27
2 ej.

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

ESTADISTICA A NIVEL MEDIO SUPERIOR

EN EL COLEGIO DE CIENCIAS Y HUMANIDADES, UNAM

T E S I S

QUE PARA OBTENER EL TITULO DE

MATEMATICO

P R E S E N T A

GUILLERMO A. PEREZ URIBE

M. EN C. JOSE ANTONIO FLORES DIAZ
ACT. MARIA DEL PILAR ALONSO REYES
BIO. BENJAMIN ALVAREZ RUBIO
FIS. GENOVEVA OLGUIN RAMIREZ
DRA. HORTENSIA GALEANA SANCHEZ

MEXICO, D.F. JULIO DE 1994

CIUDAD UNIVERSITARIA

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS
División de Estudios Profesionales
Exp. Núm. 55

M. EN C. VIRGINIA ABRIN BATULE
Jefe de la División de Estudios Profesionales
Universidad Nacional Autónoma de México.
P r e s e n t e .

Por medio de la presente, nos permitimos informar a Usted, que habiendo revisado el trabajo de tesis que realizó el pasante PEREZ URIBE GUILLERMO ARMANDO

con número de cuenta 7119275-6 con el título:
"ESTADISTICA A NIVEL MEDIO SUPERIOR EN EL COLEGIO DE CIENCIAS Y HUMANIDADES"

Consideramos que reune los méritos necesarios para que pueda continuar el trámite de su Examen Profesional para obtener el título de - MATEMATICO.

GRADO NOMBRE Y APELLIDOS COMPLETOS	FIRMA
M. EN C. JOSE ANTONIO FLORES DIAZ Director de Tesis	
ACT. MARIA DEL PILAR ALONSO REYES	
BIO. BENJAMIN ALVAREZ RUBIO	
FIS. GENOVEVA OLGUIN RAMIREZ Suplente	
DR. HORTENCIA GALEANA SANCHEZ Suplente	

Ciudad Universitaria, D.F., a 19 de ABRIL de 1994

ESTADISTICA
A NIVEL MEDIO SUPERIOR

EN EL COLEGIO DE CIENCIAS Y HUMANIDADES,UNAM

INDICE

DEDICO MI TRABAJO

A MI MADRE

INTRODUCCION

En la actualidad el mundo moderno exige la aplicación de distintas ramas de la ciencia para el mejor entendimiento y desarrollo del ser humano en general. Es por esto que a medida que las relaciones entre ellas se vuelven más vastas y complejas, es necesario emplear con mayor cuidado las distintas técnicas que la investigación científica ofrece para la comprensión de los fenómenos o experimentos que se estudien.

En particular la estadística es una rama de la ciencia que se emplea en áreas de investigación, económico-administrativas, poblacionales, etc.; por mencionar algunas aplicaciones, ésta se emplea en el levantamiento de censos de población cuyo valor es demasiado alto para el desarrollo de las naciones es muy utilizada en la economía y las finanzas de los países para observar, por ejemplo, el avance de los indicadores económicos y financieros. En toda empresa tanto gubernamental como de la industria privada, se utiliza la estadística en las áreas de producción y financieras para poder evaluar los avances de la misma.

Desde la perspectiva científica se emplea en áreas tan prioritarias como la medicina, la biología, la veterinaria, etc. En la medicina, por mencionar algunas aplicaciones, se utiliza para determinar el avance de enfermedades en una población, en la investigación de la génetica, en la evaluación de algún medicamento, etc.; en biólogia para evaluar la extinción de flora y fauna en una región determinada, para la producción de alimentos, en el mejoramiento de procesos de producción agrícolas, etc.; en veterinaria para tratar de optimizar los métodos de producción de animales de uso comercial, para calificar los procesos de alimentación de ganado, etc.

En resumen los campos de aplicación de la estadística pueden ser tan vastos y complejos como lo determinen las personas que la empleen.

OBJETIVO

El presente trabajo pretende proporcionar una guía que sirva de apoyo en la materia optativa de estadística a los alumnos del nivel medio superior del Colegio de Ciencias y Humanidades de la UNAM en los dos últimos semestres de acuerdo al plan de estudios vigente en dicha institución*. Debido a que cada plantel respeta un esquema general del temario de dicha materia pero le da su propia perspectiva, este trabajo pretende dentro de lo posible respetar el plan de estudios propuesto.

ESTRATEGIA

Para lograr el objetivo del presente trabajo se requieren los siguientes elementos:

En primer término considerar el temario del plan de estudios de la materia de estadística aprobado por la academia de matemáticas del plantel Oriente y respetar en lo más posible, en el desarrollo del trabajo, dicho temario.

En segundo término la evaluación práctica del material con un grupo piloto de estadística en el ejercicio cotidiano de la docencia en dichc plantel. Con el objeto de medir el aprendizaje del material por parte de los estudiantes, se establecerán mecanismos de evaluación cuyo análisis estadístico aunque sea básico proporcionará un resultado concreto.

En tercer término se proponen actividades complementarias para los estudiantes como investigaciones dentro de su entorno social, investigaciones en publicaciones de difusión general, etc.; y al final de cada tema se propone una evaluación de los conceptos así como las soluciones correspondientes.

METODOLOGIA

La evaluación del material contenido en este trabajo se desarrollará básicamente en 3 fases:

1a Fase) La exposición en clase, del material propuesto, tanto por el profesor como por el alumno interesado en hacerlo para obtener mayor puntuación en su evaluación individual.

2a Fase) Consistirá en el desarrollo de los ejercicios propuestos en cada tema por parte del estudiante como una forma activa de preparación para una evaluación formal al final del tema.

3a Fase) Los mecanismos de evaluación serán por medio de examenes aplicados a los estudiantes de dicho grupo piloto cuyo análisis e interpretación aparecerán al final de cada tema. Así como todas las actividades desarrolladas durante la exposición del tema.

PLAN DE ACTIVIDADES
1.- Selección del grupo de estudiantes para realizar el trabajo, que serán los integrantes de un grupo definitivo de estadística del quinto y sexto semestres en el Colegio de Ciencias y Humanidades Plantel Oriente.

2.- Capacitación y preparación de los estudiantes a través del trabajo cotidiano de las clases de la materia durante un semestre. Este punto incluye:
a) Presentación del material en clase.
b) Desarrollo de actividades de capacitación para los estudiantes, en el salón de clase y fuera de él.
c) Técnicas de recolección de datos para evaluar a los estudiantes, por ejemplo a través de la elaboración de exámenes por tema y exposiciones individuales de problemas específicos e investigaciones básicas con respecto a su problemática cotidiana y/o al tema que se este trabajando.
d) Análisis e interpretación de los resultados obtenidos.

Dentro del plantel Oriente los alumnos que escogen cursar la materia optativa de estadística son personas cuyo punto de referencia principal consiste en que en su mayoría tienen o tuvieron problemas con las materias de matemáticas del plan de estudios obligatorio, según información recabada por la academia de matemáticas del colegio, en una encuesta realizada hace algunos años y que se marca en la referencia (1). Es por este motivo, principalmente, que se trata de presentar un trabajo que sin perder el tratamiento riguroso de la matemática en los temas que así lo requieran, se muestre de forma mínima el enfoque matemático de la estadística.

Se utilizan los conceptos básicos de la aritmética, de la teoría de conjuntos, de álgebra y del cálculo como son: valor absoluto, intervalos, factoriales, variables, funciones, límites, progresiones geométricas y sucesiones de la manera más sencilla, con base a su definición y utilizandolos a través de ejemplos para facilitar su entendimiento. Estos conceptos serían los prerequisitos mínimos deseables para los estudiantes que opten por cursar la materia de estadística en los 2 últimos semestres del plan de estudios del Colegio de Ciencias y Humanidades. Por lo tanto los interesados en cursar la materia deberían de haber aprobado satisfactoriamente los cursos de matemáticas I a IV del mismo plan de estudios*.

Sin olvidar que actualmente existen diversas herramientas que facilitan y hacen más dinámico el manejo de la parte de cálculo como por ejemplo las cálculadoras y las computadoras, se pretende introducir los conocimientos básicos de la estadística a los alumnos de este nivel medio superior. La computadora es una herramienta muy eficiente en el desarrollo de gráficas ya sea a través de hojas de cálculo o paquetes de graficación, también existen paquetes estadísticos que permiten un manejo preciso y eficaz de los datos que se van a analizar para obtener conclusiones a través de ellos.

(*) Plan de estudios revisado por la Academia de Matemáticas. Colegio de Ciencias y Humanidades Plantel Oriente. 1988 . Se incluye en el índice 2 del presente trabajo.
(1) Referencia: Cuadernos del Colegio de Ciencias y Humanidades. Plantel Oriente.

TEMARIO

METODO DE INVESTIGACION DE FENOMENOS

Los fenómenos o experimentos sobre los que se puede utilizar la estadística pertenecen a muy diversas áreas, pero principalmente es en la investigación donde se ha generalizado su aplicación para encontrar una solución adecuada a los problemas. A continuación se dan de manera general algunas etapas para la aplicación de la estadística en dichos casos.

I.- PLANTEAMIENTO DEL FENOMENO.

Todo fenómeno o experimento que se pretenda investigar debe de estar bien definido y determinado por el investigador y/o por las personas que pretendan estudiarlo. La primera etapa sería lograr una buena conceptualización del fenómeno, identificando detalladamente las características principales y los factores que lo afectan.

II.- ESTABLECIMIENTO DE HIPOTESIS.

El investigador y/o personas que pretendan estudiar un fenómeno deben de establecer una o varias hipotesis de acuerdo a la concepción del mismo, es decir, se deberá de identificar los posibles resultados que se pretendan determinar con base a los datos con que se cuentan con anterioridad. Si la o las hipótesis planteadas se aceptan o se rechazan, las conclusiones en cuestión, siempre ofrecerán un nuevo conocimiento con respecto al fenómeno, que seguramente redúndara en la identificación de nuevos elementos, muy probablemente, a considerarse en futuros replantamientos del problema en cuestión. La estadística en la mayoria de sus aplicaciones es de carácter empírico y su propósito consistiría en determinar relaciones rígidas de causa-efecto entre las caracteristicas de los fenómenos utilizando modelos probabilisticos.

Cuando se establece una hipotésis con relación a un fenómeno, se puede intentar representar dicha situación mediante un modelo matemático para facilitar su comprensión, dicho modelo deberá tratar de representar las propiedades de la hipótesis planteada y del fenómeno en análisis o estudio.

III.- OBSERVACION DEL FENOMENO O EXPERIMENTACION.

Consiste en la actividad de proceder a realizar el experimento o fenómeno, para poder recabar datos y verificar si se cumplen o no la o las hipótesis establecidas. El proceso de recolección de datos debe ser estricto para tratar de suprimir cualquier desviación o mala concepción que solo produciria errores en las conclusiones del experimento.

Los datos obtenidos experimentalmente se organizarán y analizarán de tal forma que se puedan realizar pruebas de las hipótesis consideradas y con ello se obtengan conclusiones.

Estas serian de forma muy general las etapas por las que pasarian los fenómenos o experimentos que se quisieran investigar utilizando las técnicas de la estadística.

Antes de proseguir se establecerán algunas definiciones que aclaren el entorno teórico sobre el cual se trabajará. Dar una única definición de estadística no es posible, por esta razón se ofrece una, que se apega a el objetivo planteado para este trabajo.

Definición.- La **estadística** es una ciencia que estudia a los fenómenos utilizando métodos de planeación de recolección de datos, de organización y la extracción de inferencias a partir únicamente de los casos observados.

La estadística se clasifica de acuerdo a los objetivos que se persiguen en: estadística descriptiva y estadística inferencial.

Definición.- La **estadística descriptiva** se refiere principalmente a la forma de organizar y resumir los datos de un experimento, investigación o fenómeno de tal manera que se pueda observar las tendencias y normas de la población en cuestión, mediante gráficas y tablas, y el cálculo de parámetros sencillos, así como la redacción de informes de los resultados obtenidos. La información en este caso puede pertenecer a toda una población o bien a un subconjunto de ella llamada muestra.

Definición.- La **estadística inferencial** es aquella que proporciona predicciones acerca del comportamiento de un fenómeno, experimento o investigación que se efectúa sobre un pequeño grupo de elementos extraídos de la población de interés; es decir, se obtienen inferencias a cerca del comportamiento de un fenómeno en una población a partir de una pequeña parte de ella.

Debido a la diversidad de poblaciones sobre las cuales se puede efectuar una investigación y a los distintos experimentos que se pueden realizar sobre una misma población, se hace necesario establecer una definición de lo que se entenderá por población y muestra, y la forma en que se seleccionarán los elementos sobre los que se aplicará o se observará el fenómeno.

Definición.- **Población** es cualquier conjunto de elementos, -individuos, objetos, eventos, etc.-, por los cuales se tiene cierto interés o se desea conocer cierta información de ellos, también recibe el nombre de población objeto.

Definición.- **Muestra** es un subconjunto de la población. Es decir, es un conjunto de elementos de la población seleccionados de una manera adecuada, no se trata de todos los integrantes de la población.

Con relación a los fenómenos que se investigan, las poblaciones se pueden clasificar en población real o hipotética.

Definición.- Una población es **real** cuando se pueden contar a los elementos que integran la población físicamente. Por ejemplo un conjunto de alumnos en un salón de clases, un conjunto de edificios en una ciudad, el conjunto de libros en una biblioteca, etc.

Definición.- Una población es **hipotética** cuando no es real. Por ejemplo una fábrica productora de chips para componentes de computadora selecciona 1000 chips para probar el rendimiento de algunos de ellos, se pone en operación algunos de ellos y se toma el tiempo en que funcionan normalmente; utilizando el mismo équipo, materiales y método de producción la fábrica continuará produciendo chips y éstos constituirán una población hipotética.

Ejemplos:

1) Se desea investigar los nacimientos en un determinado hospital, durante una semana. Por lo tanto se tiene que durante una semana en dicho sanatorio se tendrán un número finito de nacimientos. $P = \{ x_1, x_2, ..., x_n \ / \ x_i$ es niño o niña que nació en el hospital en esa semana$\}$ donde n es finito.

P es un conjunto finito y es una población real.

2) Una empresa productora de botellas de plástico desea saber si su producto es biodegradable, para esto selecciona 25 botellas al azar de su producción de un día para someter algunas de ellas a pruebas de descomposición. Utilizando la misma maquinaria, pero modificando -si es necesario- alguno de los materiales y el método de producción; la empresa producirá las botellas de plástico biodegradables diariamente.
$P = \{ x_1, x_2, ..., x_i \ / \ i = 1,2,...,25$ y x_i es una botella seleccionada$\}$
La población es hipotética.

3) Se quiere investigar los kilogramos de arroz que se cultivan en un 1 kilómetro cuadrado de tierra de cultivo de la región huasteca.

$P= \{ x \ / \ x$ es un Kg. de arroz$\}$
P es un conjunto finito y es una población real.

4) Un estudiante empieza su jornada diaria con N$5.00 en su bolsillo, con dicha cantidad tiene que pagar los transportes de ese día, los gastos de materiales de ese día, los gastos alimenticios de ese día y los imprevistos. Si desea saber al final del día si le alcanza su presupuesto tiene que contar cuanto dinero le sobra.

$P=\{ x \ / \ x$ es la cantidad gastada por cada concepto$\}$
P es un conjunto finito y es una población real.

Uno de los aspectos más importantes en estadística es identificar el conjunto de elementos que se quiere estudiar, esto es, la población sobre la que se realizará la investigación o estudio.

En el proceso de observación de un fenómeno o experimento, se registra para cada unidad seleccionada -que pueden ser individuos, grupos de personas, porciones de ciudades, naciones, etc.- alguna característica particular de esa unidad, dicha observación constituye un dato. Los datos son los valores de varias variables obtenidos a partir de un conjunto de unidades.

Una variable se puede entender como la característica que no permanece constante en todos los miembros de una población o muestra sobre la que se esta efectuando una investigación. Por ejemplo el número de materias de matemáticas que tiene reprobadas un alumno es diferente en cada caso, aunque en algunas ocaciones pudiera coincidir; en este caso la variable sería el número de materias de matemáticas reprobadas.

Definición.-Una **variable** es una característica, atributo o cualidad que toma valores distintos de un miembro a otro de la población que se estudia, para propósitos de comparación, por lo regular siempre se le representa con una letra del abecedario.

Debido a la diversidad de características o variables sobre las que se puede desarrollar alguna investigación, por ejemplo la estatura de personas, el color de ojos de los alumnos en un plantel escolar, el sueldo de una persona, el peso de un animal, la velocidad de un galgo, etc., se hace necesario distinguir entre dos tipos de datos.

-**Datos numéricos** son los que se pueden obtener mediante mediciones numéricas ya sea auxiliados por algún instrumento o simplemente por el proceso de contar. Los datos numéricos permiten mayor flexibilidad y precisión en su organización y presentación.

-**Datos categóricos** se obtienen de la observación de una característica previamente determinada en los elementos de la población seleccionada.

Existen distintos tipos de variables para poder identificar y medir las distintas características en las poblaciones sobre las cuales se efectúan experimentos; entre ellos se tienen las variables cuantitativas y las variables cualitativas.

Definición.- Una variable **cuantitativa** es aquella que puede "medirse" y cuantificarse es decir, que la característica de interés se puede identificar mediante un número. Por ejemplo las estaturas de los adultos de sexo masculino, el peso en kilogramos de los niños la temperatura de los seres vivos, el número de personas en un concierto, el número de habitantes en una colonia, etc.

Definición.- Una variable **cualitativa** es aquella característica que sólo "califica" al individuo o elemento de una población, es decir únicamente se puede identificar que "valor" le corresponde. Por ejemplo la designación de una persona dentro de un grupo socioecónomico (clase baja, clase media, clase alta), el color de cabello de una persona (negro, castaño, cafe, rubio), el estado civil de las personas (soltero, casado, divorciado, viudo, unión libre), etc.

Las variables cuantitativas se pueden distinguir en dos tipos, aquellas que para observarse necesitan "medirse" con algún instrumento o bien aquellas para las cuales se requiere realizar un proceso de conteo únicamente. Esto permite distinguir dos tipos de variables cuantitativas continuas y discretas.

Definición.- Una variable **discreta** es aquella característica que se puede "medir" contando las ocurrencias de dicha característica en los elementos de la población finita seleccionada. Por ejemplo el número de mujeres en una familia, el número de trabajadores femeninos en una fábrica, el número de accidentes diarios en una fábrica, etc.

Definición.- Una variable **continua** es aquella característica que se puede "medir" utilizando un instrumento para ello, son cantidades como la longitud, el peso, la temperatura, la intensidad del espectro de luz, etc.

De acuerdo a la clasificación que se ha establecido para las distintas características de interés que se pueden investigar en poblaciones previamente seleccionadas, existen distintas escalas de medición que permiten ordenar los resultados que se obtengan. Al hablar de medida se esta hablando de comparación.

ESCALAS O NIVELES DE MEDICION

La medición es un proceso por medio del cual se asigna un número a una característica o propiedad de uno o varios elementos de una población con propositos de comparación.

Existen varias formas de seleccionar los resultados de un experimento efectuado en una determinada población para poder presentar sus propiedades relevantes, con éste propósito se tienen lo que se llama NIVELES O ESCALAS DE MEDICION.

Para datos categóricos se tienen las escalas:

- Escala nominal
- Escala ordinal

Para datos numéricos se tienen las escalas:

- Escala intervalar
- Escala de razón

La **escala nominal** únicamente involucra el proceso de nombrar o etiquetar los diferentes resultados que se puedan obtener de un experimento, es decir, no existe un orden natural para los distintos datos que se obtienen, solamente se clasifican si tienen o no la característica de interes. Por ejemplo sexo de las personas (masculino, femenino), estado existencial de los animales (vivo, muerto), estado de servicio de un foco (sirve, no sirve), etc.

La **escala ordinal** de medición involucra un orden natural en las categorias en las que se clasifican los datos de una investigación, es decir, aparte de nombrar o etiquetar los resultados, los ordena de acuerdo al grado o intensidad en que poseen una determinada característica. Por ejemplo el estado socioecónomico de las personas (bajo, medio, alto), posición al final de una competencia (primero, segundo, tercero, ...), posición entre hermanos (primero, segundo, tercero, ...), estado civil de las personas (soltero, casado, divorciado, viudo), etc.

La **escala intervalar** clasifica a los datos de una investigación a partir de un cero natural, sin embargo este cero no es único. Esta escala emplea unidades exactas de medición que permiten colocar cada dato dentro de uno y sólo un intervalo específico. Por ejemplo: el grado 0° de la escala del sistema Fahrenheit no es el grado 0° de la escala del sistema centigrado, un objeto que se encuentra entre el punto de ebullición y congelación del agua esta a 50° en la escala de grados centigrados [(0°+100°)÷2] pero en la escala de grados fahrenheit esta a 122°, es decir, [(32°+212°)÷2].

La **escala de razón** permite determinar proporciones entre los valores de los datos de una investigación, además identifica si hay un dato en que la propiedad que se esta midiendo no existe para asignarle el número cero. En esta escala las razones entre los valores que reflejan la propiedad del fenómeno son constantes. Por ejemplo el peso de las personas en kilogramos, si una persona pesa 200 kilogramos entonces pesa dos veces cien kilogramos. Por lo tanto si se pesan dos objetos de cien kilogramos se obtiene un objeto de 200 kilogramos.

Cualquier variable que se pueda medir en una escala intervalar o en una escala de razón se puede medir también en una escala ordinal.

En todo proceso de observación de los resultados de algún fenómeno, se distinguen dos tipos de mediciones: medición fundamental y medición derivada.

La **medición fundamental** consiste en medir la propiedad o característica en estudio en un objeto con relación a otro que tenga la misma propiedad.

La **medición derivada** consiste en medir la característica en estudio en un objeto con respecto a una ley natural aceptada como válida que se relacione con la propiedad que se va a medir. En las mediciones derivadas se corre el riesgo de que algunas conclusiones sean verdaderas, pero pueden ser no verdaderas para las propiedades reales del experimento debido a que la escala de medición no refleja con suficiente exactitud la propiedad medida. Por ejemplo para medir la variable inteligencia de una persona un sociologo emplea una especie de prueba o cuestionario cuyos resultados son los datos que se usan para "medir" la inteligencia de la persona.

El siguiente cuadro muestra en forma resumida las propiedades de las escalas de medición:

ESCALAS DE MEDICION

Escala	Operaciones básicas	Cambios permisibles en los valores de la escala	Propiedad medida	Ejemplos Valores de la escala	Características
Nominal	Determinación de igualdad. De pertenencia a una categoría.	Se puede hacer cualquier cambio en la denominación de cada categoría o grupo de igualdad	1.-Sexo 2.-Estado Civil 3.-Situación laboral	Masculino y femenino Soltero, casado, divorciado y viudo Ocupado, desocupado	No existe el cero. Se usan los números naturales para el conteo de frecuencia.
Ordinal	Determinación del grado de intensidad de la característica, sin especificar posición exacta	Se puede efectuar cualquier cambio que mantenga las relaciones de orden en los numeros asignados.	1.- Estado de un paciente con osteoartritis	1=Empeoramiento intenso 2=Empeoramiento 3=Sin cambio 4=Mejorado 5=Muy mejorado	Aparece el cero. Se usan números enteros para el conteo de frecuencia.
Intervalar	Determinación de la igualdad de intervalos o diferencias entre valores.	Se puede cambiar la unidad de medida y el origen de la medida.	1.- Temperatura 2.- Sismos	Cualquier número entero o fraccionario. Centigrados o Farenheit. Sismegrafo con escala Mercalli o con escala Richter	El cero no es único depende de la escala que se usa.
Razón	Determinación de la igualdad de relaciones o proporciones.	Se puede cambiar la unidad de medida pero no el origen	1.-Medidas antropomórficas. 2.-Concentración de bilirrubina, ácido úrico, creatinina, etc. en suero 3.-Proporción o por ciento de volumén ocupado por vacuolas en un tipo de celula	Números enteros o fraccionarios en cm, m, kg, g, l, etc. Números enteros o fraccionarios en mg por 100 cc de sangre Números enteros o fraccionarios entre 0 y 100.	El cero es único.

Ejemplos:

1) En el D.F. se realizó una encuesta para determinar si se convertia en estado y se elegia un gobernador o continuaba siendo D.F. Se recabaron 500,000 votos de los cuales 340,000 eran sí a las preguntas plantedas y los restantes eran no.

En este ejemplo la característica es una variable **cualitativa.** Se emplea una escala **nominal** para cuantificar los votos.

2) Un fábricante de velocímetros para automóvil desea probar su funcionamiento, para esto toma una muestra de 10 velocímetros y los coloca en un prototipo de automóvil y los corre a distintas velocidades durante una hora cada uno.

La característica de interés es el funcionamiento de los velocímetros por lo tanto es una variable **cualitativa** (bueno o malo) y se emplea una escala **de intervalos** para cuantificar los resultados.

3) Se realizó una encuesta entre las familias de los alumnos de un grupo de Estadística del Colegio de Ciencias y Humanidades - Plantel Oriente, para determinar el número de integrantes de cada familia considerando únicamente a familiares directos.

La característica de interés es el número de integrantes de las familias, por lo tanto es una variable **cuantitativa**. Se emplea una escala de **razón** para cuantificar los resultados.

Ejercicios:

En los siguientes ejercicios señala que tipo de variable se está utilizando y cuál sería la escala adecuada para medir las características de interés de cada investigación que se va a realizar.

1.- En la campaña anual de la cruz roja de donación altruista de sangre se presentarón un grupo de 20 alumnos de plantel oriente a la ambulancia móvil. Para donar sangre se les "tomo" el peso en kg. para saber si podian ó no donar sangre.

2.- En el colegio de ciencias y humanidades plantel oriente se tomó una muestra de 50 alumnos con los mejores promedios en su último año escolar para determinar si continuarían sus estudios superiores en la universidad nacional o en otra institución.

3.- Se realiza una encuesta sobre 100 trabajadores administrativos de una fábrica trasnacional para averiguar el dominio que tienen del idioma inglés.

4.- En una muestra aleatoria de 50 alumnos que trabajan del turno 4 del plantel oriente se desea investigar cuál es su ingreso promedio mensual en nuevos pesos.

5.- Se realiza una encuesta entre 20 amas de casa tomadas al azar en un supermercado para medir el grado de aceptación de un producto -cereal americano- de acuerdo al número de paquetes que comprarían cada una de ellas.

6.- Se desea investigar en el sindicato de la universidad, cuántos emplazamientos a huelga por incremento salarial se han declarado en un año.

7.- En el plantel oriente se tiene un registro diario por clase para los trabajadores académicos para medir su porcentaje de asistencia al colegio. Si se escogen 10 profesores al azar como se puede determinar su porcentaje de asistencia en un mes.

8.- La encargada de compras de una tienda comercial, desea saber cuáles son las tallas de ropa que tienen mayor demanda para adquirir mayor cantidad de ellas.

9.- En una encuesta estadística que se realiza al final del año escolar en el plantel oriente se reparten cuestionarios a los alumnos de cada clase para "medir" el desempeño académico de los profesores. Si se toman aleatoriamente 3 cuestionarios de un solo profesor, cómo se puede medir su desempeño académico?.

10.- En una libreria se desea conocer el número de ventas por la cantidad de libros que los clientes compran en una semana. Para esto se registran diariamente cuantos libros se venden.

ARREGLOS

En una investigación los datos se presentan en muchas ocasiones como una masa desordenada de información, y a menos que el número de observaciones sea extremadamente pequeño es improbable que proporcionen mucha información a quien los tenga si no los transforma y moldea la muestra de una manera adecuada.

Los datos obtenidos de un experimento o investigación se suelen presentar primero en forma de una tabla dentro de ella se procede a formar un arreglo ordenado, si es posible se presentarán en forma ascendente o descendente. Los datos, cuantitativos o cualitativos, comúnmente se registran en la primera columna de la izquierda y posteriormente se muestran a la derecha columnas que pueden referirse a uno o varios de los siguientes conceptos: la frecuencia absoluta, la frecuencia relativa, la frecuencia absoluta acumulada y la frecuencia relativa acumulada.

En teoría las variables continuas se pueden medir tan precisamente como sea posible, pero en la práctica la precisión en las medidas está determinada por los dispositivos de medición, los intereses de la investigación, así como también el comparar lo que se puede perder o ganar por registrar información demasiado exacta. Para este tipo de datos se suele utilizar el redondeo en las cifras. Ejemplos de características que se suelen registrar de la manera antes descrita son el peso de las personas (en kilogramos), la temperatura de un líquido (en grados centígrados o farenheit), la presión del aire (centímetros), etc.

Ejemplo: Notas de un examén de matemáticas (notas sobre 100), de 50 estudiantes.

30	55	44	60	43	72	47	65	67	40
59	58	14	32	58	46	41	35	68	50
59	21	42	45	41	48	28	47	77	60
30	57	45	49	33	48	47	52	38	61
54	42	54	42	49	51	39	60	61	63

Arreglo ordenado en forma ascendente:

14	32	40	42	46	48	52	58	60	65
21	33	41	43	47	49	54	58	60	67
28	35	41	44	47	49	54	59	61	68
30	38	42	45	47	50	55	59	61	72
30	39	42	45	48	51	57	60	63	77

Arreglo ordenado en forma descendente:

77	63	60	57	51	48	45	42	39	30
72	61	59	55	50	47	45	42	38	30
68	61	59	54	49	47	44	41	35	28
67	60	58	54	49	47	43	41	33	21
65	60	58	52	48	46	42	40	32	14

TABLA 1.

El propósito de ordenar un gran conjunto de datos obtenidos de algún experimento es que no resulte tedioso o cansado analizarlos. Una vez que los resultados ya se encuentran ordenados es mucho más sencillo observar el número de veces que se presenta el mismo valor, sin embargo a pesar de ésto en poblaciones o muestras muy extensas es muy complicado observarlos uno a uno, por esta razón se agrupan en **clases o categorías**, que permiten formar una tabla que los contenga pero de manera más resumida, a dicha tabla se le llama TABLA DE DISTRIBUCION DE FRECUENCIAS.

En una tabla de distribución de frecuencias se muestra como están distribuidos los datos en categorías, es decir cuantos de ellos están asociados con cada categoría.

Definición.- Se llama **FRECUENCIA** al resultado de contar el número de veces que se presenta la misma observación en un experimento determinado.

La frecuencia de ocurrencia se puede dar de dos maneras: absoluta y relativa.

Definición.- La frecuencia **absoluta** es el número de veces que se repite una misma observación dentro de una categoría.

Definición.-La frecuencia **relativa** es la proporción que se obtiene al dividir la frecuencia absoluta de cada categoría entre el total de observaciones, se puede presentar de dos maneras en forma de proporción o en porcentaje; para presentarlo en forma de porcentaje simplemente se multiplica la proporción obtenida por 100. La suma de las frecuencias relativas en cualquier tabla es igual a 1 cuando se presenta en forma de proporción e igual a 100% cuando se presenta en forma de porcentaje.

Para obtener un buen reporte de los datos, éstos se agrupan en clases para formar la tabla de distribución de frecuencias. Se trata de que dichas clases sean aproximadamente de igual tamaño, la elección de la longitud de la clase o intervalo y sus puntos medios respectivamente determinan los límites de las clases. El número de observaciones en un intervalo es la frecuencia de dicho intervalo.

Ejemplo con los datos de la Tabla 1:

Intervalos	Clases	Marca de Clase	Frecuencia	Frecuencia Relativa
14-30	13.5-30.5	22	5	5/50
31-46	30.5-46.5	38.5	16	16/50
47-62	46.5-62.5	54.5	23	23/50
63-78	62.5-78.5	75.5	6	6/50
		total	50	

Con el objetivo de facilitar la evaluación de los resultados de un experimento sobre una muestra, se busca que se puedan dibujar ciertas figuras o gráficas para ayudar a la toma de decisiones con base a dichos resultados y concentrados en una tabla de distribución de frecuencias.

Entre los diagramas que se pueden construir se encuentran los siguientes:

- Diagramas de barras verticales y horizontales
- Diagramas circulares o "pie"
- Series temporales
- Polígono de frecuencias
- Histograma

Estos diagramas se pueden elaborar utilizando un eje cartesiano, es decir un eje X horizontal y un eje Y vertical. Sobre el eje horizontal se colocan todos los posibles valores de la variable X - en su caso las categorías o intervalos- y sobre el eje vertical Y la frecuencia correspondiente a cada valor de la variable. Comúnmente se utilizan los diagramas de barras para datos categóricos para representar sus frecuencias absolutas o sus frecuencias relativas por medio de rectángulos separados.

Un histógrama es una gráfica que representa la frecuencia de los datos medidos con una escala continua en donde todos los intervalos son rectángulos de áreas consecutivas. En un histógrama los intervalos de clase son de igual base pero su altura es variable y está dada por la frecuencia.

Ejemplos:

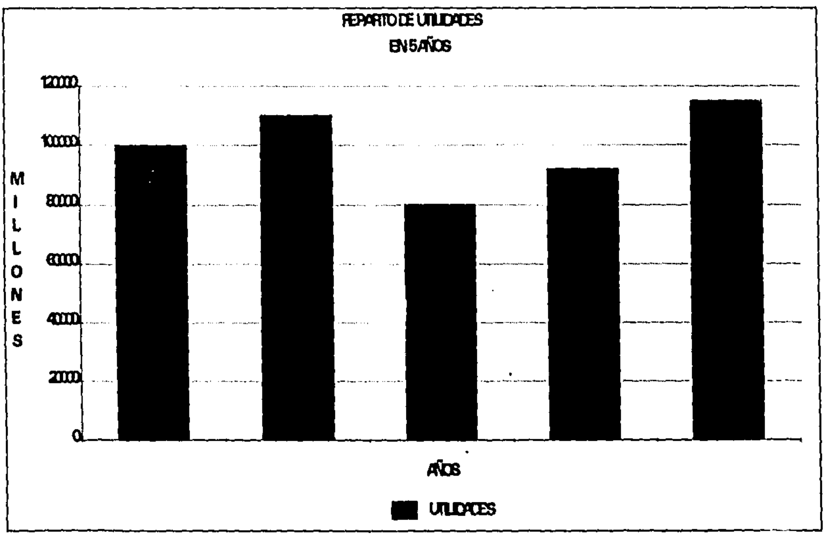
En un periodo de 5 años una compañía tuvo los siguientes niveles de reparto de utilidades. En 1985 repartió $100,000,000 entre sus 1000 empleados, en 1986 repartió $110,000,000, en 1987, $ 80,000,000, en 1988, $92,000,000 y en 1989, $115,000,000.

a) Elaborar un diagrama de barras que muestre el reparto de utilidades en esos 5 años.

b) Elaborar una serie temporal con los datos antes mencionados.

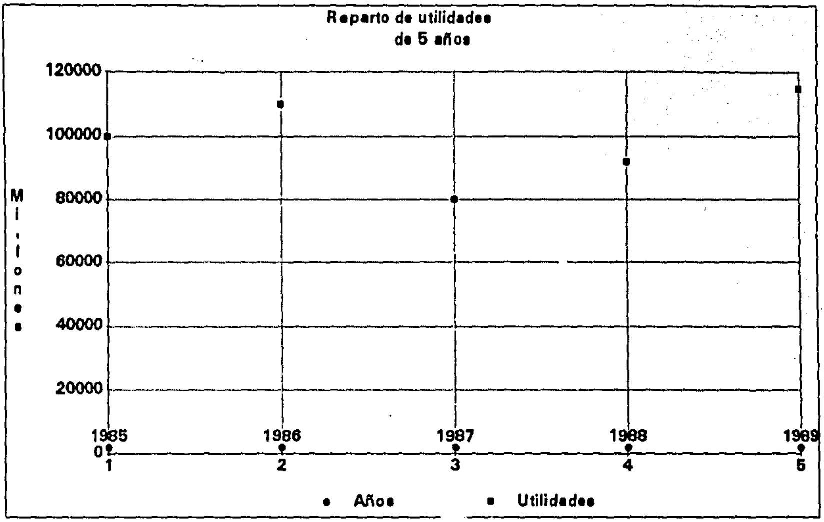
c) Elaborar un diagrama circular o "pie" del reparto de utilidades de la compañía.

d) Elaborar un polígono de frecuencias con los datos del reparto de utilidades.



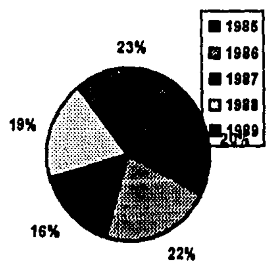
b)

Serie temporal



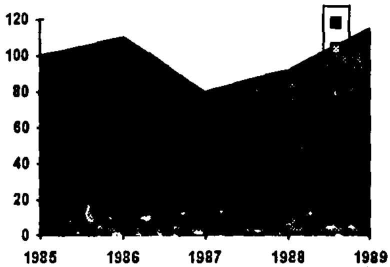
c) Diagrama circular

Reparto de utilidades en 5 años



d) Polígono de frecuencias

Reparto de utilidades en 5 años



EVALUACION PROPUESTA

1.- Describir que parte de la estadística se encarga de recolectar, organizar y resumir los datos recabados de una investigación?

2.- A qué se le llama población?:

3.- En cada uno de los siguientes ejemplos identificar de que tipo de población se trata: en un grupo de 30 alumnos mayores de 18 años se formaron las siguientes poblaciones

P={ x / x es un glóbulo rojo de 1litro de sangre de cada alumno}________________
P={ y / y es un alumno casado} ________________
P={ z / z es una materia de matemáticas reprobada por los alumnos} ____________

4.- Cuál es el nivel de medición que se identifica con el continuo lineal o recta real?

5.- A que se le llama muestra?

6.- Definir que tipo de variable se trata en cada uno de los siguientes ejemplos:
a) En una encuesta de opinión realizada con 20 alumnos del Plantel Oriente se desea conocer el número de integrantes familiares que habitan bajo el mismo techo __________
b) En una investigación ecológica en el Ajusco se desea saber el número de arboles talados en un kilómetro cuadrado de bosque en el último año __________
c) En un laboratorio de biologia se desea investigar el número de ratas hembras sacrificadas durante el ciclo escolar __________

7.- Mencionar cual nivel de medición se emplearía en cada uno de los siguientes casos para resumir los datos que se obtuvieron:

a) Se investiga el estado civil de 30 personas tomadas al azar _______________

b) La temperatura ambiente en los ultimos 20 días del presente mes _______________

c) Las calificaciones de 50 alumnos de un grupo de matemáticas _______________

Los siguientes datos representan el número de reses sacrificadas en el rastro en 30 días consecutivos:

20	75	40	70	58	78
30	20	90	75	72	64
25	30	45	60	89	62
45	60	55	53	85	80
65	20	45	22	75	86

8.- Formar la tabla de distribución de frecuencias en 4 intervalos

9.- Dibujar un diagrama de barras con los datos del ejercicio 9.

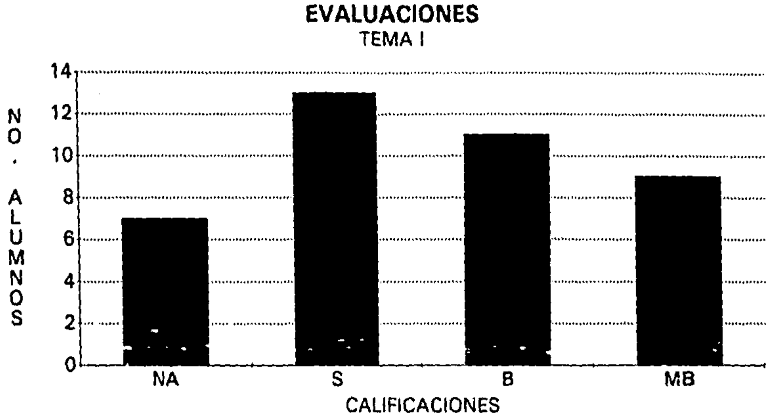
10.- Dibujar un diagrama circular utilizando los datos de la frecuencia relativa.

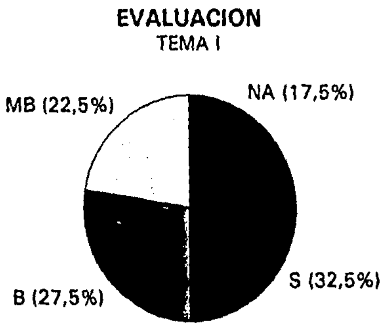
CONCLUSIONES

Los resultados que se obtuvieron al aplicar la evaluación a 40 alumnos del grupo 505 de estadística en el Colegio de Ciencias y Humanidades - Plantel Oriente se muestran en un arreglo ordenado que se resume en una tabla de distribución de frecuencias y se construye un diagrama de barras con dicha información.

4	5	6	7	8	8	9	10
4	5	6	7	8	8	10	10
4	6	6	7	8	9	10	10
4	6	6	7	8	9	10	10
5	6	6	7	8	9	10	10

Calificaciones (letras)	Frecuencia (no.alumnos)	Frec. relativa
NA	7	7/40
S	13	13/40
B	11	11/40
MB	9	9/40
total	40	





Las conclusiones que se observan con respecto al aprovechamiento de los alumnos del material correspondiente al primer tema del trabajo, muestran que más del 80% comprendieron los conceptos de la evaluación escrita.

TEMARIO

TABLA DE DISTRIBUCION DE FRECUENCIAS

Para formar una tabla de distribución de frecuencias agrupada se forma un conjunto de intervalos que no se traslapen, de tal manera que cada dato pueda colocarse en uno y sólo uno de dichos intervalos. Se debe considerar cuantos intervalos van a incluirse, usar demasiado pocos no es conveniente debido a que hay pérdida de información y si se utilizan muchos se pierde el sentido de la síntesis; un criterio general un tanto empírico es que no sean menos de diez ni más de 20. Existen varias recomendaciones para determinar el número de intervalos, entre ellas esta la fórmula o regla de Sturges:

$K = 1 + 3.322\,(\mathrm{Ln}\ N)$ donde

Ln= logaritmo natural de N
K= número de intervalos
N= número de elementos en el conjunto de datos

La segunda pregunta que surge es: ¿Cuál es la amplitud que deben tener dichos intervalos?, por lo general los intervalos deben de ser de longitudes iguales y la longitud de cada intervalo se determina de la siguiente manera:

Rango = marca mayor - marca menor

$$\text{Longitud de cada intervalo} = \frac{\text{Rango}}{\text{\# de intervalos}}$$

Ejemplo:

Arreglo ordenado en forma ascendente:

14	32	40	42	46	48	52	58	60	65
21	33	41	43	47	49	54	58	60	67
28	35	41	44	47	49	54	59	61	68
30	38	42	45	47	50	55	59	61	72
30	39	42	45	48	51	57	60	63	77

Con los datos agrupados del arreglo anterior, construir una tabla de distribución de frecuencias agrupada en 4 intervalos:

Rango = 77 - 14 = 64

$$\text{Longitud de c/intervalo} = \frac{64}{4} = 16$$

Para obtener cada uno de los intervalos se empieza con el número o dato menor que se tenga -como resultado del experimento- como límite inferior del primer intervalo y se le suma el tamaño o longitud calculada para obtener el límite superior del primer intervalo, después al límite superior del primer intervalo se le suma 1 para que sea el límite inferior del segundo intervalo y a éste se le suma el tamaño del intervalo para obtener el límite superior del segundo intervalo; se continúa de esta manera hasta que los intervalos contengan todos los datos obtenidos del experimento.

Se empieza con el dato 14 como límite inferior del primer intervalo y se suma 16 para obtener 30 que es el límite superior del primer intervalo; al límite superior 30 se le suma 1 para obtener el límite inferior del segundo intervalo que será 31 al que se le suma 16 para obtener 47 que es el límite superior del segundo intervalo, continuando de esta manera hasta que cada uno de los datos del experimento queden dentro de algún intervalo.

Para obtener los intervalos de clase o clases se resta .5 al valor de cada límite inferior y se le suma .5 al valor de cada límite superior, de ésta manera se tiene el límite real inferior y el límite real superior de cada clase respectivamente. La característica principal de las clases es que no presentan discontinuidades entre ellas y se pueden dibujar sus límites reales sobre la recta real, que entre.otras ventajas permite dibujar histogramas y polígonos de frecuencia. Además permite localizar con una mayor precisión los elementos de la población y los parámetros que se calculan. Si en el ejemplo se tuviera un dato cuyo valor fuera 30.5 exactamente dicho dato se colocaria en el primer intervalo en donde apareciera el límite 30.5, es decir, cualquier dato que sea igual a cualquiera de los límites superiores de los intervalos de clase se cuantifican para el intervalo en donde aparece por primera vez dicho valor. De esta forma se obtiene la siguiente tabla:

Intervalos	Intervalos de clase	Frecuencia absoluta	Frecuencia relativa	Marcas de clase
14 - 30	13.5 - 30.5	5	5/50	22
31 - 47	30.5 - 47.5	19	19/50	39
48 - 64	47.5 - 64.5	21	21/50	56
65 - 81	64.5 - 81.5	5	5/50	73
	Total	50		

TABLA 2

Las siguientes columnas a la derecha de los intervalos de clase se obtienen como a continuación se describe.

<u>Definición</u>.- Las **marcas de clase** de los intervalos de clase son los puntos medios de dichos intervalos.

$$\text{Marca de clase} = \frac{\text{límite real inferior} + \text{límite real superior}}{2}$$

Ejemplo:

Calcular las marcas de clase de los intervalos con los datos de la Tabla 2:

$$X_1 = \frac{13.5 + 30.5}{2} = \frac{44}{2} = 22 \qquad X_3 = \frac{47.5 + 64.5}{2} = \frac{112}{2} = 56$$

$$X_2 = \frac{30.5 + 47.5}{2} = \frac{78}{2} = 39 \qquad X_4 = \frac{64.5 + 81.5}{2} = \frac{146}{2} = 73$$

Para representar gráficamente los datos de la tabla 2 se utiliza en todos los casos un sistema cartesiano de 2 dimensiones, es decir, con dos ejes X y Y. Sobre el eje horizontal X por lo regular, se colocan los datos en los intervalos que se construyen (o categorías), y sobre el eje vertical Y la frecuencia con que se encuentra un mismo resultado o dato del experimento.

El propósito de una gráfica es obtener una figura que proporcione la mayor información posible sobre el fenómeno que se estudia de una manera visual clara y satisfactoria. Entre estas representaciones gráficas se tienen las siguientes - como se mencionaba anteriormente-:

DIAGRAMA DE BARRAS E HISTOGRAMA

Para construir un diagrama de barras horizontal se coloca cada una de las categorías sobre el eje horizontal X considerando un plano cartesiano, y sobre el eje vertical Y se establece la escala correspondiente a las frecuencias y en cada límite de cada uno de los intervalos de clase o categorías se levantan líneas verticales del tamaño de la frecuencia respectiva de cada intervalo y paralelas al eje vertical Y para formar rectángulos proporcionales a la frecuencia respectiva de cada categoría.

Para construir un diagrama de barras vertical, solamente se invierten los conceptos sobre los ejes del plano cartesiano, es decir, sobre el eje vertical Y se colocan las escalas o intervalos que contienen los datos y sobre el eje horizontal X la frecuencia absoluta de cada intervalo respectivamente.

Este tipo de diagrama de barras se puede construir cuando los datos obtenidos de un experimento se clasifican utilizando el nivel nominal u ordinal estableciendo escalas proporcionales para los resultados del experimento.

Cuando los datos obtenidos de un experimento se clasifican con el nivel intervalar, el diagrama que se puede dibujar empleando los ejes del plano cartesiano colocando las clases o escalas sobre el eje horizontal X y la frecuencia sobre el eje vertical Y se llama **histograma**. Dichas clases quedan "contiguas" es decir no hay discontinuidades - separaciones- entre los límites de cada una de ellas, sobre los límites de cada clase se levantan líneas verticales del tamaño de la frecuencia respectiva del intervalo, se unen los topes para formar rectángulos contiguos quedando definida de esta manera la gráfica. En un diagrama de barras es fácil comparar una categoría con otra observando simplemente el tamaño de las columnas.

DIAGRAMA CIRCULAR O "PIE"

Para dibujar un diagrama circular o "pie" se utilizan las categorías o intervalos que se establecen cuando los datos obtenidos de un experimento se clasifican con el nivel nominal, ordinal, intervalar o de razón. Cada una de las secciones que se forma sobre el círculo son ángulos centrales proporcionales a la frecuencia respectiva con que se presenta cada categoría o clase. En una gráfica circular es más sencillo comparar cada categoría - sección -, con respecto al total.

POLIGONO DE FRECUENCIAS

El polígono de frecuencias es una curva que se forma colocando sobre el eje X las marcas de clase correspondientes a cada uno de los intervalos de clase formados a partir de los datos obtenidos, sobre el eje Y se colocan las frecuencias absolutas considerando una escala conveniente entonces sobre cada marca de clase se levantan puntos proporcionales a la frecuencia absoluta correspondiente de cada intervalo de clase, se unen dichos puntos empezando desde el primer límite real inferior del primer intervalo de clase o categoría y terminando en el límite real superior del último intervalo de clase o categoría.

SERIES DE TIEMPO

Si los datos de una investigación se recabàn en períodos de tiempo bien determinados y sobre el eje X de un plano carteciano se colocan dichos lapsos de tiempo, y sobre el eje Y la frecuencia con la que se presentan los resultados en estos períodos y se traza un punto por cada cruce de información, a este diagrama se le llama series de tiempo.

Los datos que se resumen con vaariables cualitativas con nivel nominal u ordinal, se pueden representar con diagramas de barras y circulares; los datos que se clasifican con variables continuas con nivel intervalar o de razon se prestan para representarse con histogramas y polígono de frecuencias debido a la continuidad de las curvas.

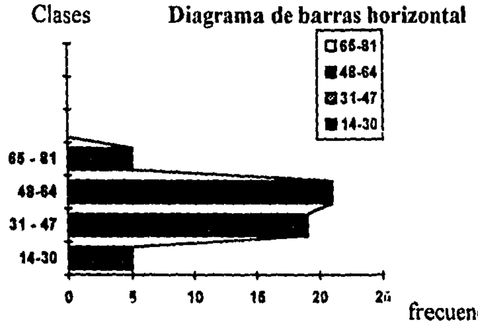
Ejemplo:

1.- Con los datos de la siguiente tabla de distribucion de frecuencias, elaborar los diagramas que se indican a continuación:

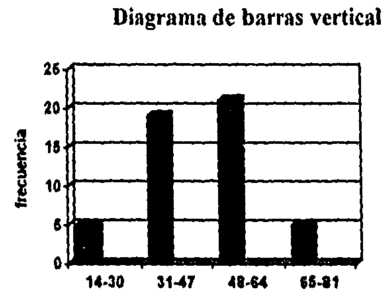
Intervalos	Intervalos de clase	Frecuencia absoluta	Frecuencia relativa	Marcas de clase
14 - 30	13.5 - 30.5	5	5/50	22
31 - 47	30.5 - 47.5	19	19/50	39
48 - 64	47.5 - 64.5	21	21/50	56
65 - 81	64.5 - 81.5	5	5/50	73
	Total	50		

a) Elaborar un diagrama de barras

Clases

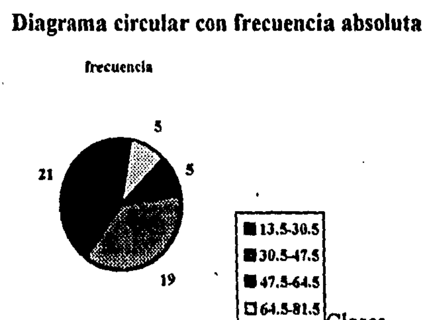


frecuencia



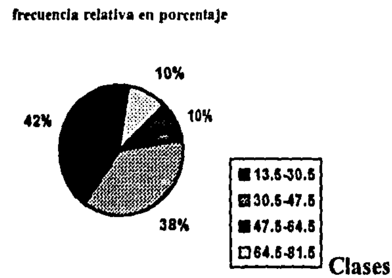
Clases

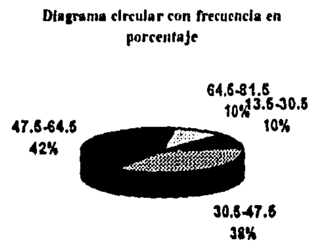
b) Diagrama circular o"pie"



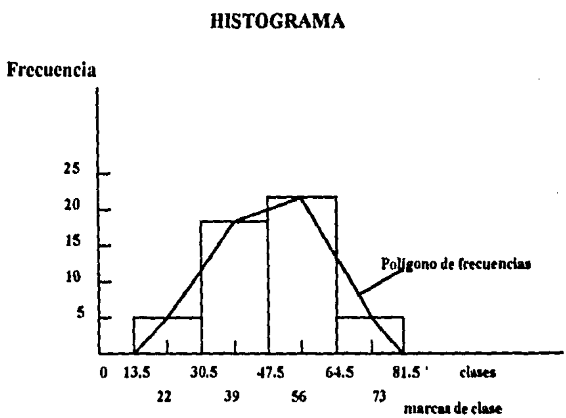
Clases

Diagrama circular con frecuencia relativa



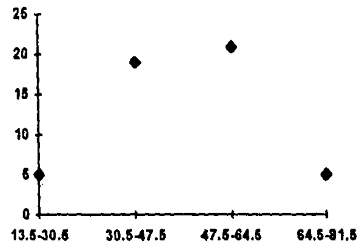


c) Histograma



d) Polígono de frecuencias

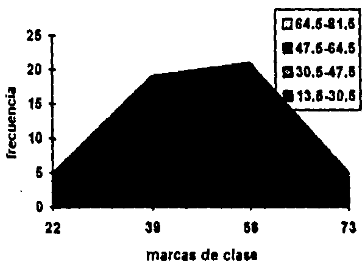
Frecuencia



Clases

Los puntos representan la frecuencia de cada clase sobre las marcas de clase.

Polígono de frecuencias



2.- En un grupo de Estadística se les preguntó su peso en Kg. a 50 estudiantes dichos datos se presentan a continuación:

60	33	85	52	65	77	84	65	57	74
71	81	35	50	35	64	74	47	68	54
80	41	61	91	55	73	59	53	45	77
41	78	55	48	69	85	67	39	76	60
84	66	88	66	73	45	65	84	86	88

a) Se desean concentrar los datos en una tabla de distribución de frecuencias con 5 intervalos.
b) Obtener las clases o intervalos de clase, la frecuencia absoluta, la frecuencia relativa, la frecuencia acumulada y las marcas de clase.
c) Hacer el histograma y el polígono de frecuencias con los datos de la tabla que se obtenga.
Método:
1°) Se ordenan los datos de menor a mayor (orden ascendente):

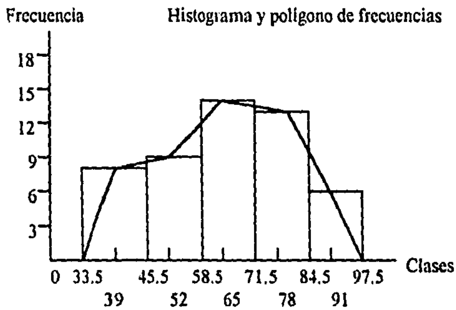
33	41	50	55	61	66	71	76	81	85
35	45	52	57	64	66	73	77	84	86
35	45	53	59	65	67	73	77	84	88
39	47	54	60	65	68	74	78	84	88
41	48	55	60	65	69	74	80	85	91

2°) Se forman los intervalos:

Rango = 91 - 33 = 58 Tamaño = $\frac{58}{5}$ = 11.6 ≈ 12

Intervalos	Intervalos de clase	Frecuencia absoluta	Frecuencia relativa	Marcas de clase	Frec. acumulada
33 - 45	32.5-45.5	8	8/50	39	8
46 - 58	45.5-58.5	9	9/50	52	17
59 - 71	58.5-71.5	14	14/50	65	31
72 - 84	71.5-84.5	13	13/50	78	44
85 - 97	84.5-97.5	6	6/50	91	50
	Total	50			

3°) Se hace la gráfica:

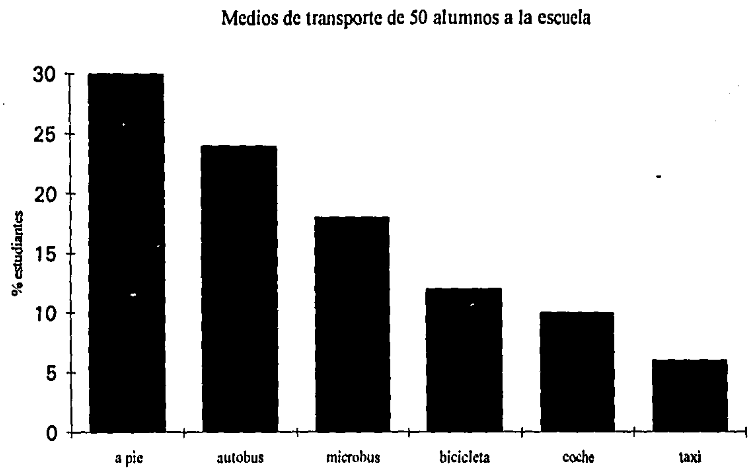


3.- A un grupo de 50 estudiantes del plantel oriente seleccionados al azar se les realizó una encuesta para conocer el medio de transporte que utilizan para llegar a la escuela. Las categorías que se establecieron para los transportes fueron las siguientes: a pie, bicicleta, microbus, coche, autobus y taxi. Los resultados se muestran a continuación:

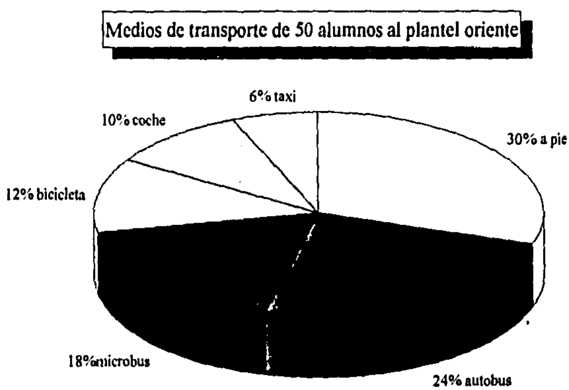
categorías	**frecuencia**	**frecuencia relativa %**
a pie	15	30%
autobus	12	24%
microbus	9	18%
bicicleta	6	12%
coche	5	10%
taxi	3	6%
total	50	100%

Usualmente se cambia la frecuencia absoluta por porcentajes, - frecuencia relativa -, para poder comparar más facilmente los resultados con respecto al total. Para esto se divide la frecuencia de cada categoría entre el número total de encuestados y se obtienen los resultados que se muestran en la columna de la derecha.

a) Elaborar un diagrama de barras



b) Diagrama circular



SUMAS

En Estadística se utiliza un símbolo que permite reducir la representación de sumas sucesivas de n-números, dicho símbolo es la letra del alfabeto griego sigma Σ y se llama **suma de n-números**. Debido a que el lenguaje matemático es un lenguaje que utiliza símbolos que permiten expresar operaciones y sus relaciones, es donde surgió el concepto para facilitar su comprensión. Para entender el término de esta suma, se tiene el siguiente conjunto de números {5,20,15,30,11} y se quiere efectuar la suma de estos números, es decir

$$5+20+15+30+11 = 81$$

Para indicar esta operación simbólicamente se procede de la siguiente manera, se toma la variable X para representar los elementos del conjunto; por ejemplo $X_1=5$, $X_2=20$, $X_3=15$, $X_4=30$, $X_5=11$, esto permite que Xi tenga distintos valores de acuerdo al conjunto de números que se considere para tomar valores, la letra i se llama un subíndice que indica cuantos elementos se tienen que considerar. La suma de dichos números se representa de la siguiente manera:

$$X_1+X_2+X_3+X_4+X_5 = \sum_{i=1}^{5} Xi = 81$$

los números situados en la parte inferior y superior del signo Σ sigma se llaman los límites de la suma, el número inferior se llama límite inferior que indica desde donde se empieza a sumar y el número superior se llama límite superior que indica hasta donde se debe terminar de sumar. En el ejemplo anterior se empieza desde X_1 y se termina en X_5, esta es una forma general de representar una suma de 5 números, si se cambian los valores del conjunto dicha forma general permanece igual, el resultado es el que cambia de valor. Por ejemplo si el conjunto esta formado por los siguientes elementos {7,9,-1,8,3} en lugar del conjunto anterior entonces la suma de estos números es:

$$\sum_{i=1}^{5} Xi = 7+9+(-1)+8+3 = 26$$

La suma generalizada es muy empleada para representar sumas de muchos números por ejemplo, la suma de n números queda representada de la siguiente manera:

$$X_1+X_2+X_3+X_4+...+Xn = \sum_{i=1}^{n} Xi$$

El índice i puede ser tanto un coeficiente, como un exponente o un divisor, se verán ejemplos de cada uno de estos casos.

El índice como coeficiente:

$$\sum_{i=1}^{5} i*4 = 1*4+2*4+3*4+4*4+5*4=4+8+12+16+20=60$$

El índice como exponente:

$$\sum_{i=1}^{3} 2^{i} = 2^{1} + 2^{2} + 2^{3} = 2+4+8=14$$

El índice como divisor:

$$\sum_{i=1}^{4} 5 / i = 5/1+5/2+5/3+5/4 =120+60+40+30=250$$

PROPIEDADES

Las sumas generalizadas cumplen con las siguientes propiedades:

1a) $\sum_{i=1}^{n} cXi = c \sum_{i=1}^{n} Xi$

2a) $\sum_{i=1}^{n} (Xi+Yi) = \sum_{i=1}^{n} Xi + \sum_{i=1}^{n} Yi$

3a) $\sum_{i=1}^{n} c = n*c$

donde c es cualquier número real, es decir $c \in R$ y Xi, Yi son variables cualesquiera

Demostración de la 1a. propiedad $\sum_{i=1}^{n} cXi = c \sum_{i=1}^{n} Xi$

$$\sum_{i=1}^{n} cXi= cX_1+cX_2+...+cX_n = c(X_1+X_2+...+X_n) = c \sum_{i=1}^{n} Xi$$

Demostración de la 2a. propiedad $\sum (Xi+Yi) = \sum Xi + \sum Yi$

$$\sum_{i=1}^{n} (Xi+Yi)=(X_1+Y_1)+(X_2+Y_2)+...+(X_n+Y_n)=(X_1+X_2+...+X_n)+(Y_1 +...+Y_n)=$$

$$= \sum_{i=1}^{n} Xi + \sum_{i=1}^{n} Yi$$

Demostración de la 3a. propiedad $\sum_{i=1}^{n} c = n*c$

$\sum_{i=1}^{n} c = c+c+...+c = n*c$ c aparece n veces

Ejemplos para mostrar el uso de estas propiedades. Sea el conjunto A={1,3,5,7} donde la variable X toma distintos valores, y sea el conjunto B={2,4,6,8} donde la variable Y toma distintos valores y c=3 una constante.

1a. propiedad:

$$\sum_{i=1}^{4} 3Xi=3*X_1+3*X_2+3*X_3+3*X_4=3*1+3*3+3*5+3*7=3*(1+3+5+7)=3\sum_{i=1}^{4} Xi$$

2a. propiedad:

$$\sum_{i=1}^{4} (Xi+Yi)=(X_1+Y_1)+(X_2+Y_2)+(X_3+Y_3)+(X_4+Y_4)=(1+2)+(3+4)+(5+6)+(7+8)=$$

$$=(1+3+5+7)+(2+4+6+8)=(X_1+X_2+X_3+X_4)+(Y_1+Y_2+Y_3+Y_4)=\sum_{i=1}^{4} Xi+\sum_{i=1}^{4} Yi$$

3a. propiedad:

$$\sum_{i=1}^{5} 3 =3*1+3*1+3*1+3*1+3*1=3*(1+1+1+1+1)= 3*5=15$$

Ejercicios:
Desarrollar las siguientes sumas:

1.- $\sum_{i=1}^{3} (i+3) =$

2.- $\sum_{i=1}^{5} 2i+2 =$

3.- $\sum_{i=1}^{4} 3^{i} =$

4.- $\sum_{n=1}^{6} (2n/2)=$

5.- $\sum_{n=1}^{4} (2n+1)+\sum_{n=1}^{4} (2n) =$

PARAMETROS DE CENTRALIZACION O TENDENCIA CENTRAL

En cualquier experimento que se lleve a cabo siempre se tiene que decidir que parámetros se van a considerar y que precisión se quiere obtener con cada uno de ellos. Estos se pueden calcular a partir de los datos obtenidos de un experimento, cada uno de ellos ayuda a describir un aspecto del él. Un promedio nos daría idea sobre el valor medio de los datos, alrededor del cual se agrupan todos los resultados del experimento.

Definición.- Un promedio es un valor que es representativo de un conjunto de datos que pertenece a dicho conjunto. A un promedio se le conoce también como **parámetro de centralización o tendencia central.**

Los parámetros de tendencia central más comunes son:

- Media
- Mediana
- Moda

Para facilitar la comprensión de dichos parámetros de tendencia central, en el presente trabajo se calculan tanto para experimentos cuyos resultados se pueden recoger y formar un conjunto finito de números o datos; como para experimentos realizados sobre una determinada población o muestra donde los resultados se concentran en una tabla de distribución de frecuencias en intervalos o clases.

MEDIA

La media es el parámetro menos complicado de calcular con base a los datos recabados de algun fenómeno o experimento que se ha efectuado. En otras palabras se puede afirmar que es el promedio aritmético de los datos obtenidos.

Definición.- La media aritmética o media de un conjunto finito de números es la suma de dichos números dividida entre la cantidad de números del conjunto, es decir, la media del conjunto de n números $X_1, X_2, X_3, \ldots, X_n$ se obtiene:

$$X = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

donde n = número de elementos del conjunto

Cuando un conjunto de datos contiene observaciones repetidas, el cálculo de la media se puede simplificar contando el número de veces que se repite dicho dato, es decir, la frecuencia con que aparece dicho dato, y se multiplica el valor por su frecuencia respectiva.

<u>Definición</u>.- La media de una frecuencia simple o **promedio ponderado** es la suma de los números por la frecuencia con que se presentan cada uno de ellos respectivamente dividida entre la suma de las frecuencias, es decir, si los números $X_1, X_2, X_3, \ldots, X_n$ se presentan con una frecuencia $F_1, F_2, F_3, \ldots, F_n$ respectivamente, la media se obtiene de la siguiente manera:

$$X = \frac{F_1X_1 + F_2X_2 + F_3X_3 + \ldots + F_nX_n}{F_1 + F_2 + F_3 + \ldots + F_n} = \frac{\sum_{i=1}^{n} F_i X_i}{\sum_{i=1}^{n} F_i}$$

Ejemplo:

Se efectúa un examen de matemáticas a 9 alumnos para conocer el nivel de conocimientos que tienen; los resultados de dicho examen son: 3,5,6,4,8,4,8,10,8. Se desea saber ¿cuál es el promedio de conocimientos de los alumnos?
Se calcula la media del siguiente conjunto de números: {3,5,6,4,8,4,8,10,8}

Método:

1) Se arregla el conjunto en orden ascendente, es decir, de valor menor a mayor: {3,4,4,5,6,8,8,8,10}

2) Se calcula la media: $\bar{X} = \frac{3+4+4+5+6+8+8+8+10}{9} = \frac{56}{9} = 6.22$

3) Otra forma de calcular la media de frecuencia simple o promedio ponderado es:

$$\bar{X} = \frac{(1*3)+(2*4)+(1*5)+(1*6)+(3*8)+(1*10)}{9} = \frac{3+8+5+6+24+10}{9} = \frac{56}{9} = 6.22$$

Ejercicios:

Calcular la media de los siguientes conjuntos de números:

1) { 3,6,4,3,2,9,3,6,5,1,11,9}

2) {4,7,12,7,4,11,15,13,11,10}

3) {5,7,3,1,2,4,8,9,10,8,3,5,7,6,7}

4.- Un padre de familia compró zapatos a sus 6 hijos. El costo de cada par de zapatos fue de N\$89.00, N\$110.00, N\$125.00, N\$150.00, N\$200.00, N\$250.00. El hombre quiere saber ¿cuánto gastó en promedio en la compra de los zapatos?

MEDIA DE UNA DISTRIBUCION DE FRECUENCIAS AGRUPADA

Cuando se tiene un conjunto de datos obtenidos utilizando un instrumento de medición o una escala continua como resultado de algún experimento, éstos se pueden agrupar en una tabla de distribución de frecuencias. Para esto los datos se clasifican dentro de intervalos de clase que se consideran coincidentes con las marcas de clase o puntos medios de cada uno de ellos, ya que ésto facilita el cálculo de todos los elementos de cada clase.

La media de los datos de una tabla de distribución de frecuencias se calcula de la siguiente manera:

$$\bar{X} = \frac{\sum_{j=1}^{n} F_j X_j}{\sum_{j=1}^{n} F_j} = \frac{F_1 X_1 + F_2 X_2 + \dots + F_n X_n}{F_1 + F_2 + \dots + F_n}$$

donde:

X_j es la marca de clase del j-ésimo intervalo de clase
F_j la frecuencia del j-ésimo intervalo de clase
$\sum F_j$ el total de las frecuencias

Al utilizar esta fórmula se ha supuesto que todas las observaciones de una clase son iguales al punto medio de dicha clase o intervalo, lo que naturalmente no es correcto. No obstante se observa que los errores introducidos de este modo tienden a compensarse unos con otros. En general, se dan a algunas observaciones de la clase un valor mayor que el verdadero y a otras uno menor, el efecto de éstas aproximaciones cuando se calcula el promedio en todas las clases es muy pequeño.

EJEMPLO:

Utilizando los datos de la Tabla 2 de distribución de frecuencias, calcular la media de la distribución:

Intervalos	Intervalos de clase	Frec. Absoluta	Frec. relativa	Marcas de clase	frec. * marca de clase
14 - 30	13.5 - 30.5	5	.1	22	5*22 = 110
31 - 47	30.5 - 47.5	19	.38	39	19*39 = 741
48 - 64	47.5 - 64.5	21	.42	56	21*56 = 1176
65 - 81	64.5 - 81.5	5	.1	73	5*73 = 365
	total	50			2392

Método:

1) Se obtiene la frecuencia y la marca de clase de cada intervalo de clase (aparece en la tabla)

2) Se multiplica la frecuencia por la marca de clase de cada intervalo (aparece en la tabla)

3) Se suman todas las multiplicaciones del inciso 2) y se dividen entre la suma de las frecuencias

$$\bar{X} = \frac{5*22 + 19*39 + 21*56 + 5*73}{5 + 19 + 21 + 5} = \frac{2392}{50} = 47.84$$

Otra forma de calcular la media:

Método:

1) Se obtiene la frecuencia relativa y la marca de clase de cada intervalo (aparece en la tabla)

2) Se multiplica la frecuencia relativa por la marca de clase de cada intervalo (aparece en la tabla)

3) Se suman todas las multiplicaciones y se divide la suma entre el total de las frecuencias relativas

Intervalos de clase	Frec. relativa	Marcas de clase	frec.rel. * marca de clase
13.5 - 30.5	.1	22	.1 *22 = 2.2
30.5 - 47.5	.38	39	.38*39 = 14.82
47.5 - 64.5	.42	56	.42*56 = 23.52
64.5 - 81.5	.1	73	.1 *73 = 7.3
total	1.00		47.84

$$X = \frac{.1*22 + .38*39 + .42*56 + .1*73}{1} = \frac{2.2 + 14.82 + 23.52 + 7.3}{1} = \frac{47.84}{1} = 47.84$$

MODA

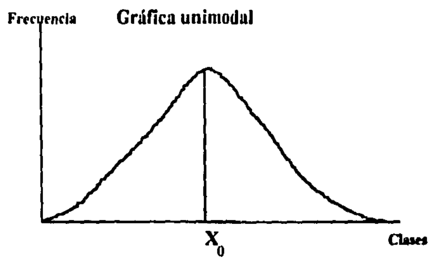
Otro de los parámetros de tendencia central que se aplica sobre los datos obtenidos de un experimento es la moda, cuando se grafican los resultados mediante un polígono de frecuencias la moda o modas -puede haber más de una- se observan como los puntos máximos de la curva de la distribución de frecuencias.

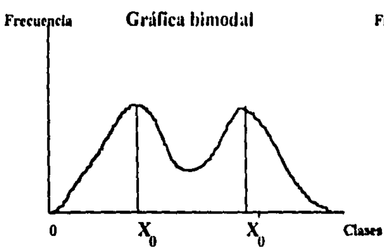
Definición.-La **moda** de un conjunto de números es el valor del conjunto que aparece con mayor frecuencia, es decir, es el valor que se presenta más número de veces.

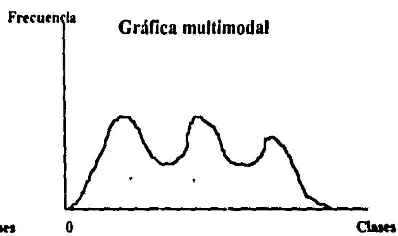
Si ningún número aparece más de una vez entonces no hay valor modal, es decir, la moda puede no existir. Por otra parte si el resultado de un experimento siempre es el mismo dato tampoco existe moda para ese conjunto de resultados.

Sin embargo si varios números aparecen con igual frecuencia y más veces que los restantes entonces estos serán modas, de esta manera es posible que un conjunto de números tenga más de una moda.

Cuando se presenta una sola moda en la distribución, a ésta se le llama unimodal. Si se encuentran dos modas se le llama bimodal y cuando hay más de dos modas se le llama a la distribución multimodal.







Ejemplo:

1.- En un consultorio médico se quería saber en promedio cuantos pacientes se atendian en una semana. De todas las citas concertadas al día, algunas se cancelaron y otras no, tomando ésto en cuenta se obtuvieron los siguientes resultados en una semana de lunes a sábado: 9,7,12,5,11,5.

Método:

1) Se ordenan los números de menor a mayor (orden ascendente): {5,5,7,9,11,12}

2) Se cuenta el número de veces que se repite un mismo dato: 5 - dos veces, 7 - una vez, 9 - una vez, 11 - una vez y 12 - una vez.

3) El número que aparece más veces es la moda, en este caso: X = 5 . Es decir en promedio el mayor número de pacientes que se atendieron fue 5 en dos días.

2.- En la materia de Estadística se efectuó un examen final a 50 alumnos para determinar la calificación mayor que obtenían en promedio los alumnos que aprobaban. Los resultados del examen se clasificaron de acuerdo a la calificación obtenida en una escala de 0 a 10, dichos resultados se encuentran en la siguiente tabla de distribución. Calcular la moda de la tabla de distribución de frecuencias que presenta las calificaciones de 50 alumnos.

Calificación (categorías)	Frecuencia No.alumnos	
10	1	
9	1	
8	6	
7	3	
6	11	*
5	6	
4	9	
3	8	
2	5	
	--	
	50	

Método:

1) Como las calificaciones ya están ordenadas en forma descendente, de mayor a menor se procede al siguiente paso.

2) Se observa que frecuencia tiene cada una de las calificaciones, es decir, el número de estudiantes que obtuvieron cada calificación.

3) Se busca la calificación que hayan obtenido mayor número de estudiantes, es decir, la que tiene mayor frecuencia. En este caso X= 6.

Ejercicios:

Calcular la moda de los siguientes conjuntos de números:

1) { 3,6,4,3,2,9,3,6,5,1,11,9}

2) {4,7,12,7,4,11,15,13,11,10}

3) {5,7,3,1,2,4,8,9,10,8,3,5,7,6,7}

MODA DE UNA TABLA DE DISTRIBUCION DE FRECUENCIAS

Cuando los datos de un experimento se presentan mediante una tabla de distribución de frecuencias, todos los valores de los resultados se encuentran dentro de unos intervalos de clase. En este caso la moda se puede definir en términos de la frecuencia de las clases, para poder calcular el valor modal de la distribución se tiene primero que localizar en que clase se puede encontrar dicho valor.

Definición.-La **clase modal** es el intervalo de clase que aparecc con mayor frecuencia, es decir, el que tiene el valor más alto de las frecuencias.

Definición.-La moda de una distribución de frecuencias agrupada es el punto medio de la clase modal, cuando está es el primer o el último intervalo de clase de la distribución; en cualquier otro caso la moda será el valor (o valores) de X correspondientes al máximo (o máximos) de la curva que representa la distribución (polígono de frecuencias), y se obtiene mediante la siguiente fórmula:

$$\text{Moda} = L_1 + \left(\frac{\delta_1}{\delta_1 + \delta_2}\right) C$$

donde:

L_1= Límite real inferior de la clase modal
δ_1 = Exceso de la frecuencia modal sobre la frecuencia de la clase contigua inferior a la clase modal
δ_2 = Exceso de la frecuencia modal sobre la frecuencia de la clase contigua superior a la clase modal
C = Longitud de la clase modal = Límite superior - límite inferior

Se vuelve a reiterar que en caso de que la clase modal fuera el primer intervalo de clase de la tabla de distribución de frecuencias, la moda es simplemente el punto medio de la clase modal, es decir, su marca de clase.

Analógamente en el caso de que la clase modal fuera el último intervalo de clase de la tabla de distribución de frecuencias, la moda es el punto medio de la clase modal, es decir, su marca de clase.

Ejemplo:

Calcular la moda de la tabla de distribución de frecuencias que se presenta a continuación:

Intervalos	Intervalos de clase	Frec.	Marcas de clase	frec. * marca de clase
14 - 30	13.5 - 30.5	5	22	5*22 = 110
31 - 47	30.5 - 47.5	19	39	19*39 = 741
48 - 64	47.5 - 64.5	21	56	21*56 = 1176
65 - 81	64.5 - 81.5	5	73	5*73 = 365
	total	50		2392

Método:

1) Se localiza la clase modal, es decir la clase que tiene mayor frecuencia: 47.5 - 64.5

La clase contigua inferior a la clase modal es 30.5 - 47.5 y su frecuencia es 19

La clase contigua superior a la clase modal es 64.5 - 81.5 y su frecuencia es 5

2) Se toma el límite real inferior de la clase modal: $L_1 = 47.5$

3) Se calcula: $\delta_1 = 21 - 19 = 2$

$\delta_2 = 21 - 5 = 16$

$C = 47.5 - 30.5 = 17$

4) Se sustituyen los valores calculados en la fórmula:

$$\text{Moda} = 47.5 + \left(\frac{2}{2+16}\right) 17 = 47.5 + 1.87 = 49.37$$

MEDIANA DE UN CONJUNTO DE NUMEROS

Una medida de tendencia central que se puede calcular para poder interpretar los resultados de un experimento para indicar aproximadamente cuál es el dato que se encuentra a la mitad de los demás valores obtenidos, es la mediana.

La mediana de un conjunto de datos ordenados es el valor que divide a los elementos de dicho conjunto a la mitad, ya que los datos en una parte del conjunto son menores o iguales que el valor medio y los datos en la otra mitad son mayores o iguales que dicho valor medio.

Definición.-La **mediana** de un conjunto de números es aquel que se encuentra en la mitad del conjunto; es decir, cuando los elementos de un conjunto están ordenados de acuerdo a su mágnitud, - orden ascendente o descendente -, se considera a la mediana como la medida de tendencia central que corta en dos partes iguales a estos elementos.

A continuación se identifican los dos casos que se pueden encontrar de acuerdo a los datos de una variable discreta, obtenidos de un experimento que sirven para integrar un conjunto.

Si se tiene un número impar de datos en el conjunto entonces la mediana será el dato que cae exactamente en la mitad del conjunto. La posición del valor de la mediana se puede localizar por inspección o se emplea la siguiente fórmula:

$$\text{Posición} = \frac{N+1}{2} \qquad \text{donde } N = \text{número impar de casos}$$

Si el número de datos en el conjunto es par entonces la mediana será la media aritmética de los dos valores medios o centrales del conjunto. Si hay varios números medios de idéntico valor entonces la mediana será el valor de dicho número.

Ejemplo:

1.- Se desea saber cual es la "mitad" de litros de leche que se consumen en un comedor estudiantil que sirve desayunos durante una quincena, para esto se consideran los litros que se consumen al día durante 12 días hábiles y se obtienen los siguientes datos: {10,7,2,9,10,18,11,12,2,9,5,9}. Calcular la mediana del conjunto de datos.

Método:
1)Se ordenan los elementos del conjunto en forma ascendente: {2,2,5,7,9,9,9,10,10,11,12,18}

2) Se localiza la posición de la mediana del conjunto:

$$\text{Posición} = \frac{12+1}{2} = 6.5$$

3) Como el número de casos es par la mediana se calcula entre los 2 valores centrales del conjunto, si es el mismo valor como en este caso, la mediana será dicho valor:

$$\text{Mediana} = \frac{9+9}{2} = 9$$

Ejercicios. Calcular la mediana de los siguientes conjuntos de números:

1) { 3,6,4,3,2,9,3,6,5,1,11,9}

2) {4,7,12,7,4,11,15,13,11,10}

3) {5,7,3,1,2,4,8,9,10,8,3,5,7,6,7}

MEDIANA DE UNA TABLA DE DISTRIBUCION DE FRECUENCIAS

Cuando los datos de un experimento se agrupan en una tabla de distribución de frecuencias, todos los valores de los resultados del experimento se encuentran dentro de unos intervalos de clase o clases. Si se desea obtener la mediana de la distribución, es decir, encontrar el valor que se encuentra aproximadamente a la "mitad" de la distribución dicho dato se encuentra en un intervalo de clase que análogamente se encuentra a la mitad de la distribución de frecuencias.

Definición.-La **clase mediana** es el intervalo de clase que contiene a la mediana, es decir, es el intervalo de la distribución bajo el cual se encuentra aproximadamente el 50% de la frecuencia y sobre el cual se encuentra aproximadamente el otro 50% de la frecuencia.

Definición.-La mediana de una distribución de frecuencias agrupada se obtiene mediante la siguiente fórmula:

$$\text{Mediana} = L_1 + \left(\frac{\frac{N}{2} - F}{F_1}\right) C$$

donde:

L_1 = Límite real inferior de la clase mediana

N = Total de las frecuencias

F = Suma de las frecuencias de todos los intervalos de clase que se encuentran antes de la clase mediana, es decir, la frecuencia acumulada de la clase inmediata inferior a la clase mediana

F_1 = Frecuencia de la clase mediana

C = Tamaño del intervalo de la clase mediana = límite superior - límite inferior

Geométricamente la mediana es el valor X sobre el eje de las abscisas -que es donde se colocan las clases- desde donde se levanta una vertical que divide a la distribución en dos partes iguales, es decir, si se gráfica el polígono de frecuencias de la distribución éste queda dividido en dos partes de igual área.

Ejemplo:

Calcular la mediana de la siguiente tabla de distribución de frecuencias.

Intervalo	Intervalos de clase	Frec.	Frecuencia acumulada	Marca de clase	frec*marca de clase
14 - 30	13.5 - 30.5	5	5	22	5*22= 110
31 - 47	30.5 - 47.5	19	24	39	19*39= 741
48 - 64	47.5 - 64.5	21	45	56	21*56=1176
65 - 81	64.5 - 81.5	5	50	73	5*73= 365
	total	50			2392

Método:

1) Se localiza la clase mediana, es decir, la clase que aproximadamente divide en dos partes iguales a la frecuencia: 47.5 - 64.5

2) Se toma el límite real inferior de la clase mediana: $L_1 = 47.5$

3) Se calcula: $N = 50$

$F = 24$

$F_1 = 21$

$C = 47.5 - 30.5 = 17$

3) Se sustituyen los valores calculados en la fórmula:

$$\text{Mediana} = 47.5 + \left(\frac{\frac{50}{2} - 24}{21}\right) 17 = 47.5 + 0.68 = 48.18$$

Ejercicios:

1.- En una encuesta realizada a 12 familias para conocer el número de hijos que tienen se obtuvieron los siguientes datos: {3,2,2,2,3,2,1,4,5,4,5,5}.

a)Calcular el número promedio de hijos de las 12 familias
b)Cuál es el número máximo de hijos de las familias? ¿es único?
c) Cuál es el número medio de hijos?
d) Seleccionar el diagrama adecuado para representar los datos obtenidos.

2.- En la siguiente tabla de distribución de frecuencias se concentran los datos de los pesos en kg. de 50 personas:

Intervalos	Intervalos de clase	Frecuencia absoluta	Frecuencia relativa	Marcas de clase	Frec. acumulada
33 - 45	32.5-45.5	8	8/50	39	8
46 - 58	45.5-58.5	9	9/50	52	17
59 - 71	58.5-71.5	14	14/50	65	31
72 - 84	71.5-84.5	13	13/50	78	44
85 - 97	84.5-97.5	6	6/50	91	50
	Total	50			

a) Calcular el peso promedio de las personas.
b) Calcular el peso máximo de las personas.
c) Calcular la mediana de los pesos de las personas.
d) Hacer el histograma y el polígono de frecuencias de la tabla de distribución.

3.- En una muestra elegida al azar de 50 personas se observó que el númer de hermanos (hombres y/o mujeres) que tenían esas personas fueron:

7	6	1	7	7	7	2	3	6	1
3	8	1	6	3	6	8	6	8	8
2	2	8	6	4	8	6	5	5	6
8	0	1	4	4	9	2	8	6	5
3	3	3	8	0	6	3	7	3	4

a) Obtener el rango
b) Dibujar los histogramas y los polígonos de frecuencia (simple y acumulada)
c) Elaborar una tabla de frecuencias con intervalos de ancho 3
d) Calcular la media, la moda y la mediana de la tabla de frecuencias.

4.- Del censo de 1990 se tomaron datos sobre 60 hombres para anotar el estado civil de cada uno de ellos. La codificación adoptada fue la siguiente: CR- casado civil y religiosamente, C- casado por lo civil, R- casado religiosamente, UL- unión libre, SO- soltero, V- viudo, S- separado, D- divorciado. Los datos se presentan en el siguiente arreglo:

CR	C	SO	V	S	CR	V	V	UL	SO	S	CR
V	CR	V	R	CR	S	CR	CR	S	V	S	CR
CR	S	CR	CR	V	UL	C	C	S	UL	S	V
SO	S	S	CR	D	CR	V	CR	SO	S	SO	S
C	SO	V	CR	CR	CR	CR	CR	SO	S	UL	C

a) Construya la tabla de frecuencias simple
b) Obtenga la media, la moda y la mediana
c) Dibuje alguna representación gráfica que muestre el comportamiento de los datos reportados

5.- En un anuario estadístico de 1992 de la Secretaría de Industria y Comercio se obtuvieron datos de 72 personas sobre causas de suicidio en la República Mexicana. De la muestra 26 fueron mujeres y 46 hombres. Las siglas y los datos fueron los siguientes. Causas de suicidio: A- amorosas, E- económicas, DF- disgustos familiares, EG- enfermedades graves e incurables, EM- enajenación mental, IA- intoxicación por alcohol, ID- intoxicación por drogas, R- remordimientos, O- otras, SI- se ignora. Los datos por sexos fueron:

HOMBRES: ID, R, O, A, DF, DF, EG, O, SI, SI, SI, IA, EM, EM, EG, DF, A, A, E, SI, SI, DF, EG, EG, EG, IA, E, SI, SI, IA, EM, DF, DF, EG, A, A, SI, E, DF, SI, SI, IA, EM, EG, EG, IA.

MUJERES: E, DF, DF, EG, SI, SI, O, R, SI, ID, IA, SI, A, EM, DF, SI, SI, A, SI, DF, EG, EM, A, IA, SI, EG.

Los incisos a) y c) realícelos para mujeres, hombres y ambos.
a) Calcular la media, la moda y la mediana.
b) Construir la tabla de frecuencias (simples y acumuladas)
c) Dibujar los histogramas y polígonos de frecuencias (por sexo)

RELACIONES ENTRE LA MODA, MEDIANA Y MEDIA

Por "tendencia central" se indica la tendencia de los datos de cualquier experimento a "concentrarse" alrededor de un valor particular, - o en una categoría en particular -, más que dispersarse en el rango de datos, - o en todas las categorías establecidas-. Para la elección de una medida de tendencia central específica influyen varios factores entre los cuales están:

a) El objetivo de la investigación
b) El nivel de medición de las variables
c) La forma de distribución de los datos

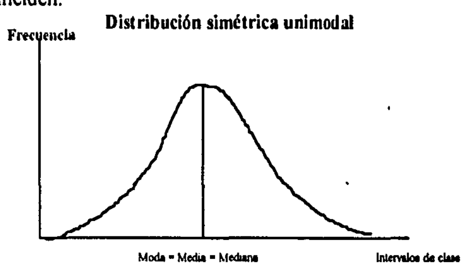
Por ejemplo para la investigación del medio de transporte utilizado por 50 estudiantes para llegar al plantel oriente:

categorías	**frecuencia**	**frecuencia relativa %**
a pie	15	30%
autobus	12	24%
microbus	9	18%
bicicleta	6	12%
coche	5	10%
taxi	3	6%
	---	------
total	50	100%

La medida de tendencia central que se utilizaría para los datos de las categorías sería la moda, "a pie" sería la categoría modal. Pero la moda sería la medida de tendencia central menos usada cuando los datos se obtuvieran con variables numéricas, en ese caso la medida del promedio que se utilizaría sería la media.

Distribuciones simétricas

La forma de la curva de una distribución de frecuencias representada por su polígono de frecuencias es un factor que puede influir en la elección de la medida de tendencia central. En una distribución unimodal perfectamente simétrica la moda, la mediana y la media serán idénticas ya que el punto de máxima frecuencia (moda) es también el punto que divide a la población en dos partes iguales (mediana), así como el "centro de gravedad" de la misma (media), es decir, los tres parámetros coinciden.



$$X_0 = \bar{X} = \tilde{X}$$

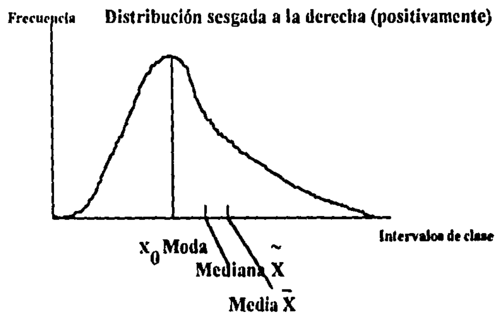
Distribuciones sesgadas

En una distribución sesgada la moda, la mediana y la media no coinciden, a pesar de que sus posiciones relativas permanecen constantes, alejándose del punto máximo de la curva (polígono de frecuencias) y acercándose al sesgo de la misma. El orden en que se presentan estas medidas de tendencia central es siempre el mismo: moda, mediana y media. Por medio de estas posiciones relativas se puede saber si la curva de frecuencias está sesgada hacia la derecha (positivamente) o hacia la izquierda (negativamente).

En las distribuciones sesgadas, la media puede no ser una medida muy descriptiva, es decir, cuando hay valores muy altos o valores muy bajos el valor que representa a la media puede dar la impresión visual de que no se encuentra en el "centro" de la curva. En general la mediana es una mejor medida de localización que la media en distribuciones muy sesgadas.

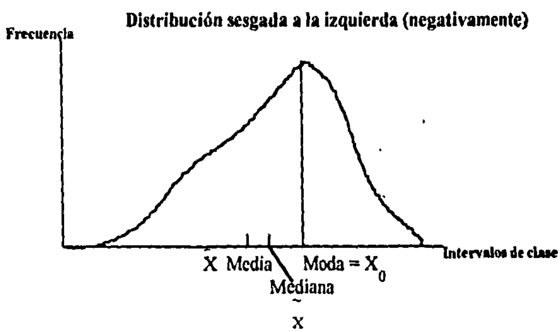
Curva sesgada hacia la derecha, es decir positivamente

Posición relativa de los parámetros: moda < mediana < media



Curva sesgada hacia la izquierda, es decir negativamente.

Posición relativa de los parámetros: media < mediana < moda



Para curvas de frecuencias unimodales que no son demasiado asimétricas, es decir, que son moderadamente sesgadas, existe una relación entre éstos tres promedios:

$$(\text{media} - \text{moda}) \equiv 3(\text{media} - \text{mediana})$$

En curvas de distribución de frecuencias multimodales, es decir, que tienen más de una moda o valores máximos, el uso de la mediana o la media podría obscurecer aspectos importantes en tales distribuciones.

Ejercicios:

1.- Los siguientes datos representan el número de puntos anotados por un alumno en un juego de basquetbol durante los 4 períodos del juego {2,3,0,5,4,3,1,3,5,2}.
Calcular la media, la moda, la mediana

2.- En una colonia del sur de la ciudad se seleccionaron 25 familias para investigar el número de vehículos que tienen, los resultados se muestran en la siguiente tabla:

Número de vehículos	0	1	2	3	4
Número de familias	1	2	5	11	6

a) Cuál es el número promedio de vehículos que tienen las familias?
b) Cuál es el número máximo de vehículos que tienen las familias?
c) Cuál es número medio de vehículos que tienen las familias?

3.- Las siguientes cifras corresponden al peso en kilogramos de 1000 bultos de papel que salen de una fábrica productora de papel bond:

Bultos	Peso	Bultos	Peso
6	57.5 - 59.5	190	69.5 - 71.5
13	59.5 - 61.5	175	71.5 - 73.5
40	61.5 - 63.5	70	73.5 - 75.5
96	63.5 - 65.5	40	75.5 - 77.5
175	65.5 - 67.5	10	77.5 - 79.5
180	67.5 - 69.5	5	79.5 - 81.5

a) Elaborar una tabla de distribución de frecuencias, un histograma y un polígono de frecuencias
b) Calcular la media, la mediana y la moda.

4.- Las siguientes temperaturas máximas y mínimas en grados centígrados durante el mes de mayo en Monterrey de 1986 fueron:

<u>Máximas</u>:	50	49	45	49	52	51	49	51	48	40	41	39	41	47	40
	51	48	43	46	38	47	38	47	46	52	52	47	43	42	49
<u>Mínimas</u>:	33	30	29	28	27	29	30	30	31	28	28	28	27	29	25
	25	27	29	30	31	31	30	32	28	27	26	27	29	28	29

De acuerdo con los datos determinar:
a) La media de las temperaturas máximas y mínimas
b) La mediana de las temperaturas máximas y mínimas
c) La moda de las temperaturas máximas y mínimas
d) Una tabla de frecuencias de las temperaturas máximas y mínimas
e) Dibujar un histograma y un polígono de frecuencias de las temperaturas máximas y mínimas

EVALUACION PROPUESTA

1.- Cuáles son las medidas de tendencia central?:
a) rango,media,mediana   b) media,desviación media,moda,mediana
c) media,mediana,moda   d)desviación media,desviación estándar,varianza

2.- Cuál es la medida de tendencia central en donde la curva de distribución alcanza un dato máximo de un experimento?
a) media   b) mediana   c)moda   d)rango

3.- Si las medidas de tendencia central se encuentran en el siguiente orden media 〉 mediana 〉 moda entonces la curva de la distribución tiene un:
a) sesgo negativo   b) sesgo positivo   c) sin sesgo   d) orden natural

4.-En una investigación realizada en 10 familias seleccionadas al azar en la región del bajío para saber el número de hijos que tienen por familia se encontraron los siguientes datos {6,2,5,8,6,4,5,6,9,10}

a) ¿Cuál es el número promedio de hijos que tienen las familias?
b) ¿Cuál es la moda del conjunto de hijos?
c) ¿Con respecto al número de hijos, que familia tenia la mitad de hijos?

5.- En una investigación realizada sobre 12 borregas se encontró que en una sola camada tienen el siguiente número de crias {2,3,1,5,7,3,6,7,3,1,7,4}.
a) Calcular el número promedio de crias que tienen las 12 borregas
b) ¿Cuál es la moda del conjunto de crías de las borregas?
c) ¿Con respecto al número de crias, cuál es la mediana del conjunto?

6.- Del censo de 1990 se tomaron datos sobre 60 hombres para anotar el estado civil de cada uno de ellos. La codificación adoptada fue la siguiente: CR- casado civil y religiosamente, C- casado por lo civil, R- casado religiosamente, UL- unión libre, SO- soltero, V- viudo, S- separado, D- divorciado. Los datos se presentan en el siguiente arreglo:

CR	C	SO	V	S	CR	V	V	UL	SO	S	CR
V	CR	V	R	CR	S	CR	CR	S	V	S	CR
CR	S	CR	CR	V	UL	C	C	S	UL	S	V
SO	S	S	CR	D	CR	V	CR	SO	S	SO	S
C	SO	V	CR	CR	CR	CR	CR	SO	S	UL	C

a) Construya la tabla de frecuencias simple
b) Obtenga la media, la moda y la mediana
c) Dibuje alguna representación gráfica que muestre el comportamiento de los datos reportados

7.- Las siguientes cifras corresponden al peso en kilogramos de 1000 bultos de papel que salen de una fábrica productora de papel bond:

Bultos	Peso	Bultos	Peso
6	57.5 - 59.5	190	69.5 - 71.5
13	59.5 - 61.5	175	71.5 - 73.5
40	61.5 - 63.5	70	73.5 - 75.5
96	63.5 - 65.5	40	75.5 - 77.5
175	65.5 - 67.5	10	77.5 - 79.5
180	67.5 - 69.5	5	79.5 - 81.5

a) Elaborar una tabla de distribución de frecuencias, un histograma y un polígono de frecuencias
b) Calcular la media, la mediana y la moda.

8.-Considerando que la siguiente tabla de distribución de frecuencias contiene la información de la estatura en centímetros de 50 niños seleccionados en un kinder:

Intervalos	Intervalos de clase	Frecuencia	Frecuencia Relativa	Frecuencia Relativa %
29 - 45	28.5 - 45.5	4	4/50	(4/50)*100
46 - 62	45.5 - 62.5	14	14/50	(14/50)*100
63 - 79	62.5 - 79.5	20	20/50	(20/50)*100
80 - 96	79.5 - 96.5	12	12/50	(12/50)*100

a) Calcular la estura promedio de los 50 niños.
b) Calcular la mayor estura de los 50 niños.
c) Calcular la estatura del niño que se encuentra aproximadamente en medio de los 50 niños.

9.- Los siguientes datos representan las edades de 30 personas encuestadas sobre los problemas de contaminación:

20	72	41	70	58	78
30	23	88	75	73	64
25	30	43	63	89	52
42	60	55	50	85	80
60	22	45	22	75	86

a) Elaborar una tabla de distribución de frecuencias en 4 intervalos, identificando frecuencia absoluta y relativa y marcas de clase.
b) Calcular la edad promedio, la edad máxima promedio y la edad que se encuentra a la mitad de las edades de las personas encuestadas.
c) Elaborar un histograma y un polígono de frecuencias de la distribución de frecuencias.

10.- Las siguientes temperaturas máximas y mínimas en grados centígrados durante el mes de mayo en Monterrey de 1986 fueron:

Máximas:	50 49 45 49 52 51 49 51 48 40 41 39 41 47 40
	51 48 43 46 38 47 38 47 46 52 52 47 43 42 49
Mínimas:	33 30 29 28 27 29 30 30 31 28 28 28 27 29 25
	25 27 29 30 31 31 30 32 28 27 26 27 29 28 29

De acuerdo con los datos determinar:

a) La media de las temperaturas máximas y mínimas
b) La mediana de las temperaturas máximas y mínimas
c) La moda de las temperaturas máximas y mínimas
d) Una tabla de frecuencias de las temperaturas máximas y mínimas
e) Dibujar un histograma y un polígono de frecuencias de las temperaturas máximas y mínimas

CONCLUSIONES

Los resultados que se obtuvieron al aplicar la evaluación correspondiente a este tema se encuentran en en la siguiente tabla de distribución de frecuencias en 4 intervalos, se obtienen las medidas de tendencia central en base a estos datos y se construye un histograma y el polígono de frecuencias correspondiente para en base a estos promedios observar la evaluación del material del tema del trabajo desarrollado.

1	3	5	6	7	7	8	9
1	3	5	6	7	8	8	9
2	3	5	6	7	8	8	9
2	4	5	6	7	8	8	10
3	5	6	6	7	8	9	10

Intervalos	Clases	Frecuencia	Frecuencia Relativa	Marca de clase	Marca * Frecuencia
0 - 2	0 - 2.5	4	4/40	1	4
3 - 5	2.5 - 5.5	10	10/40	4	40
6 - 8	5.5 - 8.5	20	20/40	7	140
9 -11	8.5 - 11.5	6	6/40	10	60
	total	40			244

$$\text{Media} = \bar{X} = \frac{244}{40} = 6.1$$

$$\text{Mediana} = \tilde{X} = 5.5 + \frac{(40/2 - 14)\,3}{20} = 5.5 + \frac{18}{20} = 6.4$$

$$\text{Moda} = X_0 = 5.5 + \left(\frac{10}{10 + 14}\right) 3 = 5.5 + 1.2 = 6.7$$

Diagrama circular de los resultados de la evaluación del tema:

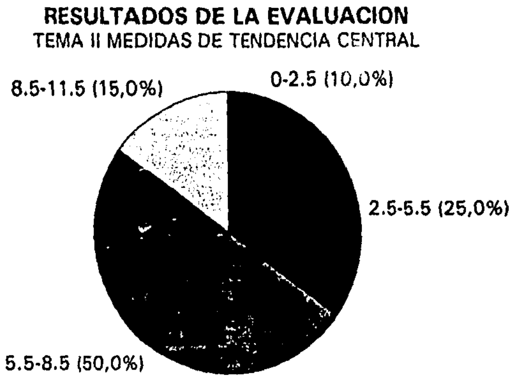
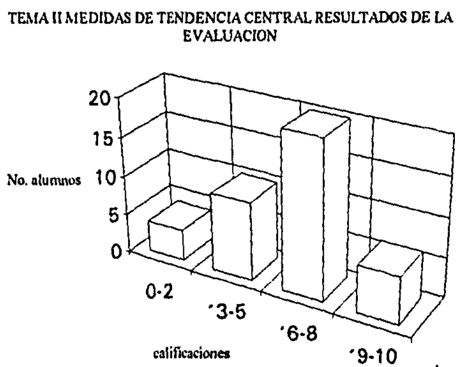
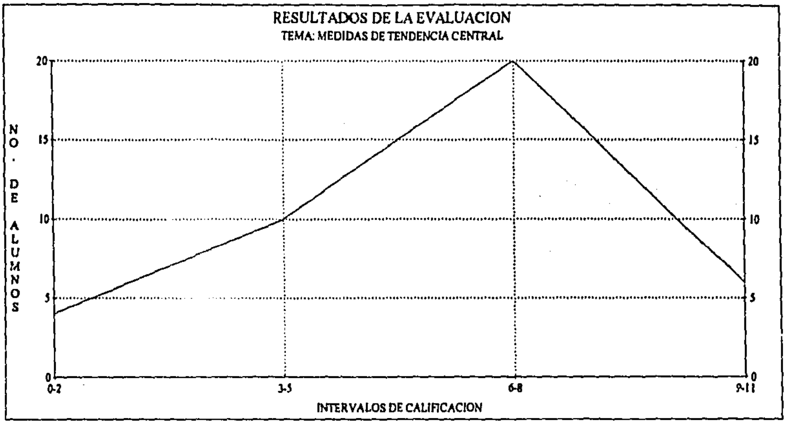


Diagrama de barras



Polígono de frecuencias



De acuerdo a los resultados se observa que más del 65% de los alumnos comprendieron los conceptos relativos al tema de medidas de tendencia central. También muestran que va en aumento el número de alumnos que no aprobaron la evaluación escrita sin embargo éstos no representan aún la mitad del grupo.

50

TEMARIO

VALOR ABSOLUTO

En el lenguaje matemáticc ..xisten simbologías que tienen distinto significado dependiendo del área de aplicación o el tema que se este tratando.

En el caso de la estadística se utilizan dos barras verticales para delimitar un valor positivo de cualquier número, es decir, para denotar el valor absoluto de ese dato númerico. Con este valor absoluto en estadística se suprimen los resultados negativos de un experimento para considerar únicamente resultados positivos.

Definición.- El valor absoluto de $a \in R$ se denota por el símbolo $|a|$, y se define como:

$$|a| = \begin{cases} a & \text{si } a \geq 0 \\ -a & \text{si } a < 0 \end{cases}$$

Es decir, si a es positivo o cero su valor absoluto es el valor númerico de a. Por otro lado si a es negativo su valor absoluto se obtiene multiplicando el valor númerico de a por menos uno.

Propiedades:

El valor absoluto tiene las siguientes propiedades que pueden ser demostradas a partir de la definición. Sean $a,b \in R$ dos números reales cualesquiera.

1) $|a| = 0$ si y sólo si $a = 0$

2) $-|a| \leq a \leq |a|$

3) $|ab| = |a|\,|b|$

4) $|a + b| \leq |a| + |b|$

Demostración de la propiedad (1):

1er caso) Si $|a| = 0$ entonces por definición se tiene que $a = 0$

2do caso) Si $a = 0$ entonces por definición tenemos que $|a| = |0| = 0$

Demostración de la propiedad (2):

1er caso) Si $a > 0$ entonces $|a| = a$ por definición, por lo tanto $-|a| = -a < a$

2do caso) Si $a < 0$ entonces $|a| = -a$ por definición, por lo tanto $-|a| = -(-a) = a$

Por lo tanto $-|a| \leq a \leq |a|$

Demostración de la propiedad (3):

1er caso) Si $a \geq 0$ y $b \geq 0$ entonces $ab \geq 0$, por lo tanto $|ab| = ab$ pero como a es positivo y b es positivo, entonces $a=|a|$ y $b=|b|$ es decir, $|ab| = ab = |a|\,|b|$

2do caso) Si $a \geq 0$ y $b<0$ entonces $|a| = a$ y $|b|=-b$ por lo tanto $|ab| = -ab = a(-b) = |a|\,|b|$

3er caso) Si $a<0$ y $b \geq 0$ la demostración es igual al caso 2.

4to caso) Si $a<0$ y $b<0$ entoces $ab > 0$ por lo tanto $|ab| = ab = (-a)(-b) = |a|\,|b|$

Demostración propiedad (4):

$|a+b| \leq |a| + |b|$

Se tiene por definición que si $a \geq 0$ entonces $a=|a|$ y que si $a<0$ entonces $a= -|a|$

por lo tanto $-|a| \leq a \leq |a|$ además $-|b| \leq b \leq |b|$

Se sigue entonces que $-(|a| + |b|) \leq a+b \leq |a| + |b|$

por lo tanto $|a+b| \leq |a| + |b|$

Ejemplos :

a) $|3| = 3$

b) $|-25| = -(-25) = 25$

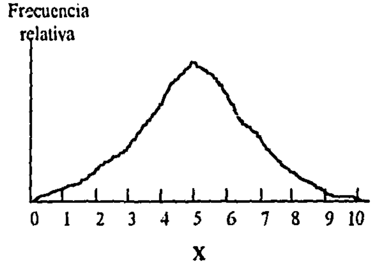
c) $|5*3|=15=5*3=|5|\,|3|$

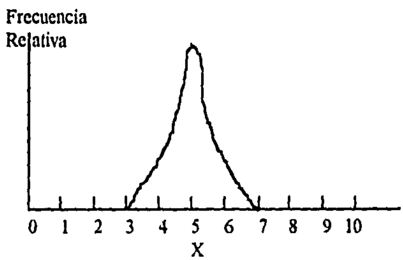
d) $|\,4+9| = |4| + |9| = 4+9 = 13$
$|\,4+9| = |13| = 13$

e) $|\,5 + (-7)| = |\,5-7\,| = |-2| = 2$
$2 = |\,5 + (-7)| \leq |5| + |-7| = 5 + -(-7) = 5 + 7 = 12$

MEDIDAS DE DISPERSION O VARIABILIDAD

Una vez localizadas las medidas de centralización o tendencia central de la distribución de los datos obtenidos de un experimento, es decir, que tan cerca se distribuyen de un valor central ("centro") de la distribución; se busca también una medida de la variabilidad o dispersión de los mismos con respecto a dicho valor central. En algunos casos la relación de los datos con la media, la mediana o la moda proporcionan igual información, en tal situación se requiere conocer "algo" más con relación a dicho conjunto de resultados. Por ejemplo en las siguientes gráficas de dos distribuciones unimodales aproximadamente simétricas, ambas se encuentran "centradas" en la media de las distribuciones X=5, pero existe una gran diferencia en la variabilidad de las observaciones con respecto a la media.





Se observa que en la primera figura la dispersión de los datos con respecto a la media (X=5) es mayor que en la segunda figura donde se concentran más al rededor de la media.

Además de una medida de tendencia central se necesita un índice de como están diseminados los resultados de un determinado experimento con respecto a una medida bien determinada sobre una población, es decir, se necesita una medida de lo que se conoce comúnmente como dispersión o variabilidad de los datos de un experimento. Las medidas de dispersión más conocidas son:

- Rango
- Desviación media
- Varianza
- Desviación estándar

Definición.-El **rango** es la diferencia entre el puntaje más alto y el más bajo de la distribución, es decir, entre los valores extremos de la distribución de los resultados de algún experimento.

El rango generalmente proporciona sólo una medida de la dispersión de los datos globalmente, no proporciona una idea precisa de la dispersión de cada dato con respecto a una medida de referencia, y en el mejor de los casos debe considerarse sólo como una medida preliminar o muy aproximada. Es muy útil en el estudio de muestras pequeñas, su gran virtud es su facilidad de cálculo ya que depende solamente de los datos más extremos.

Para obtener una medida de dispersión más confiable, se calcula "un mini-rango" más cercano al centro de la distribución. Este rango se basa en los cuartiles de la distribución, como la mediana es la medida que divide en dos partes iguales a la distribución será igual a uno de los cuartiles, ya que éstos dividen a la distribución en cuatro partes iguales como se ilustra a continuación:

	25% de las observaciones		25% de las observaciones	mediana	25% de las observaciones		25% de las observaciones	
valor mínimo		Q_1		Q_2		Q_3		valor máximo

valor mínimo______________ arreglo de datos ordenado en forma creciente ____________máximo

Para calcular cada cuartil se utiliza la misma fórmula que se utiliza para calcular la mediana:

$$\text{Mediana} = L_1 + \left(\frac{\frac{N}{2} - F}{F_1}\right) C$$

donde:

L_1 = Límite real inferior de la clase mediana
N = Total de las frecuencias
F = Suma de las frecuencias de todos los intervalos de clase que se encuentran antes de la clase mediana, es decir, la frecuencia acumulada de la clase inmediata inferior a la clase mediana
F_1 = Frecuencia de la clase mediana
C = Tamaño del intervalo de la clase mediana = límite superior - límite inferior

Simplemente para el primer cuartil Q_1 se sustituye n/4 en lugar de n/2, el segundo cuartil Q_2 es igual a la mediana y el tercer cuartil Q_3 se obtiene sustituyendo 3n/4 en la fórmula en lugar de n/2.

Ejemplo:

1.- En un salón de clases se observa la estatura en metros de 25 estudiantes, si todos tienen aproximadamente la misma edad entonces todos miden aproximadamente lo mismo y la diferencia sería mínima, pero si hay un niño que ha adelantado cursos debido a su capacidad intelectual entonces la diferencia de estatura entre éste y el más alto de los 25 alumnos no daría mucha información acerca de los estaturas restantes.

1.20	1.55	1.57	1.58	1.64
1.56	1.60	1.60	1.60	1.58
1.58	1.58	1.65	1.62	1.68
1.60	1.57	1.63	1.64	1.62
1.56	1.55	1.59	1.63	1.64

Rango = 1.68 - 1.20 = .48

el rango no proporciona información muy "verídica" acerca de la diferencia de estaturas de los 25 alumnos.

2.- En una muestra elegida al azar de 50 personas se observó que el número de hermanos (hombres y/o mujeres) que tenían esas personas fueron:

7	6	1	7	7	7	2	3	6	1
3	8	1	6	3	6	8	6	8	8
2	2	8	6	4	8	6	5	5	6
8	0	1	4	4	9	2	8	6	5
3	3	3	8	0	6	3	7	3	4

Para observar la diferencia entre el dato menor y el dato menor se ordena el conjunto de datos:

0	1	3	3	4	6	6	7	8	8
0	2	3	3	4	6	6	7	8	8
1	2	3	3	5	6	6	7	8	8
1	2	3	4	5	6	6	7	8	8
1	2	3	4	5	6	6	7	8	9

Rango = 9 - 0 = 9, el rango nos informa que la diferencia del número de hermanos es 9.

Calcular los cuartiles Q_1, Q_2, Q_3. Para poder calcular los cuartiles formamos una tabla de distribución con los datos en 3 intervalos, se usan pocos intervalos debido al rango de los datos.

Intervalos	Clases	Frecuencia	Frecuencia acumulada
0 - 3	0 - 3.5	18	18
4 - 6	3.5 - 6.5	17	35
7 - 9	6.5 - 9.5	15	50

Total		50	

$$Q_1 = 3.5 + \left(\frac{50/4 - 18}{17}\right) 2 = 3.5 + \left(\frac{-5.5}{17}\right) 2 = 3.5 - 1.28 = 2.2$$

$$Q_2 = 3.5 + \left(\frac{50/2 - 18}{17}\right) 2 = 3.5 + \left(\frac{7}{17}\right) 2 = 3.5 + .82 = 4.3$$

$$Q_3 = 3.5 + \left(\frac{150/4 - 18}{17}\right) 2 = 3.5 + \left(\frac{37.5}{17}\right) 2 = 3.5 + 4.41 = 7.9$$

DESVIACION MEDIA DE UN CONJUNTO DE NUMEROS

Si se desean conocer medidas de variación más precisas, primero se tiene que seleccionar una medida que sirva de referencia para medir la variación de los datos con respecto a ella, comunmente esta medida es la media de la distribución.

La desviación entre cualquier dato y la media de la distribución se determina como la resta entre ambos. Por lo tanto una forma sencilla de encontrar la desviación de un conjunto de datos con respecto a un estadístico -media-, es la diferencia entre cada uno de los datos y la media.

Definición.-La desviación media de una serie de **n** números $X_1, X_2, ..., X_n$ se obtiene al tomar la diferencia entre cada número X_i y la media de dichos números en valor absoluto, sumando éstas desviaciones y luego dividiendo ésta suma entre n que es el número de datos que se tienen. Es decir, se obtiene mediante la siguiente fórmula:

$$D.M. = \frac{\sum_{j=1}^{n} |X_j - \bar{X}|}{n} = \frac{|X_j - \bar{X}|}{n}$$

donde: $\bar{X}$ = media del conjunto de números

$|X_j - \bar{X}|$ = valor absoluto de las desviaciones de los distintos X_j con respecto a la media

n = número total de elementos del conjunto

Ejemplo:

1.-Un alumno que practica la natación se prepara para una competencia de 100m libres, conforme avanza en su entrenamiento desea conocer cuál es su progreso en segundos de acuerdo a su promedio de tiempo. Con este propósito tomó su progreso en segundos durante 6 días consecutivos que se muestran en el siguiente conjunto: {2,3,4,6,7,8}. Calcular la desviación media del conjunto de números.

Método:
1) Se calcula la media del conjunto de datos:

$$\bar{X} = \frac{2+3+4+6+7+8}{6} = \frac{30}{6} = 5$$

2) Se resta la media a cada uno de los elementos del conjunto y se considera la diferencia en valor absoluto:

$$|2-5| = |-3| = 3$$
$$|3-5| = |-2| = 2$$
$$|4-5| = |-1| = 1$$
$$|6-5| = |\ 1| = 1$$
$$|7-5| = |\ 2| = 2$$
$$|8-5| = |\ 3| = 3$$

3) Se suman las desviaciones absolutas y se divide la suma entre el total de elementos del conjunto:

$$D.M. = \frac{\sum_{j=1}^{6} |X_j - \bar{X}|}{6} = \frac{12}{6} = 2$$

2.- En una avenida de 9 cuadras se localizan 5 de las mejores tiendas de ropa en la 5a, 6a, 7a, 8a y 9a cuadras respectivamente. Si en la primera cuadra se encuentra la parada del "metro" y los compradores tienen que tomar un autobús para ir a las tiendas, en dondé debe hacer "parada" el autobús para que se puedan visitar todas las tiendas?

Se calcula la desviación media, en este caso la media es la cuadra 7, por lo tanto las desviaciones son:

$$D.M. = \frac{|5-7| + |6-7| + |8-7| + |9-7|}{5} = \frac{|-2| + |-1| + |1| + |2|}{5} = \frac{6}{5} = 1.2$$

el autobús tiene que parar en la tienda de enmedio, es decir, la de la cuadra 7 para que los compradores únicamente caminen en promedio cuadra y media para visitar todas las tiendas.

DESVIACION MEDIA DE UNA DISTRIBUCION DE FRECUENCIAS

Cuando los datos se presentan mediante una tabla de distribución de frecuencias todos ellos se encuentran dentro de intervalos de clase; en este caso se considera que cada intervalo queda "representado" por su correspondiente marca de clase, esto sirve tanto para calcular la media de la distribución como la desviación media de los datos.

Definición.-La desviación media de los datos que se presentan en una tabla de distribución de frecuencias se obtiene mediante la siguiente fórmula:

$$D.M. = \frac{\sum_{j=1}^{n} F_j \, |X_j - \bar{X}|}{n} = \frac{F\,|X - \bar{X}|}{n}$$

donde $\bar{X}$ = media de la distribución

X_j = marca de clase del j-ésimo intervalo de clase

F_j = frecuencia del j-ésimo intervalo de clase

$n = \sum_{j=1}^{n} F_j$ = suma total de las frecuencias

Ejemplo:

1.- Calcular la desviación media de la tabla de distribución de frecuencias cuyos datos se encuentran en la Tabla 2.

Intervalo	Intervalos de clase	Frecuencia	Frecuencia acumulada	Marcas de clase	frec*marca de clase
14 - 30	13.5 - 30.5	5	5	22	5*22= 110
31 - 47	30.5 - 47.5	19	24	39	19*39= 741
48 - 64	47.5 - 64.5	21	45	56	21*56= 1176
65 - 81	64.5 - 81.5	5	50	73	5*73= 365
		----			------
	total	50			2392

TABLA 2

Método:

1) Se calcula la media de la distribución

$$\bar{X} = \frac{2392}{50} = 47.84$$

2) Se encuentra la desviación en valor absoluto de la marca de clase de cada intervalo con respecto a la media, es decir, se resta la media a cada marca de clase y se toma el valor absoluto:

$|22 - 47.84| = |-25.84| = 25.84$

$|39 - 47.84| = |- 8.84| = 8.84$

$|56 - 47.84| = |8.16| = 8.16$

$|73 - 47.84| = |25.16| = 25.16$

3) Se multiplica cada desviación en valor absoluto por la frecuencia del respectivo intervalo de clase

5*25.84= 129.2
19*8.84 = 167.96
21*8.16 = 171.36
5*25.16= 125.8

4) Se suman dichos productos y se divide la suma entre el total de las frecuencias

$$D.M. = \frac{129.2+167.96+171.36+125.8}{5+19+21+5} = \frac{594.32}{50} = 11.88$$

2.- En una muestra elegida al azar de 50 personas se observó que el número de hermanos (hombres y/o mujeres) que tenían esas personas se observan en la siguiente tabla de frecuencias en 3 intervalos:

Intervalos	Clases	Frecuencia	Frecuencia acumulada	Marca de clase	Marca * Frecuencia
0 - 3	0 - 3.5	18	18	2	2*18 = 36
4 - 6	3.5 - 6.5	17	35	5	5*17 = 85
7 - 9	6.5 - 9.5	15	50	8	8*15 =120
		---			-----
Total		50			241

la media es $\bar{X} = \frac{241}{50} = 4.8$ es decir, en promedio las 50 familias tienen 4.8 hermanos

$$D.M. = \frac{|2-4.8| * 18 + |5-4.8| * 17 + |8-4.8| * 15}{50} = \frac{50.4+3.4+48}{50} = \frac{101.8}{50} = 2.03$$

cada familia tiene una diferencia aproximadamente de 2 hermanos con respecto al promedio de hermanos de las 50 familias.

DESVIACION ESTANDAR O DESVIACION TIPICA

La desviación media evita el problema de los números negativos que cancelan a los positivos pasando por alto los signos (+) y (-) mediante el valor absoluto, y sumando las desviaciones con respecto a la media en valor absoluto. Sin embargo existe otro procedimiento para calcular una medida de variabilidad que sea más precisa y que no tenga que calcular tales valores absolutos, esta medida de dispersión es la desviación típica o estándar que se designa por σ. En este caso también se utiliza a la media de la distribución como la medida de referencia de la dispersión de los datos.

Otra medida de variabilidad es la varianza que se denota por σ^2, que va estrechamente ligada con la desviación estándar ya que prácticamente la única diferencia entre ambas es que la varianza es la suma de las desviaciones al cuadrado de cada dato con respecto a la media entre el total de datos de la muestra, y la desviación estándar es la raíz cuadrada de esta suma, es decir, de la varianza. La varianza tiene una desventaja para propósitos de uso común y se debe a que toma el cuadrado de cada desviación, ya que por ejemplo si los datos fueran latidos del corazón por minuto la varianza consideraría latidos del corazón al cuadrado por minuto.

DESVIACION ESTANDAR DE UN CONJUNTO DE NUMEROS

<u>Definición</u>.-La desviación típica de un conjunto de N números $X_1, X_2, \ldots, X_n$ se representa por σ y se define como la raíz cuadrada de las desviaciones de cada elemento del conjunto con respecto a la media de la distribución elevadas al cuadrado, es decir:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}}$$ donde $X_i - \bar{X}$ son las desviaciones de cada número X_i de la media $\bar{X}$

La desviación estándar es la medida de variabilidad más utilizada, debido a que sumar desviaciones con respecto a la media al cuadrado se puede efectuar más rápidamente que considerando signos positivos y negativos.

Ejemplo:
1.- Calcular la desviación estándar del siguiente conjunto de números: {2,3,4,6,7,8}.

Método:

1) Se calcula la media del conjunto: $\bar{X} = 5$

2) Se obtienen las desviaciones de cada elemento con respecto a la media, y se eleva al cuadrado cada desviación:

$(2-5)^2 = (-3)^2 = 9$
$(3-5)^2 = (-2)^2 = 4$
$(4-5)^2 = (-1)^2 = 1$
$(6-5)^2 = (1)^2 = 1$
$(7-5)^2 = (2)^2 = 4$
$(8-5)^2 = (3)^2 = 9$

3) Se obtiene la suma de todos estos cuadrados y se divide entre n (total de elementos), de esta manera se obtiene lo que se conoce como VARIANZA:

$$Var(X) = \sigma^2 = \frac{\sum_{i=1}^{6} (X_i - \bar{X})^2}{n} = \frac{28}{6} = 4.66$$

4) Se extrae raíz cuadrada a la varianza para obtener lo que se conoce como desviación estándar:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{6} (X_i - \bar{X})^2}{n}} = \sqrt{4.66} = 2.16$$

DESVIACION ESTANDAR DE UNA DISTRIBUCION DE FRECUENCIAS

Cuando los datos se presentan mediante una tabla de distribución de frecuencias todos estos valores se encuentran dentro de intervalos de clase (clases).

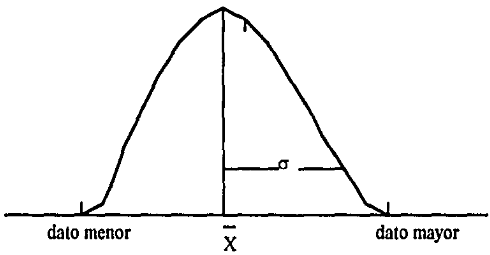
Definición.-La **desviación estándar** o típica de datos que se presentan en una tabla de distribución de frecuencias se obtiene de la siguiente manera: si X_1, X_2,..., X_n son los puntos medios de los intervalos de clase y F_1, F_2,..., F_n son las frecuencias de cada intervalo respectivamente, la desviación estándar se obtiene mediante la siguiente fórmula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} F_i (X_i - \bar{X})^2}{\sum_{i=1}^{n} F_i}}$$

donde X_i = marca de clase del i-ésimo intervalo
F_i = frecuencia del i-ésimo intervalo

$\bar{X}$ = media de la distribución

Se puede observar de forma gráfica la interpretación de la dispersión estándar, como la "abertura" de la curva de la distribución con respecto a la media.



Ejemplos:

1.- Obtener de desviación estándar y la varianza de la siguiente tabla de distribución de frecuencias:

Intervalos	Intervalos de clase	Frecuencia	Frecuencia acumulada	Marcas de clase	frec*marca de clase
14 - 30	13.5 - 30.5	5	5	22	5*22= 110
31 - 47	30.5 - 47.5	19	24	39	19*39= 741
48 - 64	47.5 - 64.5	21	45	56	21*56=1176
65 - 81	64.5 - 81.5	5	50	73	5*73= 365
	total	50			2392

Método:

1) Se obtienen las desviaciones de cada marca de clase con respecto a la media y se eleva al cuadrado cada desviación:

$(22 - 47.84)^2 = (-25.84)^2 = 658.40$
$(39 - 47.84)^2 = (- 8.84)^2 = 78.15$
$(56 - 47.84)^2 = (8.16)^2 = 66.58$
$(73 - 47.84)^2 = (25.16)^2 = 26.62$

2) Se multiplica cada desviación al cuadrado por la respectiva frecuencia de su correspondiente intervalo de clase

658.40*5 = 3292
78.15*19= 1484.8
66.58*21= 1398.18
26.62*5 = 133

3) Se obtiene la suma de éstas multiplicaciones y se divide entre la suma de las frecuencias, obteniendose de ésta manera la varianza

$$Var(X) = \sigma^2 = \frac{\sum_{i=1}^{4} F_i (X_i - \bar{X})^2}{\sum_{i=1}^{4} F_i} = \frac{6308}{5+19+21+5} = \frac{6308}{50} = 126.16$$

4) Se extrae la raíz cuadrada de la varianza para obtener la desviación estándar:

$$\sigma(x) = \sqrt{126.16} = 11$$

2.- Las siguientes temperaturas máximas en grados centígrados durante el mes de mayo en Monterrey de 1986 fueron:

<u>Máximas</u>:	50	49	45	49	52	51	49	51	48	40	41	39	41	47	40
	51	48	43	46	38	47	38	47	46	52	52	47	43	42	49

De acuerdo con los datos determinar:

a) La media de las temperaturas máximas
b) Una tabla de frecuencias de las temperaturas máximas (en 4 intervalos)
c) La desviación estándar de las temperaturas máximas

rango= 52-38=14
tamaño = 14/4 = 3.5

intervalos	clases	marca de clase	frecuencia	marca * frecuencia	marca - media	(marca-media)²
38 - 41	37.5 - 41.5	39.5	7	39.5* 7= 276.5	39.5 - 46= -6.5	(-6.5)²= 42.25
42 - 45	41.5 - 45.5	43.5	4	43.5* 4= 174.0	43.5 - 46= -2.5	(-2.5)²= 6.25
46 - 49	45.5 - 49.5	47.5	12	47.5*12= 570.0	47.5 - 46= 1.5	(1.5)²= 2.25
50 - 53	49.5 - 53.5	51.5	7	51.5* 7= 360.5	51.5 - 46= 5.5	(5.5)²= 30.25
Total			30	1381.0		

$\bar{X} = 1381 / 30 = 46$

$$\sigma^2 = \frac{42.25 * 7 + 6.25 * 4 + 2.25 * 12 + 30.25 * 7}{30} = \frac{295.75+25+27+211.75}{30} = \frac{559.5}{30} = 18.65$$

$\sigma = \sqrt{18.65} = 4.3$

APLICACIONES

El matemático ruso Tchebysheff presentó un resultado que permite encontrar un significado práctico a la desviación estándar.

<u>Teorema</u>.- Sea k un número mayor o igual a 1, se tiene un conjunto de n datos $X_1, X_2, \ldots, X_n$ de un experimento entonces al menos $(1-1/k^2)$ de los datos se encuentran dentro de k desviaciones estándar de la media.

El término "al menos" se refiere a que el resultado es un tanto conservador ya que se aplica a cualquier conjunto de observaciones.

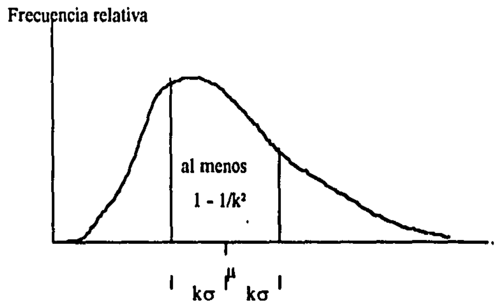


Ilustración del teorema de Tchebysheff

En el diagrama se construye un intervalo midiendo una distancia $k\sigma$ a la iquierda y a la derecha de la media de los datos μ.

Para k=1 el teorema afirma que $1-1/(1)^2 = 1-1/1 = 1-1= 0$ de los datos se encuentran en el intervalo $(\mu-\sigma, \mu+\sigma)$, en este caso no se proporciona ninguna información.

Para k=2 tenemos $1-1/(2)^2 = 1-1/4 = 3/4$ de los datos se encuentran en el intervalo $(\mu-2\sigma, \mu+2\sigma)$.

Para k=3 tenemos $1-1/(3)^2 = 1- 1/9 = 8/9$ de los datos se encuentran en el intervalo $(\mu-3\sigma, \mu+3\sigma)$.

Existen otras propuestas para medir la variabilidad de los datos que tienen una distribución normal y de distribuciones aproximadamente iguales a la normal, dicha variabilidad se describe adecuadamente mediante la siguiente regla empírica:

$(\mu-\sigma, \mu+\sigma)$ contiene aproximadamente el 68% de los datos
$(\mu-2\sigma, \mu+2\sigma)$ contiene aproximadamente el 95% de los datos
$(\mu-3\sigma, \mu+3\sigma)$ contiene aproximadamente el 99% los datos.

Ejemplo:

1.-En las escuelas de administración de la Cd. de México se requiere que los alumnos que solicitan admisión presenten un examen. El promedio de este examen es de $\mu=480$ con una desviación estándar de $\sigma = 100$. Por experiencia se ha encontrado que la curva de la distribución de las calificaciones presenta una forma normal (de campana).

a) ¿Qué porcentaje de las calificaciones se espera encontrar en el intervalo (380,580)?

De acuerdo a la regla empírica el 68% de las calificaciones se encuentra en este intervalo, ya que $\mu-\sigma = 480-100 = 380$ y $\mu+\sigma = 480+100 = 580$, por lo tanto $(\mu-\sigma, \mu+\sigma) = (380,580)$

b) Una de las escuelas con más prestigio sólo considera una solicitud si es acompañada por una calificación mayor o igual a 680. ¿Qué porcentaje de las personas que presentan el examen cumplen con este requisito?

Como $\mu = 480$ y $\sigma = 100$ entonces $\mu+2\sigma = 480+2(100) = 480+200 = 680$
y únicamente se consideran las calificaciones mayores o iguales a 680, es decir, las calificaciones no deben estar en el intervalo $(\mu-2\sigma,\mu+2\sigma) = (280,680)$ entonces solamente el 5% de los aspirantes cumplen con este requisito.

Ejercicios:

1.-Calcular la varianza y la desviación estándar de los siguientes conjuntos de datos:

a) {2,1,3,0,4,2,1,3}
b) {4,1,0,3,6,4,5,1,2}
c) {0,-2,-5,-3,2,-4,3,-1}

2.-Los préstamos hipotecarios para comprar una vivienda tienen distintas tasas de amortización dependiendo de la ubicación del tipo de vivienda, del sueldo promedio del comprador y otros factores. Una muestra de n=25 tasas de amortización para préstamos hipotecarios se presentan a continuación:

5.2	6.0	7.5	8.0	5.0
7.9	6.6	9.2	7.4	6.5
4.6	8.5	5.5	9.3	9.5
6.5	7.5	6.5	8.1	8.2

a) Construir una tabla de distribución de frecuencias en 3 intervalos. Identificando la frecuencia absoluta, la frecuencia relativa, las marcas de clase.
b) Contruir un histograma de frecuencias relativas para estos datos.
c) Calcular la media de la distribución, la mediana y la moda.
d) Calcular la varianza y la desviación estándar de la distribución.

3.- Una planta refinadora de gasolina ha realizado n=10 pruebas experimentales para comparar un aditivo nuevo con la marca regular que se encuentra ya en el mercado. Los incrementos en Km por litro en las 10 pruebas son los siguientes:

1.5, 3.4, 0.2, 1.6, 2.0, 1.8, 2.9, 3.7, 0.4, 2.6

a) Calcular la media, el aumento promedio en km por litro debido al nuevo aditivo.
b) Calcular la varianza y la desviación estándar de los datos.

4.- En la compra de una vivienda hay varios factores que se deben tomar en cuenta como: lugar de ubicación, precio, tasa de amortización, tipo de construcción y otros, esto hace que el tiempo que un comprador tarde en llegar a su decisión final sea muy variable. Los datos que se presentan a continuación representan la duración de la búsqueda (en semanas) de 25 compradores de viviendas en una cierta población:

15	17	7	15	20
5	3	19	10	3
11	10	4	8	13
9	15	6	2	8
12	1	2	13	4

a) Construir una tabla de distribución de frecuencias en 4 intervalos. Identificando la frecuencia relativa, las marcas de clase y la frecuencia absoluta.
b) Construir un histograma y un polígono de frecuencias para estos datos.
c) Calcular la media y la desviación estándar de la distribución.
d) Encontrar la proporción de datos que "caen" en los intervalos $(\mu-\sigma,\mu+\sigma)$ y $(\mu-2\sigma,\mu+2\sigma)$. Decir que opinión se tiene a cerca de si estos resultados son o no consistentes con el teorema de Tchebysheff y con la regla empírica.

5.- Un doctor desea saber a que edad promedio las mujeres empiezan a manifestar los sintomas de la menopausia, para esto realizó una encuesta entre 50 pacientes mujeres que visitarón su consultorio en un mes. Los datos de las edades en que se presentarón los sintomas estan en la siguiente tabla:

38	41	43	39	39	47	49	42	52	46
39	40	46	45	42	46	50	45	49	45
40	40	47	44	41	43	41	46	47	48
43	39	46	48	46	45	40	41	49	44
46	46	44	50	46	48	40	43	50	43

a) Elaborar una tabla de distribución de frecuencias en 3 intervalos. Identificando la frecuencia relativa y las marcas de clase.
b) Dibujar un histograma y un polígono de frecuencias de la distribución.
c) Calcular la moda, la mediana y la media
d) Calcular la varianza y la desviación estándar.
e) ¿Qué porcentaje de las mujeres presenta los sintomas de la menopausia entre los 42 años y los 46 años?

6.- En una muestra elegida al azar de 50 personas se observó que el número de hermanos (hombres y/o mujeres) que tenían esas personas se observan en la siguiente tabla de frecuencias en 3 intervalos:

Intervalos	Clases	Frecuencia	Frecuencia acumulada	Marca de clase	Marca * Frecuencia
0 - 3	0 - 3.5	18	18	2	2*18 = 36
4 - 6	3.5 - 6.5	17	35	5	5*17 = 85
7 - 9	6.5 - 9.5	15	50	8	8*15 =120
		---			-----
Total		50			241

a) calcular la desviación estándar y la varianza de la distribución

7.- Las siguientes cifras corresponden al peso en kilogramos de 1000 bultos de papel que salen de una fábrica productora de papel bond:

Bultos	Peso	Bultos	Peso
6	57.5 - 59.5	190	69.5 - 71.5
13	59.5 - 61.5	175	71.5 - 73.5
40	61.5 - 63.5	70	73.5 - 75.5
96	63.5 - 65.5	40	75.5 - 77.5
175	65.5 - 67.5	10	77.5 - 79.5
180	67.5 - 69.5	5	79.5 - 81.5

a) Calcular la varianza y la desviación estándar.

8.- Las siguientes temperaturas mínimas en grados centígrados durante el mes de mayo en Monterrey de 1986 fueron:

Mínimas: 33 30 29 28 27 29 30 30 31 28 28 28 27 29 25
25 27 29 30 31 31 30 32 28 27 26 27 29 28 29

a) Elaborar una tabla de distribución de frecuencias (en 4 intervalos)
b) Calcular la varianza y la desviación estándar de la distribución

EVALUACION PROPUESTA

1.- Para observar la dispersión de los datos de una investigación se tienen las medidas de dispersión que son:

2.- Cuál es la medida de dispersión que considera la suma de los cuadrados de las desviaciones de cada dato con respecto al promedio de la distribución ?

3.- Las medidas de dispersión miden la _______ que hay entre cada dato obtenido con respecto a la ________ de la distribución.

4.-Un nadador se prepara para una competencia de 100 m libres, conforme avanza en su entrenamiento desea conocer su progreso medido en segundos de acuerdo a su promedio de tiempo, su progreso en 6 días consecutivos fue {2,3,5,6,7,9}. Calcular la desviación media de los tiempos del nadador.

5.- Un corredor de distancia se prepara para competir en el maratón, conforme avanza en su entrenamiento aumenta el número de kilometros recorridos en 7 días que son {7,9,10,11,13,14,15}. Calcular la desviación media de los km recorridos por el corredor.

6.-Considerando la siguiente tabla de distribución de frecuencias:

Intervalos	Intervalos de clase	Frecuencia	Marcas de clase
14 - 30	13.5 - 30.5	5	22
31 - 47	30.5 - 47.5	19	39
48 - 64	47.5 - 64.5	21	56
65 - 81	64.5 - 81.5	5	73

a) Calcular la desviación media de la distribución
b) Calcular la desviación estándar y la varianza

7.- Las siguientes temperaturas mínimas en grados centígrados durante el mes de mayo en Monterrey de 1986 fueron:

Mínimas: 33 30 29 28 27 29 30 30 31 28 28 28 27 29 25
25 27 29 30 31 31 30 32 28 27 26 27 29 28 29

a) Elaborar una tabla de distribución de frecuencias (en 4 intervalos)
b) Calcular la varianza y la desviación estándar de la distribución

8.- En una muestra elegida al azar de 50 personas se observó que el número de hermanos (hombres y/o mujeres) que tenían esas personas se observan en la siguiente tabla de frecuencias en 3 intervalos:

Intervalos	Clases	Frecuencia	Frecuencia acumulada	Marca de clase	Marca * Frecuencia
0 - 3	0 - 3.5	18	18	2	2*18 = 36
4 - 6	3.5 - 6.5	17	35	5	5*17 = 85
7 - 9	6.5 - 9.5	15	50	8	8*15 =120
		---			-----
Total		50			241

a) Calcular la desviación estándar y la varianza.

9.- Las siguientes cifras corresponden al peso en kilogramos de 1000 bultos de papel que salen de una fábrica productora de papel bond:

Bultos	Peso	Bultos	Peso
6	57.5 - 59.5	190	69.5 - 71.5
13	59.5 - 61.5	175	71.5 - 73.5
40	61.5 - 63.5	70	73.5 - 75.5
96	63.5 - 65.5	40	75.5 - 77.5
175	65.5 - 67.5	10	77.5 - 79.5
180	67.5 - 69.5	5	79.5 - 81.5

a) Calcular la varianza y la desviación estándar.

10.- Tomando en consideración el teorema de Tchebysheff que proporción de datos tiene cada uno de los siguientes intervalos:
(μ-σ, μ+σ) contiene aproximadamente el ________de los datos
(μ-2σ, μ+2σ) contiene aproximadamente el ________de los datos
(μ-3σ, μ+3σ) contiene aproximadamente el ________ de los datos

CONCLUSIONES

Los resultados de la evaluación del tema de las medidas de dispersión al grupo 505 de Estadística se presentan en la siguiente tabla de distribución de frecuencias en 4 intervalos, se obtienen los promedios de los datos para identificar los avances o retrocesos del aprendizaje de los alumnos con respecto a los temas anteriores.

0	2	4	5	6	6	10	10
0	4	4	5	6	7	10	10
0	4	4	5	6	8	10	10
0	4	4	6	6	8	10	10
2	4	5	6	6	8	10	10

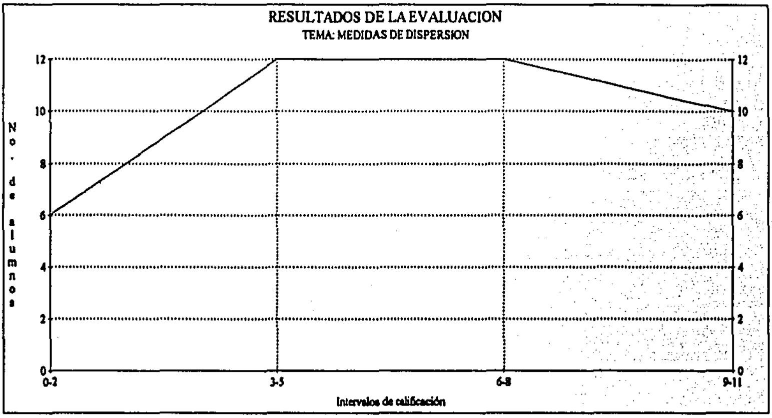
Intervalos	Clases	Frecuencia	Marcas de clase	Marcas * frecuencia
0 - 2	0 - 2.5	6	1	6* 1 = 6
3 - 5	2.5 - 5.5	12	4	12* 4 = 48
6 - 8	5.5 - 8.5	12	7	12* 7 = 84
9 - 11	8.5 - 11.5	10	10	10*10 =100
	total	40		238

$$\bar{X} = \frac{238}{40} = 5.95$$

$$\tilde{X} = 5.5 + \left(\frac{20 - 18}{12}\right) 3 = 6$$

$$X_0 = 7$$

Polígono de frecuencias de los resultados de la evaluación



Gráfica circular de los resultados:

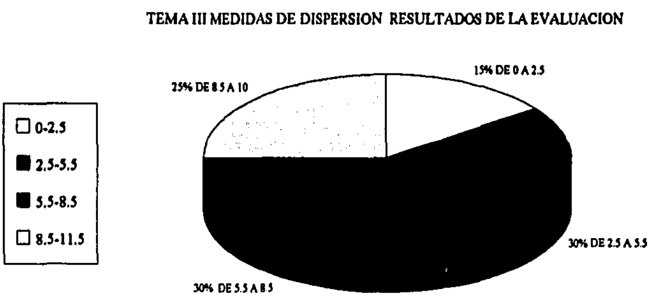
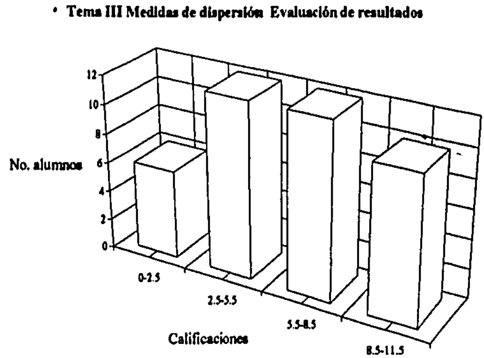


Diagrama de barras de la evaluación



La gráfica de los resultados muestra que los conceptos del tema de medidas de dispersión del trabajo que esta siendo evaluado fueron bien comprendidos y no fueron comprendido en la misma proporción, sin embargo de acuerdo a los promedios calculados se puede afirmar que en base a la evaluación escrita el 55% de los alumnnos comprende adecuadamente el ttrabajo.

Otros factores influyen en los resultados de la evaluación escrita que se tendrían que considerar como el ausentismo a clase, la falta de constancia en la elaboración del trabajo fuera del aula, etc. sin embargo para el objetivo planteado no se consideran en este trabajo.

92

TEMARIO

INTRODUCCION

Los juegos de azar han contribuido fundamentalmente con el concepto de probabilidad al pensamiento estadístico. El deseo de los jugadores de poder "predecir" los resultados de un juego de azar permitió desarrollar la teoría de la probabilidad, éste empezó en el siglo XVII debidó al interés del matemático francés Blaise Pascal por su afición a jugar dados con un amigo.

En la actualidad existe una relación muy estrecha entre la probabilidad y la estadística para poder obtener resultados de experimentos aleatorios. Cada persona tiene una idea intuitiva de lo que significa probabilidad debido a que es un concepto de uso cotidiano; sin embargo, en esta sección se trata de introducir a los estudiantes de nivel medio superior en las ideas fundamentales de la teoría de la probabilidad que es lo que la convierte en un conocimiento formal.

Los fenómenos aleatorios se caracterizan porque no se puede determinar con suficiente precisión su estado final, sino que tiene cierto grado de incertidumbre. La estadística ofrece herramientas que se utilizan tanto en el proceso de medición de resultados de dichos fenómenos, como en el establecimiento de hipótesis o leyes del comportamiento de éstos. Esto se logra por el uso de modelos matemáticos para tratar de representar las relaciones que muestran las diferentes propiedades de los fenómenos aleatorios que se estudian.

Algunas de las ideas de la probabilidad provienen de juegos de azar, por ejemplo de un paquete de 52 cartas de dos colores, rojo y negro, y cuatro "palos": tréboles (negros), diamantes (rojos), corazones (rojos) y espadas (negras); de cada "palo" hay 13 cartas as, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, reina y rey. Para empezar a jugar se "barajan" las cartas y se "cortan" para tratar de ponerlas en **orden aleatorio** para que cada carta tenga la misma posibilidad de ser escogida. Existen algunos factores importantes en la determinación de las probabilidades en un juego de cartas, por ejemplo que es un número finito de cartas por lo tanto es un número finito de posibilidades, la base del cálculo de las probabilidades en este caso es que las ocurrencias de las cartas son igualmente probables.

Otro ejemplo de probabilidad es jugar con un dado que es un cubo con 1, 2, 3, 4, 5 y 6 puntos respectivamente en cada uno de sus seis lados, la intuición sugiere que si se lanza un dado bien balanceado cada "cara" tiene 1/6 de quedar hacia arriba. Sin embargo si el dado esta "cargado" no es lógico suponer que cada cara tiene la misma probabilidad de quedar hacia arriba, en este caso para poder asignar una probabilidad a cada cara se tendría que lanzar el dado un gran número de veces y observar la frecuencia con que cada lado del dado aparece.

El estudio de la probabilidad es primordial si se desean utilizar muestras para obtener inferencias acerca de las poblaciones. En algunas ocaciones los elementos de las muestras se obtienen aleatoriamente de las poblaciones correspondientes y en otras el investigador considera el método de selección de los elementos lo más aleatoriamente posible.

ESPACIOS MUESTRALES

Es frecuente en estadística que los datos se obtengan de elementos que forman una muestra de una población de interés. Por lo tanto una muestra aleatoria de n individuos de una población de tamaño N (N>n), es uno de todos los posibles conjuntos de tamaño n que se pueden formar y que tienen la misma posibilidad de ser seleccionados; es decir, que los N elementos de la población tienen la misma oportunidad de ser uno de los n elementos incluidos en la muestra. En algunos casos una muestra se puede obtener de una población por un procedimiento totalmente aleatorio, en algunos otros mediante algún procedimiento físico o lógico más o menos aleatorio.

Si se consideran todos los posibles resultados de un experimento bien determinado y con ellos se forma un conjunto entonces a dicho conjunto se le conoce como **Espacio Muestral** del experimento.

Definición.- Al conjunto de todos los resultados posibles de un experimento se le llama **Espacio Muestral** y se denota por E.

Los espacios muestrales se dividen en: **Discretos y Continuos**.

Definición.- El espacio muestral discreto es aquel que tiene un conjunto de resultados finito. En otras palabras tiene un conjunto de resultados "numerable".

Definición.- El espacio muestral continuo es aquel que tiene un conjunto de resultados infinito, es decir, que tiene un conjunto de resultados no numerable en donde se pueden establecer escalas con números reales para cuantificarlos.

Para representar gráficamente los elementos que forman un espacio muestral se puede utilizar un diagrama de árbol, en donde cada rama representa un posible resultado del experimento. En esta sección se trabajará únicamente con espacios muestrales discretos.

Ejemplo:

Escribir el espacio muestral de cada uno de los siguientes experimentos, utilizando diagramas de árbol y/o tablas de doble entrada para obtenerlos:

1.- Se observa el sexo de los dos primeros hijos de una mujer. Como cada hijo únicamente puede ser Hombre (H) o Mujer (M), y se tienen 2 hijos entonces los posibles resultados se muestran en la siguiente tabla de doble entrada, en donde cada entrada esta formada por los posibles sexos de cada hijo:

n2/n1	H	M
H	HH	HM
M	MH	MM

E={HH,HM,MH,MM}

2.- Observar las caras que quedan hacia arriba cuando tres monedas se lanzan al aire. Con un diagrama de árbol se obtienen cada uno de los posibles resultados de este ejemplo, debido a que cada moneda únicamente puede tener dos resultados águila(a) o sol(s) se tienen los siguientes resultados:

```
    a  -aaa          a  -saa
  a/               a/
    \s -aas          \s -sas
a/               s/
 \ a  -asa        \ a  -ssa
  s/               s/
    \s -ass          \s -sss
```

Por lo tanto E={aaa,aas,asa,saa,ass,sas,ssa,sss}

3.- Observar las "caras" que quedan hacia arriba de dos dados bien balanceados que se lanzan sobre una mesa. Cada uno de los dados puede tener 6 posibles resultados del 1 al 6, por lo tanto sus posibles resultados son:

d2/d1	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

E={(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(3,1),
(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(5,1)(5,2),(5,3),
(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

EVENTOS O SUCESOS

Se dice que un experimento es **compuesto** si se puede dividir en una sucesión de experimentos independientes entre sí. Es decir, sea X un experimento y sean X_1 y X_2 dos resultados cualesquiera de dicho experimento, si el resultado X_1 no afecta de ninguna manera al resultado X_2 y viceversa se dice que los resultados X_1 y X_2 son **independientes**.

Debido a que los espacios muestrales son **conjuntos** de resultados de experimentos se pueden utilizar las propiedades del álgebra de conjuntos para definir las propiedades de los espacios muestrales.

Definición.- A cualquier subconjunto de resultados con alguna característica de interés en común de un espacio muestral de un experimento se le llama **SUCESO o EVENTO**. A cada elemento del espacio muestral de un experimento se le llama evento elemental o simple.

Los eventos se clasifican en:

a) Eventos elementales o simples
b) Eventos imposibles
c) Eventos seguros
c) Eventos complementarios
d) Eventos mutuamente excluyentes

Definición.- Se llama **evento imposible** a cualquier subconjunto del espacio muestral que sea igual al conjunto vacio ϕ, es decir, que no tenga resultados o que sus resultados no pertenezcan al espacio muestral del experimento.

Definición.- Se llama **evento seguro** al subconjunto que contenga todos los resultados del espacio muestral, es decir, que sea igual al espacio muestral.

Definición.-Sean E y F dos eventos cualesquiera. Dichos eventos se llaman **complementarios** si no ocurren simultáneamente, es decir, si ocurre E entonces no ocurre F y viceversa.

Sea A un evento en un espacio muestral E, el evento complementario de A se le denota por A' y contiene todos los resultados que están en E pero que no son resultados de A.

Definición.- Sean E y F dos eventos cualesquiera se llaman eventos **mutuamente excluyentes** a aquellos eventos que no tienen resultados en comunes, es decir, $E \cap F = \phi$.

Cuando se tiene un espacio muestral finito se considera el conjunto de todos sus subconjuntos y se le denota por E. Por otra parte si el espacio muestral es continuo (infinito) el conjunto de todos sus subconjuntos es no numerable, por esta razón se trabajará con espacios muestrales numerables.

PROPIEDADES

Los espacios de eventos cumplen con las siguientes propiedades:

a) El evento imposible es un elemento del espacio muestral, es decir, el conjunto vacío ϕ siempre es un elemento del espacio muestral.

b) Si un evento A pertenece al espacio muestral también pertenece su complemento A'.

c) Si una sucesión numerable de eventos $A_1, A_2, \ldots, A_n$ pertenecen al espacio muestral entonces también pertenece el evento unión $A_1 \cup A_2 \cup \ldots \cup A_n$ al espacio muestral.

Ejemplo:

1.- Considerar el experimento de lanzar dos dados simultáneamente sobre una mesa una sola vez. Determinar los resultados que pertenecen a los siguientes eventos:

a) Sea el evento A formado por los resultados de obtener dos números pares.
b) Sea el evento B formado por los resultados de obtener dos números impares.
c) Sea el evento C formado por los resultados de obtener un número par para el primer dado y un número impar para el segundo dado.
d) Sea el evento D formado por los resultados de obtener un número par mayor que 2 y menor que 6 para el primer dado.
e) Sea el evento F formado por los resultados de obtener un número mayor e igual a 1 y menor e igual a 6 en ambos dados.
f) Sea el evento G formado por los resultados de obtener un número menor que 1 y mayor que 7 para ambos dados.

El espacio muestral de lanzar 2 dados una sola vez esta formado por 36 resultados, como ya observó anteriormente.

d_2/d_1	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

Entonces los resultados de los eventos son:

$A=\{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}=\{(x,y)/\ x=2n,\ y=2n,\ n=1,2,3\}$

$B=\{(1,1),(1,3),(1,5),(3,1),(3,3),(3,5),(5,1),(5,3),(5,5)\}=\{(x,y)/x=2n+1,y=2n+1,n=0,1,2\}$

Los eventos A y B son mutuamente excluyentes ya que $A\cap B=\phi$
Los eventos A y B también son complementarios es decir, $A=B'$ o $B=A'$
y en este caso también se tiene que $A\cup B=E$

c) $C=\{(2,1),(2,3),(2,5),(4,1),(4,3),(4,5),(6,1),(6,3),(6,5)\}=\{(x,y)/\ x=2n,\ y=2m+1,\ n=1,2,3 \text{ y } m=0,1,2\}$

d) $D=\{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}=\{(x,y)\ /\ 2<x<6\}$

e) $F=\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(5,1)(5,2),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}=\{(x,y)\ /\ 1\le x\le 6\ ,\ 1\le y\le 6\}$
En este caso se tiene que F es el evento seguro ya que F=E

f) $G=\{(x,y)\ /\ 1>x>7\ ,\ 1>y>7\}=\phi$ evento imposible

Ejercicios:

Escribir los elementos del espacio muestral y de cada uno de los eventos que se definen para ese espacio muestral de los siguientes experimentos:

1.-Un jugador extrae dos cartas al azar de una baraja de 52 cartas de dos colores y cuatro "palos". Sea el evento A : al menos una de las cartas es un as y el evento B: las dos cartas son diamantes

2.-En su caja de juguetes un niño tiene 6 pelotas, 2 verdes y 4 azules, el niño saca dos pelotas sin verlas ni regresarlas. Si se definen los eventos A: las dos bolas son azules, B: una de las dos bolas es azul.

3.-Un ama de casa saca 3 billetes de una caja que contiene 9 billetes cuyos valores son de 1 a 9 nuevos pesos inclusive respectivamente. Sea el evento A: los 3 billetes sacados son menores que 5, B: los 3 billetes extraidos tienen valor par.

4.-Cuatro alumnos entran al salón de clases, son dos hombres y dos mujeres; si se sientan en una banca de cuatro lugares se definen los eventos A: los de mismo sexo quedan juntos, B: se alternan hombres y mujeres.

5.-Una persona escoge dos calcetines sin verlos ni regresarlos, de un cajón que contiene 6 calcetines rojos, 8 amarillos y 2 azules. Sea el evento A: los dos calcetines son del mismo color, B: los dos calcetines son amarillos.

6.-Un juego consiste en lanzar un dado y una moneda simultáneamente una sola vez. Sean los eventos
A: que se obtenga un número par y un águila
B: que se obtenga un número impar y un sol
C: que se obtenga un número entre 1 y 6 y un sol o un águila
D: que se obtenga un número menor que 1 y mayor que 7 y un 2 soles.

PROBABILIDAD

Si un evento no es imposible, ni seguro tiene que ser posible o incierto entonces ¿Cuál es su grado de incertidumbre?. La teoría de probabilidades pretende medir cuantitativamente la incertidumbre de un determinado evento. Por ejemplo si se lanza una moneda al aire los posibles resultados, águila o sol, que se pueden obtener son igualmente probables, y a éstos se les llama ocurrencias.

Definición.- Si en un espacio muestral hay **n** resultados igualmente probables y un evento está formado por **m** resultados posibles entonces la **probabilidad** del evento es **m/n**. En otras palabras la probabilidad de un evento es el número de ocurrencias del evento entre el número total de resultados del espacio muestral.

Definición.- Si el espacio muestral es un conjunto finito de casos que se estiman igualmente probables, cada uno de los n elementos del espacio muestral E tiene las mismas posibilidades de ocurrir entonces se puede asignar la probabilidad 1/n a cada uno de los n resultados:

$$P(\{X\}) = 1/n \qquad \text{para cada} \quad X \in E$$

Esto se conoce como la **función de probabilidad de Laplace.**

A esta definición se le conoce como la forma clásica de cálcular la probabilidad de un evento. Una muestra aleatoria tiene como característica de que todas las combinaciones de las posibles ocurrencias de los elementos que la forman son igualmente probables.

Ejemplo:

1.- Los alumnos de un grupo de estadística presentan un examen de 10 preguntas, cada una vale un punto, para obtener una calificación entre 0 y 10. ¿ Cuál es la probabilidad de cada alumno de obtener 10 en el examen?
Como son 10 preguntas y cada una vale un punto entonces si se define el evento A como obtener 10 respuestas correctas, la probabilidad de A es:
P(A)= 1/ 10

2.- Un jugador de cartas pide una baraja nueva, dicha baraja esta formada por 52 cartas de dos colores, y 4 "palos" distintos con 13 cartas por cada "palo". Si se define el evento A como extraer una carta de espadas, ¿Cuál es la probabilidad de que el jugador saque una carta de espadas?

Como se tienen 13 cartas de espadas en la baraja entonces la probabilidad de extraer una carta de espadas es P(A) =13/52 = 1/4

3.- Con los mismos datos del ejemplo anterior, si A es el evento de obtener un as de la baraja entonces su probabilidad es P(A) = 4/52 = 1/13

4.- Un estudiante se dedica en vacaciones a repartir envios de una tienda de casa en casa, entre ellos tiene que repartir una caja de una docena de botellas de refrescos; si se define el evento A como el número de botellas que se rompen en el trayecto, ¿ cuál es la probabilidad de que cuando llegue a su destino el estudiante solamente tenga dos botellas rotas?
Como se tienen doce botellas en la caja entonces la probabilidad de que se rompan solamente dos de ellas es:

P(A) = 2/12 = 1/6

5.- Un niño tiene en una caja 3 pelotas negras y 2 pelotas blancas, si el evento A es que el niño meta la mano a la caja y saque una pelota negra; y el evento B es que el niño saque una pelota blanca, ¿Cuál es la probabilidad de A y cuál la probabilidad de B?
Como se tienen 5 pelotas en total en la urna 3 de un color y 2 de otro entonces su probabilidad es:

P(A) = 3/5 y P(B)=2/5

Si se tienen dos eventos A y B sin ningún resultado en común y si el primero A tiene x resultados posibles y el segundo B tiene y resultados posibles entonces el evento que está formado por todos los posibles resultados de los 2 eventos A y B, tiene x+y posibles resultados. Es decir, si en un espacio muestral de n elementos se tiene un evento A que tiene una probabilidad de x/n y el evento B tiene una probabilidad de y/n entonces el evento de todos los posibles resultados de A y B tiene una probabilidad de (x+y)/n.

Definición.- Si un evento está formado por todos los posibles resultados de dos eventos A y B que no tienen ningún resultado en común entonces la probabilidad de ese evento es la suma de las probabilidades de los dos eventos A y B.

Definición.-El **evento imposible** tiene probabilidad cero. Es conveniente definir al evento imposible como el evento que no tiene ningún resultado es decir, como el conjunto vacío ϕ por lo tanto $P(\phi) = 0$.

Definición.-El **evento seguro** tiene probabilidad uno. Es decir, si A es un evento que tiene todos los resultados del espacio muestral del experimento entonces $P(A)=1$.

Todos los demás eventos, distintos al imposible y al seguro, tienen una probabilidad que son números comprendidos entre cero y uno, es decir, si A es cualquier evento que no es ni seguro ni imposible entonces $0<P(A)<1$.

Ejemplo:
En el patio juegan los estudiantes con un dado bien balanceado, se trata de lanzar un dado al aire y observar el lado que queda hacia arriba. Considerando este juego responder las siguientes preguntas:
a)¿Cuál es la probabilidad de que salga un número impar como resultado?
b)¿Cuál es la probabilidad de que salga un número par mayor que 2?
c)¿Cuál es la probabilidad de obtener un número mayor que 6?
d)¿Cuál es la probabilidad de obtener cualquier número entre 1 y 6 inclusive?

El espacio muestral del experimento está formado por {1,2,3,4,5,6}
a) Sea A el evento que salga un número impar, por lo tanto el conjunto de los posibles resultados de A es {1,3,5} por lo tanto la probabilidad de A es:

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

Otra forma de calcular la probabilidad es:

$$P(\{1,3,5\})=P(\{1\})+P(\{3\})+P(\{5\}) = 1/6 + 1/6 + 1/6 = 3/6 = 1/2$$

b) Sea B el evento que salga un número par mayor que 2, por lo tanto el conjunto de los posibles resultados de B es {4,6} por lo tanto la probabilidad de B es:

$$P(B) = \frac{2}{6} = \frac{1}{3}$$

c) Sea el evento C obtener un número mayor que 6, como no existe ningún resultado mayor que 6 entonces el conjunto de resultados del evento C es el conjunto vacio ϕ, es decir es el evento imposible y su probabilidad es: $P(C) = P(\phi) = 0$

d) Sea D el evento de obtener un número entre 1 y 6 inclusive, por lo tanto el evento D es el evento seguro y su probabilidad es:

$$P(D) = \frac{6}{6} = 1$$

AXIOMAS Y TEOREMAS

Existen dos situaciones más en las cuales el concepto de un número finito de resultados igualmente probables no es posible: cuando el número de posibles resultados es finito pero no todos los resultados son igualmente probables, -por ejemplo un dado "cargado"-; y cuando el espacio total de posibles resultados no es finito, -por ejemplo los posibles estados del agua-.

Los resultados de un evento que dependen de una variable continua no son finitos; para estos casos el cálculo de probabilidades tiene principios más generales para los métodos de obtener resultados igualmente probables. Dichas propiedades y reglas de probabilidad son válidas tanto para el caso clásico como para los casos mencionados en el párrafo anterior.

Como ya se sabe un evento A es un conjunto de posibles resultados, la probabilidad de cada evento se denota por P(A) que es la razón entre el número de resultados del evento A entre el número total de resultados del experimento.

Las probabilidades asignadas a los eventos cumplirán con las siguientes reglas o axiomas:

Axioma I: Para todo evento A : $0 \leq P(A) \leq 1$
Axioma II: Si A es el evento seguro $P(A)=1$
Axioma III: Si $A_1, A_2,\ldots,A_n$ es una sucesión numerable de eventos mutuamente excluyentes, la probabilidad de su unión es igual a la suma de sus probabilidades individuales, es decir:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n)=P(A_1)+P(A_2)+\ldots+P(A_n)$$

Se enuncian otros resultados como teoremas.

Teorema.- Sea A el evento imposible: Si $A = \phi$ entonces $P(A)=0$

Teorema.- Sea A el evento seguro: Si $A = E$ entonces $P(A) = 1$

Teorema.- Sean A y B dos eventos cualesquiera entonces $P(A)=P(A\cap B)+P(A\cap B')$

Teorema.- Sea A un evento cualesquiera entonces $P(A')=1-P(A)$

Teorema.-Sean A y B dos eventos cualesquiera entonces

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

Los resultados enunciados como teoremas se pueden demostrar a partir de la definición de probabilidad y de los axiomas enunciados.

REGLA DE LA SUMA DE PROBABILIDAD

Definición.- Sean A y B dos eventos cualesquiera, y sean P(A) y P(B) sus respectivas probabilidades. Si se denota con $A \cup B$ el evento unión de ambos eventos donde $A \cap B \neq \phi$, es decir, que tienen al menos un resultado en común entonces se define la probabilidad del evento unión como:

$$P(A \cup B)=P(A)+P(B)-P(A \cap B)$$

donde $A \cap B$ representa la probabilidad de que ambos eventos ocurran simultáneamente.

Definición.- Dos eventos A y B son **mutuamente exclusivos o disjuntos** si no tienen ningún resultado en común y se denota por $A \cap B=\phi$.

Teorema.- Si A y B son dos eventos mutuamente exclusivos, es decir, $A \cap B=\phi$ entonces la probabilidad del evento unión $A \cup B$ es: $P(A \cup B)=P(A)+P(B)$

Como cualquier evento definido en un espacio muestral finito se puede expresar como la unión de sucesos elementales independientes, y con la definición de probabilidad de Laplace se cumple que:

$$E=\{X_1, X_2, ..., X_r\}=\{X_1\}U\{X_2\}U...U\{X_r\}$$

$$P(E)=P(\{X_1, X_2, ..., X_r\}) = P(X_1)+P(X_2)+...+P(X_r) = 1/n+1/n+...+1/n = r/n$$

entonces

$$P(E)=\frac{N(E)}{N(S)}=\frac{r}{s}$$

donde: N(E)= número de elementos del suceso
N(S)= número de elementos del espacio muestral

Ejemplo:

1.- Un juego consiste en lanzar un dado una vez. En el juego se consideran los siguientes eventos:
A: se obtiene número par
B: se obtiene número impar
C: se obtiene un número entre 1 y 6

Calcular las probabilidades de $P(A \cup B)$ y $P(B \cup C)$

Los resultados de los eventos son: $A=\{2,4,6\}$ $B=\{1,3,5\}$ $C=\{1,2,3,4,5,6\}$

$P(A \cup B)=P(A)+P(B)=3/6 + 3/6 = 6/6 = 1$ ya que $A \cap B=\phi$

$P(B \cup C)=P(B)+P(C)-P(B \cap C)=3/6+6/6-3/6= 1$ ya que $B \cap C=\{1,3,5\}$

2.-En una encuesta realizada en una gasolinería, se determinó que la probabilidad de que un número de automóviles formen una fila de espera en una bomba de gasolina se muestra en el siguiente cuadro:

No. de automóviles	0	1	2	3 o más
Probabilidad	0.08	0.20	0.32	0.40

Calcular la probabilidad de que en la fila haya:

a) A: a lo más un automóvil
$P(A) = (0.08+0.20) / 1 = 0.28$

b) B: 3 o más autos
$P(B)= 0.40/1 = 0.40$

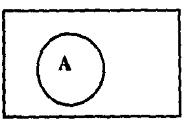
c) C: a lo más 2 autos
$P(C)=(0.08+0.20+0.32)/1 = 0.60/1 = 0.60$

Diagramas de Venn para probabilidad

En la teoría de conjuntos existe un método gráfico para representar a los conjuntos y sus propiedades, los diagramas de Venn. Como los eventos son conjuntos de resultados y tienen propiedades parecidas a los conjuntos entonces en probabilidad también se pueden utilizar los diagramas de Venn para representar sus propiedades; en este caso el espacio muestral E de cualquier experimento queda representado por un rectángulo:

Espacio muestral E

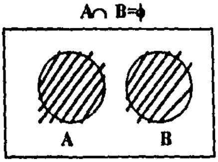
Cada uno de los eventos del espacio muestral queda representado por un círculo:



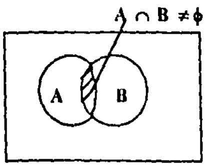
Evento A

Los diagramas de dos eventos A y B mutuamente exclusivos y no exclusivos serían:

$A \cap B = \phi$



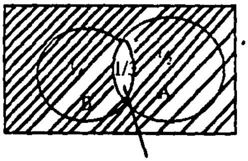
$A \cap B \neq \phi$



$P(A \cup B) = P(A)+P(B)$

$P(A \cup B)=P(A)+P(B)-P(A \cap B)$

b) Usando diagramas de Venn determinar la probabilidad de $P(A' \cup B')$



$A \cap B$

$P(A' \cup B') = P(A \cap B)' = 2/3$

Por leyes de D'Morgan se tiene que $P(A' \cup B') = P(A \cap B)'$

2.-Sea $E=\{(x,y)/1 \leq x \leq 6 \text{ y } 1 \leq y \leq 6\}$ un espacio muestral, en donde se definen los siguientes eventos:

$E_1=\{(1,1)\}$ $E_2=\{(1,2),(2,1)\}$ $E_3=\{(1,3),(2,2),(3,1)\}$ $E_4=\{(1,4),(2,3),(3,2),(4,1)\}$ $A=\{(x,y)/1 \leq x \leq 6,\ y=4\}$
$B=\{(x,y) / x=4,\ 1 \leq y \leq 6\}$

Encontrar:

a) $P(E_4) = 4/36$ ya que E tiene 36 resultados porque es un dado

b) $P(E_3 \cap E_4)=P(\{\ \})=0$ ya que $E_3 \cap E_4 = \phi$

c) $P(E_1 \cup E_2) = P(E_1)+P(E_2) = 1/36+2/36 = 3/36 = 1/12$ ya que $E_1 \cap E_2=\phi$

d) $P(A')= 1-P(A) = 1- 5/36 = 31/36$ ya que $A=\{(1,4),(2,4),(3,4),(4,4),(5,4),(6,4)\}$

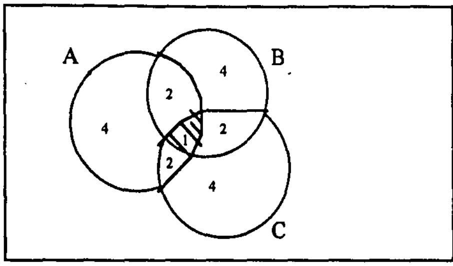
3.- En un estudio sobre un grupo de 27 alumnos, se halló que 9 sobresalían en matemáticas, 9 en música y 9 en deporte. También se halló que 3 sobresalían a la vez en matemáticas y música, 3 sobresalían a la vez en música y deportes, y 3 sobresalían a la vez en matemáticas y deporte. Sólo se encontró 1 estudiante que sobresalía en las tres asignaturas.

a) si se escogen dos estudiantes al azar, calcular la probabilidad de que sobresalgan en matemáticas y música.

b) si se escoge al azar un estudiante de los 27, cuál es la probabilidad de que sobresalga en matemáticas?

El espacio muestral E esta formado por los 27 alumnos, si se definen los eventos:

A: sobresalir en matemáticas B: sobresalir en música C: sobresalir en deporte



A : matemáticas
B : música
C : deporte

a) para calcular la probabilidad del evento D de que los 2 alumnos escogidos sobresalgan en matemáticas y música al mismo tiempo, se toman los alumnos que se encuentren en $A \cap B$ y su probabilidad sería:

$P(D) = 2/3$

b) para calcular la probabilidad del evento F de escoger un alumno de los 27 pero que sobresalga en matemáticas, se calcula:

$P(F) = 9/27 = 1/3$

PROBABILIDAD CONDICIONAL

La probabilidad se puede desarrollar como una teoría formal, es decir, como una rama de la matemática cumpliendo y comprobando todos los axiomas y teoremas enunciados.

El objetivo de la probabilidad condicional es obtener la probabilidad de un evento B dado que ya sucedió otro evento A. La probabilidad condicional esta definida en términos de las probabilidades de los eventos dados y las combinaciones de ellos.

Definición.- La probabilidad de que ocurra un evento B dado que otro evento A ya ocurrió se le llama **probabilidad condicional** de B dado A y se expresa como sigue:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \; ; \; P(A) \neq 0$$

Sean A y B dos eventos cualesquiera. Sea B un subconjunto del espacio muestral de eventos y A el evento cuya probabilidad se quiere calcular sobre el espacio de eventos de B, si n(B) es el número de resultados que pertenecen a B (cardinalidad de B) y $n(A \cap B)$ es el número de resultados que pertenecen a A y B simultáneamente entonces la probabilidad de que ocurra A si el espacio de eventos queda restringido al espacio de eventos de B está dada por:

$$P(A|B) = \frac{n(A \cap B)}{n(B)}$$

Probabilidades condicionales de dos eventos A y B:

	B	$\bar{B}$
A	$P(B\|A)$	$P(\bar{B}\|A)$
$\tilde{A}$	$P(B\|\tilde{A})$	$P(\bar{B}\|\tilde{A})$

La probabilidad condicional dado un evento específico A cumple con todas las propiedades de la probabilidad, es mayor o igual a cero porque el numerador es no negativo y el denominador es positivo y son menores o iguales a 1, porque el numerador es menor o igual que el denominador ya que "$A \cap B$" es un evento contenido en A. En particular

$$P(B|A)+P(\bar{B}|A) = 1$$

Empleando la ecuación que define la probabilidad condicional P(B|A) se puede reescribir de la siguiente manera:

$$P(A \cap B) = P(A) * P(B|A)$$

Es decir, la probabilidad de un evento "$A \cap B$" se puede calcular multiplicando la probabilidad de A por la probabilidad condicional de B dado A.

Ejemplos:

1.- Un juego consiste en lanzar dos dados simultáneamente una vez. ¿Cuál es la probabilidad de que en alguno de los dados aparezca un 5 dado que la suma de las dos caras de los dados es 6?

A: la suma de los 2 resultados es igual a 6
el evento A tiene como espacio de posibles resultados: {(1,5),(2,4),(3,3),(4,2),(5,1)}

B: aparece un 5 en una cara de los dados
el espacio de posibles resultados de B es: {(1,5),(2,5),(3,5),(4,5),(5,5),(6,5),(5,1),(5,2),(5,3),(5,4),(5,6)}

Por lo tanto la probabilidad de obtener un 5 si la suma de las dos caras es seis, esta dada por:

$$P(B|A) = P(A\cap B)/ P(A) = \frac{2/36}{5/36} = \frac{2*36}{5*36} = 2/5$$

2.- En una familia de 3 hijos se sabe que por lo menos uno de ellos es hombre. ¿Cuál es la probabilidad de que en esta familia los 3 hijos sean hombres si se sabe que por lo menos uno de ellos es hombre?

E={(hhh),(hhm),(hmh),(mhh),(hmm),(mhm),(mmh),(mmm)}

A: por lo menos un hijo es hombre, tiene como espacio de posibles resultados el siguiente conjunto {(hhh),(hhm),(hmh),(mhh),(hmm),(mhm),(mmh)}

B: los 3 hijos son hombres, el espacio de posibles resultados de este evento es {(hhh)}

por lo tanto la probabilidad de que los 3 hijos sean hombres dado que uno de ellos es hombre es:

$$P(B|A) = \frac{P(B\cap A)}{P(A)} = \frac{1/8}{7/8} = \frac{1}{7}$$

3.- En un grupo hay 50 estudiantes de los cuales 30 son casados, 15 saben hablar inglés y 10 son casados y saben hablar inglés. Si se elige uno de los estudiantes al azar ¿Cuál es la probabilidad de que el elegido hable inglés si es casado?

A: 30 estudiantes son casados
B: 15 estudiantes hablan inglés

$$P(B|A)= \frac{P(A\cap B)}{P(A)} = \frac{10/50}{30/50} = \frac{1}{3}$$

4.- Un juego consiste en lanzar una moneda dos veces.
a) si la probabilidad de obtener dos soles es de ¼, cuál es la probabilidad condicional de obtener dos soles dado que ya se obtuvo un sol ?
b) cuál es la probabilidad condicional de obtener un sol y un águila, dado que ya se obtuvo un sol?

El espacio muestral de lanzar dos veces una moneda es: {(ss),(sa),(as),(aa)}

a) $P(s|s) = \frac{1/4}{2/4} = 1/2$

b) $P(a|s) = \frac{2/4}{2/4} = 1$

REGLA DE LA MULTIPLICACION DE LA PROBABILIDAD

Cuando se tienen dos eventos cualesquiera y se desea calcular la probabilidad de la ocurrencia simultánea de los dos eventos se debe distinguir si los eventos son independientes o son eventos dependientes.

Definición.-Dos eventos son **independientes** si la probabilidad de ocurrencia de uno de ellos no afecta la probabilidad de ocurrencia del otro y viceversa.

Definición.-Dos eventos son **dependientes** si la probabilidad de ocurrencia de uno de ellos depende de la ocurrencia del otro.

Definición.-Dos eventos A y B, se dice que son independientes si y sólo si la probabilidad de ocurrencia conjunta de A y B es igual al producto de sus respectivas probabilidades. Es decir, A y B son independientes si:

$P(A \cap B) = P(A) * P(B)$

Ejemplos:

1.- Se lanza una moneda dos veces independientemente, considerando los siguientes eventos calcular su probabilidad:
A: que se obtenga águila en el primer lanzamiento
B: que se obtenga águila en el segundo lanzamiento
C: que se obtenga águila en ambos lanzamientos

E={(a,a),(a,s),(s,a),(s,s)}
P(A) = 2/4 = 1/2
P(B) = 2/4 = 1/2
P(C) = 1/4

¿Son A,B, C eventos mutuamente independientes? No, porque tienen el resultado (a,a) en común

2.- Se arroja una moneda al aire, si cae sol se vuelve a lanzar la moneda pero si cae águila se lanza un dado. Calcular la probabilidad de:
a) lanzar un dado en la 2da. tirada.
b) obtener un águila y un número par
c) un águila y un 3 o un 4
d) obtener el resultado (s,5)

El espacio muestral E se puede obtener considerando el siguiente diagrama de árbol

```
    1 -a,1
    2 -a,2
a/  3 -a,3
 \  4 -a,4
    5 -a,5
    6 -a,6
```

a -s,a
s/
\s -s,s

a) $P(A) = 1/2$ ya que la probabilidad de lanzar un dado en la 2da. tirada equivale a que caiga un águila en la 1a. tirada

b) A: obtener un águila
B: obtener un número par
$A \cap B$: obtener un águila y número par

En este ejemplo el evento B esta condicionado a la ocurrencia del evento A, por lo tanto
$P(A \cap B) = P(A)\, P(B|A) = (½)\,(3/6) = (½)(½) = ¼$

c) A: obtener un águila
B: obtener un 3 o un 4
$A \cap B$: obtener un águila y un 3 o un 4

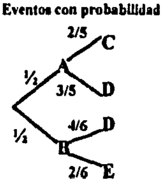
$P(A \cap B) = P(A)\, P(B|A) = (½)(2/6) = (½)(1/3) = 1/6$

d) El resultado (s,5) no puede ocurrir pues si cae águila en el 1er lanzamiento entonces se lanza un dado por lo tanto
$P(s,5) = P(\phi) = 0$

3.-Se tiene una urna No.1 que contiene 2 bolas rojas y 3 negras, y otra urna No.2 que contiene 2 bolas blancas y 4 negras. Ambas urnas tienen la misma probabilidad de ser seleccionadas. Calcular la probabilidad de:
a) sacar una bola negra de la urna No. 1
b) sacar una bola roja
c) sacar una bola negra sabiendo que se seleccionó la urna No.2
d) sacar una bola blanca sabiendo que se seleccionó la urna No.2
e) sacar una bola negra

Si se determinan los eventos por:
A: se selecciona la urna 1
B: se selecciona la urna 2
C: se extrae bola roja
D: se extrae bola negra
E: se extrae bola blanca

Eventos con probabilidad



a) Calcular la probabilidad de sacar una bola negra de la urna 1 implica calcular $P(A \cap D) = P(A)\,P(D|A)$ $= (½)(3/5) = 3/10$

b) $P(A \cap C) = P(A)\,P(C|A) = (½)\,(2/5) = 2/10 = 1/5$

c) $P(B \cap D) = P(B)\,P(D|B) = (½)(4/6) = 4/12 = 1/3$

d) $P(B \cap E) = P(B)\,P(E|B) = (½)\,(2/6) = 2/12 = 1/6$

e) $P(D) = P(A \cap D) + P(B \cap D) = P(A)\,P(D|A) + P(B)\,P(D|B) = (½)(3/5) + (½)(4/6) =$
$(3/10) + (4/12) = 3/10 + 1/3 = 9+10/30 = 19/30$

TEOREMA DE BAYES

Un procedimiento para obtener una muestra de n individuos de una población de N individuos consiste en ponerles n números aleatorios entre 1 y N y seleccionarlos para determinar la muestra. En este procedimiento algunos individuos pueden aparecer varias veces en la muestra a esto se le llama **muestra con reemplazo**, ya que se devuelve el número antes de seleccionar otro número.

Pero si se forma una muestra aleatoria de tal manera que todos los individuos en la muestra sean diferentes, es decir, que no se vuelvan a seleccionar se obtiene una muestra de n individuos de una población de tamaño N en donde ninguno de ellos se repita a esta muestra se le llama **muestra sin reemplazo**.

El teorema de Bayes en el caso simplificado se utiliza para obtener la probabilidad condicional de los eventos.

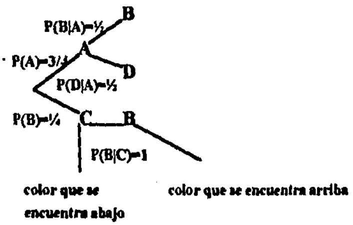
Teorema.- Sean A y B dos eventos. Se cumple que:

$$P(B|A) = \frac{P(B) * P(A|B)}{P(B) * P(A|B) + P(\bar{B}) * P(A|\bar{B})}$$

Ejemplo:

1.- Hay dos tarjetas en un cajón, una de ellas es negra por ambas caras y la otra es negra por una sola cara y blanca por la otra. Se saca una tarjeta al azar y se coloca sobre una mesa (la cara que esta hacia arriba también se ha escogido al azar). Si se dice que la cara que esta hacia arriba es negra, ¿Cuál es la probabilidad de que la cara que está hacia abajo sea negra?
Sean los eventos
A: la cara que está hacia abajo es negra
B: la cara que está hacia arriba es negra
C: la cara que está hacia abajo es blanca
D: la cara que está hacia arriba es blanca

Este es una caso en que la probabilidad se calcula después de sacar una tarjeta y observar que la cara que está hacia arriba es negra, ya que se pide la probabilidad de la que está abajo sea negra.



$$P(A|B) = \frac{P(A)\,P(B|A)}{P(A)\,P(B|A) + P(C)P(B|C)} = \frac{(3/4)(\tfrac{1}{2})}{(3/4)(\tfrac{1}{2}) + (\tfrac{1}{4})\,(1)} = \frac{3/8}{3/8 + \tfrac{1}{4}} = \frac{3/8}{5/8} = 3/5$$

Si el color que está hacia abajo es negro, el de arriba puede ser negro o blanco, pero si el color que está hacia abajo es blanco forzosamente el de arriba será negro.

P(A) = 3/4 ya que hay 3 alternativas de obtener dicho color de un total de 4.

Ejercicios:

1.- En los siguientes ejemplos se dan dos eventos A y B, identificar en cada caso que pares son mutuamente exclusivos:

a) A: ser un hijo de un abogado
   B: haber nacido en Querétaro
b) A: ser menor de 18 años de edad
   B: votar legalmente en las elecciones para presidente
c) A: tener un nissan
   B: tener un ford lincon

2.- Cuáles de los siguientes pares de eventos son independientes?

a) A: estar intoxicado
   B: tener un accidente automovilístico
b) A: llegar a tiempo a clase
   B: tener buen tiempo por la mañana

3.- Una caja contiene 9 tornillos bien construidos y 3 defectuosos. Si se escogen al azar 5 tornillos de la caja sin volverlos a regresar, calcular la probabilidad de los siguientes eventos:

a) que ningún tornillo sea defectuoso
b) sacar 3 tornillos defectuosos y 2 bien construidos
c) que los 5 tornillos esten bien construidos
d) obtener un tornillo defectuoso dado que ya se sacaron 4 bien construidos

4.- Un juego consiste en lanzar 2 dados balanceados, un jugador apuesta que su suma será 7 si pierde pagará n$1. ¿Cuánto deberá recibir si gana para que la apuesta sea pareja, es decir, para que ninguno de los jugadores tenga ventaja sobre el otro?.

5.- Dados 3 eventos cualesquiera A,B,C suponer que A y B son independientes y que B y C son mutuamente excluyentes. Sus probabilidades son respectivamente: $P(A) = .5$ $P(B) = .3$ $P(C)=.1$
Expresar los siguientes eventos en notación de conjuntos y calcular sus probabilidades:
a) B y C ocurren simultáneamente
b) al menos uno de los eventos A o B ocurre
c) B no ocurre
d) los tres eventos ocurren

6.- Los alumnos de un grupo deciden realizar una rifa a fin de recabar fondos para su graduación. Son 50 boletos y sólo 45 alumnos por lo que el profesor decidió comprar los 5 boletos restantes.
a) ¿Cuál es la probabilidad de que un alumno quien adquirió un boleto obtenga el premio?
b) ¿Cuál es la probabilidad de que el profesor gane la rifa?
c) ¿Cuál es la probabilidad de que el premio quede en manos de un alumno?

7.- El espacio muestral asociado a cierto experimento está formado por los siguientes eventos elementales X_1,X_2,X_3,X_4,X_5. Si $P(X_1) = 0.01$ $P(X_2) = 0.02$ $P(X_3) = 0.04$
¿Cuál debe ser el valor de $P(X_4)$ si se sabe que $P(X_4) = P(X_5)$?

8.- Se consultan 10 familias con respecto al programa de "hoy no circula" contra la contaminación. De ellas 7 familias se oponen al programa y 3 estan a favor o son indiferentes.
a) Cuál es la probabilidad de que 2 familias escogidas aleatoriamente se opongan al programa?
b) Cuál es la probabilidad de que una de ellas se oponga y la otra sea indiferente?
c) Cuál es la probabilidad de que las 2 familias sean indiferentes?

9.- Dos hombres h_1 y h_2 y tres mujeres m_1,m_2,m_3 intervienen en un torneo. Los del mismo sexo tienen la misma probabilidad de ganar, pero cada hombre tiene el doble de probabilidad de ganar que una mujer.
a) Hallar la probabilidad de que una mujer gane el torneo.
b) Si h_1 y m_1 son casados hallar la probabilidad de que uno de ellos gane el torneo.
c) Hallar la probabilidad de que cualquiera de los hombres gane el torneo.

10.- Tres caballos C_1,C_2,C_3 intervienen en una carrera. El caballo C_1 tiene el doble de probabilidad de ganar que C_2, y el caballo C_2 tiene el triple de probabilidad de ganar que C_3. Hallar la probabilidad de ganar de cada uno de los caballos. Hallar también la probabilidad de que el caballo C_1 no gane.

11.- En un grupo de 100 alumnos, 70 estudian matemáticas, 60 estudian fisica y 40 ambas materias. Si se selecciona un alumno al azar, calcular la probabilidad de que el alumno seleccionado:
a) estudie matemáticas pero no fisica
b) estudie matemáticas y fisica
c) no estudie matemáticas ni fisica
Nota: emplear diagramas de Venn para resolverlo

12.- Tres hombres en una tienda departamental escogen al azar cada uno de ellos una corbata de color de un gran montón en donde hay 5 corbatas verdes, 3 rojas y 2 amarillas. Cuál es la probabilidad de que los 3 hombres escojan una corbata del mismo color?

13.-Un dado tiene el número 1 en tres de sus caras, el número 2 en dos de ellas y el número 3 en la cara restante. Construir el espacio muestral asociado al juego de lanzar en dos ocasiones el dado, siendo los posibles resultados la suma de los puntos obtenidos. Calcular la probabilidad correspondiente a cada elemento del espacio muestral.

14.- Una máquina deja de funcionar si dos de sus partes se rompen, lo cual ocurre con una probabilidad de 0.1. Si la probabilidad de que se rompa la 1era. pieza es 0.3. Calcular la probabilidad de que la máquina no funcione si ya se ha roto una pieza.

15.- Un grupo de turistas que visitan Cancún esta formado por 4 ingleses, 3 escoceses y 2 irlandeses. Si se seleccionan 4 de ellos para visitar Tulúm, calcular la probabilidad de que:
a) dos de los escogidos sean ingleses
b) dos de los escogidos sean irlandeses
c) se seleccione un turista de cada nacionalidad
d) dos sean ingleses y dos sean irlandeses
e) que el cuarto turista sea inglés dado que ya se seleccionaron 3 ingleses

16.- Se lanzan dos dados una sola vez. Hallar la probabilidad de que la suma de los números sea 10 si:
a) aparece un 5 en el primer dado
b) aparece un 5 en uno de los dados por lo menos

17.- Supóngase que la probabilidad de que un hombre siga viviendo dentro de veinte años es de 4/5 y la probabilidad de que su esposa siga viviendo dentro de los mismos veinte años es de 5/6. Hallar la probabilidad de que:
a) sigan viviendo los dos
b) viva sólo el hombre
c) viva sólo la mujer
d) no viva ninguno de los dos

18.- Un examen semanal de estadística esta formado por 3 preguntas de falso y verdadero, si los alumnos no estan preparados para resolver el examen, cuál es la probabilidad de que los alumnos puedan obtener dos respuestas correctas de las tres?

19.- Un ama de casa saca su "serie" de foquitos para adornar el árbol de navidad, la serie enciende solamente si cada bulbo esta bien. La serie esta formada por 4 bulbos y la probabilidad de que cada bulbo este bien es de .9 cuál es la probabilidad de que:
a) encienda la serie completa
b) no encienda la serie completa

20.- Una familia formada por 3 integrantes va a hacerse un foto-estudio. La fotografía tiene exito si cada una de las 3 personas se ve bien en la foto, en caso contrario no sirve. Suponer que la probabilidad de que cada persona se vea bien es de .4, si cada persona es independiente:
a) cuál es la probabilidad de que la fotografia sea un exito
b) si el fotógrafo toma dos fotografias, cuál es la probabilidad de que al menos una de ellas sea un exito?

EVALUACION PROPUESTA

1.- Se escoge un número al azar entre 1 y 21. Calcular la probabilidad de que:
a) el número sea mayor que 5 pero no mayor a 10
b) que el número sea divisible entre 7 pero no por 3

2.- Cuál es el espacio muestral asociado al experimento que consiste en lanzar una moneda y un dado. Calcular la probabilidad de que:
a) aparezca un sol y un número impar
b) aparezca un águila y un número par

3.- El espacio muestral asociado a un experimento consta de los siguientes elementos
$E=\{X_1, X_2, X_3, X_4, X_5\}$ Si $P(X_1)= 0.01$ $P(X_2)=0.02$ $P(X_3)=0.04$ ¿Cuál es la probabilidad de $P(X_4)$ si se sabe que $P(X_4)=P(X_5)$?

4.- Se lanzan dos dados simultáneamente. ¿Cuál es la probabilidad de que en alguno de los dados aparezca el número 5 sabiendo que la suma de los números de las dos caras es 6?

5.- En un grupo de 50 estudiantes, 30 de ellos son casados, 15 de ellos saben hablar inglés y 10 son casados y saben hablar inglés al mismo tiempo. Si se elige uno de los estudiantes al azar ¿Cuál es la probabilidad de que el elegido sepa hablar inglés si es casado?

6.- Un grupo de alumnos realiza una rifa. Son 50 boletos y sólo 45 alumnos, por lo que el profesor decidió comprar los 5 boletos restantes.
a)Cuál es la probabilidad de que un alumno que compró un boleto,obtenga el premio?
b)Cuál es la probabilidad de que el profesor gane la rifa?

7.-Cuál es el espacio muestral asociado al experimento de lanzar una ficha de dos caras-una blanca y la otra negra- y un dado simultáneamente?
a)Cuál es la probabilidad de que aparezca una cara blanca y un numero par?
b)Cuál es la probabilidad de que aparezca una cara negra y un numero impar?

8.-Dos hombres h_1, h_2 y tres mujeres m_1, m_2, m_3 intervienen en un torneo. Los del mismo sexo tienen las mismas probabilidades de ganar, pero cada hombre tiene el doble de probabilidad de ganar que una mujer.
a)Hallar la probabilidad de que una mujer gane el torneo.
b)Si h_1 y m_1 son casados hallar la probabilidad de que uno de ellos gane el torneo?

9.-Una máquina deja de funcionar si dos de sus partes se rompen, lo cual ocurre con una probabilidad de 0.1. Si la probabilidad de que se rompa la 1er. pieza es 0.3, calcular la probabilidad de que la máquina no funcione si ya se ha roto una pieza.

10.- Se consultan 10 familias con respecto al programa de "hoy no circula" contra la contaminación. De ellas 7 familias se oponen al programa y 3 estan a favor o son indiferentes.
a) Cuál es la probabilidad de que 2 familias escogidas aleatoriamente se opongan al programa?
b) Cuál es la probabilidad de que una de ellas se oponga y la otra sea indiferente?
c) Cuál es la probabilidad de que las 2 familias sean indiferentes?

CONCLUSIONES

Los resultados obtenidos de la evaluación escrita a los alumnos del grupo 505 de estadística con respecto a los conceptos del tema de probabilidad se proporcionan en una tabla de distribución de frecuencias en 4 intervalos para poder calcular las medidas de tendencia central correspondientes y poder emitir un juicio acerca del aprendizaje por parte de los alumnos. Se proporcionan también dos gráficas una circular y una de barras para observar más claramente dichos resultados.

2	4	5	6	7	7	8	9
2	4	6	6	7	7	9	9
3	5	6	6	7	8	9	9
3	5	6	6	7	8	9	9
4	5	6	7	7	8	9	9

Intervalos	Clases	Frecuencia	Marcas de clase	Marcas* frecuencia
0 - 2	0 - 2.5	2	1	2* 1 = 2
3 - 5	2.5 - 5.5	9	4	9* 4 = 36
6 - 8	5.5 - 8.5	20	7	20* 7 =140
9 - 11	8.5 - 11.5	9	10	9*10 = 90
	total	40		268

$\bar{X} = 268 / 40 = 6.7$

$$X_0 = 5.5 + \left(\frac{11}{11+11}\right) 3 = 7$$

$$\tilde{X} = 5.5 + \left(\frac{20 - 11}{20}\right) 3 = 6.8$$

Diagrama circular:

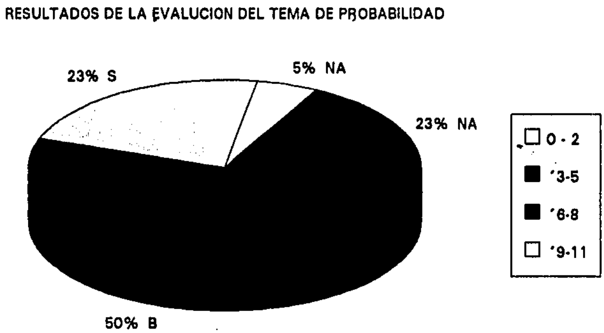


Diagrama de barras:

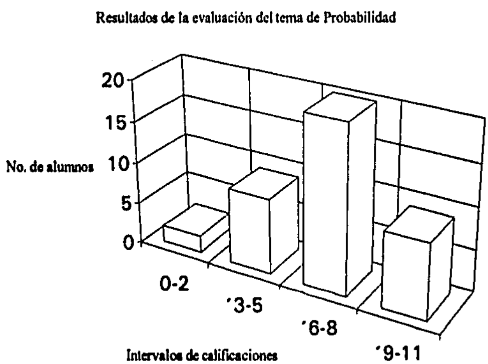
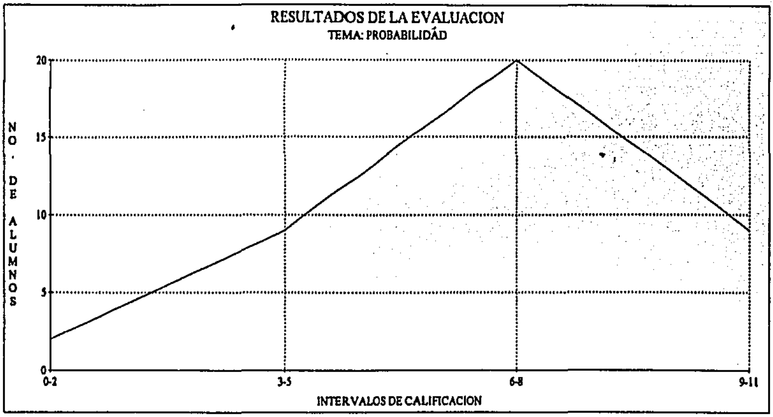


Diagrama: Polígono de frecuencias



La gráfica de los resultados obtenidos de la evaluación escrita del tema de probabilidad indican que más del 75% de los alumnos del grupo 505 de estadística comprendieron los conceptos relativos a éste. Es un resultado muy significativo con relación a que es el primer acercamiento formal a estos conceptos por parte de los alumnos.

TEMARIO

En la selección de los elementos de la población para formar una muestra aleatoria sobre la cual se aplicará un experimento, intervienen varios métodos númericos que facilitan dicha selección. Entre los conceptos básicos necesarios para aplicar los métodos se tiene el siguiente:

Definición.- El **factorial** de un número N se denota por N!, y es una sucesión de multiplicaciones que empiezan con N y van descendiendo de 1 en 1 hasta llegar a 1, es decir, se obtiene de la siguiente manera:

N! = N (N-1) (N-2) (N-3) ....(3) (2) (1)

Ejemplos:

a) 5! = 5*4*3*2*1 = 120

b) 9! = 9*8*7*6*5*4*3*2*1 = 362880

c) 3! + 4! = (3*2*1) + (4*3*2*1) = 6 + 24 = 30

d) 6! / 2! = (6*5*4*3*2*1) / (2*1) = 720/2 = 360

e) (7! - 3!) / 4! = [(7*6*5*4*3*2*1) - (3*2*1)] / (4*3*2*1) = (5040 - 6) / 24 = 209.75

f) 2! + 3! + 4! / 3! = (2+6+24) / 6 = 32 /6 = 5.3

PRINCIPIO FUNDAMENTAL DE CONTEO

Este principio se refiere al método de cómo se pueden ordenar los distintos resultados de un evento en particular considerando todos los datos del espacio muestral del experimento. Estos resultados son fáciles de obtener mediante un listado, un diagrama de árbol o un cuadro de doble entrada si el número total de resultados del espacio muestral es pequeño.

Definición.- Si un experimento1 únicamente puede realizarse de **m** formas y otro experimento 2 de **n** formas entonces el número de formas diferentes en que pueden efectuarse ambos experimentos 1 y 2 es de **m.n** formas.

En otras palabras si un evento1 de un experimento tiene n_1 posibles resultados, un evento 2 tiene n_2 posibles resultados y así sucesivamente para r eventos entonces todos los r-eventos tienen los siguientes posibles resultados: $(n_1)(n_2)(n_3)......(n_r)$

1.- El consejo administrativo de una escuela tiene que escoger un director y una directora de entre un grupo de 9 candidatos formado por 4 hombres y 5 mujeres. ¿ De cuántas formas distintas se pueden seleccionar el director y la directora?

El director se puede escoger de 4 formas distintas y la directora de 5 formas distintas, por lo tanto el consejo puede seleccionar de $(4)(5) = 20$ formas distintas.

2.- ¿Cuántos datos contiene el espacio muestral asociado al experimento de lanzar 3 monedas al aire simultáneamente?

Como cada moneda solamente tiene 2 resultados posibles - águila y sol- entonces se tiene que el número de posibles resultados de lanzar 3 monedas es: $(2)(2)(2) = 8$ resultados.

3.- El chofer de un camión puede tomar cualquiera de tres carreteras para ir de la ciudad A a la ciudad B, y puede tomar cualquiera de 4 carreteras para ir de la ciudad B a la ciudad C; finalmente tiene 3 carreteras posibles para ir de la ciudad C a la ciudad D. El desea ir de la ciudad A a la ciudad D pasando por las 3 ciudades ¿ Cuántas rutas posibles puede escoger para ir de A a D?

Sea m = número de rutas de A a B = 3
n = número de rutas de B a C = 4
o = número de rutas de C a D = 3

Por lo tanto para ir de la ciudad A a la ciudad D pasando por las 3 ciudades tiene las siguientes posibles rutas: $mno = (3)(4)(3) = 36$

PERMUTACIONES

Uno de los problemas a los que se enfrenta una persona que realiza una investigación es cuando está tratando de seleccionar los datos que respondan lo más preciso posible al objetivo específico planteado, en otras palabras si se tiene el espacio muestral de todos los datos posibles de un experimento ¿cómo se pueden seleccionar aquellos que cumplen con un evento simple en particular?.

En el tema respectivo se consideró la probabilidad clásica como el procedimiento que consiste en contar el número de elementos que forman el conjunto de resultados de un suceso en particular bien definido en el espacio muestral y el número de elementos del espacio muestral para formar el cociente de los dos números. En muchos problemas es posible encontrar estos dos números sin hacer la lista de todos los elementos que integran el espacio muestral y los del espacio del evento respectivamente, sin embargo siempre hay que decidir como formar estos dos conjuntos de resultados es decir, saber de ¿cuántas formas se pueden encontrar dichos resultados?.

A estas formas de encontrar los resultados de un experimento se les llama permutaciones y combinaciones.

Definición.- Si hay n datos diferentes y se tiene que escoger una muestra de tamaño r de datos todos diferentes donde n > r y ordenarla entonces el número posible de muestras ordenadas diferentes se llaman las **permutaciones** de n elementos tomadas de r en r y se denotan por

$$P_r^n = n(n-1)(n-2)\ldots(n-r+1) = \frac{n!}{(n-r)!}$$

Se debe de observar que sólo se puede elegir una vez cada elemento. Las muestras que se obtienen se llaman permutaciones de r elementos, y son muestras sin reemplazo. Una permutación de r datos es cualquiera de los diferentes arreglos de esos datos en un orden definido.

Definición.- Si se tienen n datos diferentes y se selecciona una muestra ordenada de tamaño r de datos diferentes donde n > r en donde es válido el reemplazamiento, es decir, en donde se elige cada dato se anota y se devuelve antes de obtener el siguiente, entonces el número posible de muestras ordenadas diferentes se llaman **permutaciones con reemplazo** de n elementos tomados de r en r y se denotan por

$n^r = n.n.n...n$ (r veces)

Ejemplo:

1.-Un saco contiene 4 bolas de distintos colores -una roja, una blanca, una azul y una verde- y hay que extraer dos bolas del saco, es decir, se van a formar muestras de tamaño 2.

a) ¿Cuántas muestras ordenadas sin reemplazo se pueden obtener?

b) ¿ Cuántas muestras ordenadas con reemplazo se pueden obtener?

La lista de todas las permutaciones de dos bolas formadas a partir de las 4 bolas de colores es:

E = {(rv),(rb),(ra),(vb),(va),(ba),(vr),(br),(ar),(bv),(av),(ab)}

este espacio se puede apreciar mejor con un cuadro de doble entrada:

$b_1 \backslash b_2$	r	b	a	v
r	rr	rb	ra	rv
b	br	bb	ba	bv
a	ar	ab	aa	av
v	vr	vb	va	vv

a) Como sólo hay una bola de cada color se pueden eliminar los pares de la diagonal del cuadro, y se obtienen los doce pares de bolas de distinto color, es decir,

$$P_2^4 = \frac{4!}{(4-2)!} = \frac{4(3)(2)(1)}{(2)(1)} = \frac{24}{2} = 12$$

b) Como el muestreo es con reemplazo esto significa que puede haber pares de la forma de la diagonal del cuadro, por lo tanto se tendrían 16 pares de bolas, es decir,

$$P_2^4 = 4^2 = 16$$

2.- Un estudiante compró 3 libros que necesitaba b_1, b_2, b_3. De cuántas maneras puede ordenar sus 3 libros en un librero si se escogen dos libros a la vez?

combinaciones	reordenaciones
$b_1 b_2$	$b_2 b_1$
$b_1 b_3$	$b_3 b_1$
$b_2 b_3$	$b_3 b_2$

$$ {}^{3}P_2 = \frac{3!}{(3-2)!} = \frac{(3)(2)(1)}{1} = 6 $$

Este resultado es fácilmente deducido de la regla m.n ya que el primer libro puede escogerse de 3 maneras distintas pero el segundo solamente de 2 maneras distintas por lo tanto se puede ordenar en un librero de (3)(2) = 6 maneras distintas.

COMBINACIONES

Definición.- Si hay n datos diferentes y se tiene que escoger una muestra de tamaño r de datos sin reemplazamiento donde n > r entonces el número posible de muestras diferentes se llaman las **combinaciones** de n elementos tomadas de r en r y el total se designa por el **coeficiente binomial** es decir,

$$ \binom{n}{r} = {}^{n}C_r = \frac{n!}{r!\,(n-r)!} $$

En una combinación no tiene importancia el orden de los datos. El número de combinaciones de n datos tomados de r en r corresponde a la elección aleatoria y sin reemplazo.

Todas las combinaciones producidas por el coeficiente binomial son igualmente probables porque representan la elección aleatoria de r datos del evento a partir de los n datos del espacio muestral del experimento.

Ejemplo:

1.-Utilizando los datos del ejemplo anterior de las 4 bolas de distinto color, calcular todas las posibles muestras sin reemplazamiento que se pueden extraer, es decir, las combinaciones que se pueden obtener:

A = { (rv),(rb),(ra),(vb),(va),(ba) }

Utilizando el cuadro de doble entrada se observa que si la primera bola que se extrae es roja, entonces la segunda bola puede ser de cualquiera de los tres colores restantes, pero como la selección es sin reemplazo entonces para las otras opciones de la primera bola de los pares ya no podemos encontrar la bola roja en la bolsa porque ya se seleccionó. Sucede exactamente lo mismo con la bola verde y la blanca.

Efectuando el cálculo da:

$$C_2^4 = \frac{4!}{2!\,(4-2)!} = \frac{4(3)(2)(1)}{2(1)\;2(1)} = \frac{12}{2} = 6$$

2.- Un fábricante de televisores desea adquirir tubos de pantalla de televisión y solamente hay 5 productores de dichos tubos especiales y su calidad varia según el productor. Si se seleccionan a 3 de los 5 productores al azar ¿ Cuál es la probabilidad de que en la selección se tengan exactamente a 2 de los 3 mejores?

Sin necesidad de hacer una enumeración de los datos del espacio muestral puede afirmarse que cada dato del evento de seleccionar a 3 de ellos, es decir, cada combinación de 3 tiene la misma probabilidad.

$$C_3^5 = \frac{5!}{3!\,(5-3)!} = \frac{(5)(4)(3)\,2!}{(3)(2)(1)\,2!} = \frac{60}{6} = 10$$

Para obtener exactamente 2 de los 3 mejores productores se calculan las combinaciones y se denotan por n:

$$C_2^3 = \frac{3!}{2!(3-2)!} = \frac{(3)(2)(1)}{(2)(1)\,(1)} = 3$$

Para calcular el número de formas en que se puede seleccionar un productor, el restante, entre los 2 peores se calculan las combinaciones y se denotan por m:

$$C_1^2 = \frac{2!}{1!\,(2-1)!} = 2$$

Por lo tanto el total de formas en las que se eligen exactamente dos de los mejores productores entre 3 de ellos es: $mn = (3)\,(2) = 6$

por lo tanto la probabilidad es:

$$P\,(A) = \frac{6}{10} = \frac{3}{5}$$

Para seleccionar los datos del espacio muestral que integren el espacio de un evento para determinar su probabilidad se tiene que tomar en cuenta que si dicha selección es con reemplazo entonces deben considerarse todas las permutaciones de los datos; pero en caso de que la selección sea sin reemplazo entonces debe considerarse sin importar el orden de los datos o si no importa el orden de ellos para utilizar las permutaciones con reemplazo o las combinaciones respectivamente.

Sugerencias para utilizar las distintas reglas de conteo

1.- Una de las 3 reglas de conteo se puede utilizar en un problema de probabilidad si los elementos de un espacio muestral son identificables por un número fijo de características.

2.- La regla **mn** puede ser aplicable si sólo se considera una característica para los datos de cada espacio distinto de evento distintos.

3.- Las combinaciones se pueden calcular si las características se toman de un sólo conjunto y el reordenamiento de las características no produce otro dato del espacio muestral.

4.- Las permutaciones se pueden calcular si las características se toman de un sólo conjunto y cada reordenamiento de ellas corresponde a un nuevo dato del espacio muestral.

Ejercicios:

1.- Un viajero puede ir de San Francisco en 3 aerolíneas distintas y cada una tiene 4 vuelos diarios y directos. Si la selección de una aerolínea y un vuelo particular representa un punto muestral.¿Cuántas características lo definen? ¿De cuántos conjuntos se extraen estas características? Utilizar la regla mn para dar el número de formas de que dispone un viajero para aerolínea-vuelo? Construir un cuadro de doble entrada para identificar las combinaciones aerolínea-vuelo.

2.- Un candado de combinación se abre sólo cuando la combinación correcta de los 3 dígitos es seleccionada. Cada dígito puede ser cualquier número entre 0 y 9. Si una combinación particular de dígitos representa un dato del espacio muestral ¿Cuántas características se estan utilizando para definirlo?. Utilizar la regla mn para encontrar las posibles combinaciones del candado.

3.- En una asociación se va a elegir el presidente, el vicepresidente, el secretario y el tesorero de entre 10 posibles candidatos. Utilizar la regla de las permutaciones para encontrar el número de formas distintas en que esos puestos pueden ocuparse por los candidatos.

4.- Un equipo de tenis tiene 10 jugadores entre los cuales se escogerán para jugar los 6 torneos de singles. Los torneos de singles se listan como singles 1, singles 2, etc. Utilizando la regla de las permutaciones encontrar el número total de asignaciones jugador-posición.

5.- Una empresa seleccionará 5 secretarias para ofrecerles contrato de entre un grupo de 20 solicitantes. Utilizar la regla de las combinaciones para encontrar el número total de formas distintas en que se pueden seleccionar las 5 secretarias.

6.- Se efectuó un estudio para determinar las actitudes de las enfermeras de un hospital frente a diversas disposiciones administrativas. Si se seleccionó una muestra de 10 enfermeras de un total de 90, ¿cuántas muestras posibles había? (nótese que el orden no importa)

7.- ¿De cuántas formas pueden seleccionarse 3 bolas con reemplazamiento de una bolsa que contiene 9 bolas negras, 4 bolas blancas y 13 bolas rojas? ¿Es importante el color? ¿Cuántos eventos simples tiene el espacio muestral?

8.- Si se seleccionan 5 cartas con reposición de un conjunto de 52 cartas, esto es se selecciona al azar la primera y se regresa al conjunto de cartas, se selecciona al azar la segunda carta y se regresa, etc. ¿Cuántas selecciones posibles hay?

9.- En un reunión se utiliza un juego de 52 cartas de 4 figuras distintas. A cada jugador se le dan 5 cartas ¿Cuál es la probabilidad de que se tengan 2 cartas de diamantes, 2 cartas de corazones y 1 carta de tréboles?

10.- Un juego de 52 cartas está dividido en partes iguales en dos colores negro y rojo. Si se extraen 4 cartas al azar ¿Cuál es la probabilidad de sacar 2 cartas negras y 2 cartas rojas?

107

TEMARIO

INTRODUCCION

La Estadística como se ha visto se emplea para poder referirse a ciertos cálculos que caracterizan de manera particular a una colección de datos y que se pueden apreciar por medio de diagramas. Sin embargo también se puede referir a los métodos empleados para coleccionar, procesar o interpretar cierta cantidad de datos sin tener que considerarlos todos; esta parte trata de los cálculos que se pueden efectuar utilizando las observaciones ya obtenidas como una base se realizan estimaciones o predicciones acerca del comportamiento de los datos, es decir, inferencias acerca de una situación o fenómeno que aún no se ha efectuado sobre dicha colección de datos.

En particular se trata en esta sección de las distintas formas en que se pueden obtener los datos de un evento en particular dentro del conjunto total de posibles resultados de un experimento, es decir, del espacio muestral del experimento. Con el objetivo de determinar la probabilidad de ocurrencia de dicho evento.

Dentro de la estadística inferencial existe una clara diferencia entre lo que se considera una población y una muestra y aún más las distintas muestras que se pueden formar de acuerdo a los distintos eventos que se determinen; ya que ésta se encarga de manejar los datos de una muestra para hacer estimaciones que afectarán a la población entera. Con este fin se introduce un nuevo concepto de variable que permite identificar las características seleccionadas de los eventos en estas muestras.

En esta sección se tratarán funciones con valores definidos sobre espacios muestrales, es decir, sobre espacios de eventos en donde a la variable que asume dichos valores se le llama variable aleatoria o estocástica.

VARIABLE ALEATORIA

Las muestras están formadas por individuos que pueden ser seres humanos, meses del año, lluvia, insectos, etc.; dichos individuos o elementos de la muestra tienen una característica en común en la cual se esta interesado, dicha característica varía de un miembro a otro. Por lo tanto, una variable es cualquier atributo o característica que permitierá distiguir entre un individuo y otro. Para poder trabajar matemáticamente dichas variables se les relaciona con un número.

Por otra parte un experimento es un proceso que termina con la toma de mediciones o datos, la mayoría de los experimentos producen mediciones númericas que desde luego varian al considerar distintos puntos muestrales y esta variación es aleatoria, a dicha medición se le llama una **variable aleatoria.**

En este caso la muestra aleatoria esta formada por un espacio de eventos y a cada uno de ellos se les asignará un número real.

Definición.-Una **variable aleatoria** X es una función definida en un espacio de eventos $E=\{e_1, e_2,\dots,e_n\}$ en donde a cada elemento de E le corresponde un número real único $X(e_i)$, $i = 1,2,\dots,n$. Es decir, $X(e_i) = x$ donde $x \in \mathbb{R}$

De esta manera la función X valuada en e_i determina un y sólo un valor de la función que se designa por x .

Es decir, a las funciones X: **E** ---> R se les llama variables aleatorias.

En otras palabras una variable aleatoria X es una variable que toma un valor númerico único para cada uno de los resultados en un espacio muestral de un evento probabilístico.

Definición.- A las variables aleatorias tales que X(E) es contable se les llama **variables aleatorias discretas**. Es decir, es aquella que toma a lo más una cantidad numerable de valores distintos.

El hecho de que la cantidad de valores que pueda tomar una variable aleatoria discreta sea **contable** quiere decir que todos los valores se pueden asociar a los números enteros.

Definición.- A las variables aleatorias donde X(E) es infinito numerable se les llama **variables aleatorias contínuas**. Es decir, es aquella que puede tomar cualquier valor de entre todos los contenidos en un intervalo de la recta.

En las variables aleatorias contínuas se establecen escalas para medir la cantidad de valores que pueden asumir dichas variables. Es decir, sus valores se asocian a los números reales.

Es importante establecer la diferencia entre ambas variables aleatorias ya que requieren modelos probabilísticos distintos, por ejemplo las probabilidades asociadas a cada valor posible de una variable aleatoria discreta suman 1, y esto no es posible para las probabilidades asociadas a una variable aleatoria contínua. En el presente trabajo se manejarán unicamente variables aleatorias discretas debido a que las variables aleatorias continuas requerirían que el alumno manejara herramientas matemáticas más avanzadas por ejemplo que estuviera familiarizado con el uso de límites e integrales.

Ejemplos de variables aleatorias discretas:

1) Considerar el experimento que consiste en lanzar dos monedas al aire y observar las caras que quedan hacia arriba. El espacio de eventos es E={aa,as,sa,ss}; si se selecciona la variable número de soles observados, los valores que puede asumir la variable son X={0,1,2}. De esta manera se ha establecido una relación en la que a cada resultado de E le corresponde un único valor de X:

X(aa)=0 X(as)=1 X(sa)=1 X(ss)=2

2) En un experimento realizado sobre las 20 familias que habitan en una cuadra que tienen en promedio 4 hijos. Si se selecciona la variable número de hijos varones que tiene cada familia, los valores que puede asumir la variable aleatoria X son X={0,1,2,3,4}. De esta manera se ha establecido una relación en la que a cada familia le corresponde un único valor de X.

Ejemplos de variables aleatorias continuas:

1) En un experimento se considera el peso en kilogramos de 1293 muchachos de 11 años de edad, para considerar todos los pesos se establecen escalas o intervalos en los cuales se pueden clasificar todos los pesos. El peso de un muchacho seleccionado aleatoriamente de entre los 1293 es una variable aleatoria continua.

2) La cotización de alguna acción determinada de una compañía en la bolsa de valores al final del día.

Ejercicios:

Identifique si las siguientes variables aleatorias son continuas o discretas:

a) número de transistores defectuosos en un embarque de 10,000 transistores

b) número de robos ocurridos en un almacén en determinado período de tiempo

c) El tiempo requerido por un empleado para terminar una tarea determinada cuando es observado en un estudio de "tiempos y movimientos"

d) La cantidad de gasolina consumida por un vehículo en una prueba de 100 kilómetros.

e) El número de accidentes de automóvil por año en el Estado de México

f) La cantidad de grano producido por acre

g) La duración de una bombilla eléctrica observada en un experimento

h) Las ventas brutas de un supermercado en un día determinado

i) El punto de fátiga en kilogramos por centímetros cuadrados de un cable de acero de 2 cm. de diámetro.

j) La demanda diaria de energía eléctrica en una determinada ciudad

FUNCION DE DISTRIBUCION DE PROBABILIDADES

En un determinado experimento una variable aleatoria siempre esta asociada a un evento porque siempre existe un evento simple o compuesto que origina un valor para dicha variable. De esta forma si un valor de la variable aleatoria está asociada a un evento simple o compuesto de E entonces dicho valor deberá tener una probabilidad de ocurrencia en E.

La distribución de probabilidades de una variable aleatoria discreta se puede representar por medio de una fórmula, tabla o gráfica que proporciona las probabilidades asociadas a cada posible valor de la variable aleatoria.

Ejemplo:

Considerar el "número de soles" que se obtienen en el experimento de lanzar dos monedas. El espacio de eventos es E={0,1,2} por lo tanto se tiene que:

$P(x=0)=\frac{1}{4}$ $P(x=1)=2/4$ $P(x=2)=\frac{1}{4}$

X	0	1	2
P(X)	1/4	2/4	1/4

Donde x=0 significa que no aparezca ningún sol, es decir, (a,a)
x=1 significa que aparezca un sol, es decir, (s,a) (a,s)
x=2 significa que aparezcan dos soles, es decir, (s,s)

En resumen la probabilidad de que la variable aleatoria X asuma el valor x se denota por $P(X=x_i)$

Definición.-Una **función de distribución de probabilidad** de una variable aleatoria discreta está determinada por las probabilidades asociadas a todos los valores que la variable aleatoria X puede asumir, es decir, es la función f(x) que toma los valores $p_1, p_2, \ldots, p_n$ cuando x toma los valores $x_1, x_2, \ldots, x_n$ en otras palabras es el conjunto de pares ordenado $\{x_i, f(x_i)\}$ donde x_i es un número real. Por lo tanto

$$f(x_i)=P(X_i)=x_i$$

$$\sum_{i=1}^{n} f(x_i)=1$$

A esta función también se le conoce como **Función de Densidad** de X

Ejemplo de función de densidad:

Continuando con el experimento de lanzar una vez dos monedas al aire, se determina la función de distribución de probabilidad de la variable aleatoria número de soles observados:

1) Se determina el espacio de eventos: E={aa,as,sa,ss}
2) Se define la variable aleatoria X : X=No. de soles observados
3) Se define la función $X(e_i)=x_i$

X(aa)=0
X(as)=1
X(sa)=1
X(ss)=2

Por lo tanto X={0,1,2}

4) Se obtienen los eventos asociados a cada valor de X en el espacio de eventos

$X_1 = 0$ cuando ocurre {aa}

$X_2 = 1$ cuando ocurre {as,sa}

$X_3 = 2$ cuando ocurre {ss}

5) Se obtienen los valores de $f(x_i) = P(X = x_i)$

$f(0)=P(X=0)= 1/4$

$f(1)=P(X=1)= 2/4$

$f(2)=P(X=2)= 1/4$

6) Se organizan los datos tabularmente:

x	0	1	2
f(x)	¼	2/4	¼

Ejemplo de función de distribución de probabilidades de una variable aleatoria contínua, es decir, con valores al infinito numerable.

El experimento consiste en tirar una moneda al aire repetidamente hasta que salga águila.

Solución:
1) Espacio de eventos E ={a,(s,a),(s,s,a),(s,s,s,a),...}
2) Variable aleatoria X: No. de intentos para obtener un águila
3) Se define la función

$f(x_1)=P(X=x_1)=1/2$

$f(x_2)=P(X=x_2)=1/4$

$f(x_3)=P(X=x_3)=1/8$

....

4) Se organizan los datos tabularmente:

X	1	2	3	4	5...	# de intentos en q'se obtuvo águila
f(X)	1/2	1/4	1/8	1/16	1/32 ...	

Si se observa cada probabilidad que se obtiene se puede percibir que:
la segunda probabilidad 1/4 = (1/2) (1/2)
la tercera probabilidad 1/8 = (1/2) (1/4)
la cuarta probabilidad 1/16 = (1/2) (1/8)
la quinta probabilidad 1/32 = (1/2) (1/16)
y así sucesivamente
A esta sucesión de valores se le llama una progresión geométrica ínfinita de razón ½ , que en este ejemplo en particular esta dada por el número de tiradas que se realizan.

FUNCION DE DISTRIBUCION ACUMULADA

Definición.- La función de distribución de probabilidades acumulada se define como el conjunto de las sumas parciales de las probabilidades f(x) correspondientes a todos los valores de la variable aleatoria X que sean menores o iguales que X. Esta función da las probabilidades de que la variable aleatoria discreta X (escalar) tome valores menores o iguales que x. para cualquier i, es decir:

$$F(X_i) = P(X_i) = \sum_{i=1}^{n} f(x_i) \quad \text{para } i=1,2,\ldots,n$$

se le llama Función de Distribución Acumulativa de X.

Ejemplo:

1) Continuando con el ejemplo anterior de lanzar dos monedas al aire, se tiene que la función de distribución de probabilidades acumulada es:

X	0	1	2
f(X)	¼	2/4	¼
F(X)	¼	3/4	1

2) Un experimento consiste en lanzar 3 monedas al aire, si se considera la variable aleatoria X como el número de soles que resultan del lanzamiento se tiene que:

a) El espacio de eventos es: {aaa,aas,asa,ass,saa,sas,ssa,sss}
b) La variable aleatoria es: X = número de soles que resultan
c) Los posibles resultados de la variable aleatoria son: {0,1,2,3}
d) Por lo tanto los resultados de la función de distribución de probabilidad y de la función acumulada de probabilidad son:

X	0	1	2	3
f(X)	1/8	3/8	3/8	1/8
F(x)	1/8	4/8	7/8	1

Ejercicios:

1.-Suponer que un estudio realizado muestra que sólo el 20% de los clientes de un supermercado se toman la molestia de leer los precios de los artículos antes de tomar la decisión de comprarlos.Si n=2 clientes entran a ese supermercado, calcular la distribución de probabilidad de la variable aleatoria X número de clientes que leen los precios para hacer una mejor compra. (Sugerencia: denotar por p(0)= probabilidad de que ninguno de los dos clientes lean las etiquetas.)

2.- Un niño tiene 5 canicas en una bolsa, 4 de las cuales son de un color y una de otro. Si selecciona dos canicas al azar de las cinco, y considera la variable aleatoria X como el número de veces en que la canica de distinto color fue seleccionada. Calcular la función de distribución de probabilidad de X.

3.- Para ahorrarse tiempo y dinero, dos compradores acostumbran inspeccionar sólo una parte de los embarques recibidos, con ello juzgan la calidad del embarque con base a la calidad observada en la porción inspeccionada. Suponga que un comprador recibe un embarque de 4 fotocopiadoras, dos de las cuales él no sabe que están defectuosas, si selecciona dos de ellas al azar para inspeccionar y la variable aleatoria X denota el número de fotocopiadoras defectuosas, encontrar la función de distribución de probabilidad de X y la función acumulada de distribución.

4.- En la mayoría de las dependencias gubernamentales el otorgamiento de contratos se hace con base a un concurso. Suponga que la experiencia que se tiene en cierta dependencia gubernamental muestra que determinado contratista gana en promedio 3 de cada 5 concursos en los que compite. Denotar por X el número de concursos en los que este contratista compite antes de ganar un contrato y suponga que los resultados de cada 10 concursos son independientes entre sí.
a) Encontrar p(1), p(2) y p(3)
b) Exhibir una fórmula para p(x), x=1,2,3,...

5.- Un embarque de 5 automóviles extranjeros incluye 2 que tienen unas ligeras manchas de pintura. Si una agencia recibe 3 de estos vehículos aleatoriamente, y la variable aleatoria X representa el número de automóviles con manchas de pintura comprados por la agencia. Encontrar la función de distribución de probabilidades de X y su función acumulada de distribución.

6.- De una alcancía que contiene 4 monedas de 1000 pesos y 2 de 500, se seleccionan 3 monedas al azar. Determinar la distribución de probabilidad para la variable aleatoria X que denota el número de monedas de 1000 que se pueden extraer. Expresar gráficamente la distribución de probabilidad como un histograma.

7.- Los alumnos de un plantel de nivel medio superior se clasifican en los siguientes 4 grupos de acuerdo a la evaluación su desempeño académico, el secretario académico determinó un promedio. 20% de los alumnos se encuentran bajo el promedio determinado, 60% de ellos cumplen el promedio, 15% se encuentran por arriba del promedio y sólo el 5% de ellos cumplen con todos los requisitos establecidos. Construir la función de distribución de probabilidades donde la variable aleatoria es aquélla en donde escribimos NA para el bajo promedio, S para el promedio, B para los que estan arriba del promedio y MB para los sobresalientes.

8.- Se realiza una investigación de mercado para determinar la preferencia de las amas de casa por un limpiador A y un limpiador B. Se visita a 4 amas de casa y se les pregunta su preferencia. Construir una función de distribución de probabilidades de X y la función de distribución acumulada de cualquiera de los dos limpiadores.

PARAMETROS DE LA DISTRIBUCION DE PROBABILIDADES

La distribución de probabilidades de una variable aleatoria se puede describir en términos de una medida de tendencia central y una medida de dispersión de forma análoga a la que se estudió en estadística descriptiva. En la distribución de una variable aleatoria también intervienen promedios que resumen la tendencia central de los distintos valores de la variable aleatoria y su variabilidad con respecto a dichos promedios.

Las medidas descriptivas más utilizadas de una distribución de probabilidades de una variable aleatoria son:

-Media o esperanza matemática
-Varianza
-Desviación estándar
-Coeficiente de variación

MEDIA DE UNA VARIABLE ALEATORIA

Definición.- Sea X una variable aleatoria discreta donde $x_1, x_2, \ldots, x_n$ son los valores que puede tomar dicha variable con una probabilidad de $p_1, p_2, \ldots, p_n$ respectivamente, y con la siguiente distribución de probabilidades:

X	x_1	$x_2 \ldots\ldots\ldots x_n$
f(X)	$p(x_1)$	$p(x_2) \ldots p(x_n)$

La **media o esperanza matemática** de la variable aleatoria X se denota por E(x) y se define como:

$$E(X) = X_1\, p(X_1) + X_2\, p(X_2) + \ldots + X_n\, p(X_n) = \sum_{i=1}^{n} X_i\, p(X_i)$$

Para variables aleatorias discretas para cada valor $x_1, x_2, \ldots, x_n$ que toma dicha variable con probabilidades $p_1, p_2, \ldots, p_n$ respectivamente, se pueden sustituir éstas por las n frecuencias relativas de ocurrencia de cada uno de los valores. De ésta manera se obtiene la media de la distribución por lo tanto se puede identificar la esperanza matemática como la media de la distribución que se denota por $\mu(x)$, es decir, $E(x)=\mu(x)$.

Ejemplo:

1) Un experimento consiste en lanzar tres monedas al aire y observar las caras que quedan hacia arriba. Sea la variable aleatoria X:"no. de águilas observadas". Considerar el juego en el que por cada águila que caiga se gana un peso. Si se participa una vez en el juego ¿cuánto se espera ganar?

El espacio de eventos asociado es: E={sss,ssa,sas,ass,aas,asa,saa,aaa}
Los valores que puede tomar la variable aleatoria son X={0,1,2,3} por lo tanto:

f(0) = 1/8 se espera ganar 0 pesos 1/8 de las veces
f(1) = 3/8 se espera ganar 1 peso 3/8 de las veces
f(2) = 3/8 se espera ganar 2 pesos 3/8 de las veces
f(3) = 1/8 se espera ganar 3 pesos 1/8 de las veces

promediando, lo que se esperaría ganar por jugar una vez sería:

$$\mu(x)=E(x)= 0*(1/8)+1*(3/8)+2*(3/8)+3*(1/8)=12/8= 1.50$$

2) Determinar la media del ejemplo de observar el número de águilas que caen al lanzar 2 monedas al aire de la siguiente manera:

X	0	1	2
f(X)	1/4	2/4	1/4
X*f(X)	0	2/4	2/4

$$\mu(x)=\sum_{i=1}^{n} f(X_i)*X_i = 0*(\tfrac{1}{4})+1*(2/4)+2*(\tfrac{1}{4})=1$$

3) Un experimento consiste en lanzar un dado ordinario, si se considera la variable aleatoria X como el resultado que se obtiene. Hallar su esperanza matemática.

El espacio de eventos E={1,2,3,4,5,6}

Como la variable aleatoria es el resultado obtenido, se tiene la función de densidad:

X	1	2	3	4	5	6
f(X)	1/6	1/6	1/6	1/6	1/6	1/6

La esperanza matemática es:

$$\mu(X)=E(X)=1*(1/6) + 2*(1/6) + 3*(1/6) + 4*(1/6) + 5*(1/6) + 6*(1/6)= 21/6 = 3.5$$

VARIANZA

El valor esperado es el promedio de todos los valores que puede tomar la variable aleatoria, es decir, es una medida de tendencia central de la distribución de probabilidades de dichos valores; de la misma manera también se necesita contar con una medida de dispersión que indique que tan dispersos están los distintos valores de la variable aleatoria del promedio, para ésto se utilizan las medidas de dispersión o variabilidad como la varianza y la desviación estándar.

<u>Definición</u>.-Sea X una variable aleatoria discreta que puede asumir n valores escalares con distribución de probabilidades f(X) y con media $E(x) = \mu$, la **varianza** de X se define como:

$$Var(x) = \sigma^2 = \sum_{i=1}^{n} [x_i - \mu]^2 * f(xi)$$

donde $f(x_i)$ es la respectiva probabilidad de cada valor x_i de la variable aleatoria

La varianza σ^2 es una medida de los alejamientos de cada valor de la variable aleatoria respecto a la media de la distribución de probabilidades, es decir, es una medida de la dispersión de los valores de la variable aleatoria con respecto a μ.

Ejemplo:

Calcular la varianza del experimento que consiste en lanzar dos monedas al aire y considerar la variable aleatoria X como el número de soles observados.

La función de distribución de probabilidades es:

X	0	1	2
f(X)	1/4	2/4	1/4

La esperanza matemática E(X) es:

$X_1 * f(X_1) = 0*(1/4) = 0$ $\quad$ $X_2 * f(X_2) = 1*(2/4) = 2/4$ $\quad$ $X_3 * f(X_3) = 2*(1/4) = 2/4$

$$E(x) = \mu = \sum_{i=1}^{3} X_i f(X_i) = 0 + 2/4 + 2/4 = 1$$

A cada valor de la variable aleatoria se le resta E(x):

$X_1 - E(X) = 0-1 = -1$ $\quad$ $X_2 - E(X) = 1-1 = 0$ $\quad$ $X_3 - E(X) = 2-1 = 1$

Se eleva al cuadrado cada alejamiento X-E(X) con lo cual se obtiene $(X - E(X))^2$:

$(X_1 - E(X))^2 = (0-1)^2 = (-1)^2 = 1$

$(X_2 - E(X))^2 = (1-1)^2 = 0$

$(X_3 - E(X))^2 = (2-1)^2 = 1$

Se multiplica cada alejamiento elevado al cuadrado por su correspondiente $f(x)$:

$(X_1 - E(X))^2 * f(X_1) = 1 * ¼ = ¼$

$(X_2 - E(X))^2 * f(X_2) = 0 * 2/4 = 0$

$(X_3 - E(X))^2 * f(X_3) = 1 * ¼ = ¼$

Se suman todos los valores obtenidos y se obtiene σ^2

$$\sigma^2 = \sum_{i=1}^{3} (X_i - \mu)^2 f(x_i) = ¼+0+¼ = 2/4 = ½ = 0.5$$

Se puede hacer uso de una tabla para realizar los cálculos correspondientes:

X	f(X)	X*f(X)	F(X)	E(X)	X-E(X)	(X-E(X))²	(X-E(X))²*f(X)
0	1/4	0	1/4	1	-1	1	1/4
1	2/4	2/4	3/4	1	0	0	0
2	1/4	2/4	1	1	1	1	1/4

$\sigma^2 = ¼+0+¼ = 2/4 = ½ = 0.5$

Otra forma equivalente para calcular la varianza es:

$\sigma^2 = E(X^2) - [E(x)]^2$

que en palabras significa que la varianza es igual al valor medio cuadrático menos la media elevada al cuadrado.

X	f(X)	X*f(X)	F(X)	X²	(X²*f(X))
0	1/4	0	1/4	0	0
1	2/4	2/4	3/4	1	2/4
2	1/4	2/4	4/4	4	4/4
		E(x)=1			6/4

$\sigma^2 = (6/4) - 1^2 = 2/4 = ½ = 0.5$

DESVIACION ESTANDAR

La **desviación estándar** σ de una variable aleatoria se define como la raíz cuadrada positiva de la varianza y está dada por la siguiente expresión:

$$\sigma = \left(\sum_{i=1}^{n} [X_i - \mu]^2 * f(X_i)\right)^{1/2} = \sqrt{E(X^2)-[E(X)]^2}$$

Ejemplo:

1) Calcular la desviación estándar σ del experimento anterior de lanzar dos monedas al aire, considerando la variable aleatoria X como el número de soles observados.

Como la varianza es igual a $\sigma^2 = 0.5$ entonces $\sigma = \sqrt{0.5} = 0.7071$

COEFICIENTE DE VARIACION

El coeficiente de variación es una medida relativa de dispersión que se define como la relación de la desviación estándar con respecto a la media, en notación matemática se tiene:

$$V(x) = \frac{\sigma}{\mu(x)}$$

El coeficiente de variación es un número sin unidades y es común especificar su valor en porcentaje.

Ejemplo:

1.- Una variable aleatoria X tiene la siguiente distribución de probabilidades:

x	-2	3	5
p(x)	0.3	0.2	0.5

Calcular la esperanza matemática, la varianza, la desviación estándar y el coeficiente de variación de la variable aleatoria

La media de la distribución es:

$\mu(x) = (-2)(0.3) + (3)(0.2) + (5)(0.5) = 2.5$

La varianza es:

$\sigma^2 = (-2 - 2.5)^2 (0.3) + (3 - 2.5)^2 (0.2) + (5 - 2.5)^2 (0.5) = 2.7+0.05+3.12 = 5.87$

La desviación estándar es: $\sigma = \sqrt{5.87} = 2.42$

El coeficiente de variación es : $V(x) = 2.42 / 2.5 = .968$

Ejercicios:

1).- Un niño tiene en una caja 4 pelotas negras y 2 verdes, selecciona al azar 3 de ellas en sucesión y sin reemplazo. Se define la variable aleatoria X como el número de pelotas verdes que se extraen, encontrar la varianza y la desviación estándar de la variable aleatoria.

2).- Una colección de discos está formada por 5 discos de jazz, 2 de música clásica y 3 de polka. Si se seleccionan 4 discos al azar y se define la variable aleatoria X como el número de discos de jazz que se pueden seleccionar. Encontrar la distribución de probabilidades de X, la media, la varianza y la desviación estándar.

3).- La distribución de probabilidad de una variable aleatoria X del número de defectos por cada 100 metros de tela sintética en rollos continuos de ancho uniforme es:

x	0	1	2	3	4
p(x)	0.41	0.37	0.16	0.05	0.01

Calcular la media del número de defectos por cada 100 metros, la varianza y la desviación estándar.

4).- En un juego de azar una persona podrá ganar N$ 500 si cuando lance 3 monedas ocurren 3 soles o 3 águilas o perderá n$ 300 si ocurren uno o 2 soles. ¿Cuánto espera ganar esta persona?

5).- Por invertir en unas acciones en particular, una persona puede obtener ganancias de n$4000 con una probabilidad de 0.3 o una pérdida de n$1000 con una probabilidad de 0.7. ¿Cuál es la ganancia que espera esta persona?

6).-Un tazón contiene 5 fichas que no pueden distinguirse una de otra, 3 de las fichas están marcadas con $2 y las restantes 2 con $4. Un jugador saca del tazón 2 fichas al azar sin reemplazo, y se le paga una cantidad igual a la suma de los valores indicados en las dos fichas. Si el costo por jugar es de $5.60 ¿es justo el juego?

EVALUACION PROPUESTA

1.- Un niño tiene 5 canicas en una bolsa, 4 de las cuales son de un color y una de otro. Si selecciona dos canicas al azar de las cinco, y considera la variable aleatoria X como el número de veces en que la canica de distinto color fue seleccionada. Calcular la función de distribución de probabilidad de X.

2.- Para ahorrarse tiempo y dinero, dos compradores acostumbran inspeccionar sólo una parte de los embarques recibidos, con ello juzgan la calidad del embarque en base a la calidad observada en la porción inspeccionada. Suponga que un comprador recibe un embarque de 4 fotocopiadoras, dos de las cuales él no sabe que están defectuosas, si selecciona dos de ellas al azar para inspeccionar y la variable aleatoria X denota el número de fotocopiadoras defectuosas, encontrar la función de distribución de probabilidad de X y la función acumulada de distribución.

3.- En la mayoría de las dependencias gubernamentales para otorgar contratos se hace con base un concurso. Suponga que la experiencia que se tiene en cierta dependencia gubernamental muestra que determinado contratista gana en promedio 3 de cada 5 concursos en los que compite. Denotar por X el número de concursos en los que este contratista compite antes de ganar un contrato y suponga que los resultados de cada 10 concursos son independientes entre sí.
a) Encontrar p(1), p(2) y p(3)
b) Exhibir una fórmula para p(x), x=1,2,3,...

4.- Un embarque de 5 automóviles extranjeros incluye 2 que tienen unas ligeras manchas de pintura. Si una agencia recibe 3 de estos vehículos aleatoriamente, y la variable aleatoria X representa el número de automóviles con manchas de pintura comprados por la agencia. Encontrar la función de distribución de probabilidades de X y su función acumulada de distribución.

5.- En una familia que tiene 4 hijos se desea saber cuántos de ellos son varones. Determinar la distribución de probabilidades , su esperanza matemática y su varianza.

6.-En un lote de 5 focos idénticos se tienen tres defectuosos (no encienden). Si se toman tres focos al azar, sólo se consideran los focos defectuosos. Determinar la distribución de probabilidades, su esperanza matemática y su varianza?

7.- Se sacan 3 bolas al azar de una caja que contiene 6 bolas azules y 4 rojas y no se devuelven. Sea x la variable aleatoria que designa el número de bolas azules entre las que se han extraido.
a) escribir en forma tabular la distribución de probabilidades de la variable aleatoria x
b) calcular la media, la varianza y la desviación estándar de la distribución

8.- En un juego de azar una persona podrá ganar N$ 500 si cuando lance 3 monedas ocurren 3 soles o 3 águilas o perderá n$ 300 si ocurren uno o 2 soles. ¿Cuánto espera ganar esta persona?

9.- Por invertir en unas acciones en particular, una persona puede obtener ganancias de n$4000 con una probabilidad de 0.3 o una pérdida de n$1000 con una probabilidad de 0.7. ¿Cuál es la ganancia que espera esta persona?

10.-Un tazón contiene 5 fichas que no pueden distinguirse una de otra, 3 de las fichas están marcadas con \$2 y las restantes 2 con \$4. Un jugador saca del tazón 2 fichas al azar sin reemplazo, y se le paga una cantidad igual a la suma de los valores indicados en las dos fichas. Si el costo por jugar es de \$5.60 ¿es justo el juego?

CONCLUSIONES

La evaluación correspondiente a este tema del trabajo se aplicó a lõs 40 alumnos del grupo 505 de estadística del CCH-O, los resultados que se obtuvieron se muestran en una tabla de distribución de frecuencias en 4 intervalos, se calcularon los promedios correspondientes a estos datos y se construyen diagramas para observarlos.

1	3	4	6	7	7	8	9
2	3	4	6	7	8	8	9
2	4	5	6	7	8	8	10
2	4	5	7	7	8	9	10
3	4	6	7	7	8	9	10

Intervalos	Clases	Frecuencia	Marcas de clase	Marcas * frecuencia
0 - 2	0 - 2.5	4	1	4* 1 = 4
3 - 5	2.5 - 5.5	10	4	10* 4 = 40
6 - 8	5.5 - 8.5	19	7	19* 7 =133
9 - 11	8.5 - 11.5	7	10	7*10 = 70
	total	40		247

$$\bar{X} = \frac{247}{40} = 6$$

$$X_o = 5.5 + \left(\frac{9}{9+12}\right) 3 = 6.7$$

$$\tilde{X} = 5.5 + \left(\frac{20 - 14}{19}\right) 3 = 5.8$$

Diagrama: Polígono de frecuencias

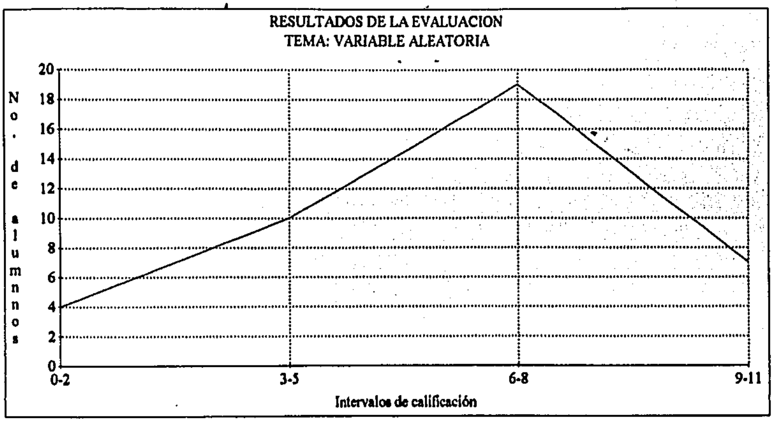


Diagrama circular:

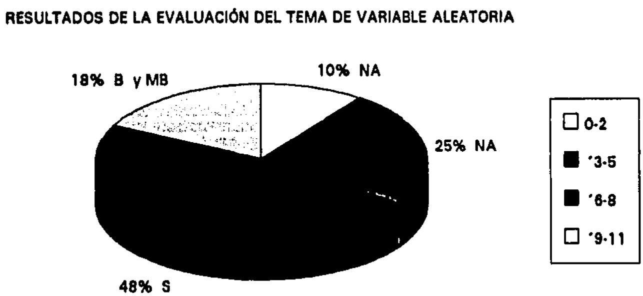
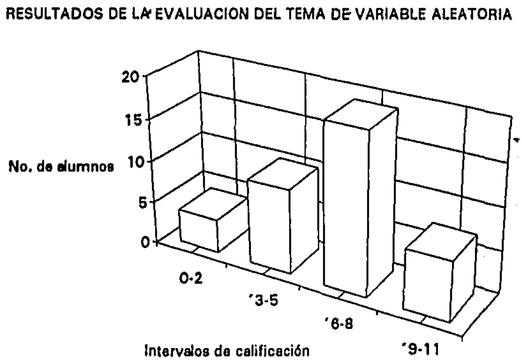


Diagrama de barras:



El diagrama circular de los resultados obtenidos de la evaluación escrita correspondiente a los conceptos del tema de variable aleatoria, muestran que aproximadamente el 65 % de los alumnos del grupo 505 de estadística los asimilaron correctamente, dichos resultados representan un avance positivo para el presente trabajo.

125

TEMARIO

FUNCIONES DE DISTRIBUCION DE PROBABILIDADES

Las funciones de distribución de probabilidades se encuentran cuando se estudia experimentalmente variables aleatorias, y se utilizan cuando se eligen modelos apropiados de probabilidades para dichos experimentos.

Utilizando las propiedades de la probabilidad se construyen las llamadas funciones de distribución que permiten determinar con qué probabilidad pueden ocurrir los diferentes valores con los que se midió una propiedad del fenómeno aleatorio. La estadística èstá encaminada a determinar cuál es la función de distribución más adecuada para representar un fenómeno aleatorio particular y comparar las distribuciones para algunas variantes de éste.

La función que da la probabilidad para cada valor de la variable aleatoria es la función de densidad para variables aleatorias discretas, en este caso se encuentran las funciones de Bernoulli, Binomial, de Poisson y Normal.

La selección adecuada de una función de probabilidad permite que el modelo resulte útil y lo más sencillo posible para representar un fenómeno aleatorio específico para encontrar las soluciones a los problemas reales que se encuentran a través del modelo para que tengan correspondencia con la realidad.

La inferencia estadística se basa en dos principios generales que son:

1) Se puede llegar a definir la medida de probabilidad con base a las observaciones repetidas del fenómeno aleatorio en estudio; como el resultado de estas repeticiones tiene carácter aleatorio y las acciones se basan en ella, es entonces que éstas ultimas resultan ser variables aleatorias. De esta manera se definen las **funciones de decisión estadística**..

2) En la búsqueda de la función de decisión óptima es inevitable que se tenga que definir una función de pérdida.

Los matemáticos como Gauss y Laplace emplearon la función de pérdida porque juzgaban que en los casos de estimación tanto la sobreestimación (error positivo) como la subestimación (error negativo) tenían igual importancia.

Por ejemplo una situación en la que subestimar fue más grave que sobreestimar fue la situación de emergencia a raíz de los temblores en Septiembre de 1985.

DISTRIBUCION BINOMIAL O DE BERNOULLI

Se llama distribución de Bernoulli en recuerdo al matemático fránces James Bernoulli (1654 - 1705) que la descubrió a finales del siglo XVII, es una función de distribución de probabilidad con relación a sucesiones de ensayos independientes de experimentos en los que se considera una variable aleatoria que sólo tienen dos resultados posibles **éxito o fracaso**. Muchos problemas en estadística se adecúan a este modelo.

Definición.- Un **experimento binomial** es tal que tiene las siguientes propiedades:

1) el experimento consiste en n ensayos idénticos.

2) cada ensayo produce uno de los dos posibles resultados únicamente, al número de aciertos se les denota por **x** y al número de fallas por n-x.

3) a la probabilidad de éxito se le denota por **p**, y es constante para los n ensayos que se efectúen. La probabilidad de falla es igual a **q** = 1-p.

4) los n ensayos son independientes.

5) x representa por lo tanto el número de aciertos en los n ensayos

Una variable aleatoria asociada a un experimento en las condiciones anteriores se llama **variable aleatoria binomial**.

Ejemplo:

En una población de un 1,000,000 de consumidores potenciales de un artículo producido por cierta empresa, se lleva a cabo un estudio de mercado. Para esto se selecciona una muestra de 1000 compradores, a cada uno de ellos se le pregunta si prefiere el producto producido por esta empresa o no. ¿Es un experimento binomial?

Método:

1.- El muestreo consiste de 1000 ensayos idénticos. Cada ensayo representa la selección de una persona del millón de compradores potenciales.

2.- Cada ensayo tiene dos resultados posibles: la persona prefiere el producto (acierto) o no (falla).

3.- La probabilidad de un acierto es igual a la proporción de compradores potenciales que prefieren el producto. Por ejemplo si del millón de compradores potenciales 300,000 prefieren el producto entonces la probabilidad de seleccionar una persona que lo prefiera es de p=0.3

4.- La probabilidad de un acierto en cualquier ensayo no se afecta por el resultado de los demás.

5.- Nos interesa el número x de personas en la muestra de 1000 que prefieren el producto.

Por lo tanto este es un experimento binomial.

Se efectuan n ensayos independientes o pruebas de un experimento en el que sólo existen dos resultados posibles éxito que se llama p y fracaso que se denota como q para cada ensayo, además se tiene que p+q=1, es decir, q=1-p. Considerando que la variable aleatoria toma sólo 2 valores 1 ó 0 para evaluar los dos resultados posibles -éxito o fracaso respectivamente-, y que a estos valores se les asocian las probabilidades p y q=1-p.

Distribución de la probabilidad binomial

Cuando se determina que se trata de un experimento binomial, el siguiente paso es determinar la distribución de probabilidades de la variable aleatoria x que representa el número de aciertos observados en n ensayos.

La distribución de probabilidad binomial de obtener x logros nos indica que tendremos n-x fracasos entre los n ensayos. Los éxitos están dados por la siguiente fórmula para el caso en el que todos los éxitos ocurren primero y todos los fracasos después:

$$P(X) = (n! \,/\, x!\,(n-x)!)\ p^{x}\ q^{n-x}$$

donde p^{x} = probabilidad de éxito en cada una de las primeras x pruebas

q^{n-x} = probabilidad de fracaso en cada uno de los restantes n-x pruebas

Pero como los éxitos y fracasos no ocurren en un orden determinado entonces pueden ocurrir en:

$\dfrac{n!}{x!\,(n-x)!}$ órdenes distintos. Cada uno con probabilidad $p^{x}\ q^{n-x}$

Cada ensayo tiene un espacio muestral E={p,q} y cada resultado posible de los n ensayos es un punto en el espacio del tipo (ppppp...qqq...pqp), es decir, algunas coordenadas son p y el resto q. El suceso p es el subconjunto de puntos de E que tiene exactamente x letras p y n-x letras q.

Ejemplo:

1.-En un estudio realizado sobre la influencia relativa de esposos y esposas en las políticas familiares de consumo, se encontró que el marido ejerce una influencia decisiva en la compra de un automóvil en el 70% de las familias. Si 4 familias han decidido comprar un automóvil, ¿cuál es la probabilidad de que en 2 de las 4 familias el marido ejerza su influencia decisiva en la selección del automóvil?

Suponiendo que las decisiones de compra de las familias son independientes y que p permanece constante de una familia a otra tenemos que:
$n = 4$
$x = 2$
por lo tanto $n-x = 4-2 = 2$
$p = 0.7$ ya que el 70% así lo determinó
$q = 1-p = 1 - 0.7 = 0.3$

$P(x) = (4!/2!\ (4-2)!)\ (0.7)^2\ (0.3)^2 = (4*3*2!\ /\ 2!\ 2!)\ (.49)(.09) \doteq (12\ /\ 2)\ (0.0441) = 6(0.0441) = 0.2646$

Ejercicios:

1.- Para determinar la efectividad de una nueva vacuna para prevenir el resfriado común se seleccionaron a 10 personas para inyectarles la vacuna y tenerlos en observación durante un año. De las 10 personas, 8 pasaron el invierno sin enfermarse de resfriado. Si se supone que cuando no se usa la vacuna la probabilidad de pasar el invierno sin enfermarse es de .5 y que además es independiente del estado de salud de cualquier persona. ¿Cuál es la probabilidad de observar 8 personas que no se enfermaron durante el invierno dado que la vacuna no tiene efecto alguno?

2.- El 20% de las ventas de automóviles nuevos en México corresponde a automóviles importados. Suponer que se seleccionan al azar 4 personas que han comprado un automóvil nuevo. Calcular la probabilidad de que las 4 personas hayan comprado un automóvil importado.

3.- Los registros de mantenimiento de una empresa de servicio revelan que solamente 1 de cada 100 máquinas de escribir de cierta marca requiere de una reparación mayor durante el primer año de uso. El gerente de una oficina ordenó la compra de 10 máquinas de esta marca.

a)Encontrar la probabilidad de que ninguna de las máquinas requiera una reparación mayor durante el primer año de uso.
b)Encontrar la probabilidad de que dos máquinas requieran una reparación mayor durante el primer año de uso.

PROPIEDADES DE LA DISTRIBUCION BINOMIAL

Es conveniente describir la distribución de probabilidad binomial mediante su media y su desviación estándar, ya que esto permite identificar valores de x (éxito). Si en un experimento binomial cuyos dos resultados x_1, x_2 ocurren con una probabilidad p_1 y p_2 respectivamente, y si se identifica el éxito como 1 y el fracaso como 0 entonces se tiene que la media de esta distribución de probabilidad es:

$\mu = p_1 * x_1 + p_2 * x_2 = p*1 + q*0 = p$

Por otra parte la varianza de dicha distribución de probabilidad es:

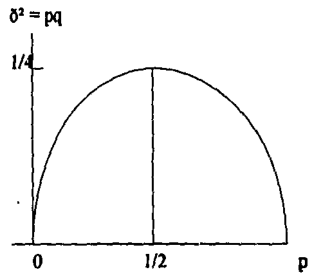
$\sigma^2 = p_1*(x_1-\mu)^2 + p_2*(x_2-\mu)^2 = p*(1-p)^2 + q*(0-p)^2 = pq^2 + q*(-p)^2 = pq^2 + p^2q = q*(q+p) = pq*1 = pq$

La distribución de probabilidad binomial es la distribución de una suma de variables aleatorias independientes $x_1, x_2, \ldots, x_n$, cada una de las cuales tiene media p y varianza pq. Por lo tanto utilizando las reglas para la media de una suma de variables independientes y para la varianza de las mismas, se tiene que:

$\mu_{x_1+x_2+x_3} = \mu_{x_1} + \mu_{x_2} + \ldots + \mu_{x_n} = p + p + \ldots + p = np$

$\sigma^2_{x_1+x_2+\ldots+x_n} = \sigma^2_{x_1} + \sigma^2_{x_2} + \ldots + \sigma^2_{x_n} = pq + pq + \ldots + pq = npq$

La varianza de una distribución es una medida de la incertidumbre de cual será el valor de un resultado aleatorio dibujado desde la curva de la distribución.



varianza de una variable Bernoulli

Las fórmulas para calcular estos parámetros son:

Media $\mu = np$
Varianza $\sigma^2 = npq$

Desviación típica $\sigma = \sqrt{npq}$

Ejemplo:

1.-Se tira una moneda 5 veces. ¿Cualés son las probabilidades de obtener 2 soles? Calcular la media y la varianza de la distribución. Dibujar un diagrama de árbol.

```
            s   -sssss                         s   -assss
          s/                                 s/
           \a   -ssssa                        \a   -asssa
        s/                                 s/
         \  s   -sssas                      \  s   -assas
          a/                                 a/
           \a   -sssaa                        \a   -assaa
     s/                                 s/
      \     s   -ssass                   \     s   -asass
          s/                                 s/
           \a   -ssasa                        \a   -asasa
        a/                                 a/
         \  s   -ssaas                      \  s   -asaas
          a/                                 a/
 /         \ a  -ssaaa             /          \a   -asaaa
 s                                 a
 \          s   -sasss             \           s   -aasss
          s/                                 s/
           \ a  -sassa                        \a   -aassa
        s/                                 s/
         \  s   -sasas                      \  s   -aasas
          a/                                 a/
           \ a  -sasaa                        \a   -aasaa
     a/                                 a/
      \     s   -saass                   \     s   -aaass
          s/                                 s/
           \ a  -saasa                        \a   -aaasa
        a/                                 a/
         \  s   -saaas                      \  s   -aaaas
          a/                                 a/
           \ a  -saaaa                        \a   -aaaaa
```

donde $n = 5$, $x = 2$, $p = 10/32 = 0.3125$, $q = 1-0.3125 = 0.6875$

$p(2) = (5! / 2! (5-2)!)\ (0.3125)^2\ (0.6875)^3 = (2/2)\ (0.098)(0.325)=0.032$

$\mu = 5*(0.3125) = 1.563$

$\sigma^2 = 5*(0.3125)*(0.6875) = 1.075$

$\sigma = \sqrt{1.075} = 1.037$

Ejercicios:

1.-Se extraen ocho cartas sucesivamente, con reemplazamiento de una baraja. Hallar la probabilidad de que 6 de ellas sean corazones.

2.-Se lanzan 8 monedas simultáneamente. Demostrar que la probabilidad de obtener por lo menos 6 soles es de 37/256.

3.-El 20 por 100 de los tornillos fabricados por una máquina son defectuosos. Si se eligen 4 tornillos al azar determinar la probabilidad de que 2 de ellos sean defectuosos.

4.-De entre 200 familias con 4 hijos cada una, ¿Cuantás se esperaría que tuvieran 2 hombres y 2 mujeres?. Suponer que $P(h)=\frac{1}{2}$.

5.-Un hombre que dispara a un blanco estima que tiene una probabilidad de un ¼ de dar en el blanco. Si dispara 5 veces, hallar la probabillidad de hacer 3 blancos.

Hallar la media y la varianza de todos los ejercicios.

DISTRIBUCION DE PROBABILIDAD DE POISSON

La distribución de Poisson se llama de esta forma en recuerdo del matemático francés D.Poisson. La función de probabilidad de Poisson es más fácil de evaluar y de manipular matemáticamente que la función binomial pero se basa en ella. Los parámetros de esta distribución se obtienen a partir de los de la distribución binomial, es decir, si en un experimento binomial el tamaño de muestra es grande y la probabilidad de éxito p es muy pequeña es frecuente usar las probabilidades de la distribución de Poisson como una aproximación de las probabilidades binomiales.

La fórmula de la distribución de probabilidad de Poisson está dada por:

$$P(x) = \frac{\mu^{x} e^{-\mu}}{x!}$$

donde $e = 2.71828....$
μ = media de la distribución
$x = 0,1,2,3,...,n$

La distribución de Poisson proporciona buenas aproximaciones cuando n número de ensayos es grande, digamos 50, y $\mu = np \leq 5$ podemos usar la función de Poisson con $\mu = np$ como aproximación de la binomial. También constituye un buen modelo para experimentos donde x representa el número de veces que ha ocurrido un evento raro en un lapso de tiempo dado, además de ser útil como aproximación de las probabilidades binomiales. Por ejemplo:

1.-El número de llamadas teléfonicas recibidas en un conmutador durante un período corto de tiempo.

2.- El número de reclamaciones contra una compañia de seguros durante una semana determinada.

3.- El número de llegadas a una estación de ferrócarril durante un minuto.

4.- El número de fallas de una máquina durante un día determinado.

5.- El número de ventas hechas por un agente de bienes raíces en un día determinado.

En cada ejemplo x representa el número de eventos raros que ocurren durante un período de tiempo en el cual se espera que un promedio μ de ellos ocurra. Las unicas suposiciones que se requieren para usar la distribución de Poisson son que los eventos ocurran en forma aleatoria e independiente unos de otros.

Ejemplo:

Una planta procesadora y enlatadora de alimentos tiene 20 máquinas enlatadoras automáticas en operación constante. Si la probabilidad de que una máquina se descomponga durante un día determinado es de 0.05, encontrar la probabilidad de que en un día determinado fallen dos máquinas. Utilizar la distribución binomial para calcular la probabilidad exacta y posteriormente calcular la aproximación de Poisson correspondiente.

El problema es un experimento binomial ya que únicamente se trata de averiguar si las máquinas fallan o no. En este caso n= 20 porque hay 20 máquinas y p= 0.05 es la probabilidad de falla.

Por lo tanto el número esperado de fallas en un día determinado es
$\mu = np = (20)(.05) = 1.0$

Por lo tanto se tiene que: $P(2) = \sum_{x=0}^{2} p(x) - \sum_{x=0}^{1} p(x) = .925 - .736 = .189$

Usando la distribución de Poisson se tiene que:

$P(2) = 1^2\ e^{-1} / 2! = .367879 / 2 = .184$

Se puede observar que la aproximación de Poisson .184 esta bastante cerca del valor de la probabilidad binomial .189.

2.- Un fabricante de podadoras de césped adquiere un proveedor de motores de dos tiempos de 1 caballo de fuerza en lotes de 1000. La experiencia indica que la probabilidad de que un motor suministrado por este proveedor este defectuoso es de 0.001, en un envio de 1000 motores ¿Cuál es la probabilidad de que ninguno salga defectuoso?. Calcular la probabilidad de tener exactamente uno, dos, tres y cuatro motores defectuosos.

En este experimento $n= 1000$ y $p=0.001$

El número esperado de defectuosos es $\mu = np = (1000)(0.001) = 1$

La probabilidad de x motores defectuosos se puede aproximar por:

$$P(x) = (\mu^{x} e^{-\mu}) / x! = (1^{x} e^{-1}) / x! = e^{-1} / x!$$

por lo tanto

$p(0) = e^{-1} / 0! = .368 / .1 = .368$
$p(1) = e^{-1} / 1! = .368 / .1 = .368$
$p(2) = e^{-1} / 2! = .368 / 2 = .184$
$p(3) = e^{-1} / 3! = .368 / 6 = .061$
$p(4) = e^{-1} / 4! = .368 / 24 = .015$

Ejercicios:

1.- Una máquina produce componentes que son defectuosos en un 10 por 100. Se elige al azar una muestra de 20 componentes. Calcular las probabilidades de que 2 componentes sean defectuosos en la muestra utilizando la distribución binomial y la distribución de Poisson y comparar los resultados.

2.- Se sabe que un promedio de 3 por 100 de las lámparas eléctricas hechas por cierta fábrica se funden cuando se prueban por primera vez. Si se prueba una muestra de 30 lámparas ¿Cualés son las probabilidades de que 2 de ellas se fundan?

DISTRIBUCION NORMAL

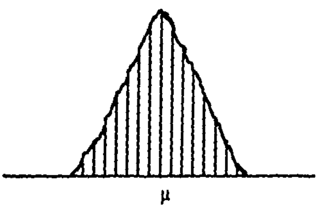
Existen varias distribuciones de probabilidades que describen la forma de las probabilidades en que ocurren los eventos contenidos en el espacio muestral que se este considerando de acuerdo al experimento que se este afectuando. Entre ellas para estudios experimentales la más importante es la distribución normal, ya que entre otras razones es simétrica respecto a la media de la muestra y la diferencia entre otras medias que se calculen es menor entre mayor sea el tamaño de la muestra que se seleccione.

Las variables aleatorias continuas son aquellas que pueden tomar cualquier valor en un intervalo de la recta; el modelo probabilístico para la distribución de frecuencias de una variable aleatoria se representa por una curva contínua que corresponde a la función de densidad de probabilidad normal que tiene una forma acampanada. Como la ecuación de la función de densidad se construye de manera que el área bajo la curva representa una probabilidad entonces el área total es igual a 1.

La distribución normal es a la que se le da más importancia en la teoría estadística, a pesar de que en la práctica es díficil que el comportamiento de una variable aleatoria sea como una distribución normal, sin embargo las distribuciones de muchas variables aleatorias en experimentos ordinarios se pueden aproximar satisfactoriamente por la curva normal. Un resultado conocido como el Teorema Central del Límite ayuda a explicar la causa de que estas distribuciones aparezcan en trabajos de Estadística tan frecuentemente, es decir, demuestra que muchas distribuciones bajo ciertas condiciones son aproximadamente iguales a la normal.

Es muy frecuente que la curva de una gráfica represente una fórmula matemática particular, en este caso que especifique una distribución de probabilidad teórica llamada distribución normal o distribución Gaussiana en honor al matemático alemán K.F. Gauss (1777 - 1855). La curva de una distribución normal es simétrica y unimodal.

Existen distintas curvas que representan a la distribución normal:



Función de densidad de probabilidad normal

La función de distribución de probabilidad normal para una variable aleatoria **X** cuya media se representa por μ y su varianza por σ^2 esta dada por:

$$N(x{:}\mu,\sigma^2) = \frac{e^{[-\frac{1}{2}(x-\mu/\sigma)^2]}}{\sigma\sqrt{2\pi}} \quad ; -\infty < x < \infty$$

donde $e = 2.7183$

$\pi = 3.1416$

μ = media

σ = desviación estándar

Si se sustituye $z = \dfrac{x - \mu}{\sigma}$ se obtiene:

$$N(x;\mu,\sigma^2) = \frac{e^{(-\frac{1}{2} z^2)}}{\sigma \sqrt{2\pi}} \qquad -\infty < z < \infty$$

que se llama **Distribución Normal Típificada**.

Si z representa el eje horizontal, y el eje vertical entonces la curva de la distribución normal es la gráfica de la fórmula:

$$y = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Teorema del límite central (TLC)

Si se extraen todas las muestras de una población cuya media es μ y cuya varianza es δ^2. La distribución muestral de la variable aleatoria x calculada a partir de todas las muestras aleatorias de tamaño n de esta población estará distribuida de forma aproximadamente normal con:

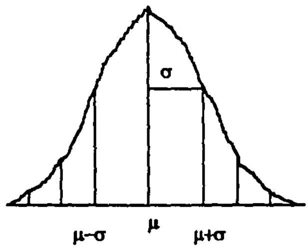
$$\text{media} = \mu_x = \mu \qquad \text{varianza} = \sigma^2_x = \sigma^2/n$$

En otras palabras el teorema central del límite establece que las medias de las mediciones aleatorias tienden a poseer una distribución acampanada en un muestreo repetitivo. Es decir, si se extraen muestras de tamaño **n** de una población con media finita μ y desviación estándar σ entonces si n es grande la media muestral tiene una distribución aproximadamente normal con media μ y desviación estándar $\sigma/\sqrt{n}$.

La media y la desviación estándar de la distribución de las distintas medias muestrales de una población están relacionadas con la media y la desviación estándar de la población así como con el tamaño de la muestra n; por lo tanto a medida que n crece la dispersión de la distribución de las medias muestrales es menor que la dispersión de la población.

Una distribución normal está definida por su media y su desviación estándar, la media μ da la localización sobre el eje horizontal y la desviación estándar σ indica que tan "estrecha" es la curva entre menor es σ más "estrecha" es la distribución. Por lo tanto μ y σ son los parámetros principales de la distribución normal. La fórmula que representa las curvas de las distribuciones normales es:

$$y = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(z-\mu)^2/(2\sigma^2)}$$



Función de densidad de probabilidad normal

Si se resta la media μ a cada una de las medidas obtenidas de un experimento, la media de las medidas transformadas es igual a 0, de la misma manera si se divide a todas las medidas por un número positivo la desviación estándar de las medidas trasformadas es igual a la desviación estándar original σ dividida por dicho número, en otras palabras la desviación estándar de las medidas transformadas es igual a 1.

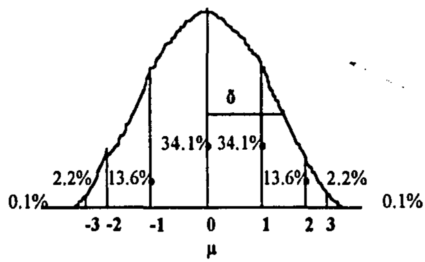
Si a la medida original x se le resta la media μ de las medidas originales, y a esta resta se le divide por la desviación estándar σ de las medidas originales, se obtienen las medidas transformadas con media igual a 0 y desviación estándar igual a 1 y la medida transformada es igual a:

$$z = \frac{x - \mu}{\sigma}$$

A esta representación de una variable aleatoria X se le llama **versión estandarizada** de X y algunas veces se denota por **z**.

Si una variable aleatoria X tiene una distribución normal con media μ y desviación estándar σ, la variable estandarizada también tiene una distribución normal estandarizada con media igual a 0 y desviación estándar igual a 1.

La distribución normal con media $\mu = 0$ y desviación estándar $\sigma = 1$ se llama **distribución normal estándar**.



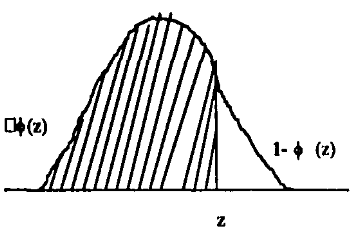
Gráfica de la distribución normal estándar

En esta gráfica el área total bajo la curva esta dividida por varias líneas verticales y los porcentajes indican el área comprendida entre ellas, así por ejemplo el 34.1% del área total se encuentra entre 0 y +1 y el mismo porcentaje entre -1 y 0. En terminos de probabilidad esto se expresa como que el valor del área comprendida entre 0 y 1 es igual a .341.

Areas bajo la curva de una distribución normal estándar

Si se denota a la función de distribución acumulada por $\phi(x)$ de una distribución normal estándar, se puede definir matemáticamente como el área a la izquierda de z bajo la curva normal estándar. Debido a que el área total es igual a 1, el área a la derecha de z es igual a $1 - \phi(z)$.

La curva normal estándar es simétrica con respecto al valor $z = 0$, que es la media de la distribución, por lo tanto el área a la derecha de z es igual al área a la izquierda de -z, es decir, $1 - \phi(z) = \phi(-z)$.



Función de distribución normal acumulada como área a la izquierda de un punto dado.

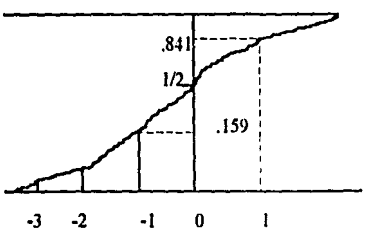
La función de distribución normal de probabilidades depende de los valores de μ y σ, de tal manera que al variar estos valores se genera una familia de distribuciones normales. Para obtener una función de distribución normal acumulada los valores se obtienen de tablas estadísticas que ya han sido calculadas en este caso para Z. Es decir, se ha formado una tabla considerando las áreas limitadas por un determinado número de desviaciones estándar de la media. Por ejemplo se sabe que aproximadamente el 68% del área total esta comprendida entre $\mu - \sigma$ y $\mu + \sigma$, el 95% entre $\mu - 2\sigma$ y $\mu + 2\sigma$ y casi toda el área bajo la curva esta entre $\mu - 3\sigma$ y $\mu + 3\sigma$.

Como la curva normal es simétrica con respecto a la media, la mitad del área bajo la curva está a la izquierda de la media y la otra mitad a la derecha de la misma. Por lo tanto en la tabla basta listar los valores del área entre la media y un determinado número z de desviaciones estándar a la derecha de μ.

En el siguiente ejemplo la tabla muestra los valores para una distribución normal estándar acumulada:

z	Probabilidad
-3	.001
-2	.023
-1	.159
0	.500
1	.841
2	.977
3	.999
4	1.000

La siguiente gráfica muestra los valores para $\phi(z)$ para valores positivos de z, y si queremos encontrar $\phi(-1.5) = 1 - \phi(1.5) = 1 - .9332 = .0668$



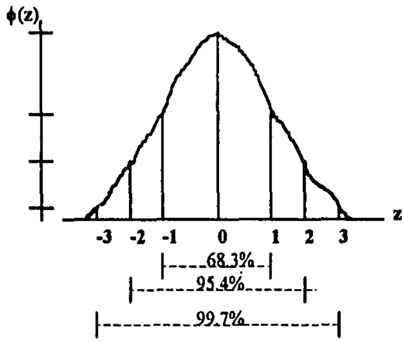
Función de distribución acumulada de una distribución normal estánda

Para toda distribución normal estándarizada el área bajo la curva normal entre $z = 0$ y un valor especificado de z - por ejemplo z_1- es la probabilidad $P(0 \leq z \leq z_1)$, por lo tanto la probabilidad entre la media y una desviación estándar (z=1.0) sobre la media es de .341, análogamente la probabilidad entre la media y una desviación estándar abajo de la media es de .341. Es decir, entre una desviación estándar y la media se localiza el 68.2 %, es decir, dos terceras partes del total de la población.

La probabilidad entre la media y dos desviaciones estándar es igual a 2(.4772) = .9544 que es el 95% del área bajo la curva.

Finalmente se puede decir que las probabilidades para variables aleatorias continuas se definen como áreas bajo la curva correspondiente a la función de densidad de probabilidad.

Distribución normal estándarizada



(a) 68.3 % de la Distribución queda entre z=-1 y z=1
(b) 95.4 % de la Distribución queda entre z=-2 y z=2
(c) 99.7 % de la Distribución queda entre z=-3 y z=3

Ejemplos:

1.- Sea y el número obtenido al lanzar un dado. Verificar que la media y la desviación estándar de y son $\mu = 3.5$ y $\sigma = 1.71$.

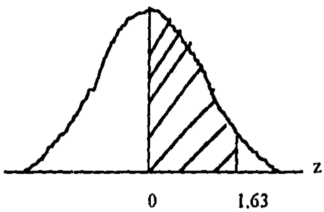
$\mu = 1(1/6)+2(1/6)+3(1/6)+4(1/6)+5(1/6)+6(1/6) = 21/6 = 3.5$

$\sigma^2=(1-3.5)^2(1/6)+(2-3.5)^2(1/6)+(3-3.5)^2(1/6)+(4-3.5)^2(1/6)+(5-3.5)^2(1/6)+(63.5)^2(1/6)$
$= (6.25)(1/6)+(2.25)(1/6)+(.25)(1/6)+(.25)(1/6)+(2.25)(1/6)+(6.25)(1/6)=$
$= 17.5/6 = 21.4$

$\sigma = \sqrt{21.4} = 1.71$

a) Repita el experimento lanzando un dado n= 5 veces y registrando las cinco observaciones. Calcular la media muestral para cada una de las 5 muestras.

2.- Calcular $P(0 \leq z \leq 1.63)$ que es el área entre la media $z = 0$ y un punto $z = 1.63$ desviaciones estándar a la derecha de la media.



Probabilidad para z = 1.63

La probabilidad del área que se quiere aparece sombreada, si se busca en las tabla correspondiente se encuentran los valores del área bajo la curva normal que están a la derecha de la media, el valor que se busca se encuentra en la intersección de la fila correspondiente a 1.6 con la columnna 0.03 y dicho valor es .4484.

3.- Sea x una variable aleatoria con distribución normal con media igual a 10 y desviación estándar igual a 2. Encontrar la probabilidad de que x este entre 11 y 13.6

Primero se encuentran los valores para $x_1 = 11$ y $x_2 = 13.6$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{11 - 10}{2} = .5$$

$$z_2 = \frac{13.6 - 10}{2} = 1.8$$

Por lo tanto la probabilidad deseada corresponde a $P(.5 \leq z \leq 1.8)$ que es el área entre z_1 y z_2. El área entre z=0 y z_1 es igual a .1915 de acuerdo a la tabla y el área entre z=0 y z_2 es igual a .4641 por lo tanto la probabilidad deseada es igual a la diferencia entre ambos valores, es decir,

$$P(.5 \leq z \leq 1.8) = .4641 - .1915 = .2726$$

4.- Algunos estudios realizados en México muestran que el rendimiento de gasolina de los autos compactos vendidos se distribuye normalmente con una media igual a 2.5 km por litro (kpl) y una desviación estándar igual a 4.5 kpl. ¿Qué porcentaje de autos compactos tienen un rendimiento de 3 kpl?

Primero es necesario encontrar el valor de z que corresponde a x=30, entonces tenemos que

$$z = \frac{x - \mu}{\sigma} = \frac{3 - 2.5}{4.5} = 0.1$$

El área a la derecha de la media correspondiente a z=0.1 es igual a .0398. Por lo tanto la proporción de autos compactos que tienen un rendimiento igual a 3 kpl es igual al área total de la derecha de la media menos el área calculada

$P(x=3) = .5 - P(0 \leq z \leq .1) = .5 - .0398 = 0.4602$

Por lo tanto el porcentaje de autos que exceden 3 kpl es 100(.4602) = 46.022%

Ejercicios:

1.- La duración de un determinado tipo de lavadoras automáticas tiene una distribución aproximadamente normal con una media igual a 3.1 años y una desviación estándar de 1.2 años. Si la lavadora esta garantizada por un año ¿Qué proporción del total de unidades vendidas tendrá que ser reemplazada?

2.- Un auditor encontró que los errores en las cuentas de crédito de una empresa que realiza ventas por correo tienen una distribución normal con media igual a 0 y desviación estándar igual a 1. Suponga que se elige una cuenta de crédito al azar de los registros de la compañía.
a) encontrar la probabilidad de que tenga un error entre 0 y 1.5
b) encontrar la probabilidad de que tenga un error de al menos 1.75
c) encontrar la probabilidad de que tenga un error entre -1.5 y 1.25

3.- Suponiendo que el salario de los contadores públicos tiene una distribución aproximadamente normal con media igual a $ 15,089 al año y desviación estándar de $ 1,035. ¿Qué proporción de los contadores públicos gana más de $17,000?

4.- Si se supone que el tiempo promedio requerido para terminar de resolver un determinado examen tiene una distribución aproximadamente normal con una media de 70 minutos y una desviación estándar de 12 minutos. ¿Cuánto tiempo debe durar un examen si se pretende que el 90% de las personas que lo tomaron lo termine?

MOMENTOS DE LA DISTRIBUCION NORMAL

El primer momento respecto al origen es la media μ. Puesto que la distribución es simétrica respecto de $x=\mu$ todos los momentos impares respecto de μ se anulan. Para calcular los momentos pares respecto de la media se utiliza:

$\mu_{2n} = (2n-1)\,\sigma^2\,\mu_{2n-2}$ donde n=1,2,...

Ejemplo:

En un examen de ciencias efectuado a un grupo de estudiantes, se encontró que las notas obtenidas seguían muy de cerca la distribución normal. La media era 58 y la desviación típica 10.
a) Representar la distribución
b) Tipificar las siguientes notas: 63, 41, 58
c) Convertir las siguientes notas típificadas en notas corrientes: 2.9, -1.5, 0
d) Si se elige al azar un estudiante del grupo ¿Cuál es la probabilidad de que su nota sea 68?

EVALUACION PROPUESTA

1.-Un vendedor de seguros vende pólizas a 5 hombres todos de la misma edad y con buena salud. De acuerdo con las tablas actuariales la probabilidad de que un hombre de esta edad viva 30 años más es de 2/3. Hallar la probabilidad de que a los 30 años vivan al menos 3 hombres.

2.-En una empresa productora de papel el 4 por 100 de los rollos producidos son defectuosos. Si se producen 100 rollos diarios, ¿Cuál es la probabilidad de encontrar 8 rollos defectuosos?.

3.-En una playa se tomaron las siguientes estadísticas, el promedio de ahogados en accidente por año es 3 de cada 100,000 personas que visitan la playa. Hallar la probabilidad de que si se reciben 200,000 visitantes se ahogen 4 personas.

4.-Una caja contiene 10 pelotas iguales de distinto color 6 rojas y 4 verdes. Una niño escoge una de ellas al azar y ve de que color es, después la regresa y las revuelve. Después extrae una segunda pelota y sigue el mismo procedimiento hasta que juega con 3 pelotas. ¿Cuál es la probabilidad de obtener 2 pelotas rojas y una verde entre las cinco extraídas? La respuesta debe estar sustentada en un diagrama de árbol.

5.-Se toman fichas que tienen 2 colores, una cara es blanca y la otra es negra. Si se lanzan 8 fichas y se considera sólo las caras blancas que se obtienen. ¿Cuál es la probabilidad de obtener 4 caras blancas?

6.-Una empresa produce 30 diskettes por hora y salen 2 por 100 defectuosos. Si se producen 200 diskettes al día cuál es la probabilidad de que se encuentren 6 diskettes defectuosos?

7.-Se sabe que un promedio de 15 por 100 de las lámparas eléctricas producidas por General Electric se funden cuando se prueban por primera vez. Si se prueba una muestra de 20 lámparas, ¿cuál es la probabilidad de que 3 de ellas se fundan?

8.-Entre las 2 y las 4 de la tarde el promedio de las llamadas teléfonicas que se reciben en la central teléfonica de una compañía por minuto es de 2.5. Hallar la probabilidad de que en un minuto haya 4 llamadas teléfonicas.

9.- En una población normal el 15 por 100 tiene unos valores de X menores que X=12, y el 40 por 100 tiene unos valores de X mayores que X= 16.2 . Calcular la media y la desviación estándar de la población.

10.- Se empaquetan cajas de chocolates automáticamente y el peso medio de 1.6 kilogramos. Si sólo el 5 por 100 de las cajas pesan menos de 1 kilogramo, calcular la desviación estándar suponiendo una población de pesos normal.

CONCLUSIONES

La evaluación escrita sobre los conceptos del tema se efectuó a los alumnos del grupo 505 de estadística y los resultados que se obtuvierón se presentan en una tabla de distribución de frecuencias en 4 intervalos de calificaciones; se obtuvierón los promedios correspondientes de los datos para tratar de identificar el avance en el aprendizaje de los alumnos, se muestran distintas gráficas de ellos.

1	3	5	6	7	8	8	9
2	4	5	6	7	8	9	9
2	4	5	6	7	8	9	10
3	4	5	6	7	8	9	10
3	4	5	7	7	8	9	10

Intervalos	Clases	Frecuencia	Frecuencia Relativa	Marca de clase	Marca * Frecuencia
0 - 2	0 - 2.5	3	3/40	1	3*1 = 3
3 - 5	2.5 - 5.5	12	12/40	4	12*4 = 48
6 - 8	5.5 - 8.5	16	16/40	7	16*7 = 112
9 -11	8.5 - 11.5	9	9/40	10	9*10= 90
	total	40			253

$$\bar{X} = \frac{253}{40} = 6.3$$

$$\tilde{X} = 6.5$$

$$X_0 = 6.6$$

Gráficas: Polígono de frecuencias

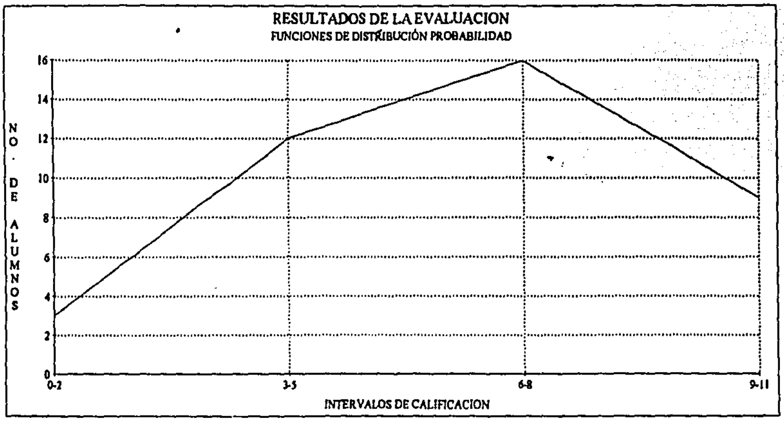


Diagrama circular:

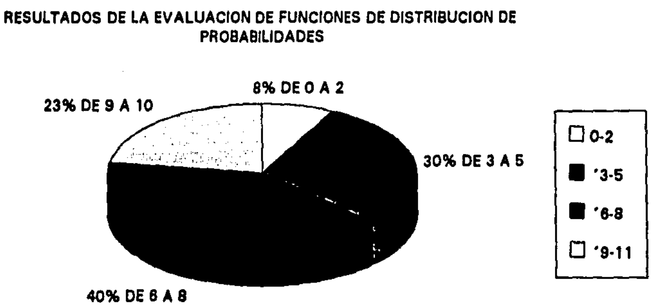
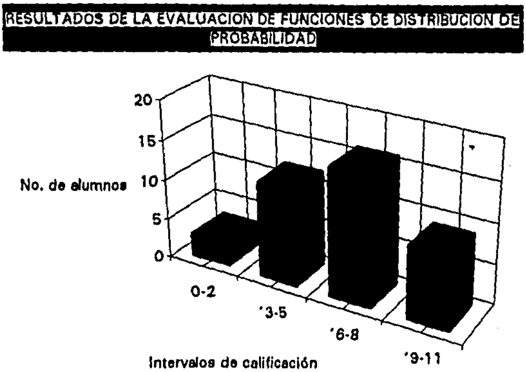


Diagrama de barras:



El diagrama correspondiente a los resultados de la evaluación del último tema funciones de distribución de probabilidades, del presente trabajo, muestra que el 62% de los alumnos del grupo 505 de estadística del CCHO aprobó la evaluación escrita de los conceptos. Este resultado se puede interpretar como positivo debido a que los tópicos de esta sección representan mayor complejidad para los estudiantes. Se puede sugerir mayor énfasis en el desarrollo de problemas de ejemplo para tratar de incrementar el porcentaje de estudiantes aprobados.

ANEXOS

PROGRAMA DE LA MATERIA DE ESTADISTICA
ACADEMIA DE MATEMATICAS CCH-O
ANEXO I

Este anexo contiene el programa aprobado por la academia de matemáticas del Colegio de Ciencias y Humanidades plantel Oriente, para las materias optativas de estadística I y estadística II del quinto y sexto semestres del plan de estudios.

PROGRAMA DE ESTADISTICA I

TEMA I.- Introducción
- 1.1 Proceso del método estadístico
- 1.2 Estadística y método científico
- 1.3 Niveles de medición

TEMA II.- Estadística descriptiva
- 2.1 Población y muestra
- 2.2 Tipos de muestra y elementos
- 2.3 Ordenación y tabulación de datos
- 2.4 Representación gráfica
- 2.5 Distribuciones de frecuencias
- 2.6 Construcción de una tabla de distribución de frecuencias y su representación gráfica
- 2.7 La suma general. Simbología y operatividad
- 2.8 Parámetros de la muestra:
  - Medidas de tendencia central
  - Concepto de interpretación
  - Medidas de dispersión
- 2.9 Relaciones entre los estadísticos de la muestra

TEMA III.- Introducción a la probabilidad
- 3.1 Conceptos básicos
  - Experimento o fenómeno aleatorio
  - Espacio muestra
  - Eventos: evento elemental, evento seguro, evento imposible
- 3.2 Algébra de eventos
  - Notación e interpretación en lenguaje común
  - Eventos mutuamente exclusivos
- 3.3 Definición de probabilidad frecuencial
- 3.4 Definición de probabilidad clásica
- 3.5 Formalización del modelo espacio de probabilidad
  - Características
  - Espacio equiprobable
- 3.6 Ley de adición de probabilidades
- 3.7 Ley de multiplicación de probabilidades
  - Concepto de independencia
- 3.8 Probabilidad condicional

PROGRAMA DE ESTADISTICA II

TEMA I.- Variable aleatoria

1.1 Definición de variable aleatoria
1.2 Variables aleatorias discretas, variables aleatorias contínuas
1.3 Distribuciones de probabilidad. Función de densidad. Función de distribución
1.4 Esperanza matemática.
1.5 Parámetros de la distribución de probabilidad: varianža, desviación estándar.

TEMA II.- Cálculo combinatorio

2.1 Nociones (tema opcional)

TEMA III.- Distribuciones especiales de probabilidad

3.1 Distribución binomial.
3.2 Distribución de Poisson
3.3 Normal

PERFIL DEL ESTUDIANTE A NIVEL MEDIO SUPERIOR PARA OPTAR POR LA MATERIA DE ESTADISTICA
ANEXO II

En el quinto y sexto semestre del plan de estudios actual del Colegio de Ciencias y Humanidades para el área de matemáticas tiene 3 opciones mutuamente excluyentes: matemáticas 5 y matemáticas 6, estadística I y estadística II, lógica I y lógica II. En una gran mayoría de estudiantes existe el concepto de optar por estadística o lógica porque no desêan cursar los "díficiles" temas de matemáticas.

En la opción de estadística se necesita tener un conocimiento básico pero definido de los cursos anteriores de matemáticas por lo que se recomienda que el estudiante que opte por esta materia debiera de cubrir el siguiente pérfil académico:

- Matemáticas I
    - Conceptos básicos de aritmética
      tema fundamental: operaciones aritméticas con fracciones comunes (quebrados)
    - Operaciones con conjuntos
    - Conceptos básicos de lógica. Métodos de deducción.

- Matemáticas II
    - Conceptos básicos de álgebra
      Manejo ágil del concepto de variable.
      Operaciones aritméticas con variables.
      Reglas de simplificación de términos semejantes.
      Ecuaciones de 1er. grado con una incógnita, métodos de solución.
      Determinantes.
      Ecuaciones de 2do. grado con una incógnita, métodos de solución.
    - Conceptos básicos de funciones.

- Matemáticas IV
    - El plano carteciano. Localización de puntos en un plano.
      Conceptos de distancia entre puntos.
      Gráficas de funciones.
      Valor absoluto.

Estos conceptos facilitarían al estudiante, la comprensión de los conceptos relativos a estadística y probabilidad.

Bibliografía

Anderson, Theodore Wilbur, *An introduction to the Statistical analysis of data*, Houghton,mifflin

Cochran, William G., *Sampling Techniques*, John wiley & sons

Daniel, Wayne W., *Bioestadística*, Limusa

Derek, Rowntree, *Statistics without tears*, Charles scribner's sons

Juárez Rodríguez, Fernando, *Prontuario de estadística I y II*, CCH-O, UNAM 1989

Levin, Jack, *Fundamentos de estadística en la investigación social*, Harla S.A. de C.V.

Mood, Alexander M., *Introduction to the theory of statistics*, McGraw-Hill International Edition

Mendenhall, William, *Estadística para administración y economía*, Grupo editorial iberoamérica

Méndez, Ignacio, *Estadística y método científico*, Comunicaciones técnicas IIMAS UNAM 1975 No.13

Méndez, Ignacio, *La estructura de la investigación y la Estadística*, Comunicaciones técnicas IIMAS UNAM, 1988 No.106

Méndez, Ignacio, *Pruebas de hipotésis estadísticas*, Comunicaciones técnicas IIMAS UNAM 1979 No.41

Méndez, Ignacio, *Comparación de medias de población*, Comunicaciones técnicas IIMAS UNAM 1976 No.17

Méndez, Ignacio, *Comentarios sobre la inferencia estadística*, Comunicaciones técnicas IIMAS UNAM 1982 No.55

Méndez, Ignacio, *La enseñanza de la estadística en México*, Comunicaciones técnicas IIMAS UNAM 1982 No.61

Phillips, John L. Jr., *Statistical thinking*, W.H. Freeman and Co.

Stewart, Ian, *Conceptos de matemática moderna*, Edit. Alianza Universidad

Turner, J.C., *Matemática moderna aplicada*, Edit. Alianza Universidad