50)}{4} = 625$$

Es decir, se debe entrevistar a 625 personas y calcular el porcentaje de éstas para obtener la estimación total de la proporción de las 750000 personas que leen la revista en cuestión.

Por otro lado, si se emplea la fórmula para universos finitos, se tiene entonces que ahora $n_0 = 625$. Sustituyendo este valor en la fórmula respectiva:

$$n = \frac{625}{1+\frac{625-1}{750000}} = \frac{625}{1{,}000832} = 624{,}48 \approx 624$$

Esto significa que se entrevistará aleatoriamente a 624 personas y se les preguntará sí leen o no la revista sujeta a estudio. Supóngase que 206 no leen la revista, entonces se tiene que $p = \frac{206}{624}(100) = 33\,\%$, es decir, el 67 % del universo sí lee la revista. Calculando el 67 % de 750000 se estima entonces que 502500 personas sí leen la revista, objeto de estudio. Y esto puede afirmarse con una seguridad del 94,45 % (nivel de confianza) y un error máximo de 4 % (error estándar).

Selección de las unidades muestrales

Una vez determinado el tamaño de la muestra, se procede la selección de los elementos que la conformarán. Para ello, debe disponerse de un listado de todos los elementos de la población investigada, o sea, un marco muestral, y a través de tablas de números aleatorios (en este capítulo encontrarás una en la página 578, programas de cómputo especializados o tómbolas, tomar los números que corresponden a las unidades muestrales seleccionadas.

10.5.4. Muestreo sistemático

Es muy similar al muestreo aleatorio simple respecto a las fórmulas para el cálculo del tamaño de la muestra, así como de la necesidad de disponer de un listado completo de los elementos poblacionales. La diferencia con el muestreo aleatorio simple, radica en los mecanismos para la selección aleatoria de las unidades muestrales, citados en el párrafo anterior. En el muestro sistemático, la estrategia de selección reduce el tiempo de ejecución de la misma.

El procedimiento consiste en, una vez que se ha determinado el tamaño n de la muestra, realizar el cociente: $M = \frac{N}{n}$ donde N es el tamaño de la población. Posteriormente, seleccionar aleatoriamente un valor k, tal que $k \in [1, m]$. Entonces se deben tomar del listado, las unidades muestrales correspondientes a las proposiciones:

$$\boxed{k, k+m, k+2m, k+3m, \ldots \ \ldots k+(n-1)m}$$

obteniéndose así, los n elementos de la muestra.

- Ejemplo:

Supóngase que el método elegido para conformar una muestra es el muestreo sistemático y que hasta el momento sus resultados fueron los siguientes:

$N = 2000$
$n = 200$

el siguiente paso es aplicar la estrategia de selección de las unidades muestrales que completarán el tamaño n de la muestra, por lo que se deben efectuar los pasos pertinentes a dicho tipo de muestreo, esto es:

1) Definir M; si $M = \frac{N}{n}$, entonces $M = \frac{2000}{200} = 10$

2) Seleccionar aleatoriamente un valor k, tal que $k \in [1, m]$; para éste caso, mediante el uso de la tecla $RAN\#$ de la calculadora, se obtuvo que $k = 5$

3) Subsecuentemente, se tienen qué tomar, del Marco Muestral, las unidades correspondientes a:

$$k, k+m, k+2m, k+3m, \ldots \ \ldots k+(n-1)m$$

que sustituida con los valores calculados, serían las siguientes:

$$5, 5 + 10, 5 + 2(10), 5 + 3(10), 5 + 4(10), 5 + 5(10) \ldots \quad \ldots 5 + (200 - 1)10$$

Así, las unidades muestrales serían las correspondientes a los números 5, 15, 25, 35, 45, 55, ...hasta completar los 200 elementos que conformarían la muestra, donde la última unidad correspondería al individuo número 1995 del marco muestral.

Algunos autores, como Padua, no consideran, desde el punto de vista de la estadística, que esta técnica sea estrictamente aleatoria. De hecho, dependiendo del objeto de estudio, este tipo de muestreo puede introducir cierto sesgo en la investigación. Por ejemplo, al investigar la opinión de las personas sobre un cierto producto comercial, estas pueden seleccionarse de manera sistemática de un directorio telefónico. Dado que éste se encuentra ordenado alfabéticamente y que hay apellidos que aparecen con mucha mayor frecuencia que otros, en la muestra tenderá a incluir a más sujetos con un apellido de común aparición y disminuirá la selección de individuos con un apellido menos común.

Como la preferencia por un producto comercial no depende del apellido de la persona, el muestreo sistemático sería confiable. Por el contrario, si se desea estimar alguna característica, por ejemplo, la proporción de individuos de una población que tienen un cierto apellido, el muestreo sistemático sería inadmisible debido al sesgo comentado en el ejemplo anterior. Así pues, el uso del muestreo sistemático depende del tipo de atributo a investigar en una población, generalmente se usa en estudios en donde el orden de aparición en el listado de los elementos poblacionales no afecta al atributo o parámetro a investigar.

Ejercicios 5

1. ¿Qué se mide mediante el muestreo de estimación por atributos?

   ----

2. ¿Qué parte del universo se denota con la expresión $(100 - p)\,\%$?

   ----

3. Señala en qué casos se utilizan las siguientes fórmulas:

   - $n = \frac{n_0}{1+\frac{n_0-1}{N}}$ ---
   - $n = z^2 \frac{p(1-p)}{e_\alpha^2}$ ---

4. Retoma el caso de la agencia de publicidad que desea conocer el porcentaje de personas que leen cierta revista de publicación periódica (página 534) y calcula el tamaño n de la muestra tanto para universos finitos como infinitos. En este caso, se desea un nivel de confianza igual a $90\,\%$ y un error estándar igual a $30\,\%$. Recuerda que no se dispone de tiempo para realizar una encuesta piloto. Finalmente, supóngase que 1325 personas de la muestra sí leen la revista: expresa en términos de porcentaje y unitarios el total de personas, pertenecientes a la población, que leen la revista.

5. Aplica la estrategia de selección de unidades muestrales, propia del muestreo sistemático, al tamaño n de la muestra para la anterior pregunta y enlista los primeros 10 números correspondientes a las respectivas unidades muestrales, así como el de la última. Para este caso, el número aleatorio seleccionado será el 2

- Corrobora que tus respuestas sean las adecuadas; para ello, cotejalas con las de la página 565

10.5.5. Muestreo estratificado

Este tipo de muestreo es apropiado cuando se desea hacer una investigación en la que la población debe ser dividida y organizada en categorías denominadas *estratos*. Esta división obedece, principalmente, a diferencias bien definidas entre los elementos de una población, respecto a sus características, tales como su nivel socioeconómico, cultural, ideología, sexo, edad y nivel educativo entre otras, ejemplos:

- Si una empresa cigarrera desea realizar un estudio de mercado sobre sus productos, debe realizar un proceso de estratificación, en el cual defina un grupo que sólo incluya personas fumadoras que tengan edad legal para adquirir cigarros. Este tipo de estudios, pueden arrojar información sobre los distintos productos de la empresa, dado que los integrantes de la N población, pueden clasificarse de acuerdo a la marca de su preferencia. Como se podrá ver, este tipo de muestreo, suele emplearse al investigar comunidades u organizaciones en las que sus integrantes puedan clasificarse de acuerdo a características comunes.

- Al estudiar un poblado, los habitantes pueden clasificarse en estratos de acuerdo a su nivel económico, educativo, ocupacional, étnico, etc.

- Otro ejemplo lo constituye un estudio a realizar entre los estudiantes de la Universidad; ésta, debido a la naturaleza misma de su organización, se puede estudiar por carreras, siendo estas categorías los respectivos estratos, ya que, la ideología, perspectivas e inclinaciones de los alumnos de cada carrera es diferente entre cada una de ellas.

Generalmente, los estratos en los que se divide una población, son de tamaño diferente. El siguiente paso, consiste en determinar a cuántos elementos de cada estrato se investigarán, es decir, seleccionar una muestra de cada estrato. Cuando la población ha sido organizada en estratos, es posible definir dos tipos de muestra, a saber:

1. Muestra estratificada proporcional. Es aquella en la que la fracción de muestreo es la misma para cada estrato. O sea se toma como tamaño de la muestra un mismo porcentaje para cada estrato.

2. Muestra estratificada no proporcional. Es aquella en la que la fracción de muestreo no es la misma para cada estrato, es decir, se establece un

número fijo de elementos de cada estrato, sin importar a qué porcentaje del estrato corresponde.

En la práctica se emplea con mayor frecuencia el muestreo estratificado proporcional, siendo empleado el otro en casos especiales. Cuando se emplea el muestro estratificado proporcional, se debe calcular, en función del parámetro poblacional a investigar, del error estándar deseado y del nivel de confianza deseado, el tamaño n de la muestra. Una vez obtenido, se determina la fracción de muestro f, con la fórmula que ya se ha revisado anteriormente:

$$\boxed{f = \frac{n}{N}}$$

Finalmente, una vez que se ha definido la proporción o fracción de muestreo, se emplea ésta, a su vez, para determinar el tamaño n de la muestra de cada estrato. Cuando esto se ha hecho, se emplea cualquiera de los mecanismos de selección de unidades muestrales del muestreo aleatorio simple. Entre las principales ventajas del muestreo estratificado se encuentra que, por lo homogéneo de la muestra, aumenta su representatividad, reduce los errores de estimación debidos a las diferencias entre estratos y permite un mejor estudio y análisis de grupos pequeños. Sin embargo, este tipo de estudios suele ser costoso e implica un amplio conocimiento de las características particulares de la población.

10.5.6. Muestreo por conglomerados

En la mayor parte de los casos, las investigaciones de carácter social a nivel gubernamental presentan un universo de estudio a gran escala, comprendiendo a la totalidad de un país, estado o conjunto de estados, los cuales pueden ser organizados en *conglomerados*. En primera instancia, es notoria una cierta similitud con el muestro estratificado, empero existen diferencias de fondo entre ambos métodos.

"El *muestreo por conglomerados* también representa un procedimiento de selección aleatoria de un conjunto de individuos (referidos ahora como *conglomerados*).
"Con el *muestreo aleatorio estratificado* comparte la característica básica de seccionar la población total en grupos, como fase previa a la extracción muestral. Si bien difiere de él en varios aspectos importantes:

"*a)* En el *muestreo estratificado* se busca la homogeneidad dentro del estrato y la heterogeneidad entre los estratos. En el *muestreo por conglomerados* es a la inversa: el *error muestral* disminuye conforme aumenta la heterogeneidad dentro del grupo (*conglomerado*). Ello se debe a la necesidad de que cada *conglomerado* constituya una representación, lo más ajustada posible de la variedad de componentes del universo.

"*b)* En el muestreo *estratificado* se selecciona aleatoriamente una *muestra* para cada *estrato*. En el *muestreo por conglomerados* lo que se extrae es una muestra aleatoria de *conglomerados*. Sus integrantes formarán la *muestra*.

"*c)* En el *muestreo estratificado*, la unidad de muestreo es el individuo. En cambio, en el muestreo por *conglomerados* es el *conglomerado* (o conjunto de individuos).

"Los *conglomerados* pueden ser las áreas geográficas que dividen a la población que se analiza (país, comunidad autónoma, municipio, distrito, áreas censales, viviendas); pero, también, organizaciones u instituciones (colegios, hospitales tribunales, centros penitenciarios)."[10]

Como se podrá notar, en general, se opta por el muestreo por conglome-

[10]op. cit. CEA, D' Ancona María Ángeles; Madrid; 1998; pp. 192-193

rados cuando la población sujeta a estudio se encuentra dispersa a lo largo de áreas geográficas de gran extensión, lo que implica un elevado costo en recursos materiales y tiempo para su investigación. A continuación se muestra el procedimiento general para el muestreo por conglomerados:

1. Se debe dividir a la población en conglomerados que sean lo más homogéneos posible. Esta conformación de conglomerados puede ejecutarse en varias etapas.

2. Cuando se han determinado los conglomerados, se selecciona aleatoriamente del primer nivel de conglomerados a una determinada proporción de ellos. Si se seleccionan a todas las unidades contenidas en estos conglomerados, se habla de una muestra de una etapa.

3. Si dentro de los conglomerados seleccionados en el paso anterior, se procede a una nueva selección de conglomerados, dentro de cada uno de los primeros (grupos extraídos de los conglomerados ya seleccionados). Al igual que en el paso anterior, si se toman todos los elementos de los estratos obtenidos en este segundo nivel, se habla de una muestra de 2 etapas.

4. Si se sigue procediendo por el mismo método al tomar a todos los elementos de los conglomerados seleccionados en un tercer o cuarto nivel, se habla de muestras *multi-etápicas.*

Debe quedar bien claro que cuando se han determinado los diversos conglomerados, se realiza la selección aleatoria entre conglomerados y, una vez hecho lo anterior, se toman de ellos todas las unidades muestrales al interior de cada conglomerado seleccionado en este nivel. Esta es la principal diferencia con el muestreo estratificado, en cual, una vez definidos los estratos, se seleccionaba proporcionalmente o no, unidades muestrales de cada uno. Por el contrario, el muestreo por conglomerados realiza un primer nivel de conglomerados, seleccionar aleatoriamente algunos de ellos y en un segundo nivel, definir nuevos estratos dentro de los primeros; este proceso, puede continuar así.

En el ejemplo en el que se indicaba que un estudio entre alumnos de una universidad puede llevarse a cabo bajo el enfoque de muestreo estratificado, puede reestructurarse y plantearse en términos de muestreo por conglomerados. Por ejemplo, un primer nivel de conglomerados lo constituyen las diferentes carreras que se imparten, dentro de ellas pueden definirse nuevos conglomerados como son turno (matutino y vespertino) y dentro de estos nuevos conglomerados es posible, a su vez, definir conglomerados adicionales, tales como el semestre que se cursa ($1°$, $2°$, $3°\ldots$).

Otro ejemplo lo constituye una investigación a nivel nacional. Un primer nivel de conglomerados lo conforman los estados del país, dentro de cada estrato o estado, pueden definirse nuevos estratos tales como delegaciones y municipios y, asimismo, dentro de cada delegación o municipio pueden establecerse estratos adicionales, que pueden ser colonias. Este proceso puede continuarse en varios niveles.

Entre las ventajas que presenta este tipo de muestreo, se encuentra que no necesita un marco muestral muy amplio, como en el caso del muestreo aleatorio simple, en el que se requiere un listado de todos los elementos de una población. En el muestreo por conglomerados, el marco muestral es mucho más breve (lista de estados, municipios, delegaciones) y por lo general ya se dispone de él, aprovechando la existencia de divisiones políticas y territoriales de un país. De igual manera, para estudios de gran escala, por ejemplo a nivel nacional, presentan un costo razonable en cuanto a recursos monetarios, temporales y técnicos.

Entre las desventajas que presenta el muestreo por conglomerados, se tiene que puede disminuir la precisión de las estimaciones por la inevitable homogeneidad de de los conglomerados.

"Sudman (1976) señala los siguientes aspectos a considerar en la elección de los *conglomerados*:

"*a)* Los *conglomerados* han de estar bien definidos y delimitados. Cada unidad de la población sólo puede pertenecer a un único *conglomerado*
"*b)* El número de elementos que componen el *conglomerado* ha de ser conocido previamente (aunque sea de manera aproximada).
c) Los *conglomerados* elegidos han de ser pocos, si realmente quieren reducirse los costes de la investigación.
"*d)* Los *conglomerados* deberían escogerse de manera que se consiguiera disminuir el aumento en el *error muestral*, generado por la agrupación (o aconglomeración).
"*e)* Los *conglomerados* no tienen por qué hallarse idénticamente definidos en todos los lugares."[11]

Ejercicios 6

- Contesta las siguientes preguntas, al finalizar compara tus respuestas con las de la página 568

1. ¿Qué son los estratos?
   ____
   ____
   ____

[11] op. cit. CEA, D' Ancona; María Ángeles; Madrid; 1998; pp. 193

2. ¿Cuántos tipos de muestra pueden ser definidos, una vez que se ha estratificado a una población?

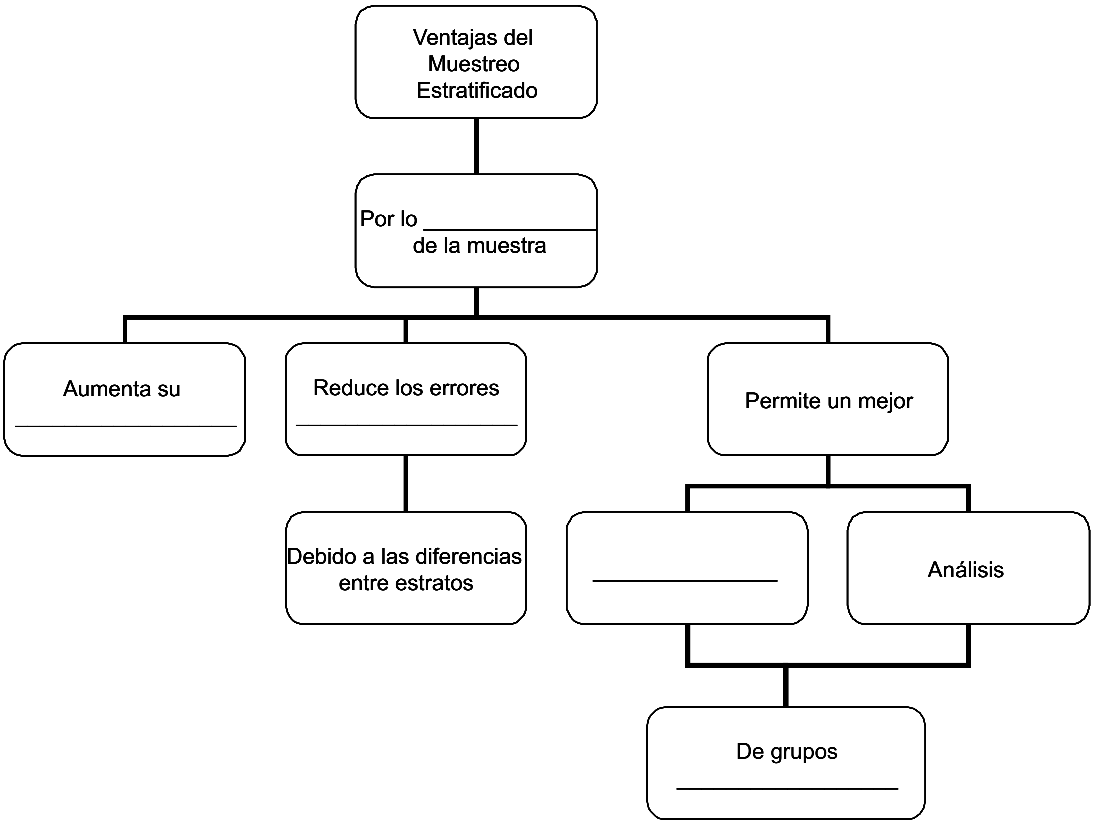
   ____

3. ¿En función de qué se debe calcular el tamaño n de la muestra?

   ____

   ____

4. Coloca en la línea correspondiente, las principales ventajas que ofrece el muestreo estratificado.



5. ¿En qué casos es recomendable el empleo del Muestreo por Conglomerados?

   ---

6. ¿Cuál es la diferencia entre el muestreo estratificado y el muestreo por conglomerados en cuanto a las unidades muestrales?

   ---

7. ¿Qué son las muestras multi-etápicas?

10.6. Autoevaluación

- *Instrucciones*: Antes de iniciar la resolución del presente apartado, revisa los ejercicios que hasta el momento has resuelto y corregido con ayuda de las respuestas que en este material se proporcionan. En igual forma, revisa tus apuntes personales y notas hechas en el texto, ellos *deben* servir para reafirmar tu asimilación del contenido y optimizar tu desempeño en la autoevaluación.

- Las preguntas y ejercicios que estás a punto de resolver, fueron seleccionadas (en función de los objetivos de la unidad) directamente de las secciones de ejercicios ya resueltas y, por tanto, el nivel personal de exigencia será mayor; mas, en definitiva, nadie como tú para imponerte un nivel de exigencia superior al ya logrado.

La única sugerencia que al respecto se te hace, es que sobre la base que te dan la lectura y repaso del contenido, así como el haber resuelto los ejercicios de la unidad, realices las actividades con toda la calma que te sea posible y que apeles al razonamiento antes que a la memoria de corto plazo para contestar.

1. ¿En qué se debe fundamentar la elección de las unidades muestrales de una investigación para poder estimar su nivel de confianza y el error?

    a) En las técnicas de muestreo.

    b) En procedimientos lógico-matemáticos.

    c) En la teoría de la causalidad.

    d) En el muestreo no probabilístico.

2. ¿En el muestreo por cuotas, de qué manera son seleccionados los individuos que forman parte de la muestra?

    a) Sistematizadamente.

    b) Investigando las características de una población con cuestionarios aleatorios.

    c) Seleccionando a n porciones del universo de forma proporcional.

    d) El encuestador selecciona a los sujetos que entrevistará.

3. ¿Qué se debe hacer para aumentar el nivel de confianza y disminuir el error muestral?

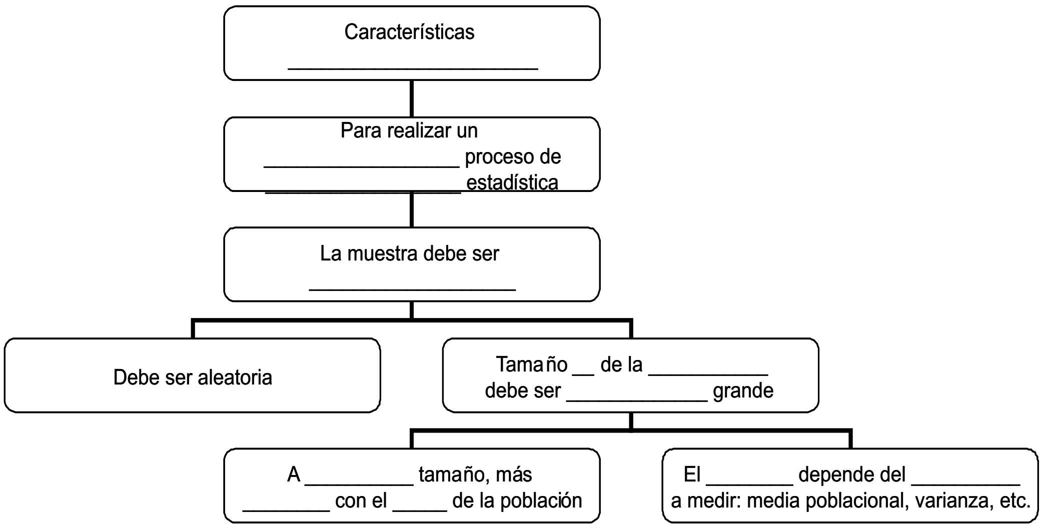
   a) Sondear el objeto-fenómeno de estudio antes de elaborar el cuestionario para recopilar los datos.

   b) Incrementar los recursos materiales y humanos.

   c) Determinar el tamaño de la población.

   d) Aumentar el tamaño de la muestra.

4. Anota en las líneas de la red conceptual, la palabra que complete correctamente el concepto.



- Completa el concepto que se te presenta, colocando en la línea correspondiente, la palabra que falta.

5. Si se ____________________ del marco muestral las unidades erróneamente incluidas en él, ya sea unidades ____________________, vacías o ______________

__________ y simultáneamente se ____________________ las omisiones, se obtiene la población objetivo. Este proceso se denomina ____________________ de marcos ____________________.

6. El término ____________________, hace referencia a la precisión buscada en la estimación. Se busca que el error estándar sea ____________________, es decir, que la precisión sea ____________________. El error estándar indica el ____________________ máximo de discrepancia que se está dispuesto a aceptar entre el ____________________ poblacional y su respectivo estimador.

7. Una vez que se ha conformado un ____________________ y que se ha depurado, se procede, en función del ____________________ que se desea investigar, a *diseñar la muestra*; esto, en primer lugar, hace necesario determinar el ______________ de la misma. Para ello, el investigador empleará la ____________________ correspondiente a dicho parámetro. Aunque las fórmulas son diferentes, todas incluyen para su cálculo, dos elementos fundamentales que son establecidos por el investigador, a saber, el ______________________________ y el ____________________ estándar.

8. ¿Qué es una encuesta piloto?
____
____
____

9. ¿Qué se debe considerar para que las preguntas de un cuestionario sean claras?
____
____
____

10. ¿Qué indican el nivel de confianza y el error estándar?

    ____

    ____

    ____

    ____

    ____

11. Se desea conocer la cantidad promedio de horas que un ama de casa ve la televisión al día en una cierta ciudad. Se observa entonces que el atributo estadístico a estimar es ______________. Para ello se realizará un estudio a través de muestreo aleatorio, por lo que es importante determinar el tamaño n de la muestra (número de __________________ que se entrevistarán). Antes de ello, se realiza al azar un muestreo piloto a 60 amas de casa, y al analizar esta muestra piloto, se obtienen los siguientes estadísticos:

    - Media: $\bar{x} = 4$ (horas de TV en promedio al día)
    - Desviación estándar: $s = 1{,}7$ (horas)

    Es decir, una ama de casa de la muestra piloto, ve $4 \pm 1{,}7$ horas la TV al día, es decir, el número x de horas que ocupa en ver televisión, se encuentra en el intervalo: ______________, que en términos más prácticos significa que en promedio, una persona de la muestra piloto, observa entre ______________ horas la televisión al día.

    Si se desea en nuestra investigación que la muestra tenga un nivel de confianza del 98,5 % y un error o precisión del 50 %, se debe determinar el tamaño n de la muestra que satisfaga las condiciones planteadas.

Como _________ se conoce el valor de la desviación estándar poblacional σ_2, se emplear la expresión:

De acuerdo a las características del ejemplo, los valores:

- $1-\alpha=$_____ (el cual puede expresarse como _______). Esto implica que z es igual a las coordenadas en la tabla normal correspondiente al valor del área $\frac{1-\alpha}{2}$, es decir a las coordenadas correspondientes al valor _________________. Las coordenadas de dicho valor son: _________________, es decir, $z=$_________.
- $s=$________ (Desviación estándar de la muestra piloto)
- $e_\alpha=$_______

Se sustituyen en la respectiva fórmula:

$n=$_______

Esto significa que con una muestra de tamaño _______, se tiene una confianza del ________________ respecto al valor estimado de la media (número promedio horas diarias que una ama de casa ve televisión al día) y el máximo error a cometer es del 50 % respecto al valor verdadero.

12. Aplica la estrategia de selección de unidades muestrales, propia del muestreo sistemático, al tamaño n de la muestra para los datos que se te proporcionan y enlista los primeros 10 números correspondientes a las respectivas unidades muestrales, así como el de la última; supóngase que:

$N = 10000$
$n = 400$
$k = 2$

____
____
____
____

13. ¿Cuál es la diferencia entre el muestreo estratificado y el muestreo por conglomerados en cuanto a las unidades muestrales?

____
____
____
____

- Respuestas en la página 570

10.7. Apéndice del capítulo

10.7.1. Respuestas a los ejercicios de la unidad

Ejercicios 1

1. ¿Cuáles son las categorías en que se puede dividir a la investigación científica desde un punto de vista metodológico?
   a) Investigación cualitativa e investigación cuantitativa.

2. ¿En cuál de estas perspectivas o categorías de la investigación científica se defiende la existencia de un solo método general a todas las ciencias?
   b) La perspectiva cientificista cuantitativa.

3. ¿En qué se debe fundamentar la elección de las unidades muestrales de una investigación para poder estimar su nivel de confianza y el error?
   b) En procedimientos lógico-matemáticos.

4. ¿Por qué las técnicas de muestreo no probabilístico tienen un bajo nivel de confiabilidad?
   c) La elección de las unidades muestrales es arbitraria.

5. ¿De qué manera(s) se pueden minimizar el sesgo y las tendencias propias del investigador en una investigación?
   d) Diseñando una muestra de tamaño adecuado y seleccionando las unidades muestrales en forma aleatoria.

6. Identifica una de las premisas más importantes para las técnicas de muestreo cuantitativo.
   b) Las muestras perfectas no existen.

7. ¿Cuál es el punto de partida de la investigación cualitativa?
   a) Hay una realidad intersubjetiva qué conocer.

8. ¿Cómo se denomina a la parte de la estadística que se encarga de analizar y describir a una parte de la población conocida como muestra?
   c) Estadística descriptiva.

9. ¿Cómo se denomina a la parte de la estadística que se ocupa de generalizar al total de la población las características obtenidas de una muestra?
   a) Estadística inferencial.

10. ¿En el muestreo por cuotas, de qué manera son seleccionados los individuos que forman parte de la muestra?
    d) El encuestador selecciona a los sujetos que entrevistará.

11. ¿Por qué se dice que una muestra es representativa de una población?
    b) Cada unidad muestreada representará las características de una cantidad conocida de unidades en la población.

12. ¿Cómo se debe expresar el proceso de generalización de las características que describen a una muestra al total de la población?
    c) En términos de probabilidad.

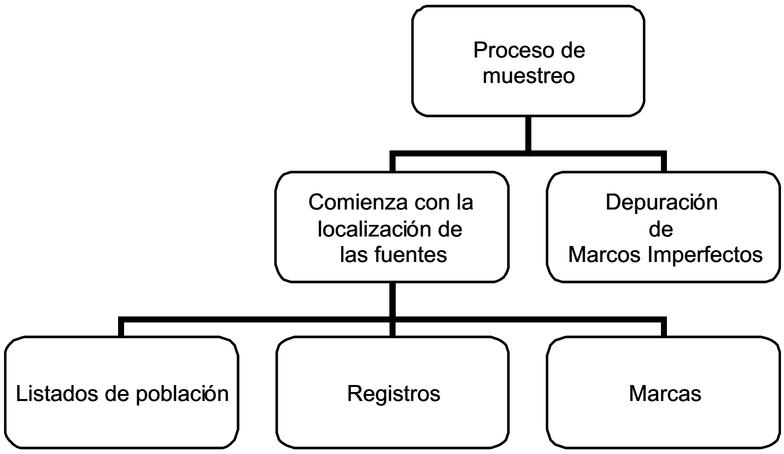
13. ¿Qué se debe hacer para aumentar el nivel de confianza y disminuir el error muestral?
    d) Aumentar el tamaño de la muestra.

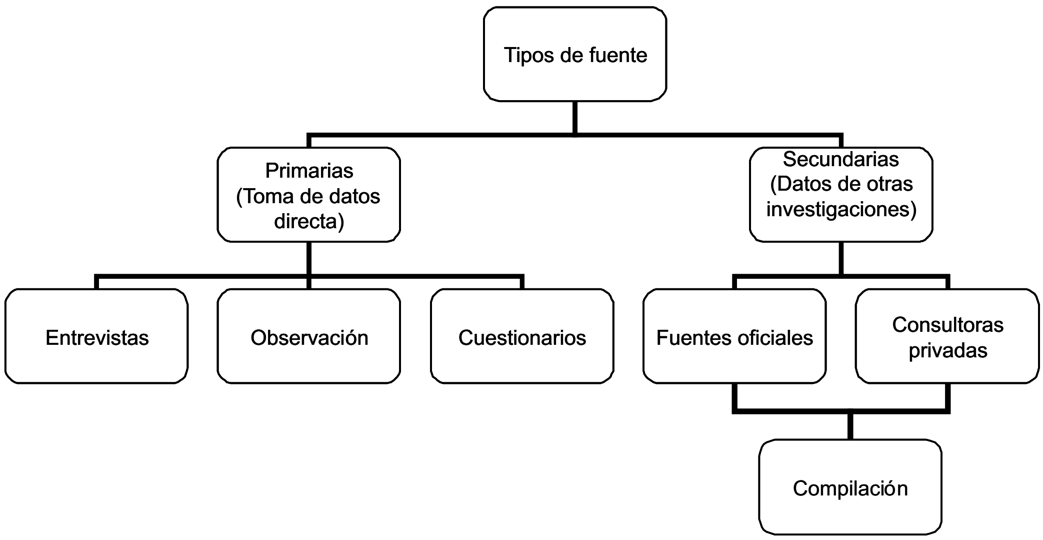
14. Señala una característica de las técnicas de muestreo no probabilístico?
    c) La muestra no aparece como representativa del Universo

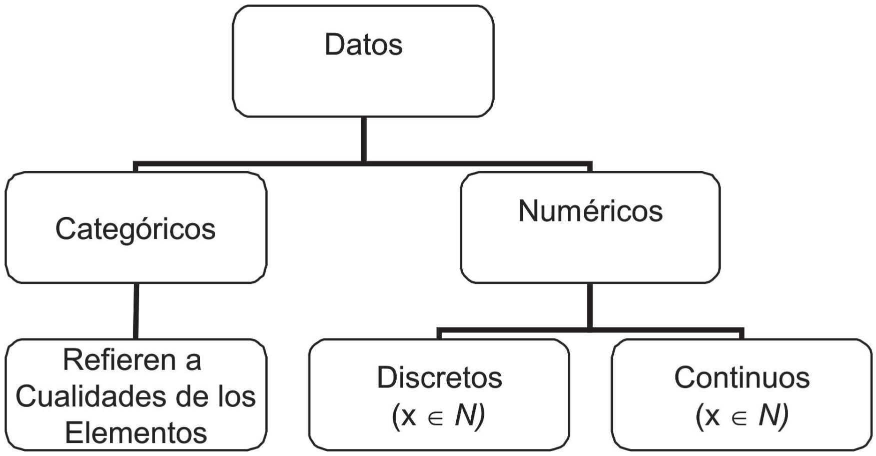
15. ¿Cuáles son las categorías, generales, en que se clasifica el muestreo no probabilístico?
    b) Muestreo Casual, Muestras Intencionales, Muestreo por Cuotas.

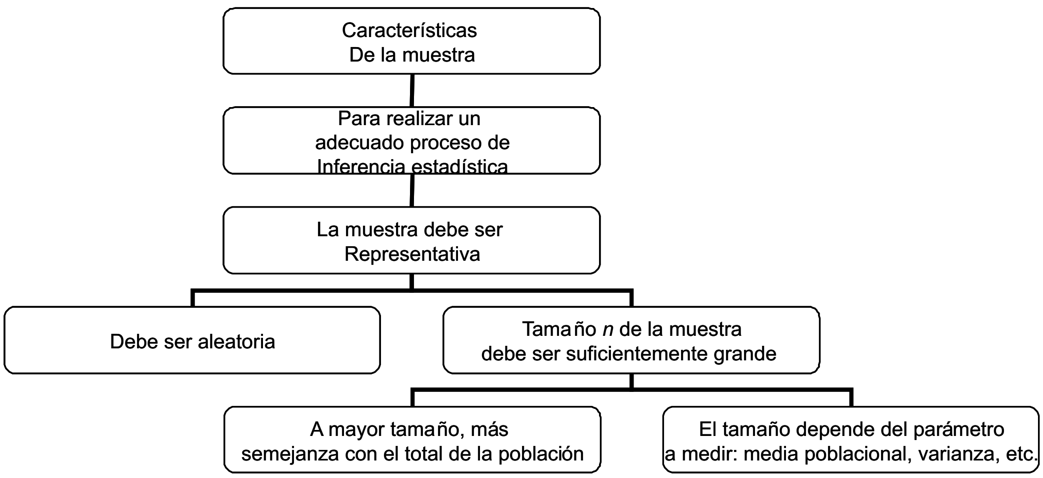
Ejercicios 2

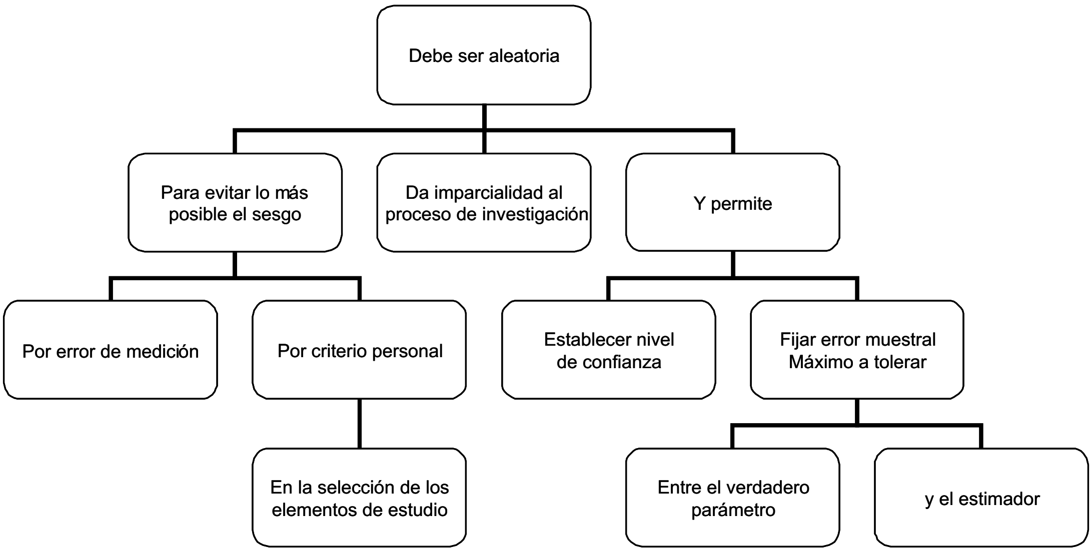
1. Anota en las líneas de la red conceptual, la palabra que complete correctamente el concepto.






Datos
Categóricos
Numéricos
Refieren a Cualidades de los Elementos
Discretos (x ∈ N)
Continuos (x ∈ N)


Características De la muestra
Para realizar un adecuado proceso de Inferencia estadística
La muestra debe ser Representativa
Debe ser aleatoria
Tamaño n de la muestra debe ser suficientemente grande
A mayor tamaño, más semejanza con el total de la población
El tamaño depende del parámetro a medir: media poblacional, varianza, etc.



- Completa el concepto que se te presenta, colocando en la línea correspondiente, la palabra que falta.

2. Para realizar un proceso de INVESTIGACIÓN, primeramente se deberá definir DETALLADA e íntegramente la POBLACIÓN sujeta a ESTUDIO. La población es la COLECCIÓN completa de elementos sobre los que se desea realizar alguna inferencia. A esta *POBLACIÓN se le denomina población objetivo.*

3. A un elemento de una población se le denomina *UNIDAD elemental de MUESTREO*, sin embargo en ocasiones es conveniente AGRUPAR a las unidades elementales de muestreo en conjuntos. En este caso se habla de una unidad de muestreo COMPUESTA, que también se conoce como *unidad de muestreo primaria.*

4. Para poder seleccionar a las unidades de muestreo, se necesita un LISTADO completo, sean elementales o compuestas, de las unidades que

constituyen la población sujeta a estudio. A este listado, se le conoce con el nombre de MARCO MUESTRAL.

5. Se denomina UNIDAD EXTRAÑA, a un elemento que pertenece al marco pero no puede participar en un proceso de muestreo.

6. Si se ELIMINAN del marco muestral las unidades erróneamente incluidas en él, ya sea unidades EXTRAÑAS, vacías o DUPLICACIONES y simultáneamente se INCORPORAN las omisiones, se obtiene la población objetivo. Este proceso se denomina DEPURACIÓN de marcos IMPERFECTOS.

7. Una vez que se ha conformado un MARCO MUESTRAL y que se ha depurado, se procede, en función del PARÁMETRO POBLACIONAL que se desea investigar, a *diseñar la muestra*; esto, en primer lugar, hace necesario determinar el TAMAÑO de la misma. Para ello, el investigador empleará la FÓRMULA correspondiente a dicho parámetro. Aunque las fórmulas son diferentes, todas incluyen para su cálculo, dos elementos fundamentales que son establecidos por el investigador, a saber, el NIVEL DE CONFIANZA y el ERROR estándar.

8. El término ERROR ESTÁNDAR, hace referencia a la precisión buscada en la estimación. Se busca que el error estándar sea BAJO, es decir, que la precisión sea ALTA. El error estándar indica el PORCENTAJE máximo de discrepancia que se está dispuesto a aceptar entre el PARÁMETRO poblacional y su respectivo estimador.

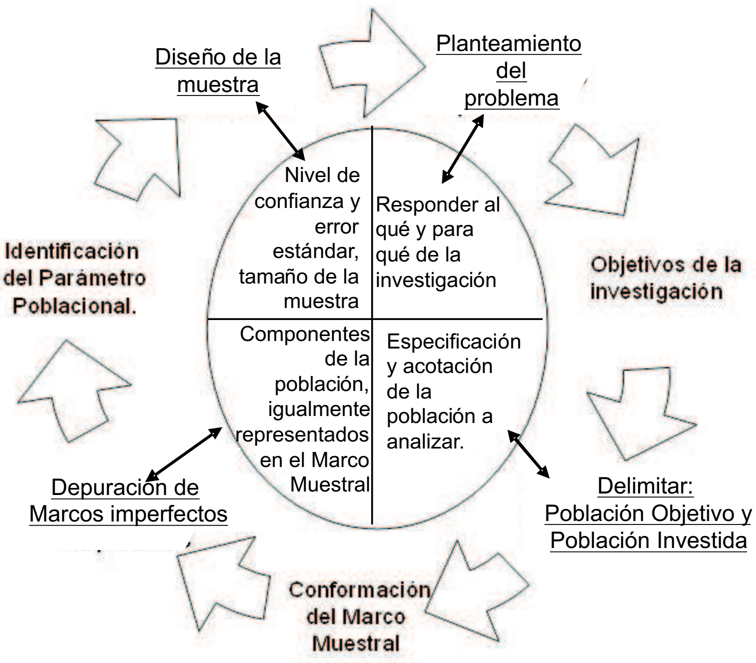
9. El ERROR estándar, al igual que el nivel de CONFIANZA, se expresa en términos de PORCENTAJE. Es razonable pensar que al realizar una investigación por muestreo, se desea un alto nivel de confianza y un bajo error estándar. Sin embargo, en la práctica NUNCA se puede llegar al CIEN POR CIENTO de confianza ni manejar un error estándar igual

a cero; debido a la naturaleza aleatoria del proceso, los resultados conllevan cierto grado de INCERTIDUMBRE (en los datos involucrados) cuantificado con métodos de la teoría de la PROBABILIDAD.

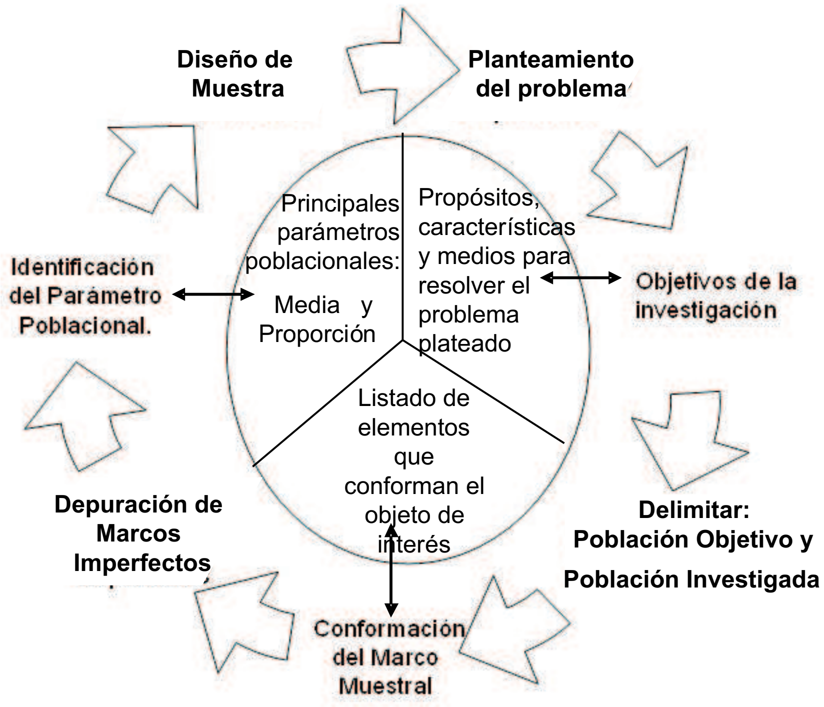
10. De igual manera, si se desea un error estándar bajo y un nivel de confianza alto, implica AUMENTAR el tamaño de la MUESTRA, con los respectivos aumentos en costo y tiempo de realización del muestreo como consecuencia. En la práctica se manejan NIVELES de CONFIANZA del 90; 95; 99 y 99,7 %.

Ejercicios 3

1. Anota los componentes que faltan en el siguiente diagrama del Muestreo Probabilístico. Apóyate en las características que al interior del círculo se presentan y señalan la con el signo ↔.



2. Ahora, coloca al interior del círculo: a)Qué indican los objetivos, b) Qué es un marco muestral y c) Principales parámetros poblacionales.



3. Investiga y anota, por lo menos, tres mecanismos para generar números aleatorios que pueden auxiliar en la selección de los elementos que integrarán una muestra.
   Tómbola; Tabla de números aleatorios; dados; la tecla *RAN#* de la calculadora y programas de cómputo especializados.

4. ¿Qué es una encuesta piloto?
   Es un mecanismo que sirve para probar el instrumento de medición o cuestionario y para obtener estadígrafos que permitan la determinación del tamaño de la muestra.

5. ¿Cuál es el mecanismo que permite extraer y cuantificar información sobre cada uno de los elementos de la muestra?
   El instrumento de medición, como puede ser un cuestionario.

6. ¿En qué caso(s), se puede afirmar que el investigador incurrió en el uso de un mecanismo de autoselección para extraer información de la muestra?
   Cuando de algún modo, se le confirió especial libertad a los involucrados, decidir si forman parte de la muestra o no

7. Proporciona dos ejemplos de mecanismos de autoselección.
   Envío de encuestas por correo (ya sea electrónico o a través del servicio postal), los sondeos realizados en los medios electrónicos, donde el auditorio es invitado a participar con su voto u opinión.

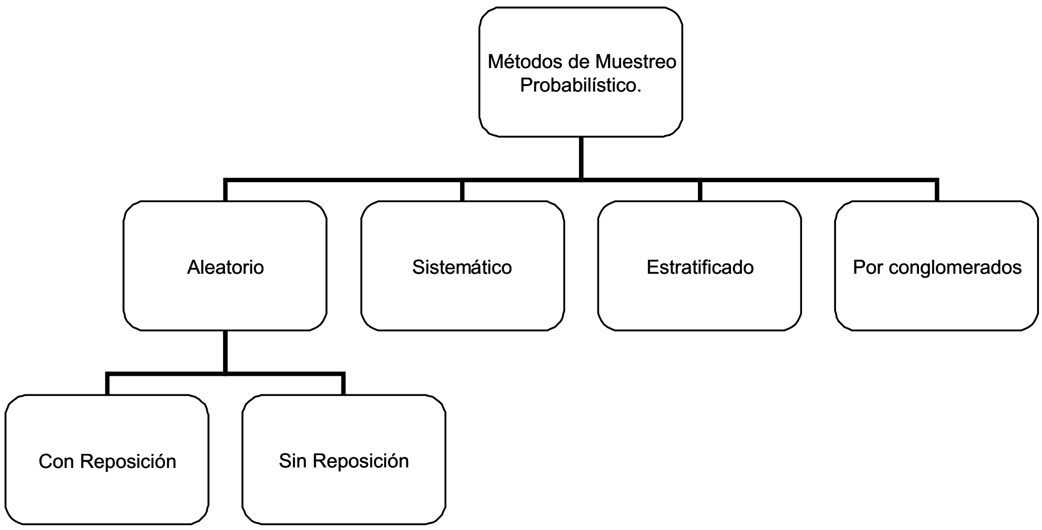
8. Observa la siguiente pregunta y señala el error en que incurre: ¿Cree usted que el acervo de esta biblioteca se debe actualizar anualmente?
   La pregunta refleja la opinión del investigador, al decir que el acervo de la biblioteca *se debe* actualizar actualmente; situación que puede inducir la respuesta de la persona investigada en este sentido.

9. ¿Qué se debe considerar para que las preguntas de un cuestionario sean claras?
   Las características socio-culturales (Ocupación, edad, escolaridad, relación con el tema, etc.) del público al que se dirigen las preguntas.

10. Enuncia una posible desventaja de las preguntas abiertas incluidas en los cuestionarios?
    Con frecuencia, éstas ocasionan que el entrevistado se desvíe del tema o divague.

Ejercicios 4

1. Anota los métodos de muestreo probabilístico que se mencionan en el texto

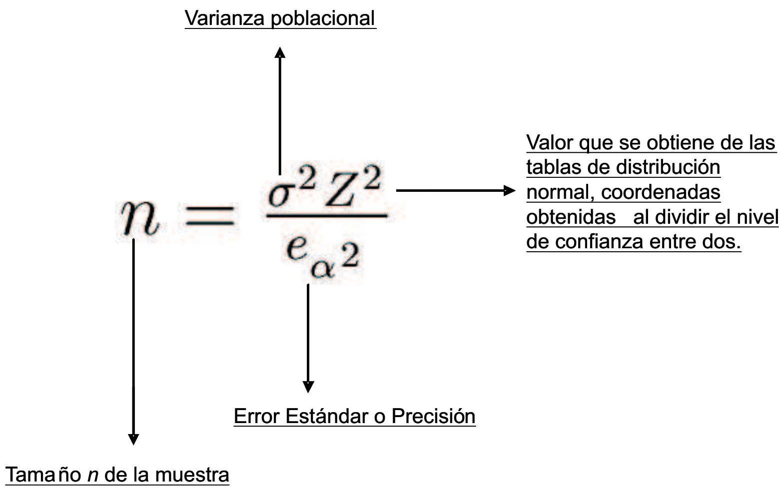


2. ¿Qué permite calcular la siguiente fórmula? $C_n^N = \begin{pmatrix} N \\ n \end{pmatrix} = \frac{N!}{n!(N-n)!}$

   Las posibles muestras de tamaño n de una población, es decir, se calculan las combinaciones de n elementos tomados de un total de N.

3. Qué indican el nivel de confianza y el error estándar?
   El primero indica la probabilidad que se desea de que el valor de la media poblacional estimada sea correcto. El segundo indica en términos porcentuales, el grado máximo que se está dispuesto a aceptar de discrepancia o diferencia entre el verdadero valor de la media poblacional y el valor estimado en la muestra.

4. Coloca el significado de las literales de la fórmula para que el tamaño mínimo n de la muestra sea representativo de la media poblacional.



5. Retoma el ejemplo de la cantidad promedio de cigarrillos (pp. 507) con los datos necesarios y calcula n para un error del 20 % y un nivel de confianza del 99 %. Compara el resultado de este ejercicio con el valor del anterior y coméntalo.

Datos:
a) $\bar{x} = 5$ es decri: 5 cigarrillos en promedio al día.
b) $s = 2{,}1$
c) $x = 5 \pm 2{,}1$ ó $2{,}9 < x < 7{,}1$ es decir: en promedio, una persona de la muestra piloto fuama entre 3 y 7 cigarrillosal día.
d) $1 - \alpha = 0{,}99$
e) $e_\alpha = 0{,}2$

Fórmulas:

$z = \frac{1-\alpha}{2}$
$n = \frac{s^2 Z^2}{e_\alpha^2}$

Sustitución y operaciones:

$z = \frac{0{,}99}{2} = 0{,}495$ que corresponde a las cordenadas 2,56 en la tabla de áreas bajo la curva normal; por tanto: $z = 2{,}56$

$n = \frac{(2{,}1^2)(2{,}56^2)}{0{,}2^2} = \frac{(4{,}41)(6{,}55)}{0{,}04} = 722{,}25 \approx 722$

Resultados:
$n = 722$

Comentarios:

Como $1 - \alpha$ aumenta de tal manera que se acerca al 100 %, $n \Rightarrow N$ y el error estándar (e_α) a tolerar con respecto a la verdadera media poblacional disminuye considerablemente, el número de unidades que conforman la muestra aumentó dramáticamente con respecto al ejemplo retomado.

Ejercicios 5

1. ¿Qué se mide mediante el muestreo de estimación por atributos?
   Mide qué *proporción* del universo posee (o no) un atributo o característica.

2. ¿Qué parte del universo se denota con la expresión $(100 - p)\,\%$?
   La parte del universo, o población, que no cumple con la característica o atributo de interés para la investigación.

3. Señala en qué casos se utilizan las siguientes fórmulas:

   - $n = \frac{n_0}{1+\frac{n_0-1}{N}}$ Universos o poblaciones finitas
   - $n = z^2 \frac{p(1-p)}{e_\alpha^2}$ Universos o poblaciones infinitas o muy grandes

4. Retoma el caso de la agencia de publicidad que desea conocer el porcentaje de personas que leen cierta revista de publicación periódica (pp. 514) y calcula el tamaño n de la muestra tanto para universos finitos como infinitos. En este caso, se desea un nivel de confianza igual a 90 % y un error estándar igual a 30 %. Recuerda que no se dispone de tiempo para realizar una encuesta piloto. Finalmente, supóngase que 1325 personas de la muestra sí leen la revista: expresa en términos de porcentaje y unitarios el total de personas, pertenecientes a la población, que leen la revista.

   Datos:
   a) $\alpha - 1 = 90\,\%$
   b) $e_\alpha = 30\,\%$
   c) $p = 50\,\%$
   d) $1 - p = 50\,\%$
   e) $N = 750000$

   Fórmulas:
   $z = \frac{1-\alpha}{2}$

$n = z^2 \frac{p(1-p)}{e_\alpha^2}$ Universos Infinitos (pp. 512)

$n = \frac{n_0}{1+\frac{n_0-1}{N}}$ Universos Finitos (pp. 513)

Sustitución y Operaciones:

$z = \frac{0{,}9}{2} = 0{,}45$ que corresponde a las coordenadas 1,64 en la tabla de áreas bajo la curva.
$z = 1{,}64$

$n = 1{,}64^2 \frac{50(100-50)}{3} = 2249{,}91 \approx 2250$

$n = \frac{2250}{1+\frac{2250-1}{750000}} = 22{,}43270189 \approx 2243$

Resultados:

El tamaño n de la muestra para *universos infinitos* es $n = 2250$

El tamaño n de la muestra para *universos finitos* es $n = 2243$; por lo que se debe entrevistar aleatoriamente a éste número de personas (2243) para saber si leen la revista o no. Si se supone que de las 2243 unidades muestrales 1325 sí leen la revista, ello indica que probablemente el 59 % *de los individuos que conforman la población* (este resultado sí se puede generaliza) leen la revista; es decir: aproximadamente unas 442500 personas.

5. Aplica la estrategia de selección de unidades muestrales, propia del muestreo sistemático, al tamaño n de la muestra para la anterior pre-

gunta y enlista los primeros 10 números correspondientes a las respectivas unidades muestrales, así como el de la última. Para este caso, el número aleatorio seleccionado será el 2

- 2
- 336
- 1004
- 1338
- 1672
- 2006
- 2340
- 2674
- 3008
- 3342
- 748830

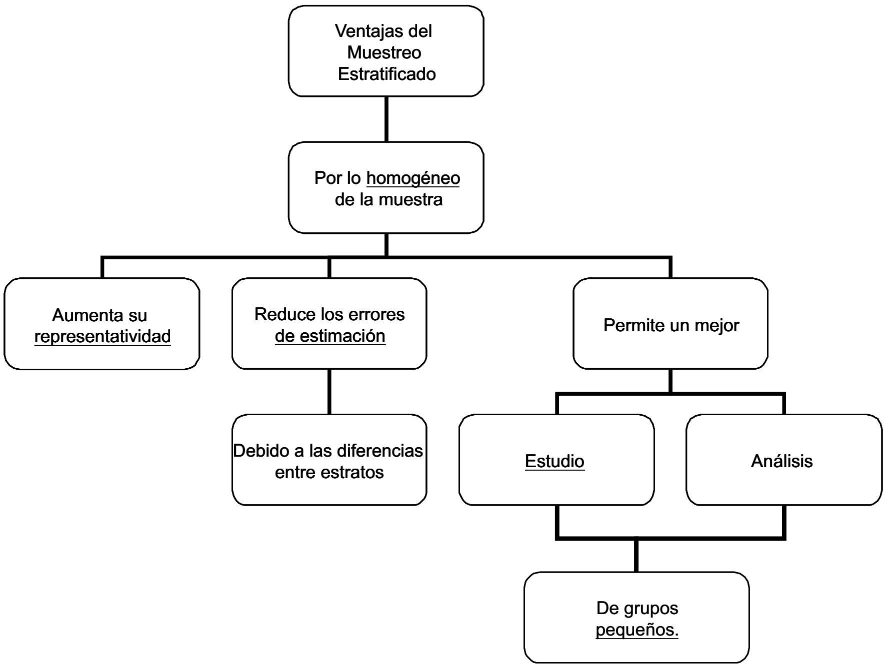
Ejercicios 6

1. ¿Qué son los estratos?
   Son categorías en que se divide y organiza a los elementos de una población, tomando en cuenta diferencias bien definidas entre éstos, tales como nivel económico, escolar, edad...

2. ¿Cuántos tipos de muestra pueden ser definidos, una vez que se ha estratificado a una población?
   Muestra estratificada proporcional y Muestra estratificada no proporcional.

3. ¿En función de qué se debe calcular el tamaño n de la muestra?
   En función del parámetro poblacional a investigar, el error estándar y nivel de confianza deseados.

4. Coloca en la línea correspondiente, las principales ventajas que ofrece el muestreo estratificado.



5. ¿En qué casos es recomendable el empleo del Muestreo por Conglomerados?
   Cuando la investigación a realizar, contempla una población o universo a gran escala (una entidad federativa, país, etc.) y ésta (la población), se encuentra dispersa a lo largo de áreas geográficas de gran extensión.

6. ¿Cuál es la diferencia entre el muestreo estratificado y el muestreo por conglomerados en cuanto a las unidades muestrales?

Que en el muestreo por estratos, la unidad muestral es un individuo y en el muestreo por conglomerados la unidad muestral es el grupo (conglomerado).

7. ¿Qué son las muestras multi-etápicas?
   Muestras obtenidas mediante la repetición del muestreo por conglomerados para elegir a las unidades muestrales de estratos previamente seleccionados.

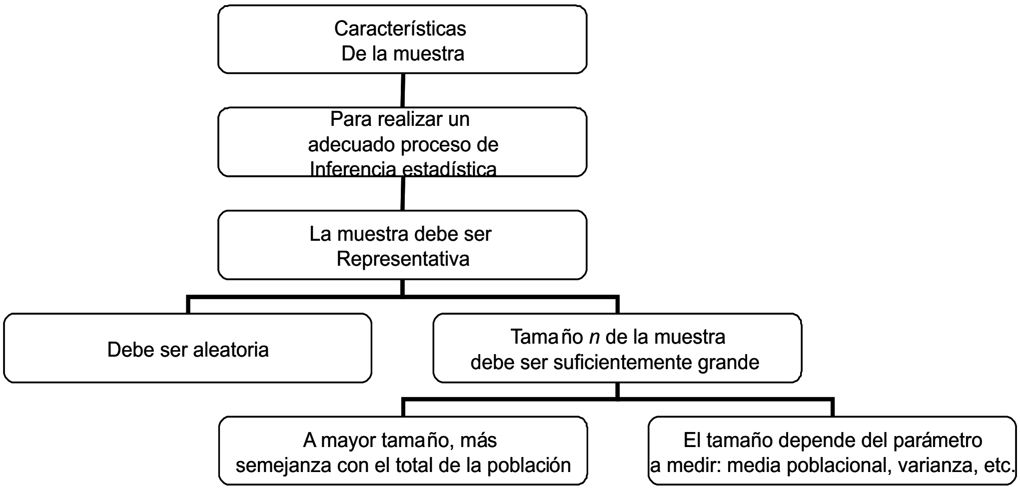
10.7.2. Respuestas de la autoevalución

1. ¿En qué se debe fundamentar la elección de las unidades muestrales de una investigación para poder estimar su nivel de confianza y el error?
   b) En procedimientos lógico-matemáticos.

2. ¿En el muestreo por cuotas, de qué manera son seleccionados los individuos que forman parte de la muestra?
   d) El encuestador selecciona a los sujetos que entrevistará.

3. ¿Qué se debe hacer para aumentar el nivel de confianza y disminuir el error muestral? *d)* Aumentar el tamaño de la muestra.

4. Anota en las líneas de la red conceptual, la palabra que complete correctamente el concepto.



- Completa el concepto que se te presenta, colocando en la línea correspondiente, la palabra que falta.

5. Si se ELIMINAN del marco muestral las unidades erróneamente incluidas en él, ya sea unidades EXTRAÑAS, vacías o DUPLICACIONES y simultáneamente se INCORPORAN las omisiones, se obtiene la población objetivo. Este proceso se denomina DEPURACIÓN de marcos IMPERFECTOS.

6. El término ERROR ESTÁNDAR, hace referencia a la precisión buscada en la estimación. Se busca que el error estándar sea BAJO, es decir, que la precisión sea ALTA. El error estándar indica el PORCENTAJE máximo de discrepancia que se está dispuesto a aceptar entre el PARÁMETRO poblacional y su respectivo estimador.

7. Una vez que se ha conformado un MARCO MUESTRAL y que se ha depurado, se procede, en función del PARÁMETRO POBLACIONAL

que se desea investigar, a *diseñar la muestra*; esto, en primer lugar, hace necesario determinar el TAMAÑO de la misma. Para ello, el investigador empleará la FÓRMULA correspondiente a dicho parámetro. Aunque las fórmulas son diferentes, todas incluyen para su cálculo, dos elementos fundamentales que son establecidos por el investigador, a saber, el NIVEL DE CONFIANZA y el ERROR estándar.

8. ¿Qué es una encuesta piloto?
   Es un mecanismo que sirve para probar el instrumento de medición o cuestionario y para obtener estadígrafos que permitan la determinación del tamaño de la muestra.

9. ¿Qué se debe considerar para que las preguntas de un cuestionario sean claras?
   Las características socio-culturales (Ocupación, edad, escolaridad, relación con el tema, etc.) del público al que se dirigen las preguntas.

10. ¿Qué indican el nivel de confianza y el error estándar?
    El primero indica la probabilidad que se desea de que el valor de la media poblacional estimada sea correcto. El segundo indica en términos porcentuales, el grado máximo que se está dispuesto a aceptar de discrepancia o diferencia entre el verdadero valor de la media poblacional y el valor estimado en la muestra.

11. Se desea conocer la cantidad promedio de horas que un ama de casa ve la televisión al día en una cierta ciudad. Se observa entonces que el atributo estadístico a estimar es LA MEDIA μ DE LA POBLACIÓN. Para ello se realizará un estudio a través de muestreo aleatorio, por lo que es importante determinar el tamaño de la muestra (número de AMAS DE CASA que se entrevistarán). Antes de ello, se realiza al azar un muestreo piloto a 60 amas de casa, y al analizar esta muestra piloto,

se obtienen los siguientes estadísticos:

- Media: $\bar{x} = 4$ (horas de TV en promedio al día)
- Desviación estándar: $s = 1{,}7$ (horas)

Es decir, una ama de casa de la muestra piloto, ve $4 \pm 1{,}7$ horas la TV al día, es decir, el número x de horas que ocupa en ver televisión, se encuentra en el intervalo: $2{,}3 < x < 5{,}7$, que en términos más prácticos significa que en promedio, una persona de la muestra piloto, observa entre dos y seis horas la televisión al día.

Si se desea en nuestro ejemplo que la muestra tenga un nivel de confianza del 98,5 % y un error o precisión del 50 %, se debe determinar el tamaño n de la muestra que satisfaga las condiciones planteadas.

Como NO se conoce el valor de la desviación estándar poblacional σ_2, se emplear la expresión:

$$n = \frac{s^2 Z^2}{e_\alpha^2}$$

De acuerdo a las características del ejemplo, los valores:

- $1 - \alpha$ (nivel de confianza) $= 98{,}5\,\%$ (el cual puede expresarse como 0,985). Esto implica que z es igual a las coordenadas en la tabla normal correspondiente al valor del área $\frac{1-\alpha}{2}$, es decir a las coordenadas correspondientes al valor $\frac{0{,}985}{2} = 0{,}4925$. Las coordenadas de dicho valor son: 2,43 es decir, $z = 2{,}43$.
- $s = 1{,}7$ (Desviación estándar de la muestra piloto)
- $e_\alpha = 0{,}5$

Se sustituyen en la respectiva fórmula:

$n = \frac{(1{,}7)^2(2{,}43)^2}{(0{,}5)^2}$
$n = \frac{2{,}89(5{,}9049)}{0{,}25}$
$n = 682{,}370244 \approx 682$
$n = 682$

Esto significa que con una muestra de tamaño 682, se tiene una confianza del 98.5 % respecto al valor estimado de la media (número promedio horas diarias que una ama de casa ve televisión al día) y el máximo error a cometer es del 50 % respecto al valor verdadero.

12. Aplica la estrategia de selección de unidades muestrales, propia del muestreo sistemático, al tamaño n de la muestra para los datos que se te proporcionan y enlista los primeros 10 números correspondientes a las respectivas unidades muestrales, así como el de la última; supóngase que:

$N = 10000$
$n = 400$
$k = 2$

1) $M = \frac{10000}{400} = 25$

2) k tal que $k \in [1, m]$ $k = 2$

3) Subsecuentemente, se tienen qué tomar, del Marco Muestral, las unidades correspondientes a:

$$k, k+m, k+2m, k+3m, \ldots \quad \ldots k+(n-1)m$$

que sustituida con los valores calculados, serían las siguientes:

$$2, 2+25, 2+2(25), 2+3(25), 2+4(25), 2+5(25), 2+6(25), 2+ 7(25), 2+8(25), 2+9(25) \ldots \ \ldots 2+(400-1)25$$

Así, las unidades muestrales serían las correspondientes a los números 2, 27, 52, 77, 102, 127, 152, 177, 202, 227... hasta completar los 400 elementos que conformarían la muestra, donde la última unidad correspondería al individuo número 9977 del marco muestral.

13. ¿Cuál es la diferencia entre el muestreo estratificado y el muestreo por conglomerados en cuanto a las unidades muestrales?
Que en el muestreo por estratos, la unidad muestral es un individuo y en el muestreo por conglomerados la unidad muestral es el grupo (conglomerado).

10.7.3. Glosario

Acotar: Término utilizado en sentído análogo a la función de las cotas, es decir, establecer límites o limitar.

Atributo(s): Características o elementos propios del objeto de estudio. Mas, desde un punto de vista pertinente a la investigación de base cuantitativa, se consideran como atributos a todas aquellas propiedades que pueden ser descritas numéricamente; así, se consideran atributos de un fenómeno u objeto de estudio, la media o promedio, su proporción y todas las medidas de tendencia central y dispersión, que lo describen.

Causalidad: "Principio que establece una relación necesaria entre el antecedente y el consecuente, la causa y el efecto."[12]

Cotas: "Sea C un subconjunto no vacío de R. El número real c es una *cota superior* de C si es mayor o igual que todos los elementos de C; si hay alguna cota superior de C se dice que ese conjunto está acotado superiormente [...] De manera análoga se dice que el número real c es una *cota inferior* de C si es menor o igual que todos los elementos de C."[13] En otras palabras, una cota es todo aquél número que limita formalmente a un conjunto de observaciones; luego entonces, una cota superior es el *límite superior* (o número mayor) de un conjunto de valores u observaciones representadas numéricamente; de forma similar, se puede inferir el concepto de cota inferior.

Paradigma(s): "Modelo o ejemplo. Platón empleó la palabra en el primer sentido en cuanto considera como P. al mundo de los seres eternos, del cual es imagen el mundo sensible. Aristóteles en la lógica usa el término en el segundo significado"[14] De ahí que en la actualidad un paradigma teórico es un una construcción *ideo-lógica* que sirve como un modelo formal o ejemplo explicativo, general, aplicable a cierto fenómeno.

Piloto: Como adjetivo en investigación, indica que cierto proceso, fenómeno o instrumento de medición, se está sometiendo a prueba; lo anterior, con la finalidad de conocer algunas características generales del fenómeno o proceso y de verificar el comportamiento (pertinencia, precisión) particular de los elementos que componen al instrumento de medición.

---

[12]Op. cit. XIRAU, Ramón; México; 1995; pp. 456

[13]CLAPHAM, Christopher; **Diccionario de Matemáticas**; Madrid; Editorial Complutense, Diccionarios Oxford-Complutense; 1998

[14]op. cit. ABBAGNANO, Nicolo; México; 1974

Sondeo: Investigación cuantitativa de carácter informal.

10.7.4. Tabla de números aleatorios

Tabla de números aleatorios

84648	48983	80914	07583	84060	61721	50069
38087	72676	53848	68180	26089	94766	56867
06652	10739	53143	32128	28854	22688	77088
45147	20907	40039	80377	69845	09308	31473
30036	58362	22312	25462	06789	70702	26707
44940	93636	91418	60322	60714	91214	00914
94647	18195	81180	72555	93945	67424	76808
78988	76089	13734	33598	72324	18876	57099
35240	87241	23404	76093	16702	15018	14541
52593	99101	26064	29833	37675	10838	16006
61846	52751	62600	54111	12800	08034	40450
20740	55071	91808	35120	35736	01044	11637
61426	41424	12780	88048	92658	40523	99845
31153	98752	56842	82116	64218	71593	27447
65017	24410	56844	53853	07023	68650	22728
76965	38070	82783	32283	14853	24145	97224
91878	28784	07441	98104	96503	44909	39249
07448	80157	32622	92166	55382	70253	59883
02253	19536	56655	26810	30794	47454	91574
73119	80960	67418	24858	15363	15188	52032
33314	25292	92732	35660	60052	16794	52185
13665	03360	03812	65125	88472	39445	53500
77148	32260	01564	53228	73474	09163	45026
50772	82229	72486	42263	23519	81367	09985
97851	75856	93505	05212	19463	57572	62227
17990	23793	87627	98098	69203	43774	42860
54991	40433	60284	91626	42590	99648	28071
38500	83929	83046	34654	31836	98866	83261
66317	35164	41992	04070	20324	42530	62944
20382	47911	41364	13194	55138	52354	68348
32636	74611	26789	88814	67637	77519	95966
00314	30435	94529	54627	11778	81489	32460
67523	88352	99738	41519	45109	49701	95715
50362	88793	19415	15282	19674	48145	44182
18933	79145	25019	77263	43189	30572	25601
77523	61149	77628	49956	50265	67259	84429
08925	31955	01660	19272	23698	81598	60915
96832	84387	57430	79651	84617	75088	50604
82820	78199	63732	10758	35678	72546	23854
75934	84524	47112	65373	43304	00419	84666

Capítulo 11

Índices

Objetivos específicos

- El estudiante practicará el cálculo de los principales *números índice*, mediante la resolución de ejercicios, para su incorporación al acervo de herramientas formales con que debe contar un comunicólogo, en el tratamiento de datos duros.

- El estudiante manipulará el *índice nacional de precios al consumidor* (INPC) para la medición, en términos porcentuales, de la pérdida del valor adquisitivo del dinero.

11.1. Preliminar

Los *índices* o *números índice*, son medidas estadísticas que resumen la información de diversos referentes en un sólo valor, y permiten cuantificar los cambios que pueden presentarse en una o más variables a través del tiempo y/o zonas geográficas (principalmente).

- Ejemplos:
  - El índice de delincuencia
  - Índice de contaminación
  - Índice de desempleo
  - índice nacional de precios al consumidor (INPC)
  - El Índice de Preciosos y Cotizaciones (IPyC) de la Bolsa Mexicana de Valores, entre otros

"Con los números índice se puede, por ejemplo, comparar el coste de alimentos u otros costes de vida en una ciudad durante un año con los del año anterior, o se puede comparar la producción de acero durante un año determinado en una parte del país con la habida en otra parte. Aunque su aplicación principal se encuentra en negocios y economías, los números índice pueden aplicarse en otros muchos campos."[1]

Como se podrá notar, la importancia que en el campo de la comunicación cobran dichos números, puede ser tan variada y útil como el quehacer del comunicólogo; el análisis de datos duros en el campo del periodismo para los medios de comunicación masiva, el estudio de los efectos de la comunicación en relación con las organizaciones y sus entornos, así como la investigación en comunicación, se pueden ver auxiliadas con un adecuado tratamiento de los

[1] op. cit. SPIEGEL, Murray R.; México; 1970; pp. 313

datos para convertirlos en información, situaciones en la cual, el conocimiento de los números índice puede contribuir significativamente. A continuación, se revisarán, algunos tipos de números índice y su aplicación en la economía, por ejemplo.

11.2. Precios relativos

El primer tipo de número índice que será revisado, es el llamado *precio relativo*, que formalmente se define como la razón, o cociente, del precio de un cierto año a, respecto de un año inicial b, denominado *año base o periodo de referencia*. Entonces el precio relativo, denotado por $P_{a|b}$ se calcula por:

$$\boxed{P_{a|b} = \tfrac{P_a}{P_b}}$$

Donde:
$P_a =$ Precio del producto en el año a.
$P_b =$Precio del producto en el año b (año base)

- **Ejemplo 1**:

  Supóngase que el precio de un litro de leche en el año 2000 era de \$7,50 y que en el año 2005 el precio es \$9,75. Calcula el precio relativo del litro de leche, tomando como año base el 2000.

  Solución:
  Del planteamiento se tienen los siguientes datos:
  $a = 2005$
  $b = 2000$
  $P_a = \$7{,}50$
  $P_b = \$9{,}75$

Entonces, sustituyendo estos datos en la fórmula correspondiente:
$P_{2005|2000} = \frac{P_{2005}}{P_{2000}}$
$= \frac{\$9{,}75}{7{,}50}$
$= 1{,}3$

Ahora, el resultado se debe multiplicar por 100 %, para obtener un referente sobre la evolución del precio (saber si subió o bajó de precio y en qué porcentaje lo hizo):
$1{,}3(100\,\%) = 130\,\%$

Como el valor obtenido es mayor que el 100 %, significa que el precio del producto aumentó. En este caso, 130 % es mayor que 100 % en 30 unidades, esto significa que el precio del litro de leche en el año 2005, se incrementó en un 30 % respecto al año 2000.

- **Ejemplo 2**:

Supóngase que en el año 1998 el precio de un cierto modelo de automóvil al salir de la agencia era de \$110000,00. Y en el 2005, su precio es de \$67000,00. Calcula el precio relativo del automóvil, tomando como año base 1998.

Solución:
Del planteamiento se tienen los siguientes datos:
$a = 2005$
$b = 1998$
$P_a = \$67000$
$P_b = \$110000$

Entonces, sustituyendo estos datos en la fórmula correspondiente:
$P_{2005|1998} = \frac{P_{2005}}{P_{1998}}$
$= \frac{\$67000}{\$110000}$
$= 0{,}609090909$
Ahora, el resultado se debe multiplicar por 100 %, para obtener un referente sobre la evolución del precio (saber si subió o bajó de precio y en qué porcentaje lo hizo):
$0{,}609090909(100\,\%) = 60{,}9090909\,\% \approx 61$

Como el valor obtenido es menor que el 100 %, significa que el valor del producto disminuyó. En este caso, 61 % significa que el valor de ese mismo automóvil en el año 2005, disminuyó en un 60.91 % respecto al año 2000.

Por otra parte, si al calcular un precio relativo respecto a un cierto año, se obtiene un valor exacto del 100 %, significa que el precio del producto *no ha cambiado respecto al año de referencia.*

11.2.1. Propiedades de los precios relativos

La fórmula empleada para el cálculo de los precios relativos, presenta algunas propiedades matemáticas importantes, como las siguientes:
Si P_a, P_b y P_c denotan los precios en los periodos a, b, c; entonces se cumplen las siguientes propiedades:

1. Propiedad de identidad: $P_{a|a} = 1$
2. Propiedad del tiempo inverso: $P_{a|b} \cdot P_{b|a} = 1$
3. Propiedad cíclica o circular: $P_{a|b} \cdot P_{b|c} \cdot P_{c|a} = 1$
4. Propiedad cíclica modificada: $P_{a|b} \cdot P_{b|c} = P_{a|c}$

- **Ejemplo:**
  Sean $P_a = 60$, $P_b = 15$ y $P_c = 30$. Verifica las propiedades de los precios relativos.

  1. Propiedad de identidad:
     $P_{a|a} = \frac{60}{60} = 1$
  2. Propiedad del tiempo inverso:
     $P_{a|b} \cdot P_{b|a} = \frac{60}{15} \cdot \frac{15}{60} = \frac{900}{900} = 1$
  3. Propiedad cíclica o circular:
     $P_{60|15} \cdot P_{15|30} \cdot P_{30|60} = \frac{60}{15} \cdot \frac{15}{30} \cdot \frac{30}{60} = \frac{27000}{27000} = 1$
  4. Propiedad cíclica modificada:
     $P_{a|b} \cdot P_{b|c} = \frac{60}{15} \cdot \frac{15}{30} = \frac{900}{450} = 2$
     y como $P_{a|c} = \frac{60}{30} = 2$, la propiedad se cumple.

Ejercicios 1

1. Calcula e interpreta los siguientes Precios Relativos.

   a) Supóngase que el hacia enero de 1998, el precio del dólar norteamericano (al menudeo) era de \$10,05 y que para diciembre del mismo año, era de \$12,90. Calcula el precio relativo de la divisa Estadounidense, tomando como periodo de referencia el mes de enero del mismo año.

   ----

b) Supóngase que en la región sureste de la República Mexicana, un jornalero percibe, en promedio \$39,50 al día; mientras que en el norte del país, un obrero con el salario mínimo tiene como ingreso diario \$78,50 por concepto de salario mínimo. Calcula el precio relativo del día laboral, tomando como base, el salario del trabajador norteño.

____

____

2. Señala qué propiedad de los precios relativos se verifica en los siguientes casos o, en su defecto, anota su expresión formal.

   a) Propiedad de identidad:____________________

   b) Propiedad del tiempo inverso:____________________

   c) $P_{a|b} \cdot P_{b|c} \cdot P_{c|a} = 1$____________________

   d) $P_{a|b} \cdot P_{b|c} = P_{a|c}$____________________

- Para comprobar que tus respuestas son correctas, pasa a la página 604

11.3. Índices aplicados en la economía

Método de agregación ponderada

Para la investigación en Ciencias Sociales y Comunicación, es importante disponer de referentes sobre el comportamiento de la economía, pues ésta es uno de los ejes sobre los que se desarrolla cualquier nación y, asimismo, suele ser origen de serias problemáticas de carácter político y social. Para ello, los gobiernos de los diversos países, a través de los respectivos departamentos de estadística, han desarrollado indicadores que resumen de manera general, en un sólo valor, el comportamiento de diversas variables económicas. Entre los indicadores más importantes se encuentran:

1. Índice de Laspeyres, denotado por L
2. Índice de Paasche, denotado por P
3. Índice ideal de Fisher, denotado por F

A continuación se muestran las fórmulas para su construcción.

Índice de Laspeyres:

$$\boxed{L = \frac{\sum_{i=1}^{n} P_n^{(i)} q_0^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_0^{(i)}}}$$

Donde:

$P_n^{(i)}$ = Precio del producto i en el año n.
$q_0^{(i)}$ = Cantidad producida del producto i en el año base.
$P_0^{(i)}$ = Precio del producto i en el año base.

Índice de Paasche.

$$\boxed{P = \frac{\sum_{i=1}^{n} P_n^{(i)} q_n^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_n^{(i)}}}$$

Donde:

$P_n^{(i)}$ = Precio del producto i en el año n.
$P_0^{(i)}$ = Precio del producto i en el año base.
$q_n^{(i)}$ = Cantidad producida del producto i en el año n.

Índice ideal de Fisher.

$$\boxed{F = \sqrt{\left[\frac{\sum_{i=1}^{n} P_n^{(i)} q_0^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_0^{(i)}}\right]\left[\frac{\sum_{i=1}^{n} P_n^{(i)} q_n^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_n^{(i)}}\right]}}$$

Donde:

$P_n^{(i)}$ = Precio del producto i en el año n.
$P_0^{(i)}$ = Precio del producto i en el año base.
$q_n^{(i)}$ = Cantidad producida del producto i en el año n.
$q_0^{(i)}$ = Cantidad producida del producto i en el año base.

Nótese que el índice de Paasche es una variante del índice de Laspeyres y, asimismo, el índice ideal de Fisher es una combinación de los dos primeros, por lo que su fórmula puede expresarse como:

$$\boxed{F = \sqrt{LP}}$$

Donde:
L=índice de Laspeyres
P=índice de Paasche

- **Ejemplo:** Considera la siguiente tabla de precios de productos y cantidades producidas de los mismos en diversos años:

Producto\Año (i)	2003	2004	2005	
Leche	8,75	9,12	9,85	
Mantequilla.	15,98	16,00	16,20	Precio en pesos
Queso.	16,35	16,95	17,00	
Producto\Año	2003	2004	2005	
Leche.	9675	9717	10436	
Mantequilla.	8117	8115	8302	Cantidades producidas.
Queso.	6700	6854	6901	

Con los datos anteriores, calcula los índices L, P y F para el año 2005

Solución

Se desean calcular los respectivos índices para el año 2005, por lo tanto, n=2005, asimismo, el año inicial consignado en la tabla es 2003, por lo que éste es el año base. Entonces:

Calculando el índice de Laspeyres:

$L = \frac{9{,}85(9675)+16{,}20(8117)+17(6700)}{8{,}75(9675)+15{,}98(8117)+16{,}35(6700)}$

$= \frac{340694{,}15}{323910{,}91}$

$= 1{,}0511814371$, que redondeado a tres decimales arroja el valor: 1,052

Calculando el índice de Paasche

$P = \frac{9{,}85(10436)+16{,}20(8302)+17(6901)}{8{,}75(10436)+15{,}98(8302)+16{,}35(6901)}$

$= \frac{354604}{336812{,}31}$

$= 1{,}052823752$, que redondeado a tres decimales arroja el valor: 1,053

Finalmente, se calcula el índice ideal de Fisher:
$F = \sqrt{1{,}052(1{,}053)}$
$= 1{,}052499881$, que redondeado a tres decimales, arroja un valor de 1,052.

Los índices calculados en el ejemplo anterior se utilizan en la construcción de otros índices a nivel gubernamental, como lo es el índice nacional de precios al consumidor (INPC).

Ejercicios 2

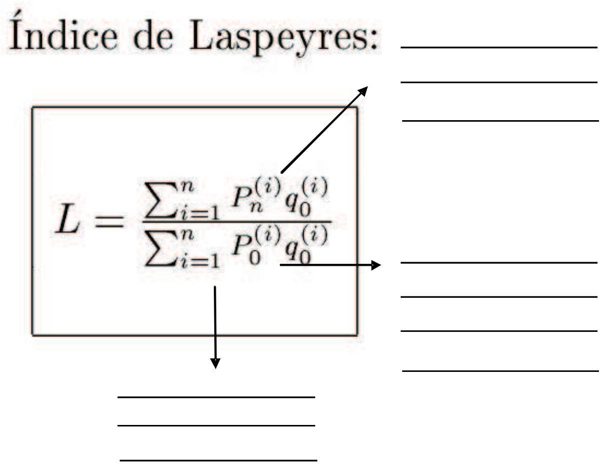
1. ¿Cómo se denomina a los indicadores que resumen de manera general, en un solo valor, el comportamiento de diversas variables económicas?
   ____

2. Coloca sobre la línea correspondiente, el nombre y la literal de esta expresión:

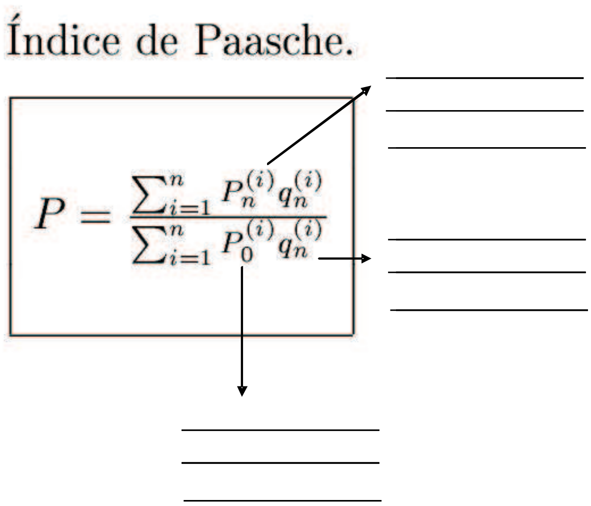
   ____

$$\boxed{\_\_\_\_\_ = \sqrt{\left[\frac{\sum_{i=1}^{n} P_n^{(i)} q_0^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_0^{(i)}}\right]\left[\frac{\sum_{i=1}^{n} P_n^{(i)} q_n^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_n^{(i)}}\right]}}$$

3. Anota en el esquema, el significado de las literales que intervienen en la expresión.



4. Anota en el esquema, el significado de las literales que intervienen en la expresión.



5. Anota la expresión corta del índice de Fisher

- Ahora confronta tus respuestas con las de la página 606

11.4. Índice de precios al consumidor

El idice ncional de pecios al consumidor (INPC) es un indicador económico de gran importancia, cuyo objetivo es la medición, a través del tiempo, de la variación de los precios de una canasta de bienes y servicios representativa del consumo de los hogares. Se considera al INPC como el instrumento estadístico por medio del cual se mide el fenómeno económico que se conoce como *inflación* (pérdida del valor adquisitivo del dinero o el crecimiento continuo y generalizado de los precios de los bienes y servicios que se expenden en una economía).

El Gobierno Federal, a través del Banco de México toma decisiones económicas en función de diversas variables, especialmente, respecto a la inflación. La medición de la inflación es un proceso complejo sobremanera.

- Por el número tan grande de precios que existe en una economía moderna.

- Por la necesidad de tener una cobertura lo más amplia posible de los gastos que realizan los agentes económicos.

- Porque los bienes y servicios se expenden a todo lo largo y ancho del territorio nacional.

- Porque los precios no cambian simultáneamente, ni evolucionan todo el tiempo al mismo ritmo.

Dado que es imposible valuar en su totalidad los precios de los bienes y servicios que se consumen, la construcción del INPC y sus cálculos se realizan con base en procedimientos muestrales.

La importancia del INPC radica en el hecho de que permite conocer cuál es la inflación promedio en el país durante un período específico y como se comentó líneas arriba, de acuerdo con el comportamiento del INPC, el Banco de México diseña la política monetaria orientada a procurar la estabilidad del poder adquisitivo de la moneda nacional; contratos de diversa índole, como alquileres de inmuebles y pensiones, se revisan con base en las variaciones del INPC.

Una de las razones fundamentales por las que se lleva a cabo una medición lo más precisa posible de la inflación es porque se trata de un fenómeno económico nocivo ya que:

- Daña la estabilidad del poder adquisitivo de la moneda nacional
- Afecta el crecimiento económico al hacer más riesgosos los proyectos de inversión
- Distorsiona las decisiones de consumo y ahorro
- Propicia una desigual distribución del ingreso
- Dificulta la intermediación financiera

De aquí la importancia de procurar que el país tenga una moneda sana. Los principales componentes del INPC se agrupan en ocho categorías, de acuerdo con la forma en que los consumidores distribuyen su gasto, a saber:

1. Alimentos, bebidas y tabaco
2. Ropa, calzado y accesorios
3. Vivienda
4. Muebles, aparatos y accesorios doméstico
5. Salud y cuidado personal
6. Transporte
7. Educación y esparcimiento
8. Otros servicios

Para la elaboración del INPC se hace un seguimiento continuo de los precios de productos específicos. Sin embargo, para fines de cálculo del INPC estos específicos se agrupan para formar conjuntos aproximadamente homogéneos de bienes y servicios que se denominan *genéricos*. Estos últimos, constituyen la menor unidad de ponderación dentro del INPC. En la práctica, cada mes se recopilan 170,000 cotizaciones de productos específicos que se agrupan en 313 conceptos genéricos.

- Por ejemplo: a) Genérico refrescos: Para conformar el genérico *refrescos*, se investiga permanentemente el precio de cada una de las marcas (específicos) con sus respectivas presentaciones (aluminio, cristal y plástico en sus diferentes tamaños).

La información para medir la importancia relativa de los genéricos dentro de la canasta de INPC se obtiene a partir de una encuesta que levanta el Instituto Nacional de Estadística Geografía e Informática INEGI, en los hogares y que tiene cobertura nacional. A este estudio se le conoce con el nombre de Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH). Así, el gasto que la ENIGH asocia con cada bien o servicio se compara contra el gasto total de las familias mexicanas. De esta comparación se obtiene un cociente mediante la siguiente fórmula:

$\omega =$Gasto en el bien i de todas las familias mexicanas, dividido entre el gasto total de las familias mexicanas

Estos cocientes indican la importancia relativa o el peso de cada satisfactor dentro de los gastos de las familias mexicanas. A su vez, cada uno de esos porcentajes tiene una correspondencia con los genéricos del INPC. Por tanto, los pesos así estimados determinan el impacto que tendrá un cambio en el precio de un genérico dentro del presupuesto familiar.

Como ejemplo, un aumento en el precio de la sal tiene menor incidencia sobre el presupuesto familiar que un aumento en el precio de los refrescos. Ello, aún cuando el de la sal fuese de 30 por ciento y el de los refrescos de tan sólo 5 por ciento. A cada uno de estos pesos relativos se le conoce como *ponderación*. Así, al conjunto de todos esos pesos relativos se le denomina *estructura de ponderaciones*.

La fórmula del cálculo del INPC es la de ponderaciones fijas de Laspeyres. Esta fórmula utiliza una canasta de artículos y una estructura de ponderaciones fijas, que representan los bienes y servicios producidos en el país en periodo base. La variación promedio de los precios de los bienes y servicios de la canasta del índice en el periodo (t) respecto al periodo base (0) se calcula mediante la fórmula de Laspeyres, revisada con anterioridad:

$$\boxed{L = \frac{\sum_{i=1}^{n} P_n^{(i)} q_0^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_0^{(i)}}}$$

Para garantizar la representatividad de los precios que intervienen en el cálculo del INPC se realiza una selección de fuentes de información en cada una de las 46 ciudades de cotización.

Estas fuentes son normalmente tiendas, comercios y prestadores de servicios ampliamente preferidos por los consumidores. Una vez elegidos los establecimientos se lleva a cabo una investigación de marcas y presentaciones para seleccionar los productos específicos de cuyo precio se hace un seguimiento recurrente.

La selección de los servicios se realiza con base en la opinión de los informantes que escogen a los más representativos. Los precios que se recaban pasan por un proceso de revisión y, si es necesario, de verificación. Así, se garantiza que las cotizaciones que intervienen en el cálculo del INPC son los precios vigentes en el mercado.

El sistema del INPC se integra de 46 ciudades y áreas metropolitanas agrupadas en siete regiones. A su vez, por su tamaño, las ciudades se clasifican en pequeñas, medianas y grandes. De este modo se calculan índices de precios para cada una de las siete regiones en que se divide el territorio nacional y para las 46 ciudades que conforman el sistema, así como para cada tamaño de localidad. Cabe señalar que al menos en una ciudad por estado se recoge información para el cálculo del INPC. De esta manera se asegura la representatividad espacial del INPC. Como se observa, para el cálculo del INPC intervienen técnicas de muestreo estadístico, tema que se ha revisado en su respectiva sección.

El Banco de México publica el nivel del INPC en el Diario Oficial de la Federación los días 10 y 25 de cada mes o, en su caso, el día hábil inmediato anterior. Un día previo a esta publicación, la información se difunde en la página electrónica de la Institución: www.banxico.org.mx.
Adicionalmente a la publicación quincenal del INPC, también se cuenta con una mensual, que se divulga el día 10 de cada mes. Cabe señalar que el dato del INPC mensual y del quincenal es un promedio del periodo respectivo.

11.5. Aplicación del INPC

En la práctica, el INPC permite al investigador social y de la comunicación, medir en términos porcentuales, la pérdida del valor adquisitivo del dinero.

- **Ejemplo**: Considerar el caso de un profesionista de la publicidad y la mercadotecnia. Suponiendo que esta persona ganaba en el año 2001 la cantidad de $144000,00 anuales y, que en el año 2004 ganaba $200400,00. Responder: ¿Aumentó o disminuyó el poder adquisitivo de su salario?

  Solución: Evidentemente la persona percibe un salario mayor en el año 2004 que en el año 2001; sin embargo, esta información no es suficiente para determinar si aumentó o disminuyó su poder adquisitivo. Para ello es necesario disponer del índice nacional de precios al consumidor, emitido por el Banco de México en los respectivos años a considerar (2001 y 2004). Del Diario Oficial de la Federación se tiene que el INPC para el año 2001 y 2004 fueron respectivamente de 0,14 y 0,21. Con la información anterior y el sueldo de la persona, se puede construir la siguiente tabla:

	2001	2004
Salario	\$144000,00	\$200400,00
INPC	0,14	0,21

De aquí pueden definirse los siguientes conceptos:
Año base = Año inicial a considerar en el análisis.
Año n = Año actual a considerar en el análisis.
I_0 = INPC del año base.
I_n = INPC del año n.
S_0 = Salario en el año base.
S_n = Salario en el año n.

Entonces, para determinar si se ha perdido el valor adquisitivo o no del dinero, primero se debe determinar el *salario ideal* denotado por S_I, que debería percibirse para no perder el poder adquisitivo del mismo. Éste se determina con la siguiente fórmula:

$$\boxed{S_I = \frac{I_n \cdot S_0}{I_0}}$$

Ahora, de la tabla se tiene entonces que:
$I_0 = 0{,}14$
$I_n = 0{,}21$
$S_0 = \$144000{,}00$

$S_n = \$200400{,}00$

Sustituyendo los valores en la respectiva fórmula:

$$S_I = \frac{0{,}21(144000)}{0{,}14} = \frac{30240}{0{,}14} = 216000$$

Es decir, el salario ideal, -el salario que debería percibirse para conservar el poder adquisitivo del dinero- es de \$216000,00 anuales. Como puede observarse, el salario percibido (\$200400,00) es menor que el salario ideal, lo que significa que se ha perdido el poder adquisitivo. Se debe entonces proceder a cuantificar dicha pérdida en términos porcentuales. Para ello, primero se calcula la pérdida en términos *brutos*:

$$\boxed{\text{Pérdida} = \text{Salario ideal } (S_I) \text{ - Salario Real } (S_I)}$$

En nuestro caso:

$$\text{Pérdida} = \$216000{,}00 \text{ - } \$200400{,}00 = \$15600{,}00$$

Una vez hecho lo anterior, se expresa dicha pérdida en términos porcentuales, dividiendo la pérdida entre el salario ideal y multiplicando por cien:

$$\text{Pérdida porcentual}{=}\frac{1560}{200400}100 = 0{,}072222222(100) = 7{,}2222222$$

Redondeando a dos decimales, se tiene que el salario de la persona perdió su poder adquisitivo en el año 2004 en un 7.22 %

Ejercicios 3

1. ¿Qué significan las siglas INPC?

   ____

2. ¿Cuál es el objetivo del INPC?

   ____

   ____

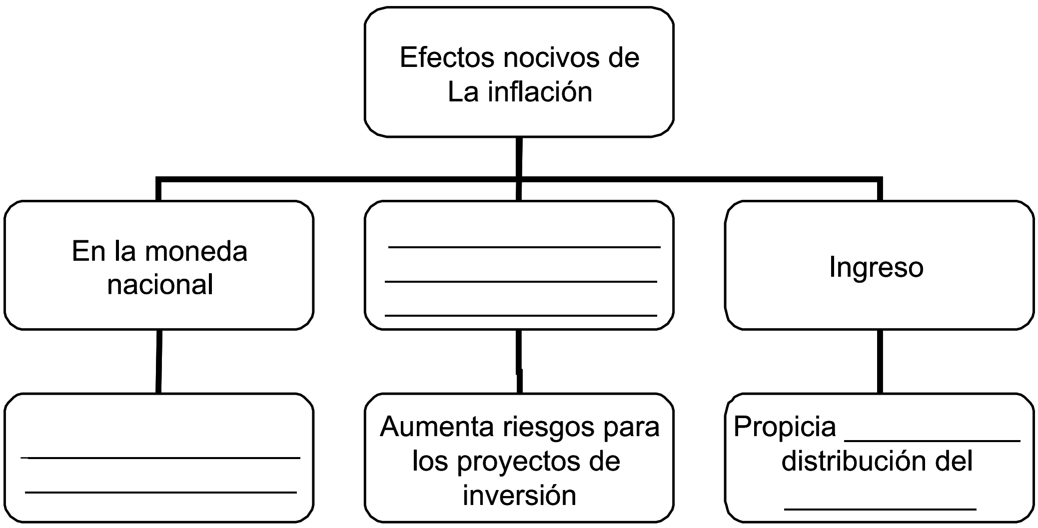
3. ¿Qué mide el INPC?

   ____

4. ¿Con base en qué se construye el INPC?

   ____

5. Complementa adecuadamente la red conceptual.



6. Enlista, por lo menos, tres categorías en las que se agrupan los principales componentes del INPC

   ____

   ____

   ____

   ____

7. ¿Qué son los genéricos en la construcción del INPC?

   ____

   ____

8. ¿Con qué fórmula se calcula el INPC?

   ____

   ____

9. Coloca el significado de las literales que intervienen para el cálculo del salario ideal.

   a) $I_0 =$ ____________________

   b) $I_n =$ ____________________

   c) $S_0 =$ ____________________

   d) $S_n =$ ____________________

10. Escribe la fórmula para determinar el salario ideal.

    ____

- Las respuestas a esta serie de ejercicios se encuentran en la página 607

11.6. Autoevaluación

1. Considera la siguiente tabla de precios de productos y cantidades producidas de los mismos en diversos años:

Producto\Año (i)	2003	2004	2005	
Leche	8,75	9,12	9,85	
Mantequilla.	15,98	16,00	16,20	Precio en pesos
Queso.	16,35	16,95	17,00	
Producto\Año	2003	2004	2005	
Leche.	9675	9717	10436	
Mantequilla.	8117	8115	8302	Cantidades producidas.
Queso.	6700	6854	6901	

Con los datos anteriores, calcula L, P y F, para el 2004.

a) $L =$

b) $P =$

c) $F =$

2. Considera el caso de un obrero de la industria textil. Suponiendo que éste, ganaba \$2107 quincenales en el 2001, y que para el 2004 tenía un sueldo de \$2507. Responder: Aumentó o disminuyó el poder adquisitivo de su salario?
   Datos:

	2001	2004
Salario	\$2107,00	\$2507,00
INPC	0,14	0,21

   Fórmula

   ____

   ____

   ____

- Respuestas a la autoevaluación en la página 609

11.7. Apéndice de la unidad

11.7.1. Respuestas a los ejercicios de la unidad

Ejercicios 1

1. Calcula e interpreta los siguientes precios relativos.

   a) Supóngase que el hacia enero de 1998, el precio del dólar norteamericano (al menudeo) era de \$10,05 y que para diciembre del mismo año, era de \$12,90. Calcula el precio relativo de la divisa Estadounidense, tomando como periodo de referencia el mes de enero del mismo año.

   Solución:
   Del planteamiento se tienen los siguientes datos:
   $a =$ Diciembre
   $b =$ Enero
   $P_a = \$12{,}9$
   $P_b = \$10{,}05$

   Sustituyendo los datos en la fórmula correspondiente:
   $P_{Diciembre|Enero} = \frac{P_{diciembre}}{P_{Enero}}$
   $= \frac{\$12{,}90}{10{,}05}$
   $= 1{,}28$
   $1{,}28(100\,\%) = 128\,\%$

   Esto significa que el precio del dólar en Diciembre, se incrementó en un 28 % respecto al inicio del año (Enero de 1998).

   b) Supóngase que en la región sureste de la República Mexicana, un jornalero percibe, en promedio \$39,50 al día; mientras que en el norte del país, un obrero con el salario mínimo tiene como ingreso diario \$78,50 por concepto de salario mínimo. Calcula el precio relativo del día laboral, tomando como base, el salario del trabajador Norteño.

Solución:
Del planteamiento se tienen los siguientes datos:
$a =$ Sur
$b =$ Norte
$P_a = \$39{,}50$
$P_b = \$78{,}50$

Sustituyendo los datos en la fórmula correspondiente:
$P_{Sur|Norte} = \frac{P_{Sur}}{P_{Norte}}$
$= \frac{\$39{,}50}{\$78{,}50}$
$= 0{,}503184713$
$= 0{,}503184713(100\,\%) = 50{,}3184713\,\% \approx 50$

Esto significa que el trabajador del norte gana un 50 % más que el trabajador sureño.

2. Señala qué propiedad de los precios relativos se verifica en los siguientes casos o, en su defecto, anota su expresión formal.

    a) Propiedad de identidad: $P_{a|a} = 1$

    b) Propiedad del tiempo inverso: $P_{a|b} \cdot P_{b|a} = 1$

    c) $P_{a|b} \cdot P_{b|c} \cdot P_{c|a} = 1$ Propiedad cíclica o circular

    d) $P_{a|b} \cdot P_{b|c} = P_{a|c}$ Propiedad cíclica modificada:

Ejercicios 2

1. ¿Cómo se denomina a los indicadores que resumen de manera general, en un solo valor, el comportamiento de diversas variables económicas? Números Índice.

2. Coloca sobre la línea correspondiente, el nombre y la literal de esta expresión:

Índice Ideal de Fisher.

$$\boxed{F = \sqrt{\left[\frac{\sum_{i=1}^{n} P_n^{(i)} q_0^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_0^{(i)}}\right]\left[\frac{\sum_{i=1}^{n} P_n^{(i)} q_n^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_n^{(i)}}\right]}}$$

3. Anota en el esquema, el significado de las literales que intervienen en la expresión.

Índice de Laspeyres:

Precio del producto *i* en el año *n*

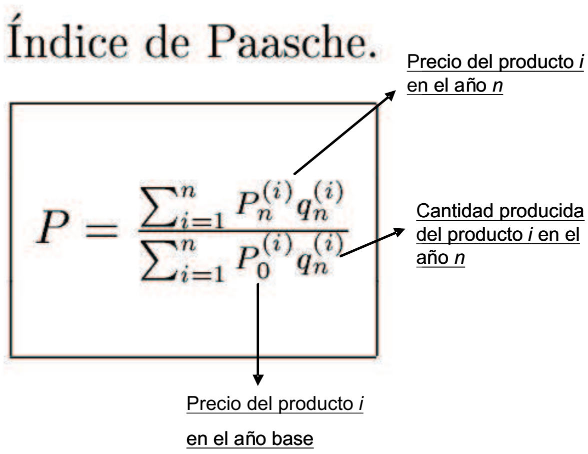
$$\boxed{L = \frac{\sum_{i=1}^{n} P_n^{(i)} q_0^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_0^{(i)}}}$$

Cantidad producida del producto *i* en el año base.

Precio del producto *i*

en el año base.

4. Anota en el esquema, el significado de las literales que intervienen en la expresión.



5. Anota la expresión corta del índice de Fisher
   $F = \sqrt{LP}$

Ejercicios 3

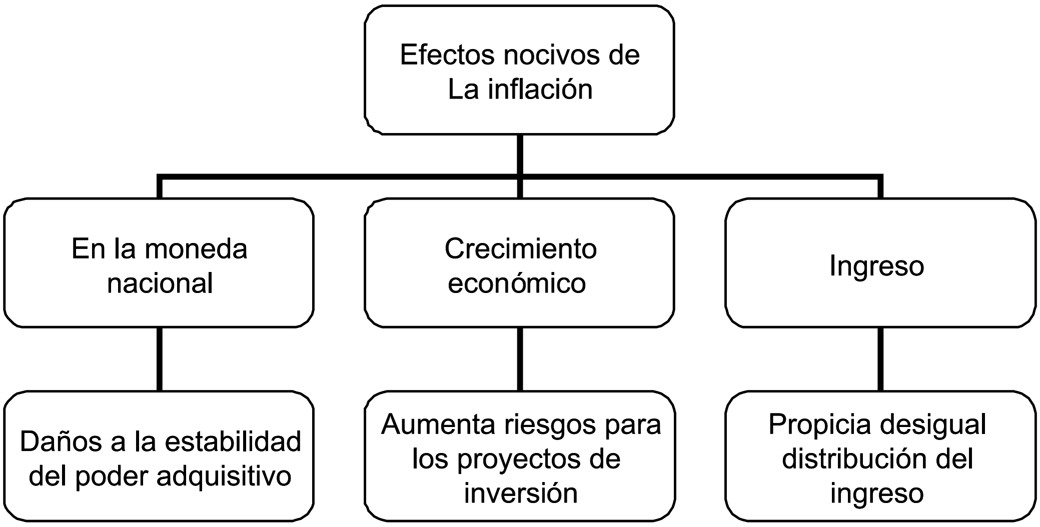
1. ¿Qué significan las siglas INPC?
   índice nacional de precios al consumidor

2. ¿Cuál es el objetivo del INPC?
   Medir, a través del tiempo, la variación de los precios de una canasta de bienes y servicios representativa del consumo de los hogares.

3. ¿Qué mide el INPC?
   El fenómeno económico llamado Inflación.

4. ¿Con base en qué se construye el INPC?
   Procedimientos muestrales.

5. Complementa adecuadamente la red conceptual.



6. Enlista, por lo menos, tres categorías en las que se agrupan los principales componentes del INPC.

   a) Alimentos, bebidas y tabaco

   b) Ropa, calzado y accesorios

   c) Vivienda

   d) Muebles, aparatos y accesorios doméstico

   e) Salud y cuidado personal

   f) Transporte

   g) Educación y esparcimiento

   h) Otros servicios

7. ¿Qué son los genéricos en la construcción del INPC?
   Conjuntos aproximadamente homogéneos de bienes y servicios que se denominan, que constituyen la menor unidad de ponderación dentro del INPC

8. ¿Con qué fórmula se calcula el INPC?
   La de ponderaciones fijas de Laspeyres, que es $L = \frac{\sum_{i=1}^{n} P_n^{(i)} q_0^{(i)}}{\sum_{i=1}^{n} P_0^{(i)} q_0^{(i)}}$

9. Coloca el significado de las literales que intervienen para el cálculo del salario ideal.

   a) I_0 = INPC del año base.

   b) I_n = INPC del año n.

   c) S_0 = Salario en el año base.

   d) S_n = Salario en el año n.

10. Escribe la fórmula para determinar el salario ideal.
    $S_I = \frac{I_n \cdot S_0}{I_0}$

11.7.2. Respuestas de la autoevaluación

1. Considera la siguiente tabla...
   ...
   Con los datos anteriores, calcula L, P y F, para el 2004.
   Solución:
   Se desean calcular los respectivos índices para el año 2004, por lo tanto, n=2004, asimismo, el año inicial consignado en la tabla es 2003, por lo que éste es el año base. Entonces:
   Calculando el índice de Laspeyres:

   $L = \frac{9{,}12(9675)+16(8117)+16{,}95(6700)}{8{,}75(9675)+15{,}98(8117)+16{,}35(6700)} = \frac{331673}{323910{,}91}$

$= 1{,}23963657$, que redondeado a tres decimales arroja el valor: 1,24
Calculando el índice de Paasche

$P = \frac{9{,}12(9717)+16(8115)+16{,}95(6854)}{8{,}75(7917)+15{,}98(8115)+16{,}35(6854)}$

$= \frac{334634{,}34}{326764{,}35}$

$= 1{,}024084604$, que redondeado a tres decimales arroja el valor: 1,024

Finalmente, se calcula el índice ideal de Fisher:
$F = \sqrt{1{,}024(1{,}024)}$
$= 1{,}024$, que redondeado a tres decimales, arroja un valor de 1,052.

a) $L = 1{,}024$

b) $P = 1{,}024$

c) $F = 1{,}024$

2. Considera el caso de un obrero de la industria textil. Suponiendo que éste, ganaba $2107 quincenales en el 2001, y que para el 2004 tenía un sueldo de $2507. Responder: Aumentó o disminuyó el poder adquisitivo de su salario?
Datos:

	2001	2004
Salario	$2107,00	$2507,00
INPC	0,14	0,21

Fórmula:
$S_I = \frac{I_n \cdot S_0}{I_0}$

Sustitución: Ahora, de la tabla se tiene entonces que:

$I_0 = 0{,}14$

$I_n = 0{,}21$

$S_0 = \$2107{,}00$

$S_n = \$2507{,}00$

$$S_I = \frac{0{,}21(2107)}{0{,}14} = \frac{44247}{0{,}14} = 3160{,}50$$

Si $S_I = 3160{,}50$ y gana 2507, entonces la pérdida del poder adquisitivo para su salario, viene dada por $3160{,}50 - 2507 = 653{,}5$ que expresada en términos porcentuales es igual a 20.68 %

Capítulo 12

Conclusiones

En la presente investigación se ofrecen cuatro tipos de conclusiones: Teóricas, metodológicas, técnicas y temáticas.

La Teoría Cognoscitivista aportó los elementos necesarios para comprender las implicaciones psicológicas intrínsecas en el proceso de aprendizaje y por tanto necesarias para el quehacer docente. Como ya se mencionó, en el proceso de aprendizaje ni se parte de cero, ni los estudiantes lo hacen desde el mismo punto.

Las interrelaciones de los cambios de los elementos tanto al interior como al exterior de la macro estructura llamada medio ambiente, afectan en distintas formas y grados a la subestructura o ente cognoscente (el sujeto). Dichas afectaciones, inciden en la formación de las microestructuras que determinan las características de la información que puede o no ser asimilada en determinado momento. Por ello es fundamental, en el proceso formal de aprendizaje (en la escuela), buscar la formación de estructuras que le faciliten la asimilación de las nuevas ideas expresadas simbólicamente.

Lo anterior, se logrará con la propuesta y aplicación de estrategias que incluyan modelos con formas similares a las microestructuras anteriormente enunciadas. Con el uso de las redes conceptuales, se incorpora dicho modelo que contiene la información necesaria para asimilar adecuada y ordenadamente las nuevas ideas a la estructura mental, de forma que no sean aprendidos los símbolos de-forma textual (al pie de la letra), sino de manera sustancial y no arbitraria; es decir, afianzar la posesión de nuevos conocimientos mediante relaciones abstractas con aspectos relevantes a la vida académica y profesional del estudiante-profesional de la Comunicación.

En cuanto a la teoría estadística desarrollada y sistematizada en el presente trabajo, se puede afirmar que pese y debido a su enfoque curricular no se debe ni puede limitar al ámbito de la investigación en comunicación. Una de las características de la Estadística es su independencia y adaptabilidad; es decir, no se puede "privatizar" a la Estadística para la Investigación en Comunicación, como tampoco para la Economía, o la Física –por citar algunos casos–, pues la estadística formalmente "no tiene apellidos" pero sí aplicaciones específicas y elementos útiles para diferentes ciencias y disciplinas. En todo caso, el presente material didáctico, puede fungir como guía no sólo para el estudio de la Comunicación, sino para el abordaje cuantitativo de diversos fenómenos sociales.

El apropiado estudio de estadística, brindará al alumno una importante herramienta formal que fortalecerá su formación crítica, metodológica y epistemológica. Esto permitirá al estudiante la aplicación de los conocimientos necesarios para la resolución de problemas planteados en diversas asignaturas que cursará como parte de su formación metodológica. Independientemente de que el estudiante se dedique a los géneros periodísticos, la comunicación organizacional, persuasiva o la investigación y docencia, los conocimientos en estadística le darán solidez metodológica para el futuro ejercicio de su profesión. En el caso del periodismo impreso o vía los medios electrónicos, el estudiante-profesional de la comunicación poseerá grandes ventajas en el campo de trabajo y el mercado laboral, dado que su formación le dará la posibilidad de ejercer un periodismo de profundidad, el llamado periodismo de investigación y/o de precisión. Para el trabajo del comunicólogo en las organizaciones, el bagaje formal del presente trabajo, le permitirá tomar decisiones importantes con fundamentos científicos.

Metodológicamente hablando, el material está sistematizado de tal manera que facilite el proceso de aprendizaje. La eficiencia del material, no es (totalmente) responsabilidad de quien lo diseño, pues su estructura no implica que el estudiante no deba esforzarse en comprender los contenidos de la asignatura. Cada sección contiene las estrategias que se consideran suficientes para la interacción con el estudiante; pero sin hábitos de estudio apropiados, el fracaso es más que una posibilidad. El esfuerzo no sólo de debe pensar en el términos de lectura y resolución de ejercicios; los apuntes personales y la innovación, serán el plus que el estudiante deba conferir a su proceso de aprendizaje para acceder al éxito deseado. Cabe aclarar que por innovación se debe entender el proceso mediante el cual el estudiante, una vez que ha comprendido algún tema estudiado con ayuda del texto, debe ejercitarlo mediante la búsqueda y análisis de situaciones reales (que a éste le interesen), para su estudio mediante la aplicación del punto en cuestión.

Otro elemento pertinente a la metodología que envuelve aspectos técnicos del Material Didáctico, es su perfectibilidad. Si bien es cierto que dicho material hasta el momento se considera concluido en cuanto a su diseño y forma; también lo es el hecho de que su mejora es una obligación, y una parte importante para ello es su aplicación experimental en el ámbito académico para el cual fue diseñado. Su amabilidad para con el público al que se dirige, es ideal en tanto que éste no lo ha probado; su aplicación debe conllevar los cuestionamientos críticos de la comunidad universitaria. La formación de un grupo piloto y otro de control, permitiría –en primera instancia– conseguir la esperada crítica e incluso propuestas, así como la verificación de su eficacia y potencial en el estado que el material didáctico actualmente guarda.

Como conclusiones temáticas se tiene lo siguiente: El presente material didáctico obedece a la lógica institucional plasmada en el Programa de la asignatura. Se respetó en la medida de lo posible tal planteamiento, pero no por ello el contenido se limitó a la propuesta curricular y quien sustenta (en colaboración con su asesor) propuso algunos elementos extracurriculares. El primero de ellos se ubica en el capítulo inicial; en éste, dentro de la evaluación diagnóstica se encuentran reactivos concernientes a la teoría de sistemas. Con seguridad esto causará cierto recelo, pues dicho tema es propio de asignaturas superiores en la estructura curricular como son las correspondientes a la Teoría de la Comunicación, que se comienzan a cursar a partir de 3^{er} semestre. Sin embargo, ello es producto del planteamiento *teórico-metodológico* (y sus implicaciones antes mencionadas); pues finalmente lo que se pretende analizara a partir de la metodología cuantitativa son, principalmente, fenómenos acaecidos al interior del sistema social que a su vez, pueden erigirse como subsistemas entre los que se encuentra la comunicación humana.

En el segundo capítulo, puesto que la asignatura –y por tanto el aprendizaje del contenido allí propuesto– requiere de una aceptable formación matemática, se propuso como un curso propedéutico para recordar y aprender los conceptos y elementos matemáticos que se consideran necesarios para el estudio ininterrumpido de la Estadística. Si el estudiante no posee los suficientes elementos matemáticos para el estudio y aprendizaje de la estadística, el quehacer comprensivo se dificultará y el aprendizaje significativo no se dará.

En este tenor, en el programa se contemplan elementos de Estadística Descriptiva que son de gran utilidad, pero éstos deben ser complementados con la inclusión, de temas fundamentales para el desarrollo de una investigación confiable y la correcta interpretación de la información, es decir, con temas de Aritmética y Estadística Inferencial. Los temas que, desde el punto de vista del sustentante, deben ser manejados por los estudiantes de la licenciatura son: Tasas, Razones, Proporciones y Variación Porcentual (incluidos en el capítulo dos), debido a la frecuencia con que se presentan en el ámbito de la comunicación tanto en medios de comunicación masiva como en el campo de la investigación social.

Con el ánimo de respetar en la mayor medida el orden de los temas, sugerido por el programa de la asignatura, en los capítulos 3, 4 y 5 se encuentran los elementos propuestos en los puntos 2, 3 y 4 del temario. Sin embargo, fue preciso sustituir el tema 5 del temario (Números Índice) por el que se colocó en el capítulo 6 (Momentos), ya que éste complementa el análisis de los datos pertinentes a una muestra obtenida en forma aleatoria. De colocar el estudio de los Momentos en un capítulo posterior a los planteados por el temario, se fracturaría la continuidad en el aprendizaje de los elementos con que se pueden estudiar los datos de una muestra. Así el tema correspondiente a los Números Índice, tuvo que ser trasladado a la parte final del texto.

Otros elementos teóricos de importancia en el capítulo 6, son los correspondientes a los temas de sesgo y Curtosis. Estas medidas permitirán analizar el comportamiento de los datos de una muestra, determinando la presencia de sesgo, cuantificando el mismo y analizando sus posibles fuentes. Asimismo, la determinación del coeficiente de curtosis permitirá al estudiante analizar la forma de de la distribución de un conjunto de datos muestrales y le proporcionará un indicador que le permitirá interpretar y comparar la forma de diversas distribuciones de datos.

Otras de las propuestas temáticas para el programa, son las concernientes a los capítulos 8 y 9 *Introducción a la Probabilidad y Distribución Normal*, respectivamente. Sin ellas, el tema de Muestreo (capítulo 10). Corre el riesgo de presentarse como algo difícil, y de entenderse parcialmente, pues en dichos temas se encuentran los elementos *fundamentales* de un correcto diseño y elección de la muestra y sus unidades. Por tanto, para lograr una sólida formación metodológica, es indispensable la comprensión de la teoría de la probabilidad y la distribución normal. Así, el tema de distribución normal, considerando como subtema en el curso de estadística, se debe abordar como un tema al cual es preciso dedicar más tiempo y esfuerzos que los dedicados a un subtema, pues dará al estudiante un sustento teórico para desarrollar el importante tema del muestreo, siendo éste, uno de los de mayor utilidad en el campo de la investigación en Comunicación.

Finalmente, cabe enfatizar que la propuesta curricular de la asignatura, se enfoca principalmente a la estadística descriptiva, lo cual es de gran utilidad para el estudio de los fenómenos que interesan a los comunicólogos; no obstante, debido a limitaciones en los recursos materiales, se tiene que incluir sólo dos conceptos de estadística inferencial (Regresión y Correlación simples), dejando fuera la mayor parte de esta rama de la estadística, especialmente los estimadores y las pruebas de hipótesis. Este punto debe considerarse especialmente, ya que, como se sabe bien, la estadística inferencial es la que permite general los resultados del estudio de una muestra al total de la población, para el importante proceso de toma de decisiones. Si bien tales cuestiones no son subsanadas con la propuesta del presente material didáctico, éste sí permite al estudiante adquirir y preservar en un soporte físico los elementos necesarios, para la posterior comprensión de los elementos teóricos más importantes de la estadística inferencial.

Bibliografía

[1] ABBAGNANO, Nicolo; **Diccionario de Filosofía**; México; Fondo de Cultura Económica, 2^a edición; 1974; pp. 1180

[2] AGUILAR, Vieyra Griselda; **Cuadernos de Educación y Comunicación. Un Modelo General de Diseño de Investigación, en Comunicación (Propuesta)**; México; Junio de 1999, número 4, Volumen 1 año 2

[3] ANDER-EGG, Ezequiel, **Técnicas de Investigación Social**; Argentina; Editorial Hvmanitas, 21; 1990; pp. 424

[4] ÁVILA, Raúl; **La Lengua y los Hablantes**; México; Editorial Trillas, 3^a edición; 1997; pp. 157

[5] BALDOR, Aurelio; **Álgebra**; Madrid; 1981; Ediciones y Distribuciones Códice, S. A. Madrid; pp. 575

[6] BUNGE, Mario; **La Ciencia, su Método y su Filosofía**; Buenos Aires; Ediciones Siglo Veinte; 1975; pp. 125

[7] BUNGE, Mario; **La Investigación Científica**; México; Siglo XXI Editores; 2000; pp. 805

[8] CLAPHAM, Christopher; **Diccionario de Matemátias**; Madrid; Editorial Complutense, Diccionarios Oxford-Complutense; 1998; pp. 403

[9] CEA, D' Ancona María Ángeles; **Metodología Cuantitativa. Estrategias y Técnicas de Investigación Social**; Madrid; Proyecto Editorial Síntesis Sociológica; 1998; pp. 415

[10] COLMENARES, Ismael; et. al. **De la Prehistoria a la Historia**; México; Ediciones Quinto Sol, 4^a edición; 1994; pp. 511

[11] DEL CAMPO, Salustiano (Director); **Diccionario de las Ciencias Sociales. Vol. 1**; Madrid; Instituto de Estudios Políticos; 1975; pp. 1186

[12] FERRATER, Mora José; **Diccionario de Filosofía Vol 1**; Argentina; Editorial Sudamericana Buenos Aires, 5^a edición; 1975; pp. 1005

[13] FERRATER, Mora José; **Diccionario de Filosofía Vol 2**; Argentina; Editorial Sudamericana Buenos Aires, 5^a edición; 1975; pp. 1072

[14] FRAGOSO, Almaraz Cruz Elvira **Opinión Pública del personal de *Mexalite Industrial S. A. de C. V.* Planta Santa Clara, Sobre el Sistema de Calidad**; Seminario - Taller extracurricular; México; UNAM-ENEP Acatlán; 2001; pp. 173

[15] GORDON, Scott; **Historia y Filosofía de las Ciencias Sociales**; Barcelona; Ariel Referencia; 1995; pp. 731

[16] GUTHRIE, William, K. C; **Los Filósofos Griegos. De Tales a Aristóteles**; México; Fondo de Cultura Económica, 2^a edición; 1995; pp. 189

[17] GUTIERREZ, Saenz Raúl; **Introducción a la Lógica**; México; Editorial Esfinge; 1997

[18] HERNÁNDEZ, Sampieri; et. al; **Metodología de la Investigación**; México; McGraw Hill; 2002. pp. 487

[19] JESSEN, Johan; **Teoría del Conocimiento**; México; Editores Unidos Mexicanos; 1999; pp. 183

[20] JOSÉ, Sametband Moisés; **Entre el Orden y el Caos. La Complejidad**; México; FCE, SEP, CONACYT, 2^a edición; 1999; pp. 161

[21] KIESS, Harold O.; **Statistical Concepts for the Behavioral Sciencies**; Allyn and Cacon Inc.; 1989; pp. 673

[22] KLAUSMEIER, Herbert J.; et. al.; **Psicología Educativa**; México; Editorial Harla; 1971; pp. 527

[23] KLEIMAN, Ariel; Et. al. **Conjuntos. Aplicaciones Matemáticas a la Administración**; México; Editorial Limusa, Biblioteca Didáctica de Matemáticas; 1975; pp. 197

[24] LEHMANN, H. Charles; **Geometría Analítica**; México; Editorial Limusa; 1980; pp. 494

[25] LOHR, Sharon; **Muestreo: Diseño y Análisis**; México; Thomson Learning, 2000

[26] MARTÍN, Serrano Manuel; et. al. **Teoría de la Comunicación I. Epistemología y Análisis de la Referencia**; México; Universidad Nacional Autónoma de México, 2^a edición; 1993; pp. 228

[27] MARTÍNEZ, Sergio F.; **De los Efectos a las Causas. Sobre la Historia de los Patrones de Explicación Científica**; Paidós; México; 1997; pp. 189

[28] MATTELART, Armand et. al.; **Historia de las Teorías de la Comunicación**; Barcelona; Paidós; 1997; pp. 142

[29] MOORE, David S.; en ARTHUR, Steen Lynn (editor); **La Enseñanza Agradable de las Matemáticas**; México; Limusa Noriega Editores, Colección Textos Politécnicos, Serie Matemáticas; 1998; pp. 241

[30] PADUA, Jorge. **Técnicas de Investigación Aplicadas a las Ciencias Sociales**; México; El Colegio de México, Fondo de Cultura Económica; 2002; pp. 360

[31] PEÑA de la, Ricardo; et. al.; **Cómo Acercarse a la Sociología**; México; Limusa Noriega; 1991; pp. 126

[32] PEÑA, Daniel; et. al. **Introducción a la Estadística para Ciencias Sociales**; Madrid; McGraw-Hill/Interamericana, S. A. V.; 1999; pp. 428

[33] PESTALOZZI; Juan Enrique; **El Canto del Cisne**; México; 1996; Editorial Porrua; pp. 162

[34] PONTÓN, Gonzalo (Director); **Grijalbo Diccionario Enciclopédico**; Madrid; Grijalbo, 1986; pp. 2601

[35] REVILLA, Basurto Mario A. **Introducción a la Teoría de la Comunicación**; México ; S y G Editores; 1997; pp. 70

[36] ROSAS, y Novelo María de Lourdes; **La Ciencia y el Método Experimental**; México; Universidad Nacional Autónoma de México; 1997; pp. 175

[37] SEBEOK, Thomas A. **Signos: una Introducción a la Semiótica**; México; Paidós; 1996; pp. 163

[38] SILLS, L. David; **Enciclopedia Internacional de las Ciencias Sociales**; Tomo 4; Madrid; 1974; Aguilar S. A. de Ediciones; pp. 801

[39] SPIEGEL, Murray R.; **Estadística**; México; McGrawg Gil, Serie Schaum; 1970; pp. 357

[40] STEWART, Ian; **¿Juega Dios a los Dados? La Nueva Matemática del Caos**; Barcelona; Crítica; 2001; pp. 443

[41] SUPPES, Patrick; et. al; **Introducción a la Lógica Matemática**; México; Editorial Reverté 4^a edición; 1997; pp. 278

[42] SZASZ, Ivonne; et. al.; **Para Comprender la Subjetividad: Investigación Cualitativa en Salud Reproductiva y Sexualidad**; México; El Colegio de México; 1999; pp. 256

[43] WILLOUGHBY, Stephen S; **Probabilidad y Estadística**; México; Publicaciones Cultural S. A.; 1977; pp. 215

[44] XIRAU, Ramón; **Introducción a la Historia de la Filosofía**; México; Universidad Nacional Autónoma de México, 12^a edición; 1995; pp. 497

Otras Fuentes

[45] Biblioteca de Consulta Microsoft Encarta 2005. © 1993-2004 Microsoft Corporation.

[46] CRUCIANELLI, Sandra; **Curso Matemáticas para Periodistas. Apuntes Electrónicos**; Texas University 2005

[47] http://www.mor.itesm.mx/~logica/logyprob/page3.html; Jueves 29 de septiembre de 2005; 14:40 PM

[48] ROCÍO DEL, María Ávila; **Conferencia: La Investigación sobre Infraestructura Física Educativa**; México; FES-Acatlán; Octubre de 2005

[49] ROMO, Martínez Socorro Leticia; **Matemáticas 1**; en *http ://www.universidadabierta.edu.mx/SerEst/Apuntes/Romo Socorro_MateI.doc*; México Ciudad Bugambilias, Zapopan, Jal.; 2003; pp. 4; Consultada: 26/08/2005; 17:23 PM

[50] ULLOA, Arellano Víctor Manuel; **Apuntes del Curso 2003 - 2**; Universidad Nacional Autónoma de México - Facultad de Estudios Superiores Acatlán; 2003