UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

CONCEPTOS BÁSICOS DE ESTADÍSTICA PARA LA ESCUELA NACIONAL PREPARATORIA

T E S I S
PARA OBTENER EL TITULO DE:
A C T U A R I A

P R E S E N T A:

GISELLE OCHOA HOFMANN

DIRECTOR DE TESIS
DR. JAVIER PÁEZ CÁRDENAS

2006

Agradecimientos

Dedico esta tesis a mis hijas Eli y Gisellita
por ser los seres que me impulsan cada día de mi vida.
A mi esposo Jorge
por su apoyo amor y comprensión.
A mi madre Yolanda
por toda su paciencia y apoyo en todas las situaciones de mi vida.
A mi hermano Alfonso
por estimularme a ser una persona mejor y por todo su apoyo, cariño y comprensión.
A mi amiga Gabriela
por toda su ayuda y amistad para revisar esta tesis.
A mi director de Tesis el Dr. Javier Páez
por toda su ayuda tanto en mi carrera como para poderla culminar.
A mi amiga Cecilia Conde
por toda la asesoría brindada en este trabajo.
A mis asesores y amigos por su apoyo.

INDICE

I N T R O D U C I Ó N

El enorme edificio de las matemáticas ha sido construido sobre los pilares del número y la forma. Del número surgió la aritmética, llevando al número a un concepto abstracto y general, ajustándolo a una mentalidad más compleja del desarrollo del pensamiento matemático, dando origen al surgimiento del cuerpo de las doctrinas matemáticas, por ejemplo: álgebra, geometría, cálculo, estadística, probabilidad, etc., las cuales han sido utilizadas para desarrollar y fundamentar el conocimiento científico.

La Estadística surge con los grandes imperios de la antigüedad. Se han encontrado vestigios de la forma como el hombre ha manifestado su necesidad de contabilizar el número de personas, bienes, fuentes de ingreso, censos de tierras, etc.

En el continente americano, por ejemplo, los incas desarrollaron un sistema de conteo en donde todos los datos relacionados con las actividades económicas y demográficas se guardaban en “quipus”.

Hacia el siglo XVI, en Europa, junto con las monarquías absolutas, Jean Bodin en Francia (1530-1596) resalta la importancia de los censos y 50 años más tarde Galileo presenta por primera vez todos los datos en tablas de distribución.

Así, las técnicas que se han utilizado y se utilizan para cuantificar todos los aspectos de la vida se les llamarán técnicas estadísticas.

El New Collegiate Dictionary de Webster define la Estadística como: "una rama de las matemáticas que trata de la recopilación, el análisis, la interpretación y presentación de una gran cantidad de datos numéricos" [1].

Se pueden encontrar una serie de definiciones, pero todas tienen algunos elementos en común: cada una hace referencia a una recopilación de datos que tienen ciertas características comunes o a la selección de un subconjunto de una gran colección de datos (muestra) para deducir o inferir el comportamiento de un conjunto total (población); para hacer juicios y tomar decisiones.

Por lo anterior, se puede decir que hoy en día la Estadística, junto con el cálculo de probabilidades, constituyen una rama de las matemáticas con aplicaciones en casi todas las actividades humanas: física, psicología, biología, agricultura, demografía, entre otras. En todas estas áreas se hacen predicciones, encuestas, gráficas, etc., pero también se aplica a fenómenos no medibles como la lingüística y la literatura.

Por lo anterior, el estudio de la Estadística y la Probabilidad es importante en la enseñanza de esta materia dentro del Sistema Educativo de la Escuela Nacional Preparatoria como parte del Sistema Educativo Mexicano y del ciclo del bachillerato en U.N.A.M.

Es conocido que la Escuela Nacional Preparatoria, tiene la siguiente misión:

> Educar mujeres y hombres para que obtengan una formación integral que les permita contar con:
> - ✓ Una amplia cultura.
> - ✓ Los conocimientos sólidos necesarios para cursar con éxito estudios superiores.

[1] Mendelhall,Sheaffer y Wackerly. *Estadística Matemática con Aplicaciones*. p.2

- ✓ Una mentalidad analítica, dinámica y critica que les permita ser conscientes de su realidad y comprometidos con la sociedad.
- ✓ La capacidad de obtener por si mismos nuevos conocimientos, destrezas y habilidades, que les posibilite enfrentar los retos de manera positiva y responsable.[2]

El fundamento primordial es entonces el de formar al estudiante de una manera integral, sentando las bases necesarias para que pueda continuar su formación profesional e ingrese a las Escuelas o Facultades correspondientes a su área de estudio. Pero en algunos casos ésta etapa es terminal, por lo que se hace necesario, que el alumno tenga y adquiera los instrumentos teórico- metodológicos básicos para comprender con mayor objetividad su realidad y tenga la posibilidad de tomar decisiones y asumir una posición critica frente a sus problemas cotidianos.

El interés de desarrollar esta investigación como tema de tesis surge de mi experiencia como profesor de la asignatura de Estadística y Probabilidad en la Escuela Nacional Preparatoria de la U.N.A.M. Plantel 5. “José Vasconcelos” , trabajo que desempeño desde hace 5 años.

Mi trabajo docente y la búsqueda del entendimiento de este conocimiento, me llevó al análisis y a la propuesta de un criterio sobre la forma que debe tener la enseñanza de la Estadística Descriptiva en el nivel de enseñanza media superior.

En mi opinión, la Estadística Descriptiva es necesaria en la formación del bachiller, ya que le brinda los elementos teóricos-metodológicos para la comprensión de su entorno, objetivos explícitos de la Escuela Nacional Preparatoria de la U.N.A.M.

[2] Plan de desarrollo de la Escuela Nacional Preparatoria. 2002-2006. p. 11

En el plan de estudios de la Escuela Nacional Preparatoria de la U.N.A.M. la asignatura de Estadística y Probabilidad se imparte en el sexto año, como materia optativa de las áreas:

1. Físico-matemáticas y de las ingenierías
2. Ciencias biológicas y de la salud;
3. Ciencias sociales;
4. Humanidades y artes;

No sólo es importante la selección de las materias que deben enseñarse dentro de un plan de estudios, sino lo es también el programa de cada una de ellas, así como el qué y el cómo se van a enseñar. Así, la construcción de programas de estudio es el eje de articulación entre el profesor y el alumno, lo que proporciona los elementos teóricos que enmarcan el proceso de enseñanza- aprendizaje.

El programa de estudios de Estadística y Probabilidad se divide en tres partes, que a continuación se desglosan:

I. Datos de Identificación
II. Presentación
III. Contenido del Programa

I.- Datos de Identificación.

El programa de estudios de la asignatura de Estadística y Probabilidad está inserto en el Colegio de Matemáticas en sexto año, con categoría de optativa y de carácter teórico, de tres horas semanarias, con un total estimado de 90 horas anuales, estimadas. En el currículo tiene un total de 12 créditos.

II.- Presentación

Este apartado se divide en 4 puntos:

a) Exposición de motivos y propósitos generales del curso.

b) Características del curso o enfoque disciplinario.

c) Principales relaciones con las materias antecedentes, paralelas y subsecuentes.

d) Estructuración listada del programa.

a) Exposición de motivos y propósitos generales del curso:

Este curso es importante, debido a que el estudiante, en algún momento, se va a enfrentar a problemas en cualquier área del conocimiento que elija, que pueden encontrar una solución con el uso de estadística y la probabilidad.

El curso tiene como propósito dar un acercamiento y despertar el interés hacia la estadística, por medio de problemas sencillos que requieren el uso de la Estadística Descriptiva.

El curso considera la estructura y secuencia de los contenidos, y sigue un enfoque metodológico orientado hacia la solución de problemas, por medio del cual, el alumno conocerá, comprenderá y aplicará la Estadística Descriptiva. Mediante esta metodología el alumno aprende a realizar el planteamiento de problemas simples que irán aumentando su complejidad en el desarrollo de un mismo tema. Así, la tendencia de este programa, pretende constituirse en una etapa intermedia del desarrollo curricular de la enseñanza de las matemáticas en el bachillerato y de tránsito progresivo de una enseñanza lineal y rítmica a una enseñanza de construcción, más acorde con el nivel esperado en la licenciatura. Para lograr lo anterior se efectúa un seguimiento y evaluación del programa de manera colegiada por parte de los profesores, lo que permite nuevas estructuraciones y dosificación que sean más funcionales para los propósitos del curso.

b) Características del curso o enfoque disciplinario:

Aquí se pretende, en los tres años de enseñanza de las matemáticas, que el alumno adquiera conocimientos indispensables para la demanda a nivel superior. En el 4° año se considera como Etapa de Introducción (Matemáticas IV), en el 5° año se conoce como Etapa de Profundización (Matemáticas V), y el 6° año se considera como Etapa de Orientación. Cada asignatura es la base de la inmediata superior, los conectivos entre estos tres años son básicamente las funciones;

c) Principales relaciones con las materias antecedentes, paralelas y subsecuentes:

El curso de Estadística y Probabilidad tiene como antecedentes a Matemáticas IV y Matemáticas V, que proporcionan las herramientas, el lenguaje y las operaciones básicas para acceder a este curso.

Materias antecedentes de otras áreas son: Física III. Química III, Biología IV y Educación para la salud, y constituyen un apoyo didáctico al ofrecer problemas de aplicación, de algunos conceptos básicos del curso aquí descrito.

Son materias paralelas Matemáticas VI, Física IV, Química IV, Biología V, Dibujo, Cosmografía, Geografía Política, Geografía Económica, Sociología, Informática aplicada a la Ciencia y a la Industria, Temas Selectos de Biología, Físico-Química, Geología y Mineralogía, y Psicología, todas ellas pueden considerarse como herramientas de apoyo al curso de Estadística y Probabilidad Las materias subsecuentes son los diferentes cursos de matemáticas que se imparten en las diversas carreras a nivel de enseñanza superior.

d) *Estructuración listada del programa*: El programa esta estructurado en tres unidades cuyos contenidos son:

Unidades:

a) Estadística descriptiva.
b) Conjuntos
c) Probabilidad

III. Contenido del Programa. [3]

Este apartado se presenta como anexo en el presente trabajo, en donde se desglosan los temas de cada unidad.

Este trabajo sólo cubre los temas relacionados con la primera unidad del curso de Estadística y Probabilidad del nivel de Enseñanza Media Superior de la Escuela Nacional Preparatoria de la U.N.A.M., teniendo en mente a los alumnos que por primera vez estudian el tema, aunque puede ser usado por estudiantes de nivel licenciatura e incluso por personas interesadas en conocer el tema.

Por lo anterior, se plantean los siguientes objetivos para este trabajo:

- Introducir al alumno a un primer acercamiento a la Estadística
- Presentar la teoría de cada tema en forma sencilla, por medio de un lenguaje cotidiano, sin la formalización estricta del lenguaje matemático.

[3] vid. Programa de Estudios de la Asignatura de Estadística y Probabilidad de la ENP. de la U.N.A.M. 2003-2004

- Darle al alumno, a través de ejemplos e ilustraciones, más de un método para resolver problemas.
- Que el alumno maneje un lenguaje sencillo y claro, para adquirir este conocimiento a través de la forma como se presentan los temas, de lo simple a lo complejo.
- Que relacione la seriación temática de un curso de Estadística, así como entender su utilidad.

El presente trabajo de tesis, basado en el contenido de la primera unidad del programa de la materia de Estadística y Probabilidad en el bachillerato Preparatoriano, se desarrolla en 3 capítulos, más la Introducción, Conclusiones, Bibliografía, Hemerografía, y Anexo.

En el primer capítulo, se presenta la forma como debe de manejarse los conceptos básicos de la Estadística, el tratamiento de los datos por medio de tablas de frecuencia, y las principales representaciones básicas de los datos.

Con base en el capítulo anterior, el segundo capítulo presenta la forma de cómo se deben de obtener las medidas que describen las bases de datos y que se conocen como las medidas de tendencia central (o de posición). Se explican cuáles son, cómo son, cómo se obtienen esas medidas a partir de una población o muestra y su importancia en la toma de decisiones.

Existen otras medidas que indican como varía la población o muestra; estas medidas se analizan en el tercer capítulo, se explica cuáles son las medidas de dispersión o variabilidad, cómo son, su importancia y cómo se obtienen a partir de una población o muestra. En este mismo capítulo se menciona el Teorema de Chevyshev, para hacer inferencias de un grupo de datos. Además, se hace una breve explicación de la distribución normal de frecuencias con el uso de variables aleatorias para invitar al estudiante al conocimiento de la probabilidad.

Al final se presenta la Bibliografía y Hemerografía consultada para la realización de este trabajo, conclusiones y un anexo.

Con base en lo anterior, este trabajo de investigación está fundamentado en el proceso de enseñanza- aprendizaje de la Estadística para el nivel de Enseñanza Media Superior en la Escuela Nacional Preparatoria de la UNAM, con la finalidad de que sirva como una propuesta pedagógica, que ha sido probada por la autora en los cursos impartidos en años anteriores.

I. FRECUENCIAS Y GRAFICAS

Existen varias posturas para definir el concepto de estadística, pero nadie duda de la aplicación de ésta en la vida cotidiana, ya que es una herramienta necesaria para tomar una o varias decisiones importantes en nuestra vida diaria, como lo muestra la siguiente cita:

LA ESTADÍSTICA ¿Herramienta confiable?
Es probable que hasta ahora no hayamos reflexionado de la gran cantidad de información estadística a la que estamos expuestos cotidianamente. Por ello es conveniente conocer más acerca de esta área del conocimiento y averiguar: ¿Qué es esta disciplina?, ¿Qué tipo de trabajo se realiza en ella?, ¿Qué hace un estadístico?, ¿Dónde trabaja? Y todo lo relacionado con ella.

> Una fuente común de información estadística son los deportes transmitidos por televisión. Así, por ejemplo, durante la transmisión de un partido de fútbol, al medio tiempo y al final, se presenta la sección titulada las “estadísticas del partido” en las que se intenta dar un resumen numérico a través de datos como tiros a gol, faltas, amonestados, expulsados, etc. Este tipo de información no es exclusiva de este deporte, en la mayoría de ellos, es común encontrar un despliegue de datos de diversa índole sobre los equipos y sus jugadores. Por ejemplo, en el básquetbol se maneja la siguiente información: promedio de puntos anotados y recibidos por equipo, porcentaje de tiros encestados (canastas) y de tiros libres anotados por jugador.

Otra fuente de datos estadísticos usual es el periódico, pródigo en información de esta naturaleza en secciones como economía y finanzas. Aquí se describen principalmente los comportamientos como índice de inflación y precios al consumidor, las fluctuaciones del peso con respecto al dólar, el resumen de las empresas que cotizan en la bolsa de valores, etc. Una característica particular de este tipo de información es que se despliega como gráfica.
También en época de elecciones, los medios de comunicación se encargan de desplegar a la opinión pública con información acerca de las preferencias electorales de la ciudadanía, tanto previo a la elección (encuestas preelectorales) como durante (encuestas de salida, conteos rápidos) y después de ella (resultados finales). La información se despliega en forma de porcentajes de electores que favorecen a algún partido, además de gráficas que refuerzan de manera visual lo dicho por las cifras. También cuando ocurren fenómenos naturales como terremotos, tormentas o huracanes, los medios proporcionan gran cantidad de cifras sobre el número de muertos,

damnificados, desaparecidos, casas destruidas y pérdidas económicas, entre otros.

> En general, podríamos definir a la estadística como una disciplina que se basa en la recopilación sistemática y ordenada de datos cuantitativos (no necesariamente numéricos). Esto puede referirse lo mismo a individuos que a grupos, que a hechos o situaciones concretas. A partir de esa información, la estadística analiza y deduce su significado en un contexto determinado y prevé patrones de comportamiento a futuro.[5]

La Estadística Matemática trata de la teoría y aplicación de métodos para obtener datos, analizarlos y hacer deducciones a partir de ellos. Los datos pueden consistir en números arreglados de forma tabular (tablas) y/o de representaciones gráficas.

Se pueden obtener datos de diferentes fuentes, sin embargo, tienen un rasgo en común, que se deriva de circunstancias que están afectadas por la casualidad, son resultado de factores que no pueden ser controlados y, con frecuencia, enumerados.

Definiciones de Estadística:

"Rama de las matemáticas que trata la recopilación, análisis, interpretación y presentación de una gran cantidad de datos numéricos."[6]

Otra acepción de Estadística la define como: "La ciencia que utiliza como instrumento a las matemáticas y al cálculo de probabilidades para estudiar fenómenos que dependen del azar (aleatorios) y no están regidos por leyes físicas."[7]

---

[5] Téllez Rojo, Martha María y Zamora Muñoz, Salvador. "Estadística herramienta confiable". En:<u>¿Cómo Ves?</u> No. 16

[6] Mendelhal, op.cit., p. 1

[7] Ibid., p. 2

I.1. Población y Muestra.

La población es el conjunto de objetos o sujetos de los cuales se van a obtener datos que sirven para describir o inferir características de estos.

Una muestra es un subconjunto de la población en estudio.

Por ejemplo:

- Se necesita saber cuántos niños en el Distrito Federal pueden contraer la varicela. Para ello se hacen estudios de los niños que asistían a los hospitales de Centro Médico y la Raza, que habitan en la entidad.

  Población: Todos los niños del Distrito Federal.

  Muestra: Los niños que asistían a los hospitales de Centro Medico y la Raza, que habitan en el Distrito Federal.

- En un bachillerato con 1500 alumnos los cuales cursan Estadística, se quiere saber si sus maestros en la materia dan a conocer el programa de actividades de la materia. Para ello se seleccionar al azar 50 estudiantes para la entrevista.

  Población: Alumnos del bachillerato que cursan Estadística.

  Muestra: 50 estudiantes del bachillerato.

Recabar datos de una población completa, sin duda, nos da un conocimiento exacto de ella. Sin embargo muchas veces es casi imposible, pues ésta puede ser muy grande así, la obtención de estos datos puede ser muy tardado y con un costo económico muy alto. En cambio, obtener información de la población a partir de una muestra es rápido a un costo económico bajo, pero la información de la población va a ser aproximada.

Una muestra puede ser grande o pequeña, en función del número de elementos que contenga y con referencia a la población en estudio y de la información que se quiera obtener. El tamaño de una muestra se suele representar por la letra "*n*".

I.2. Ramas de la Estadística.

La Estadística se divide en dos grandes ramas:

1. La **Estadística descriptiva** es la que, a partir de datos de una población o muestra aleatoria o representativa (es decir un subconjunto de datos de la población no necesariamente obtenida por métodos aleatorios) la, obtiene valores característicos que la describen para conocer los aspectos más relevantes a ésta.

2.- La **Estadística inferencial** analiza los datos obtenidos de una muestra aleatoria o probabilística, para conocer a partir de ellos características aproximadas de la población sin necesidad de tener los datos de toda la población.

I.3. Frecuencias: absoluta y relativa.

Cuando se trabaja con estadística es inevitable hablar de la variable o las variables, pero ¿que son éstas?

Al hacer una encuesta a una población de 10 mujeres, que su edad fluctúa entre 20 y 30 años, se les realiza la siguiente pregunta: ¿qué peso tiene?, las respuestas obtenidas fueron las siguientes:

Persona	1	2	3	4	5	6	7	8	9	10
Peso (Kg)	52.5	48.0	50.5	54.0	49.0	53.5	52.0	53	49.5	50.0

En este caso el peso es la variable aleatoria o de interés, ya que el peso cambia en función de cada persona.

Por tanto, podemos decir que una **variable estadística** es una propiedad o característica de cada elemento de una población que puede ser medida o clasificada (peso, estatura, color de piel, nacionalidad, tiempo de útil de un foco, etc.). Como podemos ver existen diversos tipos de variables:

Es importante decir que una ***variable estadística***, es simplemente una variable de interés que puede ser *deterministica* si se refiere a valores conocidos de una población, o bien *aleatoria*, si se refiere a valores tomados de una muestra aletoria o probabilística

Clasificación de las variables aleatorias o de interés.	**Cuantitativas (numéricas o escalares).** Expresan en forma numérica las características de los sujetos en estudio	**Discretas** Toman su valor en los números enteros (No. de personas, de bacterias, etc) Sirven para contar.
		Continuas Toman su valor en los números reales. Como el tiempo de vida de una persona o de algún aparato eléctrico) Sirven para medir
	Cualitativas o categóricas Expresan una categoría no numérica (Tipo de piel, de cabello, etc)	**Nominales** Expresan una condición de los sujetos en estudio como su domicilio, nacionalidad, genero, raza, color, etc. No son cuantificables.
		Jerarquizadas u ordinales. Denotan una característica de tamaño u orden. (chico, mediano, grande) ó (malo, regular o bueno).

Para poder entender de manera más simple el concepto de frecuencia y sus diferentes tipos se tomará el siguiente ejemplo:

Un maestro de matemáticas quiere observar las calificaciones que obtuvieron los alumnos de uno de sus grupos y realizo la siguiente tabla.

Calificación	1	2	3	4	5	6	7	8	9	10
No. de alumnos que obtuvieron la calificación	3	2	2	2	3	12	8	8	6	4

En la primera fila se observa la variable calificación y en la segunda fila el número de alumnos que obtuvieron esa calificación (o el número de veces que se repite la calificación en el grupo). A esto se le conoce como **frecuencia absoluta.**

La frecuencia absoluta es el número de veces que aparece un valor x en una muestra. Esta se representa por f(x).

En muchas ocasiones la información que muestra la frecuencia absoluta puede ser insuficiente, ya que en el ejemplo anterior 12 personas obtuvieron seis de calificación, pero ¿son muchos o pocos?, esto depende del total de alumnos. Por lo que, podemos precisar que 12 de 50 alumnos obtuvieron seis de calificación.

Al número que resulta de dividir la frecuencia absoluta entre el tamaño de la muestra (n) o de la población (N) se le llama frecuencia relativa.

La frecuencia relativa es el resultado de dividir cada frecuencia absoluta para el dato x entre el tamaño de la muestra (n) o de la población (N).

Como las frecuencias absolutas se obtuvieron contando el número de veces que aparece el mismo dato en la muestra o en la población, se puede deducir que la suma de las frecuencias absolutas es igual al tamaño de la muestra.

La suma de las frecuencias absolutas es igual al tamaño de la muestra o de la población según el conjunto de datos que se estudie.. En simbología matemática se expresa como:

$\sum_{i=1}^{m} f(x_i) = n$ (Si el conjunto de datos que se analiza es una muestra)

ó

$\sum_{i=1}^{m} f(x_i) = N$ (Si el conjunto de datos que se analiza es una población)

donde:

$x_1, x_2, ..., x_m$ *son todos los valores de la muestra y n es el tamaño de la muestra.*

Las frecuencias relativas siempre están entre cero y uno, ya que son una proporción del total de datos*. La suma de las frecuencias relativas es igual a 1.*

Las frecuencias relativas también se pueden escribir como un porcentaje del total de la muestra o de la población y se obtiene únicamente multiplicando por 100 la frecuencia relativa. A este número en algunos textos se le conoce como la *frecuencia relativa porcentual,* y es la forma en la que usualmente se presenta la información estadística en noticias o periódicos.

Ahora, si se quiere conocer la reprobación del grupo, se tendrá que sumar los alumnos que obtuvieron uno, dos, tres, cuatro y cinco. A esto se le conoce como: **frecuencia absoluta acumulada.**

Si para cierto valor x sumamos todas las frecuencias absolutas correspondientes a los valores menores o iguales a x, se obtendrá la frecuencia absoluta acumulada al valor x.

La frecuencia absoluta acumulada también se puede comparar con el tamaño de la muestra o de la población, dando lugar al concepto de **frecuencia relativa acumulada**.

Otro tipo de frecuencia es la **frecuencia complementaria absoluta**, ésta se obtiene de restarle al tamaño de la muestra o de la población la frecuencia acumulada al valor x. Como en los casos anteriores, se puede obtener la frecuencia complementaria relativa, dividiendo la frecuencia complementaria absoluta entre el tamaño de la muestra.

A continuación se da una tabla con los diferentes tipos de frecuencias a partir de los datos que obtuvo el maestro de matemáticas.

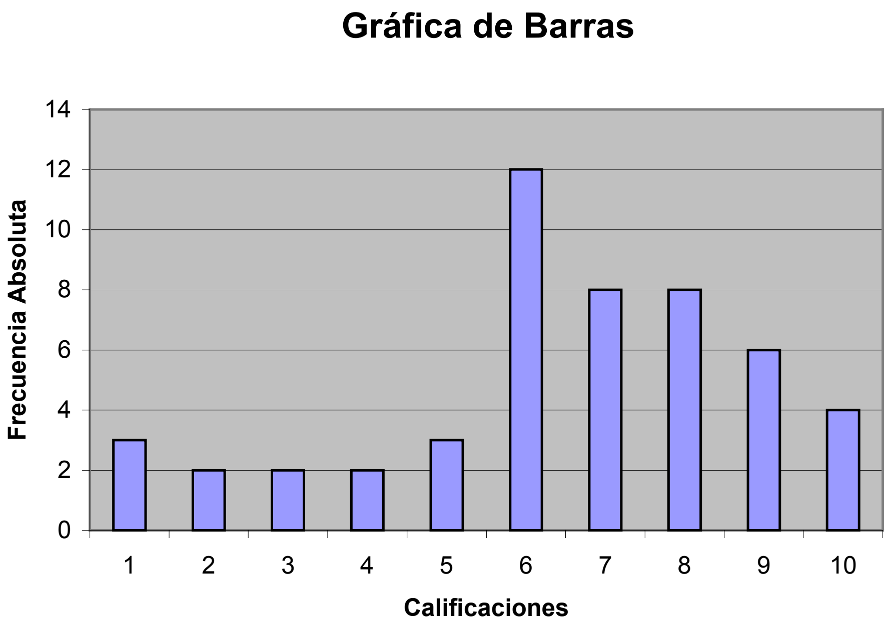
Variable Calificación	Frecuencia absoluta	Frecuencia Relativa	Frecuencia Relativa Porcentual	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada	Frecuencia Relativa Acumulada Porcentual	Frecuencia Complementaria
1	3	3/50	06%	3	3/50	06%	47
2	2	2/50	04%	5	5/50	10%	45
3	2	2/50	04%	7	7/50	14%	43
4	2	2/50	04%	9	9/50	18%	41
5	3	3/50	06%	12	12/50	24%	38
6	12	12/50	24%	24	24/50	48%	26
7	8	8/50	16%	32	32/50	64%	18
8	8	8/50	16%	40	40/50	80%	10
9	6	6/50	12%	46	46/50	92%	4
10	4	4/50	08%	50	50/50	100%	0
Totales	50	1	100%				

Cuadro I.1 Frecuencias de Calificaciones en Matemáticas

Esta representación de la información se da en tablas, pero también, se presenta por medio de gráficas. Existen varios tipos de gráficas.

I.3.1 Representaciones Gráficas

I.3.1.1. Gráfica de barras

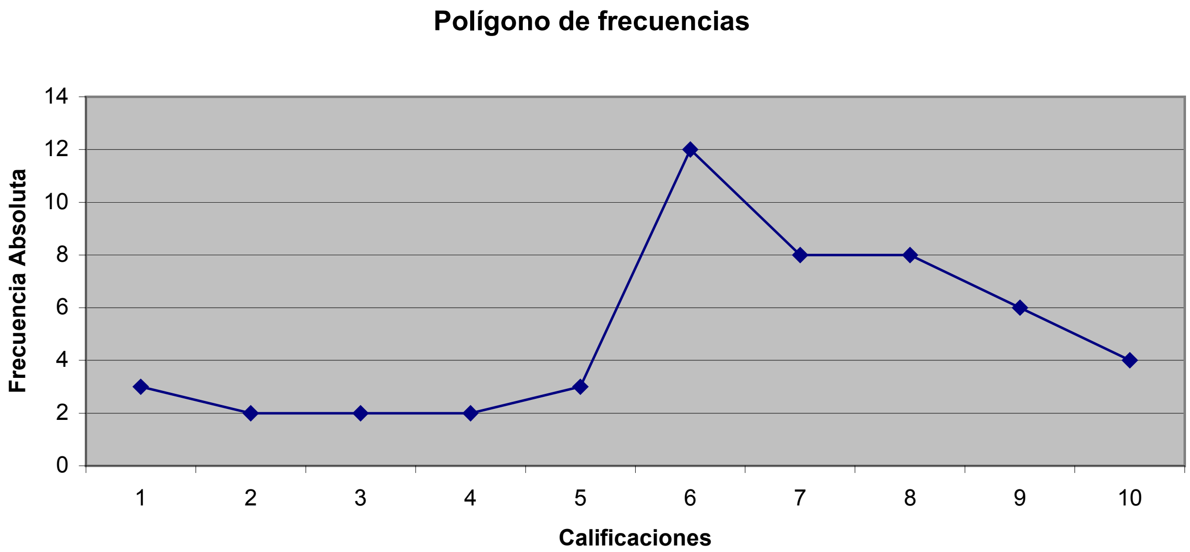


Las gráficas de barras se representan en el eje “x “u horizontal por las variables y el eje “y” o vertical representa la frecuencia absoluta para cada valor de la variable. También, se pueden obtener graficas de barras en base a la frecuencia relativa.

I.3.1.2. Polígono de Frecuencias

Al igual que la gráfica de barras, el eje “x” representa los valores que toma la variable y el eje “y” representa la frecuencia absoluta o relativa para cada valor de la variable. En este caso se traza un punto a cada valor de la variable y después se une por líneas rectas.

En los dos tipos de gráficas, se pueden dar representaciones a escala, dependiendo del valor que toma la variable o la frecuencia absoluta.

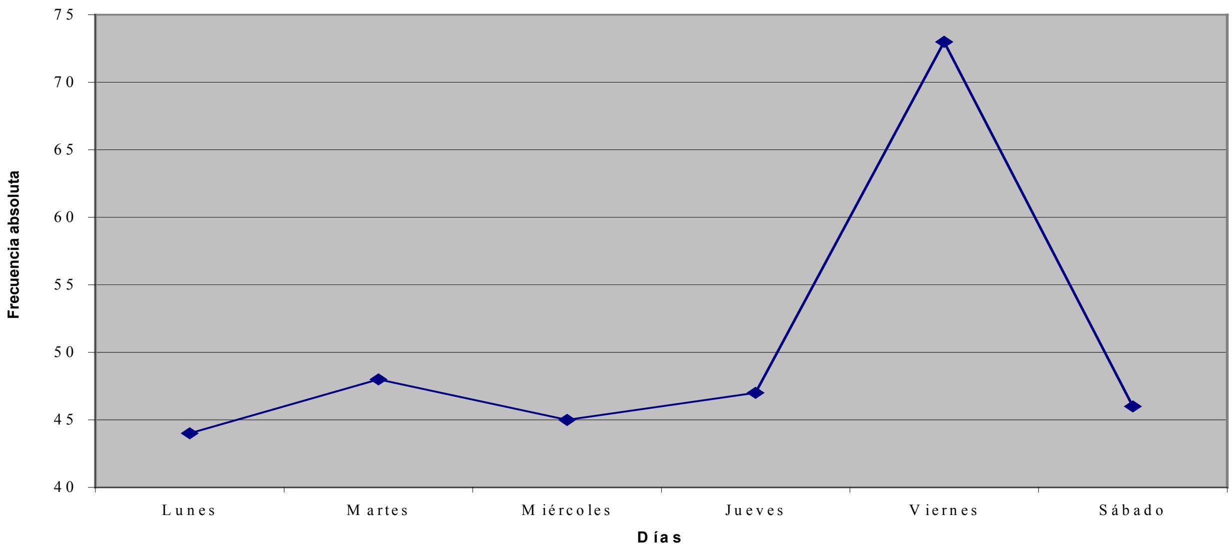


Es importante la escala que se utiliza en las gráficas ya que éstas pueden ser muy engañosas, aunque estas representen los mismos datos. Veamos el siguiente ejemplo:

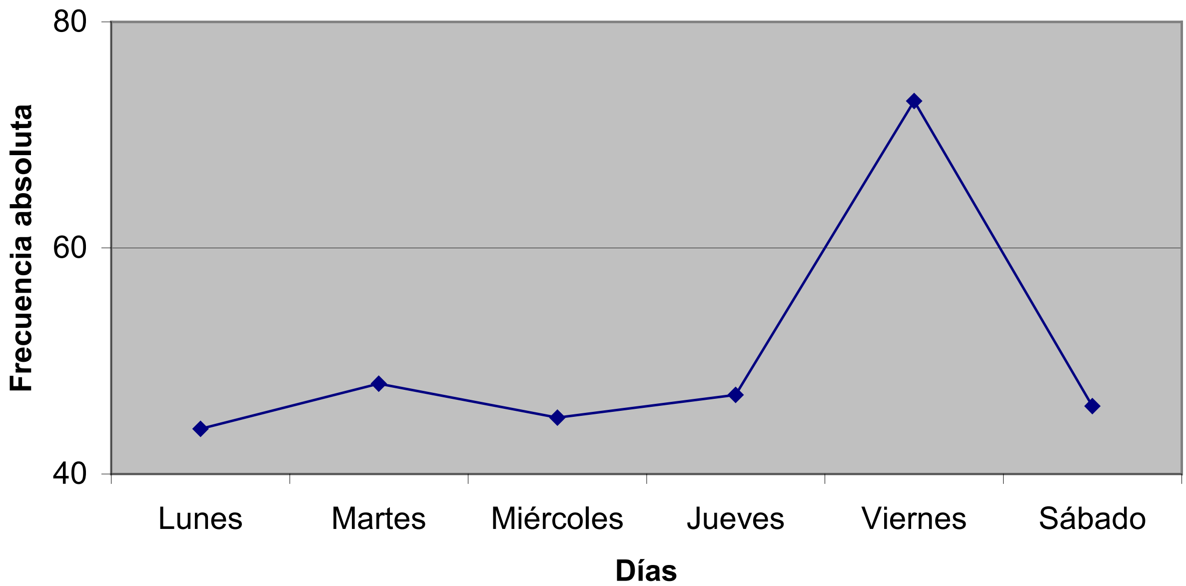
Una fábrica de suéteres quiere graficar su producción diaria, los datos obtenidos son los siguientes:

Día	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
No. De sueteres	44	48	45	47	73	46

Polígono de frecuencias (1)



Polígono de frecuencias (2)



Se observa en el polígono de frecuencias (1) la escala inicia en 0 y varía de cinco en cinco, lo que permite visualizar los cambios. En cambio, en el polígono de frecuencias (2) la escala empieza en cero y varía de 20 unidades, podríamos pensar que los dato correspondiente al viernes, no es tan extremo como se visualiza en la primer gráfica.

Estos cambios, dependen de lo que se quiera observar. Por ejemplo, si se requiere observar la variación del dólar en una semana, tal vez la unidad en la que varían los datos sería en décimas, pero si se maneja tamaño de poblaciones de países, la unidad de variación sería en miles o tal vez millones de habitantes.

I.3.1.3. Gráfica de Sectores

Otro tipo de gráfica utilizada comúnmente es la gráfica de sectores, también llamada gráfica de pastel o circular. Esta gráfica necesita un trato especial, ya que representa la proporción de los datos para un valor x (frecuencia relativa de x) como un sector de un círculo. Así, para determinar el ángulo central que corresponda a cada uno de estos sectores basta multiplicar la frecuencia relativa por 360°.

$$\boxed{\textit{Medida del ángulo central} = \textit{frecuencia relativa} \times 360°}$$

En la siguiente tabla se muestra la distribución del número de habitantes por continente en el año 2000 (los datos se redondearon):

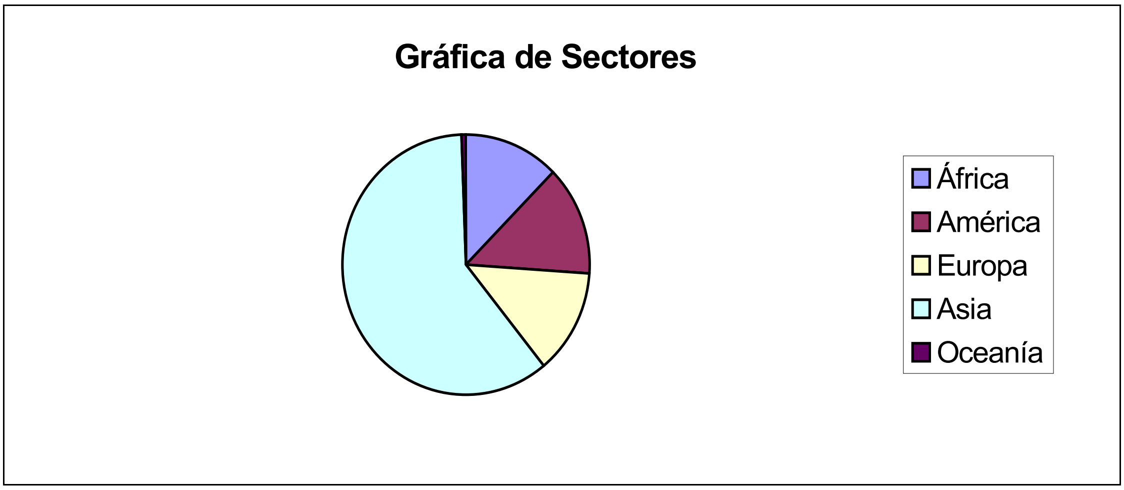
Continentes	Número de habitantes (millones de personas) año 2000 (Frecuencia absoluta)	Frecuencia Relativa	Medida del ángulo central
África	708	0.0125	45°
América	763	0.0135	49°
Europa	726	0.128	46°
Asia	3403	0.604	218°
Oceanía	28	0.005	2°
Total	5630		

Cuadro I.2 Distribución de la población mundial en el año 2000

Por ejemplo, para obtener la frecuencia relativa del continente Africano se operó de la siguiente manera:

$$Frecuencia\ Relativa\ de\ África = \frac{Número\ de\ habitantes\ de\ África}{total\ de\ habitantes} = \frac{708}{5630} = 0.0125$$

$$Medida\ del\ ángulo\ central = Frecuencia\ Relativa\ de\ África \times 360° = 0.0125 \times 360° = 45°$$



I.4. Datos Agrupados

Cuando se desea presentar la información obtenida en un censo o encuesta realizado a una población muy numerosa (personas, países, objetos, bacterias, etc.) y los valores de la variable estadística varían en un rango muy amplio, las gráficas o tablas de frecuencias no resultan muy adecuadas.

Por otra parte, la mayoría de los análisis estadísticos incluyen una gran cantidad de datos, los cuales sería casi imposible utilizar si no se les compactará mediante un sencillo procedimiento que es conocido como: *Tabla de Distribución de Frecuencias.*

Para poder obtener una distribución de frecuencias, primero hay que determinar el número de clases o categorías en que se agruparán los datos.

Para obtener las diferentes clases o categorías, en que se agruparán los datos, se debe de establecer lo que se conoce como escalas de medición.

Los tipos de escalas de medición son:

Nominal	Simplemente etiqueta a las clases en las que se agruparon los datos y no suele haber ninguna escala de medición aunque se considera una de estas.
Ordinal	Corresponden a una escala ordenada en la que de alguna manera estamos calificando las categorías de los datos. Por lo general esta escala de medición suele ser muy subjetiva.
Intervalos	Se utiliza cuando se analizan datos cuantitativos, en los cuales tiene sentido calcular diferencia entre valores.

Cuadro 1.3 Tipos de escalas de medición

Por ejemplo las edades se pueden agrupar en:

Clase o categoría	Intervalo	
	Límite inferior (años)	Límite superior (años)
Bebés	0	1
Niños	2	12
Adolescentes	12	16
Jóvenes	17	25
Adultos	25	60
Adultos mayores	61	-----

Esta es una escala *ordinal* porque ordena (por edad) las categorías de los datos. No se puede precisar cuántos años de diferencia hay entre un cierto niño y un adolescente.

Los alumnos según sus calificaciones se pueden clasificar en:

Clase o categoría	Intervalo	
	Límite inferior (Calificaciones)	Límite superior (calificaciones)
Muy buenos	9.1	10.0
Buenos	8.1	9.0

Regulares	7.1	8.0
Malos	6.0	7.0
Muy malos	0	5.9

Esta es una escala *ordinal* porque ordena los datos de acuerdo a la calificación.

Un tipo de medición de acuerdo a una escala *nominal* es:

Principales religiones que se practican en el mundo (% de la población mundial)	
Budismo	6 %
Cristianismo	34%
Islamismo	20%
Hinduismo	13%
Otras religiones	13%
Sin religión	14%

Total de la población 6000 millones.Fuente: National Geographic, agosto 1999.

Las escalas de medición por *intervalos* se usan con datos cuantitativos, son las más utilizadas en la estadística. El método consiste en establecer diferentes rangos de variación (intervalos) con la propiedad de que cualquiera de los datos tiene que pertenecer a uno de éstos.

Estos rangos de variación se establecen por lo general de acuerdo a una característica de los datos.

Las escalas de medición por intervalos se pueden clasificar de la siguiente manera:

	Características
Uniformes	Todos son del mismo tamaño
Desiguales	Algunos o todos los intervalos tienen diferente tamaño
Abiertos	Existe por lo menos un intervalo que no se puede determinar su tamaño

Cuadro 1.4 Clasificación de la escala de medición por intervalos

Antes de presentar ejemplos de esta escala de medición, es importante destacar la importancia de ordenar los datos.

"Los datos para una distribución de frecuencias pueden opcionalmente, ordenarse de menor a mayor (o de mayor a menor). Al hacer esto resultan más fáciles los cálculos subsecuentes, aunque la clasificación en sí representa un esfuerzo considerable. Por otra parte, el orden numérico o alfabético de los datos puede ser, en ocasiones, el único tratamiento que deba dárseles" [8]; por ejemplo, las listas de los alumnos de un grupo, si éstas se presentan en desorden son de poca utilidad para el profesor o para los administrativos que llevan los registros del grupo.

Una vez establecidos los conceptos necesarios, podemos regresar al problema de elaborar una tabla de distribución de frecuencias para datos cuantitativos.

Se elaborará una tabla de distribución de frecuencias, con una escala de medición de *intervalos uniformes*.

En una comunidad de profesionistas, se le pregunta el salario quincenal a 25 personas y las respuestas obtenidas son las siguientes:

$2,100	$9,000	$3,900	$ 900	$5,800
$4,500	$3,500	$4,850	$3,300	$7,100
$1,000	$2,500	$8,500	$2,900	$6,500
$1,350	$3,250	$2,300	$3,100	$3,400
$8,000	$4,800	$1,500	$2,400	$2,500

Obsérvese que el tamaño de la muestra es de 25 datos. N =25

Ordenando los datos de menor a mayor.

$ 900	$2,300	$3,100	$3,900	$6,500
$1,000	$2,400	$3,250	$4,500	$7,100
$1,350	$2,500	$3,300	$4,800	$8,000
$1,500	$2,500	$3,400	$4,850	$8,500

[8] Sánchez, Octavio. *Probabilidad y Estadística*. p. 37

$2,100	$2,900	$3,500	$5,800	$9,000

Es importante destacar que si los datos que se analizan son numéricos, el número de clases más común es de 5 a 12 clases para muestras que contengan menos de 125 datos, ya que esto permite segmentar muy bien la información, aunque puede ser menor o mayor como las circunstancias lo exijan.

Este mecanismo, para establecer el número de clases, puede cambiar a criterio de cada persona que trabajo con datos cuantitativos. Por ejemplo, Octavio Sánchez (1996) dice lo siguiente:

"En caso de que se desconozca, el número de clases puede determinarse en función de la raíz cuadrada del número de datos. Esta fórmula es conveniente si el número de datos es menor a 200 datos. A partir de ese número de datos, el número de clases resulta muy alto, por lo que podría utilizarse otra fórmula como la raíz cúbica del número de datos, o bien determinarlo en función del detalle que se desee."[9]

En este problema no se específica el número de clases para determinar intervalos uniformes por lo que, procederemos a utilizar el criterio de Octavio Sánchez (1996):

$$Número\ de\ clases = \sqrt{tamaño\ de\ \ la\ muestra}$$

$$Númerodeclases = \sqrt{n}$$

$$Númerodeclases = \sqrt{25}$$

[9] Ibid., p. 38

Lo que da por resultado 5 intervalos uniformes.

Determinemos la amplitud del intervalo correspondiente a cada clase:

$$Ancho\ de\ clase = \frac{Límite\ \sup erior\ de\ la\ muestra - Límite\ \inf erior\ de\ la\ muestra}{Número\ de\ clases}$$

$$Ancho\ de\ clase = \frac{9000 - 900}{5} = 1620$$

Se determinan los intervalos para cada clase con referencia a lo siguiente:

El límite inferior de la primera clase será el límite inferior de la muestra.

Los siguientes límites se calcularán de la siguiente manera:

Límite inferior de la clase = Límite inferior de la clase anterior + Ancho de clase

Por lo que:

Límite inferior de la clase A =900

Límite inferior de la clase B = 900 +1620 =2520

Límite inferior de la clase C =2520 + 1620 =4140

Límite inferior de la clase D = 4140 + 1620=5760

Límite inferior de la clase E = 5760 +1620 = 7380

Y para obtener el límite superior aplicamos la siguiente fórmula:

Límite superior de la clase = Limite inferior de la clase + Ancho de clase

Límite superior de la clase A = 900+1620=2520

Límite superior de la clase B = 2520 + 1620 =4140

Límite superior de la clase C = 4140 + 1620=5760

Límite superior de la clase D = 5760 +1620 = 7380

Límite superior de la clase E = 7380+1620=9000

Clase	Limite Inferior del intervalo	Limite superior del intervalo	Intervalo	Otra forma de representar el intervalo es
A	900	2520	900≤ x< 2520	[900,2520)
B	2520	4140	2520 ≤ x<4140	[2520,4140)
C	4140	5760	4140 ≤ x<5760	[4140,5760)
D	5760	7380	5760 ≤x <7380	[5760,7380)
E	7380	9000	7380≤x ≤9000	[7380,9000]

Cuando se construye una distribución de frecuencias hay que verificar que las clases no se traslapen, de modo que cada uno de los datos pertenezca exactamente a una de las clases. Se debe de incluir todos los casos, incluso aquellos que tengan una frecuencia cero. Se debe de procurar utilizar todos los intervalos con el mismo ancho de clase, aunque en muchas ocasiones, es imposible evitar los intervalos con finales abiertos, como en el ejemplo anterior, o como ocurre con en el Censo de 85 años o más.

*La **frecuencia absoluta** en datos agrupados se define como el número de elementos "x" que se encuentran dentro de cada clase.*

Frecuencia absoluta de la clase "A" = 9

Frecuencia absoluta de la clase "B" = 7

Frecuencia absoluta de la clase "C" = 3

Frecuencia absoluta de la clase "D" = 3

Frecuencia absoluta de la clase "E" = 3

En los datos agrupados, la suma de las frecuencias absolutas de cada clase es igual al tamaño de la muestra.

Observemos que la suma de las frecuencias absolutas para cada clase es:

9+7+3+3+3 =25

y sabemos que 25 es el número total de datos de la muestra o el tamaño de la muestra.

*La **frecuencia relativa de la clase "x"** es el resultado de dividir la frecuencia absoluta de la clase "x" entre el tamaño de la muestra; si se quiere reflejar en porcentaje sólo se multiplica cada frecuencia relativa por cien y se obtiene la **frecuencia relativa porcentual.***

Frecuencia relativa = Frecuencia Absoluta / Tamaño de = 9/25 =0.36
de la clase "A" de la clase A la muestra

Frecuencia relativa = Frecuencia Absoluta / Tamaño de = 7/25 =0.28
de la clase "B" de la clase B la muestra

Frecuencia relativa = Frecuencia Absoluta / Tamaño de = 3/25 =0.12
de la clase "C" de la clase C la muestra

Frecuencia relativa = Frecuencia Absoluta / Tamaño de = 3/25 =0.12
de la clase "D" de la clase D la muestra

Frecuencia relativa = Frecuencia Absoluta / Tamaño de = 3/25 =0.12
de la clase "E" de la clase E la muestra

Obsérvese que la suma de las frecuencias relativas es:

0.36+0.28+0.12+0.12+0.12 = 1.00

La suma de las frecuencias relativas es igual a 1 y, si son porcentuales, la suma de ellas da como resultado el 100%.

Frecuencia relativa porcentual = Frecuencia relativa X 100=0.36*100 =36%
de la clase "A" de la clase "A"

Frecuencia relativa porcentual = Frecuencia relativa X 100=0.28*100 =28%
de la clase "B" de la clase "B"

Frecuencia relativa porcentual = Frecuencia relativa X 100=0.12*100 =12%
de la clase “C” de la clase “C”

Frecuencia relativa porcentual = Frecuencia relativa X 100=0.12*100 =12%
de la clase “D” de la clase “D”

Frecuencia relativa porcentual = Frecuencia relativa X 100=0.12*100 =12%
de la clase “E” de la clase “E”

Si para una clase “x” sumamos todas las frecuencias absolutas menores o iguales al límite superior de la clase “x” se obtiene la ***frecuencia absoluta acumulada hasta x.***

Frecuencia absoluta = Frecuencia absoluta =09
acumulada de la clase “A” de la clase “A”

Frecuencia absoluta =Frecuencia absoluta +Frecuencia absoluta =9+7=16
acumulada de la clase “B” de la clase “B” acumulada de la clase “A”

Frecuencia absoluta= Frecuencia absoluta+ Frecuencia absoluta =16+3=19
Acumulada de la clase “C” de la clase “C” acumulada de la clase “B”

Frecuencia absoluta =Frecuencia absoluta+ Frecuencia absoluta =19+3=22
Acumulada de la clase “D” de la clase “D” acumulada de la clase “C”

Frecuencia absoluta = Frecuencia absoluta+ Frecuencia absoluta =22+3=25
Acumulada de la clase “E” de la clase “E” acumulada de la clase “D”

La ***frecuencia relativa acumulada hasta x*** *es el resultado de dividir la frecuencia absoluta de una clase “x” entre el tamaño de la muestra. En forma de porcentaje se multiplica la frecuencia relativa acumulada por cien y ésta se conoce como la* ***frecuencia relativa acumulada porcentual hasta x.***

Frecuencia relativa = Frecuencia Absoluta / Tamaño de la = 9/25 =0.36

acumulada de la clase "A" acumulada de la clase "A" muestra

Frecuencia relativa acumulada de la clase "B" = Frecuencia Absoluta acumulada de la clase "B" / Tamaño de la muestra = 16/25 =0.64

Frecuencia relativa acumulada de la clase "C" = Frecuencia Absoluta acumulada de la clase "C" / Tamaño de la muestra = 19/25 =0.76

Frecuencia relativa acumulada de la clase "D" = Frecuencia Absoluta acumulada de la clase "D" / Tamaño de la muestra = 22/25 =0.88

Frecuencia relativa acumulada de la clase "E" = Frecuencia Absoluta acumulada de la clase "E" / Tamaño de la muestra = 25/25 =1.00

La **frecuencia relativa porcentual** tiene el siguiente cálculo:

Frecuencia relativa % acumulada de la clase "A" = Frecuencia relativa acumulada de la clase "A" x 100 = 9/25 X 100 =36 %

Frecuencia relativa % acumulada de la clase "B" = Frecuencia relativa acumulada de la clase "B" X 100 = 16/25 X 100 =64 %

Frecuencia relativa % acumulada de la clase "C" = Frecuencia relativa acumulada de la clase "C" muestra X 100 = 19/25 X 100 =76 %

Frecuencia relativa % acumulada de la clase "D" = Frecuencia relativa acumulada de la clase "D" X 100 = 22/25 X 100 =88 %

Frecuencia relativa acumulada de la clase "E" = Frecuencia relativa acumulada de la clase "E" X 100 = 25/25 X 100 =100 %

*La **frecuencia absoluta complementaria** de la clase "x" se obtiene de restar la frecuencia absoluta acumulada de la clase "x" al tamaño de la muestra.*

Frecuencia absoluta complementaria de la clase "A" = Tamaño de la muestra – Frecuencia absoluta acumulada de la clase "A" =25-9 =16

Frecuencia absoluta complementaria de la clase "B" = Tamaño de la muestra – Frecuencia absoluta acumulada de la clase "B" =25-16 =9

Frecuencia absoluta complementaria de la clase "C" = Tamaño de la muestra – Frecuencia absoluta acumulada de la clase "C" =25-19 =6

Frecuencia absoluta = Tamaño de – Frecuencia absoluta =25-22 =3

complementaria de la la muestra acumulada de la clase “D”
clase “D”

Frecuencia absoluta = Tamaño de – Frecuencia absoluta =25-25 =0
complementaria de la la muestra acumulada de la clase “E”
clase “E”

*La **frecuencia relativa complementaria de una clase x** se define como la frecuencia absoluta complementaria entre el tamaño de la muestra o como la diferencia de uno menos la frecuencia relativa acumulada.*

Frecuencia relativa = 1 – Frecuencia relativa =1-(9/25) =16/25
complementaria de la acumulada de la clase “A”
clase “A”

Frecuencia relativa = 1 – Frecuencia relativa =1-(16/25) =9/25
complementaria de la acumulada de la clase “B”
clase “B”

Frecuencia relativa = 1 – Frecuencia relativa =1-(19/25) =6/25
complementaria de la acumulada de la clase “C”
clase “C”

Frecuencia relativa = 1 – Frecuencia relativa =1-(22/25) =3/25
complementaria de la acumulada de la clase “D”
clase “D”

Frecuencia relativa = 1 – Frecuencia relativa =1-(25/25) =0
complementaria de la acumulada de la clase “E”
clase “E”

*La **frecuencia relativa complementaria porcentual de una clase x** se obtiene al multiplicar la frecuencia relativa complementaria por 100.*

Frecuencia relativa = 100 x Frecuencia relativa =(16/25)*100= 64%
complementaria %de la complementaria de
clase “A” la clase “A”

Frecuencia relativa % = 100 x Frecuencia relativa =(9/25)*100= 36%
complementaria de la complementaria de
clase “B” la clase “B”

Frecuencia relativa % = 100 x Frecuencia relativa =(6/25)*100= 24%

complementaria de la clase “C” complementaria de la clase “C”

Frecuencia relativa % complementaria de la clase “D” = 100 x Frecuencia relativa complementaria de la clase “D” =(3/25)*100= 12%

Frecuencia relativa % complementaria de la clase “E” =100 x Frecuencia relativa complementaria de la clase “E” =(0/25)*100= 0%

Un elemento importante en la distribución de frecuencias es la **marca de clase**, ya que se puede tomar como un valor “representativo” de la clase en cuestión.

La marca de clase se define como el punto medio del intervalo de clase

$$\boxed{Marca\ de\ Clase = \frac{Lím.\ \mathrm{sup}.\ del\ \mathrm{int}ervalo\ de\ la\ clase\ en\ cuestión + Lím.\ \mathrm{inf}.\ de\ la\ clase\ encuestión}{2}}$$

$$\text{Marca de la clase A} = \frac{2520+900}{2} = 1710$$

$$\text{Marca de la clase B} = \frac{4140+2520}{2} = 3330$$

$$\text{Marca de la clase C} = \frac{5760+4140}{2} = 4950$$

$$\text{Marca de la clase D} = \frac{7380+5760}{2} = 6570$$

$$\text{Marca de la clase E} = \frac{9000+7380}{2} = 8190$$

Tabla de distribución de frecuencias para la muestra del ingreso quincenal de 25 personas.

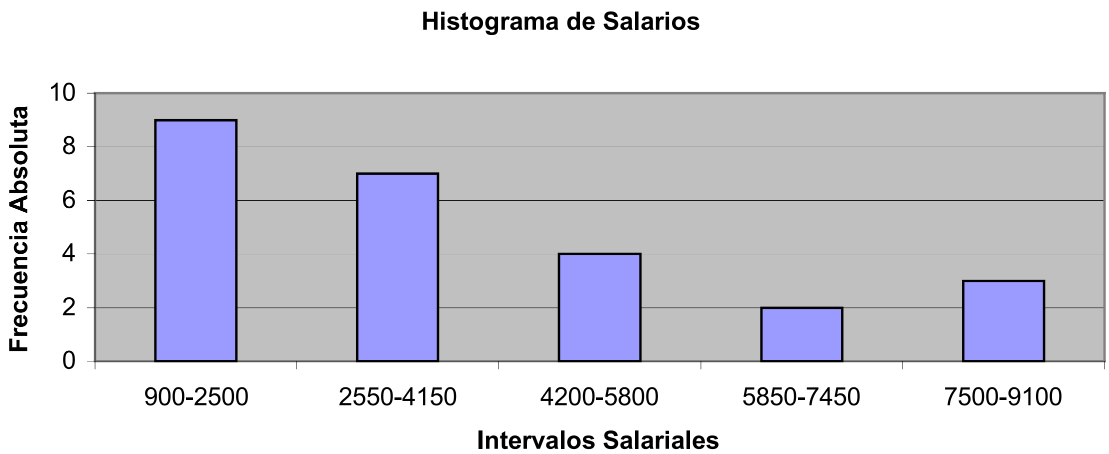
.		*Frecuencia Absoluta*	*Frecuencia Relativa*	*Frecuencia Relativa (%)*	*Frecuencia Absoluta Acumulada*	*Frecuencia Relativa Acumulada*	*Frecuencia Relativa Acumulada (%)*	*Frecuencia absoluta* Complementaria	Frecuencia Relativa Complementaria	Frecuencia Relativa Complementaria (%)	*Marca de clase*
A	[900,2520)	9	9/25	36%	9	9/25	36%	14	14/25	64%	1710
B	[2520,4140)	7	7/25	28%	16	16/25	64%	9	9/25	36%	3330
C	[4140,5760)	3	3/25	12%	19	19/25	76%	3	6/25	24%	4950
D	[5760,7380)	3	3/25	12%	22	22/25	88%	3	3/25	12%	6570
E	[7380,9000]	3	3/25	12%	25	25/25	100%	0	0	0	8190

Cuadro I.5 Distribución de salarios

En este cuadro se resumen todos los cálculos de las diferentes frecuencias

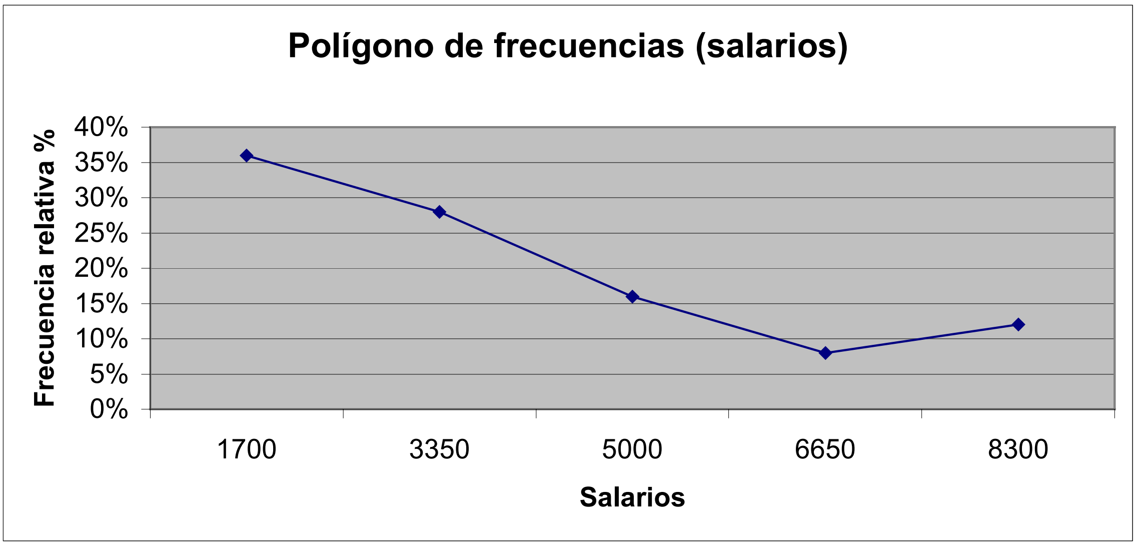
1.4.1. Representaciones Gráficas

Un histograma es una gráfica de barras para datos agrupados, éste representa, a escala, el número de elementos que comprende cada una de las clases de la distribución de frecuencias. El histograma se construye siguiendo el orden estándar: una línea horizontal (o eje “x”) a lo largo de la cual se marcan los valores de las variables, y una línea vertical (o eje “y”) que representa las frecuencias. Las marcas de clase se ordenan en el eje “x”. Obsérvese que entre más alta es la barra mayor es la frecuencia. Esta gráfica es ampliamente usada en la investigación social.



Por ejemplo en esta gráfica se observa que: 9 personas ganan entre 900 y 2500, ya que la altura de cada barra representa el número de personas que obtienen ingresos entre los límites determinados por el intervalo.

Otra gráfica representativa es el **polígono de frecuencias**. Esta gráfica puede acomodar una gran diversidad de variables, tiende a darle una continuidad a valores discretos, a lo largo de una escala, y por tanto es particularmente útil para representar escalas ordinales y de intervalos. Esto se debe a que las frecuencias se indican por medio de puntos colocados sobre la marca de cada clase. Los puntos adyacentes se conectan mediante una línea recta que cae sobre la línea base en uno y en otro extremo.

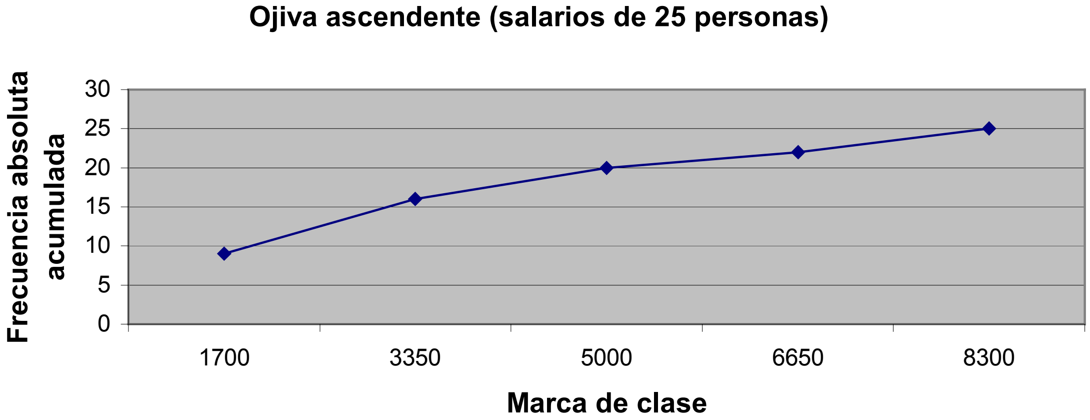


Dado que la frecuencia acumulada representa la suma de las frecuencias menores o iguales que una variable “x”; su gráfica permite analizar preguntas como: ¿Cuántos elementos de la población (o muestra) tienen menos de cierto nivel de la variable que se analiza? o ¿qué porcentaje de la frecuencia total corresponde a los que tienen menos de cierta cantidad?.

La representación gráfica de las frecuencias acumulada o complementaria recibe el nombre de ojiva; la ojiva que en el eje “y” representa las frecuencias acumuladas (absoluta o relativa) recibe el nombre de ojiva ascendente u ojiva “menos que”, ya que esta siempre nos dice un valor menor a la variable.

Como sabemos la frecuencia complementaria representa la diferencia del total de las frecuencias menos las frecuencias acumuladas, con ella podemos saber cuantos datos son mayores o iguales a la variable “x”.

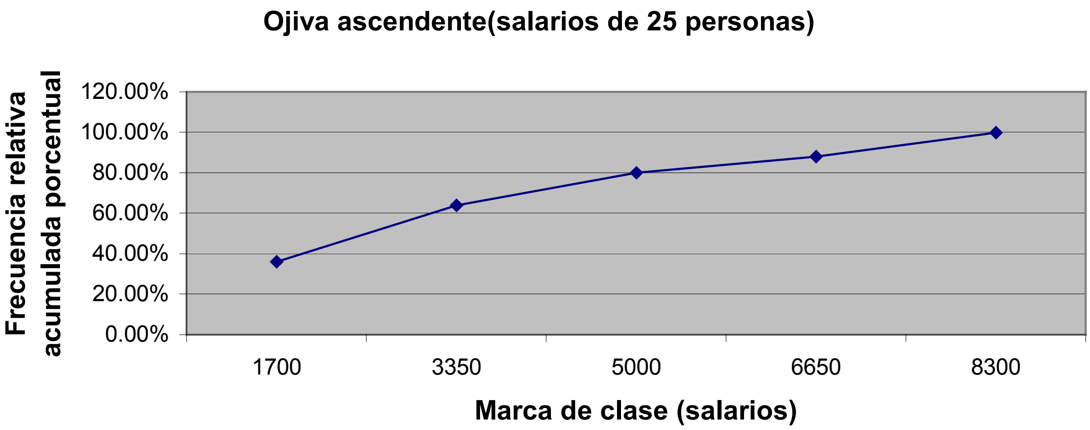
Ojiva (1)



En ésta gráfica podemos observar que 20 personas ganan menos de $ 5000 quincenales.

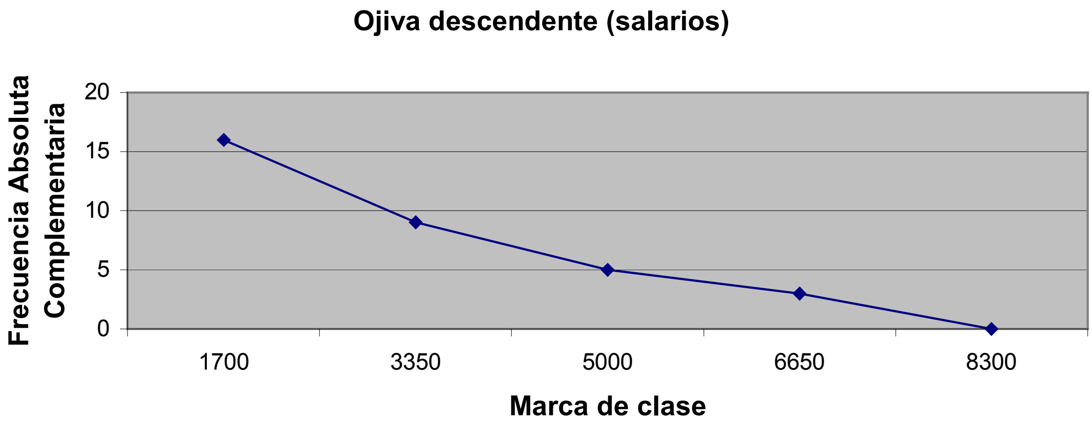
Ojiva (2)

En ésta gráfica se puede observar que el 80% de nuestra muestra gana menos de $5,000 quincenales.



Observemos que las ojivas (1) y (2) representan la misma curva y lo único que cambia en las gráficas es la representación de las frecuencias. La representación de las frecuencias va a depender de las necesidades de nuestro trabajo estadístico.

La gráfica de la frecuencia complementaria se llama **ojiva descendente** u **ojiva "más que"**, ya que permite saber ¿cuántos elementos tienen un valor mayor que cierto nivel de la variable que se analiza? O, ¿qué porcentaje de la frecuencia total corresponde a los que tienen más de cierta cantidad?



Al igual que la gráfica anterior se puede observar que 5 personas ganan más de $5,000

I.4.1.1.Gráfica de sectores

La gráfica de sectores para datos agrupados no presenta mayor problema ya que al igual que la de datos no agrupados, únicamente se multiplica la frecuencia relativa por 360° y este dato será el ángulo central.

Ejercicios Propuestos para este capitulo

Escribe en el paréntesis la letra que corresponda a la respuesta correcta.

( )	Rama de las matemáticas que trata de la recopilación, análisis, interpretación y presentación de una gran cantidad de datos numéricos.	k)	Conjunto que se analiza (muestra o población)
( )	Es un subconjunto de la población en estudio.	l)	Ojiva "menos que"
( )	Rama de la estadística que obtiene sus datos para obtener valores característicos que la describen y de estos se derivan los aspectos más relevantes.	m)	Estadística inferencial
( )	Son las variables que toman sus valores en los números enteros.(No. de personas, bacterias, etc.)	y)	Barras
( )	Son las variables que expresan una característica de tamaño u orden (chico, mediano, grande).	n)	Variables discretas
( )	La suma de las frecuencias absolutas es igual a:	j)	Muestra
( )	La representación grafica de la frecuencia acumulada se llama:	f)	Variables cualitativas jerarquizadas
( )	Es la gráfica que indica la proporción de datos en un círculo.	h)	Estadística
		w)	Media aritmética
		a)	Ojiva "más que"
		s)	Estadística descriptiva
		p)	Variables continuas
		x)	Variables cualitativas nominales.
		b)	Tamaño de la muestra
		z)	Uno
		e)	De sectores

De los siguientes datos:

65	35	26	39	95
67	97	26	75	35
46	67	24	31	57
38	46	18	36	62
26	35	19	85	37

De estos datos obtén un cuadro de frecuencias que contenga lo siguiente:

Clase	Lim. Inf.	Lim sup	Marca de clase	Frecuencia absoluta	Frecuencia acumulada	Frecuencia complementaria	Frecuencia relativa (absoluta%)

Grafica un histograma con frecuencias relativas

Gráfica una de sectores.

II. MEDIDAS DE TENDENCIA CENTRAL

II.1.Propiedades de la sumatoria (Σ)

Las sumas con frecuencia son utilizadas en la vida cotidiana, y por lo tanto, tienen gran relevancia en la Estadística debido a que siempre se trabaja con datos cuantitativos, que en muchas ocasiones queremos agrupar.

La sumatoria es un símbolo para expresar de manera abreviada una suma de números y ésta se representa con la letra griega Σ (sigma).

Supóngase que un alumno del bachillerato obtiene las siguientes calificaciones parciales y quiere conocer su calificación final del período.

Las calificaciones son: 8, 7.5, 5.8, 7.6 y 9.2. Para poder obtener su promedio debe de sumar todas las calificaciones parciales, es decir, $\sum_{i=1}^{5} x_i$ donde x_1 es la primera calificación obtenida; x_2 es la segunda calificación; y, así sucesivamente con las otras calificaciones. Cabe señalar que el símbolo $\sum_{i=1}^{5} x_i$ es una representación de la suma de estos datos. Así, la $\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5 = 8 + 7.5 + 5.8 + 7.6 + 9.2 = 38.1$. Dado este dato, podemos obtener el promedio como la suma de todos los datos entre el número de datos, de aquí que el promedio de este alumno es 38.1/5=7.62

Ahora, si se quisiera saber la suma de los primeros 10 naturales, es decir, del 1 al 10 se expresa de la siguiente manera:

$$\sum_{i=1}^{10} i = 1+2+3+4+5+6+7+8+9+10 = 55$$

La i se le conoce como el índice de la sumatoria, siempre es un número entero y en este ejemplo i=1 significa que este índice empieza en 1, pasando por dos, tres, y así sucesivamente hasta llegar a 10.

Hablando de sumatorias recordaremos una de las anécdotas más bellas de un matemático famoso en la historia de las matemáticas.

¿Eran pizarroncitos?
Der Kleine Friedrich
Esa tarde de invierno de 1786 hacía un frío que calaba los huesos. Afuera el viento hacía bailar las ramas desnudas de los árboles bajo un cielo de plomo. A lo mejor por eso los niños habían estado insoportables, al punto de que el viejo profesor de segundo grado de la escuela primaria de Braunschweig, aldea perdida en algún rincón de Alemania, decidió castigarlos.
Permanecerían ahí al terminar la clase; deberían sumar todos los números del uno al cien, y no podrían irse a sus casas mientras no terminaran.
Pusieron manos a la obra, contrariados por un castigo que se antojaba demasiado severo, pero deseosos de terminar cuanto antes. 1+2 =3, 3+3=6, 6+4=10...
Alguno, más vivillo, ya iba por el 9: 36+9=45. Ahí estaban todos, calladitos y con la cabeza gacha sobre el papel-¿o usarían pizarroncitos?- suma y vuelve a sumar. Todos menos Kleine Frederich, que así se llamaban los Federiquitos allá.
Al buen Friederich se le ocurrió -¿por qué? ¿de donde provienen las ocurrencias?- que no tenía que sumarlos obligatoriamente en orden. Así que sumó primero los extremos 1+100=101. Después sumo el segundo con el penúltimo: 2+99=101, y se dio cuenta -¿cómo? ¿de qué manera se da uno cuenta de esas cosas?- que si seguía sumando así, de uno en uno hacia delante del principio, y de uno en uno hacia atrás en la cola, siempre obtendría 101, hasta que llegara al 50+51. De tal suerte, tendría 50 sumas parciales con el mismo resultado y los habría sumado todos. De tal manera que nuestro Fefé dejó de sumar y multiplicó: 101 x 50 =5050. Esa multiplicación hasta él a sus siete años, se la sabía, y ese era el resultado.
Levanto la mano respetuoso y dijo al maestro que había terminado. Este enfurecido, lo increpó: ¡Cómo te atreves a tomarme el pelo, mentecato! ¡Por quién me tomas¡ No vio el resultado que además, ni debía de conocer; sólo se fijó en que el niño había garabateado apenas cuatro o cinco operaciones y su cólera creció. ¡Ahora sabrás quien soy yo! ¡Me vas a sumar del uno al mil!¡Aunque tengamos que pasar aquí toda la noche! ¡Te voy a enseñar yo a hacer trampas!.
Federico agacho la cara, roja de vergüenza, ante las risas burlonas de sus compañeros. El maestro bufando, retomó su lugar. Resulta sorprendente -pensó-, este niño siempre se porta bien. Es de familia muy humilde pero tiene buenos modales. Quién sabe que mosca le picó para pretender...Pero sus

pensamientos se vieron interrumpidos porque el pequeño Friederich levantaba una vez más la mano. Ya había terminado.¡Esto era demasiado!. Tanto, que se obligo a contenerse y a pedir explicaciones. Y las obtuvo ciertamente, no las que él esperaba. Ahí estaba escrito sin más ni más: 1+1000 =1001, 1000/2=500, 1001 x500 =500500. El razonamiento del pequeño era inobjetable. Acababa de ser establecida por primera vez, la fórmula de las progresiones aritméticas.
Der Kleine Friedcrich tomó sus útiles, se despidió cortés y salió a la ventisca. El maestro se sentó con la mirada atónita aún fija en el papel -¿o era pizarroncito?- con la cifra mágica 500500. Al niño se le olvido escribir su nombre, así que se le puso él: Gauss. Nunca sospechó -¿o sí?- a quién tenía entre sus pupilos, insoportables ese día.
La noche caía. El mayor matemático de la historia caminaba contra el viento, con pasitos cortos pero rápidos, la mochila al hombro y la cara tapada por la bufanda raída. Estaba preocupado; tendría que explicar a sus padres por qué llegaba tarde. En el salón otros niños continuaban en silencio:703 +38=741, 741+39=...Sólo se oía el rasgar de los lápices -¿o eran gises? [10]

Bueno después de esta bella anécdota de las matemáticas continuemos.

Si se tuviera la suma de los primeros 10 pares se representa así:

$$\sum_{i=1}^{10} 2i = 2+4+6+8+10+12+14+16+18+20$$

También se puede expresar como:

$$\sum_{i=1}^{10} 2i = 2(1+2+3+4+5+6+7+8+9+10)$$

el 2 es un factor común a todos los miembros de la suma y esto lo podemos escribir como:

$$2\sum_{i=1}^{10} i = 2(55) = 110$$

Por lo tanto:

[10] Perelló, Marcelino. "Eran pizarroncitos" . En: *Ciencia y Desarrollo*. Enero-Febrero de 2000. No. 150 p. 90

$$\sum_{i=1}^{10} 2i = 2\sum_{i=1}^{10} i$$

La suma de los primeros 10 pares es igual a 2 por la suma de los primeros 10 naturales.

De aquí surge una de las propiedades de la sumatoria:

Cuando en una suma, de cada uno de los sumandos se puede obtener un factor común, éste se puede factorizar y obtener el resultado de una constante o factor común multiplicado por la suma.

Existe una manera coloquial de referirse a esta propiedad: las constantes "salen" de la sumatoria.

La expresión formal de esta propiedad es:

$$\sum_{i=1}^{n} ci = c\sum_{i=1}^{n} i \qquad \text{c} = \text{constante}$$

Considera la siguiente suma:

$$\sum_{i=1}^{10} i = 1+2+3+4+5+6+7+8+9+10=55$$

Observe que:

$$\sum_{i=1}^{5} i = (1+2+3+4+5) = 15 \quad y \quad \sum_{i=6}^{10} i = (6+7+8+9+10) = 40$$

De modo que:

$$\sum_{I=1}^{10} i = 55 = 15 + 40 = \sum_{i=1}^{5} i + \sum_{i=6}^{10} i$$

Este ejemplo ilustra la siguiente propiedad de la sumatoria:

Una suma de n números se puede realizar como dos sumas parciales, una que se lleve desde 1 hasta m (donde m<n) y la siguiente desde m+1 hasta n.

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{m} x_i + \sum_{i=m+1}^{n} x_i \quad donde \ m \leq n$$

La tercera propiedad se refiere a sumar dos series de números. Identificando la primera serie como x_i y la segunda serie como y_i . A esta propiedad se le conoce como la *propiedad asociativa de la suma*.

Sea la serie x_i los primeros cinco números naturales 1,2,3,4 y 5. y y_i los primeros cinco números primos 2,3,5,7,11.

$$\sum_{i=1}^{5} x_i + \sum_{i=1}^{5} y_i = (1+2+3+4+5)+(2+3+5+7+11) = 15+28 = 43$$

$$pero \ 43 = (1+2)+(2+3)+(3+5)+(4+7)+(5+11) = 3+5+8+11+16 = \sum_{i=1}^{5}(x_i + y_i)$$

Esta propiedad establece que la suma de dos series de números del mismo tamaño puede realizarse sumando el primer dato de la primera serie con el primer dato de la segunda serie, el segundo dato de la primera serie con el segundo dato de la segunda serie y así sucesivamente, y después sumar todas estas sumas. La expresión de esta propiedad es:

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

La cuarta propiedad se deriva de la primera y la tercera .Esta dice que:

$$\sum_{i=1}^{n}(cx_i + ky_i) = \sum_{i=1}^{n} cx_i + \sum_{i=1}^{n} ky_i = c\sum_{i=1}^{n} x_i + k\sum_{i=1}^{n} y_i \qquad c, k = constantes$$

La primera igualdad se deduce de la propiedad 3 (que dice $\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (x_i + y_i)$) y la segunda igualdad se sigue de la propiedad uno (ya que las sumatorias "sacan" constantes: $\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$).

A esta propiedad se le conoce como la *propiedad distributiva de la suma*.

En resumen, las propiedades básicas del símbolo de sumatoria son las siguientes:

Propiedades de las sumatorias

$\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i \qquad \text{c = constante}$
$\sum_{i=1}^{n} x_i = \sum_{i=1}^{m} x_i + \sum_{i=m+1}^{n} x_i \qquad \text{donde } 1 \le m < n$
$\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (x_i + y_i)$
$\sum_{i=1}^{n}(cx_i + ky_i) = c\sum_{i=1}^{n} x_i + k\sum_{i=1}^{n} y_i \qquad c,\ k = constantes$

Cuadro II.1 Propiedades de la sumatoria

II.2. Medidas de Tendencia Central.

¿Cuántas veces no se ha utilizado el término “promedio”?, ¿cuál es tú “promedio” de secundaria?, o cuando vas al doctor te pregunta: ¿fumas?, ¿Cuántos cigarrillos al día te fumas en “promedio”?.

También, cuando vas al banco y abres una cuenta de cheques, se tiene que tener un saldo “promedio” mensual de $2,000, si no se cobra una comisión.

Una forma útil de describir un grupo de datos es encontrar un número que lo represente. Esos valores que tratan de describir un grupo de datos se les conoce como medidas de tendencia central.

Las medidas de tendencia central son las siguientes:

Media aritmética o promedio

Mediana

Moda

Rango medio

Promedio ponderado

Media Geométrica

Media Armónica

Fractiles (cuartiles, deciles y percentiles)

II.2.1. Medidas de Tendencia Central para Datos Individuales.

Las medidas de tendencia central se pueden trabajar para datos individuales o para datos agrupados. Lo que se va a tratar en esta sección es el manejo de las medidas de tendencia central para datos individuales.

¿Cuántas veces no se ha obtenido un dato "promedio" de una manera sencilla? Por ejemplo cuando un profesor de licenciatura quiere saber la edad "promedio" del grupo, haciendo sólo dos preguntas ¿Qué edad tiene el alumno más grande? y ¿Qué edad tiene el alumno más chico? Supóngase que el alumno más grande tiene 22 años y el más chico 18; sabiendo esto el profesor lo que hace para obtener un cálculo aproximado de la edad promedio del grupo es sumar la edad del estudiante más grande más la edad del estudiante más chico y dividirla entre dos. Por lo que el profesor calcula en su cerebro (22 +18)/2 = 40/2=20. Por lo tanto el profesor concluye que tiene un grupo con una edad "promedio" de 20 años. De aquí surge la primera medida de tendencia central que es el *rango medio*.

El **rango medio** es la medida de tendencia central muy fácil de obtener, ya que sólo se obtiene con la diferencia del límite superior menos el límite inferior de la muestra entre dos.

$$Rango\ medio = \frac{Limite\ superior\ de\ la\ muestra\ \text{-}\ límite\ inferior\ de\ la\ muestra}{2} = \frac{U - L}{2}$$

ó

$$Rango \text{ medio} = \frac{\text{Limite superior de la población - límite inferior de la población}}{2} = \frac{U - L}{2}$$

II.2.2. La Moda

Un muchacho quiere regalarle a su novia un disco de música, pero no sabe qué tipo de música o qué disco escoger para darle la sorpresa a su novia. Por lo que, se le ocurrió preguntarles a sus amigas para ver cuál era el disco de su preferencia, al encuestarlas obtuvo las siguientes respuestas:

Flores de Alquiler de la Quinta Estación.

Amar sin mentiras de Marc. Anthony.

Metamorphosis de Hilary Duff.

Elephunk de Black eyed Peace.

Encorn de Eminem.

Ya obteniendo esas 5 opciones empezó a preguntar a sus conocidos ¿cuál de estos cinco discos era su favorito? Y obtuvo las siguientes respuestas:

Nombre del Disco	Cantante	Votos
Flores de Alquiler	Quinta Estación	IIIII IIIII IIIII IIIII III
Amar sin mentiras	Marc. Anthony	IIIII IIIII IIIII III
Metamorphosis	Hilary Duff.	IIIII IIIII IIIII IIIII IIIII IIIII III
Elephunk	Black eyed Peace	IIIII IIIII IIIII IIIII
Encorn	Eminem.	IIIII IIIII IIIIII IIIIII IIIII IIIII IIIII IIIII III

Al contabilizar sus votos observó:

Nombre del Disco	Cantante	Votos
Flores de Alquiler	Quinta Estación	23
Amar sin mentiras	Marc. Anthony	17
Metamorphosis	Hilary Duff.	33
Elephunk	Black eyed Peace	20
Encorn	Eminem.	43

Entonces dedujo que a su novia le regalaría el disco Encorn de Eminem, porque es el que esta de “moda” y es el favorito de más personas.

Por ello, la **moda,** en Estadística, como en la vida cotidiana, es lo que se usa con más frecuencia. *La moda en estadística es el valor más frecuente de un conjunto de datos*. La moda se suele calcular fácilmente por inspección más que por medio de un cálculo.

Algunos conjuntos de datos pueden tener dos a más modas. Es decir, dos o más datos tienen la misma frecuencia en nuestra muestra. Cuando sucede esto, si existen dos datos con la misma frecuencia, se dice que el conjunto de datos es bimodal, si tienen tres modas es trimodal, si tiene más de tres modas se le conoce como multimodal.

¿Cuántas ocasiones se ha mencionado que un artículo de arreglo personal, está de moda?, por ejemplo: la moda es tener muchas perforaciones en la cara, o los pantalones de campana están de moda, etc.

Ahora, calcúlese la moda del siguiente grupo de datos:

Datos	**Frecuencia absoluta**
2	**5**
4	**8**
7	**3**
9	**8**
10	**3**

El dato 4 y el dato 9 tienen mayor frecuencia, lo cual indica que las modas son 4 y 9; por lo tanto, ésta muestra es bimodal.

Para obtener la(s) moda(s) sólo se requiere de un conteo de frecuencias, suele aplicarse cuando nuestras variables son nominales, es decir, sólo se caracterizan por el nombre, como la religión más practicada en el mundo o en las elecciones sólo se requiere contar los votos para obtener el partido político ganador.

II.2.3.Media Aritmética o promedio

¿Cuántas veces se ha obtenido el promedio?, por ejemplo: ¿Cuánto se gasta en promedio diario en la alimentación de tu casa?

Para ello, la (el) encargada(o) de la compras en tu casa, debe de anotar lo que gasta según el día de la semana. El lunes, compra la carne para toda la semana, gasta \$300; el martes, como es día de tianguis compra la fruta y la verdura gasta \$150; el miércoles, compra las carnes frías para preparar el lunch de los niños, gasta \$100; el jueves, compra los quesos y la crema, por que por su casa pasa un señor vendiendo productos oaxaqueños y paga \$150; el viernes, va a surtir la despensa al centro comercial, donde gasta \$400; el sábado, por lo general, comen en la calle y en promedio se gasta \$250; y, el domingo, siempre se come con la familia donde se ahorra el gasto.

Ahora si calculemos el gasto promedio diario de esta familia.

Para ello, sumamos lo que se gastó en toda la semana y lo dividimos por siete y de ahí, se obtiene el gasto promedio diario.

$$\frac{300+150+100+150+400+250+0}{7}=\frac{1350}{7}=192.57$$

Así, esta familia tiene un gasto promedio de \$192.57 diarios.

La media aritmética o promedio, es la medida de tendencia central más utilizada. La media aritmética se expresa con la letra griega μ ó con una x con una barra horizontal en la parte superior.

La media aritmética es entonces la suma de los valores de los datos entre el tamaño de la muestra o población.

$$\text{Media aritmética} = \bar{x} = \mu = \sum_{i=1}^{n} \frac{x_i}{n}$$

Calcúlese la media del siguiente grupo de datos, que representa la edad de un grupo de 10 personas:

16, 18, 20, 24, 30, 29, 15, 38, 27, 20.

La media es:

$$\mu = \frac{16+18+20+24+30+29+15+38+27+20}{10} = \frac{237}{10} = 23.7$$

Características de la media:

1. La suma de las diferencias de cada valor con respecto a la media es igual a cero.

$$\sum_{i=1}^{N}(x_i - \mu) = 0$$

Obsérvese que esto se verifica en el ejemplo anterior, donde el valor de la media fue 23.7

(16-23.7)+(18-23.7)+(20-23.7)+(24-23.7)+(30-23.7)+(29-23.7)+(15-23.7)

+(38-23.7)+(27-23.7)+(20-23.7)=-7.7-5.7-3.7+0.3+6.3+5.3-8.7+14.3+3.3-3.7=0

2. La media puede utilizarse para determinar la suma de los datos del total de la población.

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

Es decir: $\mu \cdot n = (23.7)(10) = 237 = \sum_{i=1}^{n} x_i$

3. La media se afecta sustancialmente si al menos uno de los datos se cambia por un valor, “mucho más grande o más pequeño respecto que la mayoría de los datos”, en el ejemplo de las edades, en lugar de tener una persona de 16 tenemos una de 61, que es un dato sustancialmente mayor que el dato más grande de la muestra original, así, la media sería:

$$\mu = \frac{61+18+20+24+30+29+15+38+27+20}{10} = \frac{282}{10} = 28.2$$

II.2.4. La mediana

Un grupo de estudiantes del bachillerato tenía una tarea de morfología para determinar si todas las compañeras de su colegio padecían bulimia o anorexia. Se escogió una muestra de 25 compañeras de 15 años, con una estatura promedio de 1.55 mts. Los resultados con respecto al peso de sus compañeras fueron los siguientes:

49,52,36,45,51,40,53,47,50,52,50,49,46,33,51,54,43,38,50,53,48,54,35,49, 48

Estos datos se ordenan de menor a mayor para poder trabajar mejor con ellos.

33	43	48	50	52
35	45	49	50	53
36	46	49	51	53
38	47	49	51	54
40	48	50	51	54

Una vez obtenidos los datos calcularon la media aritmética y obtuvieron el siguiente resultado:

$$\mu = \frac{\begin{array}{l}33+35+36+38+40+43+45+46+47+48+48+49+49+49+50+50+50+51+51\\+51+52+53+54+54\end{array}}{25}$$

$$\mu = \frac{1175}{25} = 47$$

Los estudiantes quieren comparar el dato obtenido, con sus notas de morfología, donde se especifica que el peso promedio de una persona de 15 años con una estatura de 1.55 es de 49 a 51 kilogramos.

Con este dato podrían inferir que todas las compañeras de su colegio padecían alguna de estas enfermedades.

Los estudiantes de la clase de morfología vieron que algo raro estaba pasando y se preguntaron: ¿por qué?, la respuesta se dedujo de manera inmediata al analizar los datos, debido a que en estos se encontraban dos personas con un peso muy pequeño. Otro estudiante sugirió que en lugar de tomar el promedio de los datos se tomara el peso que se encontraba exactamente en medio de todos los datos. Para lo que aplicó la siguiente fórmula:

$$\boxed{\frac{\text{tamaño del conjunto de datos que se analiza} + 1}{2} = \frac{n+1}{2} = \frac{25+1}{2} = \frac{26}{2} = 13}$$

Al ubicar el dato número 13, se da cuenta que es 49Kg. y este dato concordaba con lo escrito en sus apuntes. Al consultar con su profesor de Estadística les dijo que habían hecho lo correcto, puesto que el peso “adecuado” de una persona (que tiene 15 años y mide 1.55mts) se obtiene a través de la mediana y no de la media aritmética, porque el peso promedio de un grupo de personas calculado a través de la media aritmética puede verse afectado seriamente si se encuentra una persona muy obesa o una persona muy delgada. Por esta razón, las tablas de estatura – peso se calculan a través de la mediana.

La mediana es la medida de tendencia central que determina el elemento que ocupa la posición central de los datos individuales y son ordenados de menor a mayor (o viceversa), marca la mitad de los valores mayores que y la mitad de valores menores que el elemento que ocupa la posición central, es decir, está a la mitad, con el 50% de valores a su derecha y el 50% de valores a la izquierda.

El cálculo de la mediana se obtiene según el número de datos de la muestra. Si el número de datos es un número impar o par.

Para calcular la mediana en el caso de que haya un número impar de datos se necesita:

a) Ordenar los datos de menor a mayor o viceversa.

b) Calcular la posición de la mediana.

$$\boxed{\text{Posición de la mediana} = \frac{\text{tamaño de la muestra} + 1}{2} = \frac{n+1}{2}}$$

c) Determinar el dato de la posición central.

Nótese que cuando el tamaño de la muestra es impar, la posición de la mediana es única.

En el caso de que el número de datos sea un número par:

a) Se requiere ordenar los datos de menor a mayor o viceversa.

b) Se sabe que cualquier número par puede ser representado como 2k donde k es un número natural. Por lo que, la mediana se calcula como promedio de los valores que ocupan las posiciones intermedias (k y (k+1)).

Cualquier observación escogida al azar puede ser mayor o menor que la mediana. El valor de la mediana sólo se ve afectado por el número de datos, no por la magnitud de estos.

Dado que la mediana no se ve afectada por la magnitud de los datos, es recomendable utilizarla cuando nuestros datos contengan valores muy grandes o muy pequeños.

Ejemplo: Determínese la mediana de los siguientes datos:

16,29,30,49,47,38,56,73,47,29,10,08,27,50,57

Obsérvese que el tamaño de la muestra es 15 (número impar):

1. Ordenemos nuestra muestra de menor a mayor:

   08,10,16,27,29,29,30,**38**,47,47,49,50,56,57,73.

2. Calculemos la posición de la mediana.

$$Posición\ de\ la\ mediana = \frac{n+1}{2} = \frac{15+1}{2} = \frac{16}{2} = 8$$

La mediana ocupa el octavo lugar en los datos ordenados. La mediana es 38.

Obsérvese, que si el tamaño de la muestra es impar, la mediana coincide con uno de los datos.

Si en la muestra, en lugar de tener 15 elementos se le añade uno más quedaría:

16,29,30,49,47,38,56,73,47,29,10,08,27,50,57,34

Ordénese la muestra de menor a mayor:

08,10,16,27,29,29,30,**34,38**,47,47,49,50,56,57,73.

Obsérvese que el tamaño de la muestra es16 (número par). En este caso se puede representar a 16 como 2(8)=2k, lo que implica que k=8; entonces la mediana es el promedio del octavo y el noveno lugar en los datos ordenados. La mediana es el promedio de 34 y 38. Sería:

$$\frac{34+38}{2}=\frac{72}{2}=36$$

La mediana en este grupo de datos es 36.

Si el número de datos es par, la mediana no necesariamente coincide con uno de los datos.

Cabe señalar, que la mediana, es utilizada en Estadística para la obtención de talla y peso promedio; de edad promedio de vida en mujeres y en hombres; promedio de vida de otro ser vivo; para calcular el "promedio de vida" de una lavadora o de cualquier artefacto eléctrico, etc.

II.2.5. Promedio ponderado

En muchas ocasiones se ha experimentado la necesidad de no sólo considerar el promedio de los datos, sino que, a cada valor se le da un cierto "peso". Por ejemplo:

Cuando se llega por primera vez a una clase, el profesor da los métodos de la forma de calificar, por ejemplo:

Exámenes	60%
Asistencia y puntualidad	10%
Participación	10%
Tareas	20%
Total	100%

En el primer período, la calificación del examen de un alumno fue 6. Dado que el alumno nunca faltó, pero en ocasiones llegaba tarde, el maestro le asignó un 9, por no entregar una tarea obtuvo 8 y como es muy tímido no participó activamente en clase, por lo que, se le asignó un 7, ¿cuál fue el promedio de este alumno en este período?

El alumno al obtener su promedio hizo lo siguiente:

$$\frac{6+9+8+7}{4}=\frac{30}{4}=7.5$$

El profesor le dijo: tu calificación es 6.8 y no 7.5, pero el alumno quedó desconcertado y al ver su cara el profesor le explicó, en este caso no sirve sumar todas las calificaciones y dividirlas entre el total, debido a que:

$$\frac{6+9+8+7}{4}=\frac{1}{4}(6)+\frac{1}{4}(9)+\frac{1}{4}(8)+\frac{1}{4}(7)=0.25(6)+0.25(9)+(0.25)8+0.25(7)=7.5$$

A todas las calificaciones les estas dando un mismo "peso" y cada calificación tiene un "peso" diferente.

Para obtener tu calificación tienes que hacer lo siguiente:

Multiplica la calificación de tu examen por el decimal que representa el porcentaje asignado en este rubro, es decir, 6(0.6)=4.9, después suma la calificación de tu asistencia y multiplica por el factor de ponderación (porcentaje asignado) 9(0.1) =0.9 además debe sumar la calificación de tus tareas y multiplicar por el factor de ponderación que es 8(0.2)=1.6. y por último, debes sumar la calificación de tu participación y multiplicarlo por su factor de ponderación 7(0.1)=0.7 Finalmente tu calificación es $6(0.6)+9(0.1)+8(0.2)+7(0.1)=3.6+0.9+1.6+0.7=6.8$

$$\text{Promedio ponderado} = \textit{Suma cada valor } (x_i) \times \textit{factor de ponderación}$$

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

Es importante hacer notar que la suma de todos los factores de ponderación debe ser igual a 1 o si estos están dados en porcentajes debe ser el 100%.

Otro ejemplo es:

Un empresario tiene 4 empresas y desea premiar a los gerentes de cada empresa por su desempeño. Solicitó a cada uno de los gerentes un reporte de ventas y de utilidad anual descrito en el siguiente cuadro:

Empresa	Ventas	Utilidad Neta
A	585,736	27,687
B	769,864	72,713
C	675,890	132,567
D	1,567,896	354,879
Total	3,599,386	587,846

Al ver sus utilidades decidió repartir $100,000 de éstas a sus gerentes. En un principio pensó en repartir la cantidad en partes iguales, pero al razonarlo vió que no era justo, debido a que, todas las empresas tenían volúmenes de ventas diferentes y por lo tanto, sus utilidades eran distintas.

El empresario hizo un porcentaje de las utilidades con respecto a las ventas de cada empresa, bajo la siguiente fórmula:

$$Porcentaje\ de\ utilidad\ de\ cada\ empresa = \frac{Utilidad\ neta\ de\ cada\ empresa}{total\ de\ ventas\ de\ cada\ empresa} \times 100$$

Empresa	Ventas	Utilidad	
		Neta	Porcentaje
A	585,736	27,687	4.72%
B	769,864	175,346	22.77%
C	675,890	132,567	19.61%
D	1,567,896	354,879	22.63%
Total	3,599,386	690,479	19.18%

Finalmente, decidió repartir el dinero de la siguiente manera:

Al gerente que obtuviera el primer lugar en utilidades con respecto a las ventas le correspondería el 40% del bono.

Al gerente que obtuviera el segundo lugar le correspondería el 30% del bono.

Al gerente que obtuviera el tercer lugar en utilidades le correspondería el 20% del bono.

Al gerente que obtuviera el cuarto lugar le correspondería el 10% del bono.

Por lo tanto:

Al gerente de la empresa B le correspondió 100,000(0.4)=40,000.

Al gerente de la empresa D le correspondió 100,000(0.3) =30,000.

Al gerente de la empresa C le correspondió 100,000(0.2) =20,000.

Al gerente de la empresa A le correspondió 100,000(0.1)=10,000.

II.2.6. Media Geométrica.

Para comprender la media geométrica se va hacer referencia a la leyenda del tablero de ajedrez.

El ajedrez es un juego antiquísimo. Cuenta muchos siglos de existencia y por eso no es de extrañarse que estén ligadas a él diferentes leyendas, cuya veracidad es difícil de comprobar debido a su antigüedad. Precisamente quiero contar una de esas leyendas. Para comprenderla no hace falta saber jugar ajedrez; basta simplemente saber que el tablero donde se juega está dividido en 64 escaques (casillas negras y blancas, dispuestas alternativamente).

1.
El juego de ajedrez fue inventado en la India. Cuando el rey hindú Sheram lo conoció, quedo maravillado de lo ingenioso que era y de la variedad de posiciones que en él son posibles. Al enterarse que el inventor era uno de sus súbditos, el rey lo mando llamar con objeto de recompensarle personalmente por su acertado invento.
El inventor, llamado Seta, se presentó ante el soberano. Era un sabio vestido con modestia que vivía gracias a los medios que le proporcionaban sus discípulos.
__Seta, quiero recompensarte dignamente por el ingenioso juego que has inventado__dijo el rey.
El sabio contesto con una inclinación.
__Soy bastante rico como para poder cumplir tu deseo más elevado__continuó diciendo el rey__. Di la recompensa que te satisfaga y la recibirás.
Seta continuó callado.
__No seas tímido__le animó el rey__.Expresa tú deseo. No escatimaré en nada para satisfacerlo.
__Grande es tu magnanimidad, soberano. Pero concédeme un corto plazo para meditar la respuesta. Mañana, tras maduras reflexiones, te comunicaré mi petición.
Cuando al día siguiente Seta se presentó de nuevo ante el trono, dejo maravillado a rey con su petición, sin precedente por su modestia.
__Soberano__dijo Seta__, manda que me entreguen un grano de trigo por la primera casilla del tablero de ajedrez.
__¿Un simple grano de trigo?__contestó admirado el rey.

__Sí, soberano. Por la segunda casilla, ordena que me den dos granos; por la tercera casilla, 4; por la cuarta, 8; por la quinta, 16; por la sexta,32…
__Basta__le interrumpió irritado el rey__.
Recibirás el trigo correspondiente a las 64 casillas del tablero de acuerdo con tu deseo; por cada casilla doble cantidad por la precedente. Pero has de saber que tu petición es indigna de mi generosidad. Al pedirme tan mísera recompensa, menosprecias, irreverente, mi benevolencia. En verdad que, como sabio que eres, deberías haber dado mayor prueba de respeto ante la bondad de tu soberano. Retírate. Mis servidores te sacarán con el trigo que solicitas.
Seta sonrió, abandonó la sala y quedo esperando a la puerta del palacio.

2.
Durante la comida, el rey se acordó del inventor de ajedrez y envió para se enteraran si habían entregado al irreflexivo Seta su mezquina recompensa,
__Soberano, tu orden se está cumpliendo__fue la respuesta__. Los matemáticos de la corte calculan el número de granos que le corresponde.
El rey frunció el ceño. No estaba acostumbrado a que tardarán tanto en cumplir sus órdenes.
Por la noche al retirarse a descansar, el rey preguntó de nuevo cuánto tiempo hacia que Seta había abandonado el palacio con su saco de trigo.
__Soberano__le contestaron__, tus matemáticos trabajan sin descanso y esperan terminar los cálculos al amanecer.
__¿Por qué van tan despacio en este asunto?__grito iracundo el rey.__Que mañana antes de que me despierte, hayan entregado a Seta hasta el último grano de trigo. No acostumbro a dar dos veces una misma orden.
Por la mañana comunicaron al rey que el matemático mayor de la corte solicitaba audiencia para presentarle un informe muy importante.
El rey mando que se le hiciera entrar.
__Antes de comenzar tu informe__le dijo Sheram__, quiero saber si se ha entregado por fin a Seta la mísera recompensa que ha solicitado.
__Precisamente para eso me he atrevido a presentarme tan temprano__contestó el anciano__. Hemos calculado escrupulosamente la cantidad total de granos que desea recibir Seta. Resulta una cifra enorme…
__Sea cual fuere su magnitud__le interrumpió con altivez el rey__mis graneros no empobrecerán. He prometido darle esa recompensa y, por lo tanto, hay que entregársela.
__Soberano, no depende de tu voluntad el cumplir semejante deseo: En todos tus graneros no existe tal cantidad de trigo que exige Seta. Tampoco existe en los graneros del todo el reino. Hasta los graneros del mundo son insuficientes. Si deseas entregar sin falta la recompensa prometida, ordena que todos los reinos de la Tierra se conviertan en labrantíos, manda a desecar los mares y océanos, ordena fundir el hielo y la nieve de los lejanos desiertos del Norte. Que todo el espacio se totalmente sembrado de trigo, y toda la cosecha obtenida en estos campos ordena que se entregada a Seta. Sólo entonces recibirá su recompensa.
El rey escuchaba lleno de asombro las palabras del anciano sabio.

__Dime, cuál es esa tan monstruosa__dijo reflexionando.
__¡Oh soberano! Diez y ocho trillones cuatrocientos cuarenta y seis mil setecientos nueve millones quinientos cuarenta y un mil seiscientos quince.
3.- Esta es la leyenda. No podemos asegurar que haya sucedido en realidad lo que hemos contado; sin embargo, la recompensa de que habla la leyenda debe expresarse por ese número; de ello pueden convencerse, haciendo ustedes mismos el cálculo. Si se comienza por la unidad, hay que sumar las siguientes cifras: 1, 2, 4, 8, etc. El resultado obtenido tras 63 duplicaciones sucesivas nos mostrará la cantidad correspondiente a la casilla 64, que deberá recibir el inventor. Operando se tiene que:
1+2+4+8+16+64+….
No es difícil observar que:

1=1
2=1+1
4=(1+2)+1
8=(1+2+4)+1
16=(1+2+4+8)+1
32= (1+2+4+6+8+16)+1, etc.

Se observa que cada uno de los números de esta serie es igual al conjunto de todos los anteriores sumados más una unidad.. De aquí podemos observar el número total de granos. Pero también se puede observar que el cociente entre 64 y 32 es dos, y de 128/64 =2 y así sucesivamente entre el término x_i de la sucesión y el termino siguiente x_{i+1} su cociente es dos. Es decir el cálculo se reduce simplemente a multiplicar 64 veces seguidas la cifra dos.
Para hacernos una idea de la inmensidad de está cifra tan grande, calculemos aproximadamente la magnitud que debería tener un granero capaz de almacenar tal cantidad de trigo. Es sabido que un metro cúbico de trigo contiene cerca de 15 millones de granos. En ese caso la recompensa del inventor de ajedrez debería ocupar 12 000km^3 . Si el granero tuviera 4 m. de alto y 10 m. De ancho, su longitud debería ser de 300 000 000 Km, o sea, el doble de la distancia que separa la Tierra del Sol.[11]
A partir de esta leyenda es importante destacar que el comportamiento geométrico de una sucesión de números, se da cuando entre dos valores consecutivos, (llamémosles $x_i \ y \ x_{i+1}$) el cociente de x_{i+1} y x_i es una valor constante (k). Es decir, una sucesión geométrica tiene la característica que:

$$\frac{x_{i+1}}{x_i} = K$$

[11] Perelman Y. "*Matemáticas recreativas*" p.119 -125.

La media geométrica es un valor promedio de los datos que suele ser muy útil para datos que tienen un comportamiento geométrico, como es el caso de los intereses generados en una tarjeta de crédito, el crecimiento de la población de humanos o de algunas bacterias, etc.

Pongamos por caso: Una persona invierte \$50,000 en una cuenta de cheques a un interés del 3% mensual. Si la persona no hizo ningún movimiento en su cuenta durante un año: a) ¿Cuánto dinero tiene al finalizar el año?, b) ¿Cuál es la tasa de interés anual?

La mayoría de las personas pensarían que la segunda pregunta es más fácil de responder, puesto que sólo se multiplica el 3% por los meses del año (12) lo que da el resultado 3%(12)=36% de interés anual.

Con este resultado, obtienen los intereses generados durante el año multiplicando la tasa anual de interés por su capital inicial ((50,000)(0.36)=18,000) que da por resultado \$18,000 de intereses durante al año. Finalmente, se suma esta cantidad al capital inicial y se obtiene \$68,000. (50,000+18,000=68,000).

Sin duda, este sería un procedimiento muy fácil, aunque no sea el correcto. Para comprobar lo que se dice, analícese que pasa mes a mes con la inversión.

1° mes.

$$\underset{\text{Capital Inicial}}{\underset{\uparrow}{50{,}000}} \;\underset{+}{+}\; \underset{\text{intereses generados del 1° mes}}{\underset{\uparrow}{(50{,}000\times 0.03)}} = \underbrace{50{,}000(1+0.03)}_{\text{Factorizando 50,000}} = 51{,}500$$

2° mes.

$$51,500 + (51,500\times 0.03) = \underbrace{51,500(1+0.03)}_{\text{Factorizando } 51,500} = 53,045$$

Capital Final del 1° mes + intereses generados del 2° mes

Nota: Hay que considerar que el capital que se vuelve a invertir no es el inicial, si no el capital generado al final de l mes.

3° mes.

$$53,045 + (53,045\times 0.03) = \underbrace{53,045(1+0.03)}_{\text{Factorizando } 53,045} = 54,636.35$$

Capital Final del 2° mes + intereses generados del 3° mes

4° mes.

$$54,636.35 + (54,636.35\times 0.03) = \underbrace{54,636.35(1+0.03)}_{\text{Factorizando } 54,636.35} = 56,275.4405$$

Capital Final del 3° mes + intereses generados del 4° mes

5° mes.

$$56,275.4405 + (56,275.4405\times 0.03) = \underbrace{56,275.4405\ (1+0.03)}_{\text{Factorizando } 56,275.4405} = 57,963.70372$$

Capital Final del 4° mes + intereses generados del 5° mes

6° mes.

$$57{,}963.7038+(57{,}963.7038\times 0.03) = \underbrace{57{,}963.70372\ (1+0.03)}_{\text{Factorizando } 57{,}963.7038}=59{,}702.6149$$

Capital Final del 5° mes + intereses generados del 6° mes

7° mes.

$$59{,}702.6149+(59{,}702.6149\times 0.03) = \underbrace{59{,}702.6149\ (1+0.03)}_{\text{Factorizando } 59{,}702.6149}=61{,}493.6933$$

Capital Final del 6° mes + intereses generados del 7° mes

8° mes.

$$61{,}493.6933+(61{,}493.6933\times 0.03)=\underbrace{61{,}493.6933+(1+ 0.03)}_{\text{Factorizando } 61{,}493.6933} = 63{,}338.5041$$

Capital Final del 7° mes + intereses generados del 8° mes

9° mes.

$$63{,}338.5041+(63{,}338.5041\times 0.03) = \underbrace{63{,}338.5041\ (1+0.03)}_{\text{Factorizando } 63{,}338.5041}=65{,}238.6592$$

Capital Final del 8° mes + intereses generados del 9° mes

10° mes.

$$\underbrace{65{,}238.6592}_{\substack{\text{Capital}\\ \text{Final}\\ \text{del}\\ 9^{\circ}\text{ mes}}}+\underbrace{(65{,}238.6592\times 0.03)}_{\substack{\text{intereses}\\ \text{generados}\\ \text{del}\\ 10^{\circ}\text{ mes}}} = \underbrace{65{,}238.6592\ (1+0.03)}_{\substack{\text{Factorizando}\\ 65{,}238.6592}}=67{,}195.819$$

11° mes.

$$\underbrace{67{,}195.819}_{\substack{\text{Capital}\\ \text{Final}\\ \text{del}\\ 10^{\circ}\text{ mes}}}+\underbrace{(67{,}195.819\times 0.03)}_{\substack{\text{intereses}\\ \text{generados}\\ \text{del}\\ 11^{\circ}\text{ mes}}} = \underbrace{67{,}195.819\ (1+0.03)}_{\substack{\text{Factorizando}\\ 67{,}195.819}}=69{,}211.6935$$

12° mes.

$$\underbrace{69{,}211.6935}_{\substack{\text{Capital}\\ \text{Final}\\ \text{del}\\ 11^{\circ}\text{ mes}}}+\underbrace{(69{,}211.6935\times 0.03)}_{\substack{\text{intereses}\\ \text{generados}\\ \text{del}\\ 12^{\circ}\text{ mes}}} = \underbrace{69{,}211.6935\ (1+0.03)}_{\substack{\text{Factorizando}\\ 69{,}211.6935}}=71{,}288.044$$

En conclusión, al final de 12 meses se tiene un capital de $71,288.04434. En definitiva, la cantidad que se obtuvo al final de 12 meses de 71,288.04434 difiere del primer resultado (68,000).

El error del primer proceso es que no se consideró que mes a mes se reinvierte el capital más los intereses generados y solamente se consideraba la inversión inicial.

Ahora, véase el proceso de reinvertir el capital mes con mes en un solo paso.

1° mes

$$\underbrace{50,000}_{\text{Capital Inicial}} + \underbrace{(50,000\times 0.03)}_{\text{intereses generados del 1° mes}} = \underbrace{50,000(1+0.03)}_{\text{Factorizando } 50,000}=51,500$$

2° mes

$50,000(1+0.03)+(50,000(1.03)\times 0.03)=$

$$\underbrace{50,000(1+0.03)}_{\text{Capital al finalizar el 1° mes}}\underbrace{(1+.03)}_{\text{Intereses del 2°mes}}=\underbrace{50,000(1+0.03)}_{\text{Factorizando } 50,000(1+0.03)}{}^{2}=53,045$$

Análogamente, se obtiene el 3° mes.

$50,000(1+0.03)^2+(50,000\times 0.03)= 50,000(1+0.03)^2\,(1+0.03)= 50,000(1+0.03)^3$

$=54,636.35$

Al final de 12 meses se tiene que:

$50,000(1+0.03)^{12}=71,288.04434$

Obsérvese que esta última operación nos puede proporcionar la tasa de interés anual. Dado que $(1+0.03)^{12}=1.42576$, y además:

$$50,000(1+03)=50,000\times 1.42576=50,000\times(1+0.42576)$$

$$=(50,000\times 1)+(50,000\times 0.42576)$$

Como $50,000\times 1$ es el capital inicial y $50,000\times 0.42576$ son los intereses generados en el año, la tasa de interés anual es 42.576% y no 36% como se pensaba en un principio.

En este ejemplo ya sabemos que la tasa mensual fue del 3%, pero si sólo se supiera la tasa de interés anual de 42.576%,la tasa de interés mensual, no se obtiene al dividir la tasa de interés anual entre 12, dando esto, un resultado de 3.5475% que no corresponde al interés del 3% mensual estipulado en el problema.

Obsérvese que las cantidades que se tienen mes a mes forman una sucesión geométrica de números.

Mes	Capital	**Razón de crecimiento** (es el cociente que resulta de dividir el capital del mes en curso entre el capital del mes anterior)
Inicial	50,000	
1° mes	51,500	$\frac{51,500}{50,000} = 1.03$
2° mes	53,045	$\frac{53,045}{51,500} = 1.03$
3° mes	54,636.35	$\frac{54,636.45}{53,045} = 1.03$
4° mes	56,275.405	$\frac{56,275.405}{54,636.45} = 1.03$
5° mes	57,963.70372	$\frac{57,963.70372}{56,275.405} = 1.03$
6° mes	59,702.61483	$\frac{59,702.61483}{57,963.70372} = 1.03$
7° mes	61,493.69327	$\frac{61,493.69327}{59,702.61483} = 1.03$
8° mes	63,338.50407	$\frac{63,338.50407}{61,493.69327} = 1.03$
9° mes	65,238.65919	$\frac{65,238.65919}{63,338.50407} = 1.03$
10° mes	67,195.81897	$\frac{67,195.81897}{65,238.65919} = 1.03$

11° mes	69,211.69354	$\frac{69{,}211.69354}{67{,}195.81897}=1.03$
12° mes	71,288.04434	$\frac{71{,}288.04434}{69{,}211.69354}=1.03$

Si se multiplicara la razón de crecimiento mes a mes se obtendría $(1.03)^{12}$ de tal manera que, para reencontrar la tasa de interés mensual bastaría con obtener la raíz doceava de este número y restarle 1.

$$\sqrt[12]{(1.03)^{12}}=1.03$$

Restándole 1 se tiene:

1.03 -1=0.03
Multiplicando por 100 para obtener en forma de porcentaje,(0.03×100=3%) para obtener el 3% la tasa de interés mensual o la razón de crecimiento original.

La media geométrica se define como: el valor promedio de las razones de crecimiento que se obtienen de los cocientes que resultan de dividir el capital del mes en curso entre el capital del mes anterior.

Otra situación se da cuando no se conocen las tasas de interés mensual y además no son iguales mes a mes, para ello analizaremos el siguiente problema:

Una persona adquiere una deuda de $1,000, pero por omisión no la paga en seis meses y al ver sus estados de cuenta ve como incrementaba su deuda mes a mes:

Mes	Deuda al final del mes
Inicial	\$1,000
1° mes	\$1,020
2° mes	\$1,091.4
3° mes	\$1,145.97
4° mes	\$1,180.3491
5° mes	\$1,321.99
6° mes	\$1,573.168

Está persona está interesada en saber cuales son las tasas de interés mensuales que está cobrando el banco, por lo que recurre a la solución más simple para resolver el problema: restó al saldo final de la tarjeta (1,573.168)el saldo inicial (1,000), resultando 573.168, después aplicó una regla de tres:

$$\frac{1{,}000}{573.168}=\frac{100\%}{x}\quad por\ \tan to\ x=57.31\%$$

Por último dividió 57.31% entre 6 meses y obtuvo 9.55%, creyendo que esa cantidad es la tasa de interés mensual. Pero al recalcular la deuda obtiene lo siguiente:

1°mes	1,000(1.0955)=1,095.5	\$1,095.5
2° mes	1,095.5 (1.0955)=1,200.12	\$1,200.12
3° mes	1,200.12 (1.0955)=1,314.73	\$1,314.73
4° mes	1,314.73 (1.0955)=1,440.27	\$1,440.27
5° mes	1,440.27 (1.0955)=1,577.82	\$1,577.82
6° mes	1,577.82 (1.0955)=1,728.50	\$1,728.50

La tasa del 9.55% no podía ser, porqué él había pagado \$573.168 de intereses y no \$728.50. Por lo tanto ésta es una manera inadecuada de obtener el interés mensual.

Después trató de obtener el interés de la siguiente forma: deuda al final de cada mes entre la deuda inicial, y los resultados fueron:

Mes	Deuda al final de cada mes entre deuda inicial.	Interés
1°	$\frac{1{,}020}{1{,}000} = 1.02$	(1.02-1)×100=2%
2°	$\frac{1{,}091.4}{1{,}000} = 1.0914$	(1.0914-1)×100=9.14%
3°	$\frac{1{,}145.97}{1{,}000} = 1.14597$	(1.14597-1)×100=14.597%
4°	$\frac{1{,}180.3491}{1{,}000} = 1.18034$	(1.1803491-1)×100=18.73%
5°	$\frac{1{,}321.99}{1{,}000} = 1.32199$	(1.32199-1)×100=32.199%
6°	$\frac{1{,}573.1681}{1{,}000} = 1.5731681$	(1.573168-1)×100=57.32%

Ahora, bajo estas tasas de interés, calcula el interés mensual por medio de la media aritmética.

$$\frac{2\% + 9.14\% + 14.597\% + 18.73\% + 32.199\% + 57.32\%}{6} = \frac{133.986}{6} = 22.31\%$$

Indudablemente, este no puede ser un buen resultado, por que si con el 9.55% la deuda se incrementaba, con el 22.31% la deuda crece de manera inconcebible.

Esta aproximación a la tasa de interés mensual es muy lejana a la real, debido a que tomaba tasas de interés sobre la deuda inicial y no sobre la deuda del mes anterior.

Finalmente, obtuvo las tasas de interés como el resultado de dividir el capital del mes en curso entre el capital del mes anterior:

Mes	Saldo final del mes entre saldo inicial.	Interés
1°	$\frac{1{,}020}{1{,}000} = 1.02$	(1.02-1)×100=2%
2°	$\frac{1{,}091.46}{1{,}020} = 1.07$	(1.07-1)×100=7%
3°	$\frac{1{,}145.97}{1{,}091.46} = 1.05$	(1.05-1)×100=5%
4°	$\frac{1{,}180.3491}{1{,}145.97} = 1.03$	(1.03-1)×100=3%
5°	$\frac{1{,}321.99}{1{,}180.3491} = 1.12$	(1.12-1)×100=12%
6°	$\frac{1{,}573.1681}{1{,}321.99} = 1.19$	(1.191)×100=19%

Obsérvese, que con estás tasas de interés si reconstruimos la deuda de manera exacta queda:

Mes	Deuda del mes anterior por uno más el interés del mes en curso	Interés
1°	1000(1.02)=1,020	2%
2°	1020(1.07)=1,091.4	7%
3°	1091.4(1.05)=1,145.97	5%
4°	1,145.97(1.03)=1,180.3491	3%
5°	1,180.3491(1.12)=1,321.99	12%
6°	1,321.99(1.19)=1,573.169	19%

En este caso las tasas de interés varían mes a mes. De aquí surge la pregunta: ¿Cuál es la tasa de interés promedio mensual?

Una posibilidad es obtener el promedio aritmético de las tasas de interés.

$$\frac{2\% + 7\% + 5\% + 3\% + 12\% + 19\%}{6} = 8\%$$

Reconstruyamos la deuda con esta tasa de interés mensual:

1°mes	1,000(1.08)=1,080	$1,080
2° mes	1,080 (1.08)=1,166.4	$1,166.4
3° mes	1,166.4 (1.08)=1,259.712	$1,259.712
4° mes	1,259.712 (1.08)=1,360.49	$1,360.49
5° mes	1,360.49 (1.08)=1,469.33	$1,469.33
6° mes	1,469.33 (1.08)=1,586.87	$1,586.87

Esta, es una mejor aproximación, pero los montos finales todavía difieren un poco, debido a que el monto final, en la deuda original, es $1,573.168 y el obtenido en esta ocasión es $1,586.87.

Para obtener una tasa promedio más adecuada, es útil observar que la deuda final se puede obtener de la deuda inicial de la siguiente manera:

1,573.168=1,000× (1.02) × (1.07) ×(1.05)×(1.03)×(1.12)×(1.19)

En el ejercicio anterior se ve que el capital final se obtiene multiplicando el capital inicial por $(1.03)^{12}$ y si a este valor se le extrae la raíz doceava, se puede decir cual fue la tasa de interés promedio mensual restando 1.

Puede ser natural pensar que, a este número por el que multiplico mi deuda inicial para obtener la deuda final, también, se le extraiga raíz sexta y se le reste uno para obtener una estimación de la tasa promedio de interés mensual.

Con este método, el interés promedio mensual sería de:

$$\sqrt[6]{1.02\times1.07\times1.05\times1.03\times1.12\times1.19}-1=\sqrt[6]{1.57316928}-1=1.07843-1=0.07844=7.844\%$$

Por último, reconstruyamos la deuda con esta tasa de interés:

1°mes	1,000(1.07844)=1,080	$1,078.44
2° mes	1,078.44 (1.07844)=1,166.4	$1,163.033
3° mes	1,163.033(1.07844)= 1,254.261	$1,254.261
4° mes	1,254.261 (1.07844)=1,352.65	$1,352.65
5° mes	1,352.65 (1.07844)=1,458.75	$1,458.75
6° mes	1,458.75 (1.07844)=1,573.17	$1,573.17

Nótese que con este método se obtiene una tasa de interés promedio precisa.

Si hacemos:

$x_1 = (1+0.02) = (1.02)$
$x_2 = (1+0.07) = (1.07)$
$x_3 = (1+0.05) = (1.05)$
$x_4 = (1+0.03) = (1.03)$
$x_5 = (1+0.12) = (1.12)$
$x_6 = (1+0.19) = (1.19)$

Para obtener, esta tasa de interés precisa se realizaron las siguientes operaciones.

Se multiplica:

$$x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6$$

se extrae la raíz sexta:

$$\sqrt[6]{x \cdot_1} x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6$$

y se le resta 1:

$\sqrt[6]{x \cdot_1} x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6$ -1

A este número $\sqrt[6]{x \cdot_1} x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6$ se le llama la ***media geométrica*** de estos números.

En general, si se tienen n números de $x_1 \cdot x_2 \cdot x_3 \cdots x_n$ la ***media geométrica*** se define como *la raíz n-ésima del producto de estos números*.

Media geométrica = $\sqrt[n]{x_1 x_2 x_3 \cdots x_n}$

Es decir:

Media geométrica = $\sqrt[n]{\prod_{i=1}^{n} x_i}$

Un ejemplo diferente es:

El número de enfermos de SIDA en el año 2000 al año 2005 es:

Año	Número de enfermos de SIDA
2000	15,000
2001	15,750
2002	17,640
2003	21,168
2004	26,460
2005	34,398

Calcúlese la tasa anual promedio de crecimiento del SIDA, para así poder estimar el número de enfermos de SIDA en el año 2006.

Solución:

Para calcular la razón anual promedio de crecimiento del SIDA se debe calcular la media geométrica de las tasas de crecimiento anual más uno y al final restarle uno.

Año	Número de enfermos de SIDA	Razón de crecimiento (número de enfermos del año actual entre el número de enfermos del año anterior)
2000	15,000	
2001	15,750	$\frac{15{,}750}{15{,}000} = 1.05$
2002	17,640	$\frac{17{,}640}{15{,}750} = 1.12$
2003	21,168	$\frac{21{,}168}{17{,}640} = 1.20$
2004	26,460	$\frac{26{,}460}{21{,}168} = 1.25$
2005	34,398	$\frac{34{,}398}{26{,}460} = 1.30$

La razón anual promedio de crecimiento es:

$$\sqrt[6]{1.05 \times 1.12 \times 1.20 \times 1.25 \times 1.30} - 1 = \sqrt[6]{2.2932} - 1 = 1.1434 - 1 = 0.1434 = 14.34\%$$

El número de enfermos de SIDA que se espera en el año 2006 es: 34,398×1.1434=39,330 enfermos de SIDA

.

II.2.7. Media Armónica

Para entender la media armónica se explicará mediante el siguiente problema:

A dos mecanógrafas se les encarga que copien un informe, la que escribía rápido hace el encargo en dos horas; la otra, en tres horas.

Con estos datos diríamos que el tiempo promedio que tarda una secretaria en hacer el informe es $\frac{2+3}{2}=\frac{5}{2}$, lo que significa que tiene una tasa de eficiencia de: $\frac{1}{\frac{5}{2}}=\frac{2}{5}\ por\ hora$, es decir, hace dos quintas partes del trabajo en una hora.

Un problema diferente es, ¿conocer en cuánto tiempo copiarían ambas ese informe, si se distribuyen el trabajo?

Sea x el tiempo en horas que se tardan en terminar el trabajo ambas secretarias, podemos determinar qué parte del trabajo realiza en una hora cada mecanógrafa (tasa de eficiencia), es decir, la mecanógrafa más rápida en una hora realiza la mitad $\left(\frac{1}{2}\right)$ del trabajo, mientras que la otra, en una hora sólo realiza una tercera parte $\left(\frac{1}{3}\right)$ del mismo. Si trabajaran juntas realizarían $\left(\frac{1}{x}\right)$ del trabajo en una hora, lo que deriva a la siguiente ecuación:

$$\frac{1}{x}=\frac{1}{2}+\frac{1}{3}=\frac{3+2}{6}=\frac{5}{6}$$

De aquí se deduce:

$$\frac{1}{x}=\frac{5}{6}$$

$$1=\frac{5}{6}x$$

$$1\times 6=5x$$

$$6=5x$$

$$\frac{6}{5}=x=1\frac{1}{5}=1.2\ horas$$

Esto quiere decir que las secretarias tardarían 1.2 horas, es decir, 1 hora + [60 min×(1.02) = 12 min], por lo tanto, tardarían 1 hora 12 minutos en terminar juntas el trabajo.

La práctica demuestra que trabajando juntas, la secretaria lenta se vuelve más rápida y la secretaria rápida se vuelve más lenta. El problema, ahora, es cómo calcular una nueva tasa de eficiencia y el tiempo en el que con esta nueva tasa realizaría cada secretaría el mismo trabajo.

Sabemos que juntas las secretarias realizan el trabajo en x tiempo, entonces una sola hace la mitad del trabajo en el tiempo x. De ahí que, una sola secretaría tardaría el doble del tiempo en realizar este mismo trabajo (2x).

Se sabe que:

$$x = 1.2 = \frac{6}{5} \; entonces \quad 2x = 2 \cdot (1.2) = 2\left(\frac{6}{5}\right) = \frac{12}{5}$$

En conclusión, una sola secretaría con esta nueva tasa de eficiencia tarda: $\frac{12}{5} = 2.4$ hrs., o bien 2 horas 24 minutos.

Si llamamos:

X_1 = Tiempo que tarda la secretaria rápida en realizar el trabajo sola.

X_2 = Tiempo que tarda la secretaria lenta en realizar el mismo trabajo sola.

n = Número de secretarias (2)

Las operaciones que se utilizaron para determinar el tiempo en el que cada secretaria tardaría en realizar el trabajo con una nueva tasa de rendimiento son:

Se obtuvo x como:

$$\frac{1}{x}=\frac{1}{x_1}+\frac{1}{x_2}=\frac{1}{2}+\frac{1}{3}=\frac{5}{6} \quad entonces\ x=\frac{1}{\frac{1}{x}}=\frac{1}{\frac{1}{x_1}+\frac{1}{x_2}}=\frac{1}{\frac{5}{6}}=\frac{6}{5}$$

Se multiplico x por n=2

$$2x=n[x]=n\left[\frac{1}{\frac{1}{x}}\right]=n\left[\frac{1}{\frac{1}{x_1}+\frac{1}{x_2}}\right]=\left[\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}}\right]=\left[\frac{2}{\frac{1}{x_1}+\frac{1}{x_2}}\right]=\left[\frac{2}{\frac{1}{2}+\frac{1}{3}}\right]\frac{2}{\frac{5}{6}}=\frac{12}{5}=2.4$$

A este número $\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}}$ con *n* =2 se le llama la media armónica de x_1 y x_2.

Si se tienen n números $x_1, x_2, \ldots, x_n$, la ***media armónica*** de estos números se define como:

$$Media\ \ armónica\ = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}=\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

La ***media armónica*** de define como el recíproco de la media de los recíprocos, esto lo podemos comprobar también por otra vía, calculando el recíproco del promedio aritmético de las tasas de eficiencia. Y se prueba mediante la siguiente forma:

1. Calculando la tasa de eficiencia de cada una de las secretarias por separado se tiene que:

   Dado que la secretaria más rápida tarda un tiempo x_1=2 hrs. en realizar el trabajo, y su tasa de eficiencia es $\frac{1}{x_1}=\frac{1}{2}$, es decir, realiza la mitad del trabajo en una hora y sabemos que la secretaria más lenta tarda un tiempo x_2=3 hrs. en hacer el mismo trabajo y su tasa de eficiencia es

$\frac{1}{x_2}=\frac{1}{3}$ significa que hace una tercera parte del trabajo en una hora.

Las tasas de eficiencia de cada una de las secretarias son: $\frac{1}{x_1}=\frac{1}{2}$ y

$\frac{1}{x_2}=\frac{1}{3}$,

2. Con estos datos obtenemos ***un promedio aritmético*** de las tasas de

   eficiencia (t) y se tiene $\frac{\frac{1}{x_1}+\frac{1}{x_2}}{2}=\frac{\frac{1}{2}+\frac{1}{3}}{2}=\frac{\frac{5}{6}}{2}=\frac{5}{12}$, este resultado quiere

   decir que en promedio su tasa de eficiencia (t) es $\frac{5}{12}$=0.4166, por lo

   tanto, cada secretaria en promedio elabora el 41.66% del trabajo en

   una hora.

3. Finalmente, queremos conocer el tiempo que se tardaría una secretaria

   en realizar esta tasa de eficiencia, es suficiente con obtener el

   reciproco de este promedio, pero veamos por qué.

Apliquemos una regla de tres:

$$\frac{\text{Proporción del trabajo realizado } 41.66\%}{\text{Todo el trabajo } 100\%}=\frac{\text{Tiempo 1 hora}}{\text{x tiempo en horas}}$$

$$x=\frac{100\%\times 1\ hora}{41.66\%}=2.4\ horas$$

Ahora, este mismo cálculo en forma de fracción es:

$$\frac{\text{Proporción del trabajo realizado } \frac{5}{12}}{\text{Todo el trabajo } 1} = \frac{\text{Tiempo 1 hora}}{x \text{ tiempo en horas}}$$

$$x = \frac{1 \times 1}{\frac{5}{12}} = \frac{1}{\frac{5}{12}} = \frac{12}{5} = 2.4\ horas = \frac{1}{\Pr omedio\ de\ las\ tasas\ de\ eficiencia}$$

Sí llamamos:

X_1 = Tiempo que tarda la secretaria rápida en realizar el trabajo sola.

X_2 = Tiempo que tarda la secretaria lenta en realizar el mismo trabajo sola.

n = Número de secretarias (2)

t_1 = Tasa de eficiencia de la secretaria rápida.

t_2 = Tasa de eficiencia de la secretaria rápida

Calculando la tasa de eficiencia de cada una de las secretarias por separado se tiene:

$t_1 = \frac{1}{x_1} = \frac{1}{2}$ y $t_2 = \frac{1}{x_2} = \frac{1}{3}$,

Con estos datos obtenemos ***un promedio aritmético*** de las tasas de eficiencia (t) . $t = \frac{t_1 + t_2}{2} = \frac{\frac{1}{x_1} + \frac{1}{x_2}}{2} = \frac{\frac{1}{2} + \frac{1}{3}}{2} = \frac{\frac{5}{6}}{2} = \frac{5}{12}$, cada secretaria en promedio elabora t partes del trabajo en una hora.

Finalmente, se obtiene el **reciproco de este promedio**:

Aplicando una regla de tres.

Proporción del trabajo realizado

$$\frac{t}{\text{Todo el trabajo} \atop 1} = \cdot \frac{\text{Tiempo 1 hora}}{\text{x tiempo en horas}}$$

$$x = \frac{1 \times 1}{t} = \frac{1}{t} = \frac{1}{\frac{5}{12}} = \frac{12}{5} = 2.4\ horas = \frac{1}{\Pr omedio\ de\ las\ tasas\ de\ eficiencia}$$

Sabemos que:

$$x = \frac{1}{t} = \frac{1}{\frac{t_1 + t_2}{2}} = \frac{2}{t_1 + t_2} = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$$

Obsérvese que, también se obtiene la media armónica de x_1 y x_2.

La ***media armónica*** determina el tiempo promedio de cada variable que es inversamente proporcional a otra variable bajo una tasa de eficiencia, (es decir, el trabajo realizado por cada secretaria), como es el caso de la productividad en el tiempo y la velocidad respecto al tiempo.

Ejemplo:

Una constructora quiere saber ¿cuál es el rendimiento representativo de sus obreros de esta compañía para colocar 10 metros cuadrados de piso?

Un obrero tarda 4 días, el segundo obrero tarda 3 y el último obrero tarda 6 días. Calculemos la media armónica:

Conocemos que n=3

$$Media\ \ \text{armónica} = \frac{3}{\sum_{i=1}^{3} \frac{1}{x_i}} = \frac{3}{\frac{1}{4} + \frac{1}{3} + \frac{1}{6}} = \frac{3}{\frac{6+8+4}{24}} = \frac{3}{\frac{18}{24}} = \frac{3}{\frac{3}{4}} = \frac{12}{3} = 4$$

El rendimiento representativo por obrero es de 4 días.

II.2.8. Fractiles o Cuantiles (cuartiles, deciles y percentiles)

Los *fractiles* pueden ser ubicados, según su estudio como medidas de localización o posición o como una medida de dispersión. El estudio de estos depende de la información que se necesite extraer de la muestra o población.

Partiendo de la misma idea que la mediana (dividir los datos ordenados de menor a mayor en bloques de igual número de datos) podemos introducir nuevas medidas de tendencia central llamados *fractiles o cuantiles*.

Los fractiles pueden ser:

1. **Cuartiles**

   Los cuartiles son valores (no necesariamente un dato de la muestra) que dividen al conjunto de datos ordenados de menor a mayor en cuatro grupos con el mismo número de elementos.

   El primer cuartil, $\mathbf{Q_1}$, es el valor que marca una cuarta parte de los datos antes que él y tres cuartas partes de los datos después de él, es decir, el 25% de los datos de la muestra se encuentran a la izquierda (o por debajo) de él y el 75% de los datos se encuentran a la derecha (o por arriba) de él.

   El segundo cuartil, $\mathbf{Q_2}$, es el valor que marca la mitad de los datos antes que él y la mitad de los datos después de él. Es decir, es el valor (que no necesariamente es un dato de la muestra) que se encuentra en la posición central. Este valor coincide con la mediana.

   El tercer cuartil, $\mathbf{Q_3}$, es el valor que marca tres cuartas partes de los datos antes que él y una cuarta partes de los datos después de él, es decir, el

75% de los datos de la muestra se encuentran a la izquierda (o por debajo) de él y el 25% de los datos se encuentran a la derecha (o por arriba) de él.

2. Deciles

Los deciles son valores que dividen a la muestra ordenada de menor a mayor en 10 bloques con el mismo número de datos.

El primer decil, $\mathbf{D_1}$, es el valor que marca el 10% de los datos ordenados de la muestra por debajo(a la izquierda) de él y el 90% de los datos de la muestra por arriba (a la derecha de él).

El segundo decil, $\mathbf{D_2}$, es el valor que marca el 20% de los datos ordenados de la muestra por debajo(a la izquierda) de él y el 80% de los datos de la muestra por arriba (a la derecha de él).

Así, sucesivamente el noveno decil, $\mathbf{D_9}$, es el valor que marca el 90% de los datos ordenados de la muestra por debajo(a la izquierda) de él y el 10% de los datos de la muestra por arriba (a la derecha de él).

3. Percentiles

Los percentiles son valores que dividen a la muestra (ordenada de menor a mayor) en 100 partes con el mismo número de datos.

Los percentiles se nombran del 1°, 2°,…,99° (del primero al nonagésimo noveno). El 100° percentil no se considera ya que es el total de los datos.

El primer percentil, $\mathbf{P_1}$, es el valor que marca el 1% de los datos ordenados de la muestra por debajo(a la izquierda) de él y el 99% de los datos de la muestra por arriba (a la derecha de él).

El segundo percentil, **P_2**, es el valor que marca el 2% de los datos ordenados de la muestra por debajo(a la izquierda) de él y el 98% de los datos de la muestra por arriba (a la derecha de él).

Así, el nonagésimo noveno percentil, **P_{99}**, es el valor que marca el 99% de los datos ordenados de la muestra por debajo(a la izquierda) de él y el 1% de los datos de la muestra por arriba (a la derecha de él).

Observemos que el segundo cuartil, el quinto decil y el quincuagésimo percentil equivalen al lugar de la mediana. También, el primero, segundo, etc. decil equivalen al décimo, vigésimo, etc. percentiles.

Al igual que la mediana, los fractiles tienen la fórmula para localizar su posición y después se ubican dentro de la muestra ordenada. Las fórmulas de localización de los fractiles se dan a conocer en el siguiente cuadro:

Cuartiles: Se expresan con la letra Q y el subíndice es el cuartil que representan.
$Posición\ de\ Q_1 = \frac{n+1}{4}$ $\quad Posición\ de\ Q_2 = \frac{2(n+1)}{4}$ $\quad Posición\ de\ Q_3 = \frac{3(n+1)}{4}$
Deciles: Se expresan con la letra D y el subíndice es el decil que representan
$Posición\ de\ D_1 = \frac{n+1}{10}$ $\quad Posición\ de\ D_2 = \frac{2(n+1)}{10}$ $\quad Posición\ de\ D_3 = \frac{(3n+1)}{10}$ $Posición\ de\ D_4 = \frac{4(n+1)}{10}$ y así, sucesivamente para los siguientes deciles.
Percentiles: Se expresan con la letra P y el subíndice es el percentil que representan
$Posición\ de\ P_1 = \frac{n+1}{100}$ $\quad Posición\ de\ P_2 = \frac{2(n+1)}{100}$ $\quad Posición\ de\ P_3 = \frac{3(n+1)}{100}$ $Posición\ de\ P_4 = \frac{4(n+1)}{100}$ y así, sucesivamente para los siguientes percentiles

Cuadro II.2 Localización de los fractiles o cuantiles

II.2.8.4. Cálculo de los Fractiles.

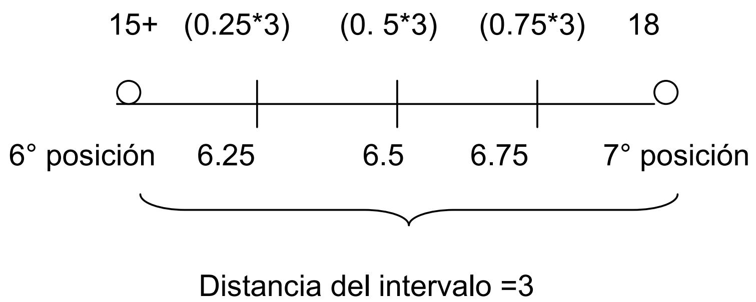
Determinaremos el valor del primer cuartil, el octavo decil, y sexagésimo sexto percentil para la siguiente muestra de datos ordenados: 3, 4, 7, 9, 13, 15, 18, 19, 22, 25, 26, 26, 29, 31, 35, 39, 42, 46, 47, 49, 49, 49, 52, 55.

El tamaño de la muestra es n = 24.

La posición del 1°cuartil es:

$$\textit{Posición}\text{ de }Q_1 = \frac{1 \times (24+1)}{4} = \frac{25}{4} = 6.25$$

Para comprenderlo mejor, trácese en una recta los valores 15 y 18 que ocupan la 6° y 7° posición de la muestra (que tiene 24 elementos), respectivamente.



La distancia entre la 6° posición y la 7° posición es 18-15=3

Para obtener el valor que corresponde a la posición 6.25 (de acuerdo a los datos y que está entre la 6° y la 7° posición) a 25 centésimos de la distancia de la 6° posición (15) y a 75 centésimos de la 7° posición (18), debe sumársele a la sexta posición (15) 25 centésimas de la diferencia entre la 6° y la 7° posición (18-15=3), es decir, 15+(0.25*3)=15+0.75=15.75 Por lo tanto, el valor del primer cuartil es 15.75. (**Q_1=15.75**)

La posición del 8° decil es:

$$Posición\ \text{de}\ D_8 = \frac{8 \times (24+1)}{10} = \frac{8 \times 25}{10} = \frac{200}{10} = 20$$

El valor que corresponde a la 20° posición (49), es un valor de la muestra, debido a que es un valor entero. Por lo tanto, el valor del 8° decil es 49.(**D_8=49)**

La posición del 66° percentil es:

$$Posición\ \text{de}\ P_{66} = \frac{66 \times (24+1)}{100} = \frac{66 \times 25}{100} = \frac{1650}{100} = 16.5$$

El valor que corresponde a la posición 16.5 (posición que está entre la 16° y la 17° posición, a 5 décimos de la distancia de la 16° posición (39) y a 5 décimos de la 17° posición (42)). Debe sumarse a la 16° posición (39), 5 décimas de la diferencia entre la 16° y la 17° posición (42-39=3), es decir, 39+(0.5*3)=39+1.5=40.5, por lo tanto, el valor del 66° percentil es 40.5(**P_{66}=40.5**)

II.2.8.5. Uso e interpretación de los fractiles.

Para poder entender el uso y significado de los fractiles analicemos el siguiente problema.

En una comunidad de profesionistas se les pregunta el salario quincenal a 25 de ellos y las respuestas obtenidas son las siguientes:

$2,100	$9,000	$3,900	$ 900	$5,800
$4,500	$3,500	$4,850	$3,300	$7,100
$1,000	$2,500	$8,500	$2,900	$6,500
$1,350	$3,250	$2,300	$3,100	$3,400
$8,000	$4,800	$1,500	$2,400	$2,500

Obsérvese que el tamaño de la muestra es de 25 datos. N =25

La tabla de frecuencias de estos datos se tiene que:

Salario	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa acumulada Porcentual
$900	1	1	4%
$1,000	1	2	8%
$1,350	1	3	12%
$1,500	1	4	16%
$2,100	1	5	20%
$2,300	1	6	24%
$2,400	1	7	28%
$2,500	2	9	36%
$2,900	1	10	40%
$3,100	1	11	44%
$3,250	1	12	48%
$3,300	1	13	52%
$3,400	1	14	56%
$3,500	1	15	60%
$3,900	1	16	64%
$4,500	1	17	68%
$4,800	1	18	72%
$4,850	1	19	76%
$5,800	1	20	80%
$6,500	1	21	84%
$7,100	1	22	88%
$8,000	1	23	92%
$8,500	1	24	96%
$9,000	1	25	100%

La tabla diría que el cuarto decil se ubica entre $2,900 y $3100 y que $2,900 es el menor dato que corresponde al 40%. Con base en esto, se puede concluir que el

40% de los profesionistas tiene un salario menor o igual a $2,900 ($2,900 es el primer dato menor o a la izquierda del valor del 4° decil).

Mediante la observación de esta tabla se puede determinar la distribución de los diferentes niveles de ingreso (pero no se obtienen los valores precisos de los fractiles) Como por ejemplo: ¿qué porcentaje de las personas ganan por lo menos $5,800?

La tabla indica que el 80% de las personas ganan por lo menos $5,800.

Cuando se trata de ubicar los diferentes niveles de ingreso (2° caso), bajo un porcentaje se llama **rango percentil**, que se emplea para indicar el tanto por ciento de los datos que quedan por debajo de un cierto valor. En el primer caso sólo se ubico cuantas personas se encuentran en un fractil.

Observemos que no es necesario tener una tabla de frecuencias de ingresos para poder determinar que valor ocupa el 40% de los datos, o simplemente ubicar el 4° decil.

La ubicación del 4° decil es:

$$D_4 = \frac{4(n+1)}{10} = \frac{4(25+1)}{10} = \frac{4 \times 26}{10} = \frac{104}{10} = 10.4$$

Ordenando los datos de menor a mayor:

$ 900	$2,300	$3,100	$3,900	$6,500
$1,000	$2,400	$3,250	$4,500	$7,100
$1,350	$2,500	$3,300	$4,800	$8,000
$1,500	$2,500	$3,400	$4,850	$8,500
$2,100	$2,900	$3,500	$5,800	$9,000

El valor que corresponde a la posición 10.4 (posición imaginaria que está entre la 10° y 11° posición) a 4 décimos de la distancia de la 10° posición (2,900) y a 6 décimos de la 11° posición (3,100), debe sumársele a la 10°

posición (2,900), 4 décimas de la diferencia entre la 10° y la 11° posición (3100-2900=300), es decir, 2900+(0.4*300)=2,900+ 120= 3020, por lo tanto, el valor del cuarto decil es 3020(**D_4=3020**).

Esto quiere decir que el 40% de las personas ganan por lo menos de $3,020 (Nótese que, este valor no corresponde a ningún dato, pero el dato por debajo más cercano a este valor corresponde a $2,900).

En conclusión, no es incorrecto decir que el 40% gana por lo menos $2,900, pero, es más preciso establecer que el 40% gana por lo menos $3,020.

Si se quiere ubicar qué porcentaje de las personas ganan menos de $5,800 no tenemos un proceso eficiente para localizar el porcentaje, esto sólo se puede obtener mediante una tabla de datos.

Al obtener los deciles se evitaría tener que analizar todos los datos, en este caso, sólo se analizan 9 valores y de manera fácil se ubicaría qué porcentaje de la población tiene menos de x ingreso.

Datos de menor a mayor:

$ 900	$2,300	$3,100	$3,900	$6,500
$1,000	$2,400	$3,250	$4,500	$7,100
$1,350	$2,500	$3,300	$4,800	$8,000
$1,500	$2,500	$3,400	$4,850	$8,500
$2,100	$2,900	$3,500	$5,800	$9,000

Ubicación del 1° decil:

$$Posición\ de\ D_1 = \frac{1(n+1)}{10} = \frac{1(25+1)}{10} = \frac{1\times 26}{10} = \frac{26}{10} = 2.6$$

El valor que corresponde a la posición 2.6 (posición que está entre la 2° y 3° posición, a 6 décimos de la distancia de la 2° posición (1,000) y a 4 décimos de la3° posición (1,350), debe sumársele a la 2° posición (1,000), 6 décimas de la diferencia entre la 2° y la 3° posición (1,350-1,000=350), es decir, 1,000+(0.6*350)=1,000+ 210= 1,210, por lo tanto el valor del primer decil es 1,210.

Esto quiere decir que el 10% de las personas ganan por lo menos de $1,210.

Ubicación del 2° decil:

$$Posición\ de\ D_2 = \frac{2(n+1)}{10} = \frac{2(25+1)}{10} = \frac{2\times 26}{10} = \frac{52}{10} = 5.2$$

El valor que corresponde a la posición 5.2 (posición que está entre la 5° y 6° posición, a 2 décimos de la distancia de la 5° posición (2,100) y a 2 décimos de la 6° posición (2,300), debe sumársele a la 5° posición (1,000), 2 décimas de la diferencia entre la 5° y la 6° posición.(2,300-2,100=200), es decir, 2,100+(0.2*300)=2,100+ 60= 2,160, por lo tanto, el valor del 2° decil es 2,160.

Esto quiere decir que, el 20% de las personas ganan por lo menos de $2,160, así sucesivamente se obtienen los siguientes valores de los deciles, como lo muestra el siguiente cuadro:

Decil	***Valor***
Primero	1,210
Segundo	2,160
Tercero	2,500
Cuarto	3,020
Quinto	3,300
Sexto	3,740
Séptimo	4,810
Octavo	6,360
Noveno	8,200

Al ver este cuadro se puede deducir que el 80% de las personas ganan por lo menos $ 6,360 pesos.

II.2.9. Datos agrupados.

Es importante destacar que muchas estadísticas tienen un rango de datos muy amplio y estos datos se compactan en grupos con fines de presentación e interpretación y en muchas ocasiones, la información de la cual se quieren obtener estas medidas, como la mediana, media aritmética, moda, etc., se presentan en tablas de frecuencias para datos agrupados, por ejemplo el censo, se presenta en grupos quinquenales de edad; el ingreso de la población se presenta en intervalos, como un grupo de personas que ganan de 3 a 5 salarios mínimos, etc.

Con base en esto, es importante poder obtener las medidas de tendencia central cuando los datos no se presentan de manera individual. La media, mediana, moda, etc. conservan la misma definición que para datos individuales, pero el cálculo para obtenerlas es un poco más complejo y su exactitud es sólo aproximada en comparación con el cálculo obtenido en datos individuales.

Para explicarlo con mayor claridad analizaremos el siguiente cuadro de distribución de frecuencias:

Clase	Límite inferior	Límite superior	Marca de clase (Lim.sup.-lim inf)/2	Frecuencia absoluta	Frecuencia acumulada
A	106	282	194	6	6
B	283	459	371	5	11
C	460	636	548	7	18
D	637	813	725	4	22
E	814	990	902	8	30

La **media aritmética** es:

$$\text{Media aritmética} = \frac{\text{Marca de clase} \times \text{Frecuencia absoluta de cada clase}}{\text{Tamaño de la muestra}}$$

$$= \frac{\sum_{i=1}^{n} M_i \times f_i}{n} = \frac{\sum_{i=1}^{n} M_i \times f_i}{\sum_{i=1}^{n} f_i}$$

M_i = Marca de clase.
f_i= frecuencia absoluta de cada clase
n= tamaño de la muestra

$$Media = \frac{194 \times 6 + 371 \times 5 + 548 \times 7 + 725 \times 4 + 902 \times 8}{30} = \frac{16971}{30} = 565.7$$

y la **mediana** se obtiene de la siguiente manera:

$$\text{Posición de la mediana} = \frac{\text{Tamaño de la muestra} + 1}{2} = \frac{n+1}{2}$$

Para obtener la mediana se utiliza la siguiente fórmula:

$$Mediana = L + \frac{(n/2) - F}{f} \times w$$

L = Límite inferior de la clase que, según la frecuencia absoluta acumulada, contiene a la posición de la mediana
f = frecuencia absoluta de la clase que, según la frecuencia absoluta acumulada, contiene a la posición de la mediana
F =Frecuencia absoluta acumulada de la clase que, según la frecuencia absoluta acumulada, contiene a la posición de la mediana
w = ancho del intervalo de la clase que, según la frecuencia absoluta acumulada, contiene a la posición de la mediana
n = tamaño de la muestra

En el ejemplo anterior, el tamaño de la muestra es: n=30.

$$Posición\ de\ la\ mediana = \frac{30+1}{2} = \frac{31}{2} = 15.5$$

Por lo tanto 15.5 se encuentra en la clase C que comprende desde el 11 hasta el 18 según la frecuencia acumulada.

$$\text{Mediana} = 460 + \left(\frac{(30/2) - 18}{7} \right) \times 176 = 460 + (-75.4285) = 384.5715$$

L = Límite inferior de la clase que, según la frecuencia absoluta acumulada, contiene a la posición de la mediana = 460

f = frecuencia absoluta de la clase que, según la frecuencia absoluta acumulada, contiene a la posición de la mediana = 7

F =Frecuencia absoluta acumulada de la clase que, según la frecuencia absoluta acumulada, contiene a la posición de la mediana = 18

W = ancho del intervalo de la clase que, según la frecuencia absoluta acumulada, contiene a la posición de la mediana =176

n = tamaño de la muestra = 30

La mediana es igual a 384.5715

Finalmente falta obtener la **moda**, la cual se obtiene de la siguiente forma:

$$\boxed{Moda = L + \frac{d_1}{d_1 + d_2} w}$$

L =Límite inferior de la clase modal

d_1 =Diferencia entre la frecuencia de la clase modal y frecuencia de la clase anterior.

d_2 =Diferencia entre la frecuencia de la clase modal y frecuencia de la siguiente clase

w = Amplitud del intervalo

Se entiende por *clase modal*, la clase que tiene una mayor frecuencia absoluta, en este caso la clase modal es E, así:

L =Límite inferior de la clase modal = 814

d_1 =Diferencia entre la frecuencia de la clase modal y frecuencia de la clase anterior = 8-4 = 4

d_2 =Diferencia entre la frecuencia de la clase modal y frecuencia de la siguiente clase =8-0 = 8

w = Amplitud del intervalo = 990-814 = 176

$$Moda = 814 + \frac{4}{8+4} \times 176 = 814 + 58.67 = 872.67$$

Si hubiera dos clases cuya frecuencia absoluta es máxima, se tiene que obtener la moda para cada una de ellas, y en este caso nuestra muestra sería bimodal, como sucedía en el caso de datos individuales.

La **media geométrica** para datos agrupados, se obtiene de la siguiente forma:

$$Media\ \ geométrica = \sqrt[n]{x_1^{f1}\, x_2^{f2}\, x_3^{f3} \cdots x_n^{fn}} \text{ en donde :}$$

x_i =Marca de clase
f_i = frecuencia absoluta de cada clase
n = tamaño de la muestra

Por ejemplo: Sí se desea saber ¿cuál es la tasa promedio de la inflación en el siglo XX? Se dan las siguientes tasas y el número de años que tuvieron ese comportamiento.

Tasa de inflación	Número de años en el que se mantuvo esta tasa de inflación (Frecuencia Absoluta)	Marca de clase
1 - 5%	18	3%
6-10%	24	8%
11-15%	30	13%
16-20%	12	18%
21-25%	16	23%

Nota: las tasas de inflación se presentaron en números enteros

El tamaño de la muestra se determina con la suma de las frecuencias absolutas, por lo tanto n = 100 años.

$$\text{Media geométrica} = \sqrt[100]{(1.03)^{18} \times (1.08)^{24} \times (1.13)^{30} \times (1.18)^{12} \times (1.23)^{16}} = 1.12$$

Restándole 1 se tiene: 1.12-1=0.12, que es equivalente al 12%.

En conclusión la tasa promedio de inflación del siglo pasado es de 12%.

La **media armónica** para datos agrupados se obtiene mediante la siguiente fórmula:

Media armónica

$$a = \frac{n}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

x_i = marca de clase
f_i =Frecuencia de clase

Por ejemplo: En una maratón de 40 personas se quiere calcular el promedio de las velocidades de los participantes. Se sabe lo siguiente:

Velocidad promedio en la competencia	Número de participantes que adquirieron la velocidad promedio en la coompetencia. (Frecuencia Absoluta)	Marca de Clase de la Velocidad en la competencia.
6-10 km. por hora.	8	8 km. por hora.
11-15 km. por hora.	12	13 km. por hora.
16-20 km. por hora.	14	18 km. por hora.
21-25 km. por hora.	6	24 km. por hora.

Su Media armónica es

$$a = \frac{n}{\sum_{i=1}^{n} \frac{f_i}{x_i}} = \frac{40}{\frac{8}{8} + \frac{12}{13} + \frac{14}{18} + \frac{6}{24}} = \frac{40}{1 + \frac{12}{13} + \frac{7}{9} + \frac{1}{4}}$$

$$= \frac{40}{\frac{468 + 432 + 364 + 117}{468}} = \frac{40}{\frac{1381}{468}} = \frac{40 \times 468}{1381} = \frac{18720}{1381} = 13.55$$

Se puede concluir que el promedio de las velocidades de los participantes en este maratón fue de 13.55 kilómetros por hora.

Los **fractiles** para datos agrupados se calculan de la siguiente manera:

Cuartiles	
Posición de $Q_1 = \frac{n+1}{4}$	$Q_1 = L_i + \frac{\frac{n}{4} - F_{i-1}}{f_i} \times w$
Posición de $Q_2 = \frac{2(n+1)}{4}$	$Q_2 = L_i + \frac{\frac{2n}{4} - F_{i-1}}{f_i} \times w$
Posición de $Q_3 = \frac{3(n+1)}{4}$	$Q_3 = L_i + \frac{\frac{3n}{4} - F_{i-1}}{f_i} \times w$
Deciles	
Posición de $D_1 = \frac{(n+1)}{10}$	$D_1 = L_i + \frac{\frac{n}{10} - F_{i-1}}{f_i} \times w$
Posición de $D_2 = \frac{2(n+1)}{10}$	$D_2 = L_i + \frac{\frac{2n}{10} - F_{i-1}}{f_i} \times w$
Posición de $D_3 = \frac{3(n+1)}{10}$	$D_3 = L_i + \frac{\frac{3n}{10} - F_{i-1}}{f_i} \times w$
Percentiles	
Posición de $P_1 = \frac{(n+1)}{100}$	$P_1 = L_i + \frac{\frac{n}{100} - F_{i-1}}{f_i} \times w$
Posición de $P_2 = \frac{2(n+1)}{100}$	$P_2 = L_i + \frac{\frac{2n}{100} - F_{i-1}}{f_i} \times w$
Posición de $P_3 = \frac{3(n+1)}{100}$	$P_3 = L_i + \frac{\frac{3n}{100} - F_{i-1}}{f_i} \times w$

Cuadro II.3 Posición de los fractiles o cuantiles para datos agrupados

En donde:

L_i =Límite inferior de la clase que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado.

F_{i-1} =Frecuencia acumulada de la clase anterior a la que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado.

f_i = Frecuencia de la clase que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado.

n= tamaño de la muestra =Suma de todas las frecuencias

w= ancho del intervalo que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado.

Por ejemplo: Determinemos el tercer cuartil, el cuarto decil y el vigésimo cuarto percentil de la siguiente tabla de distribución de frecuencias:

Clase	Límite inferior	Límite superior	Marca de clase	Frecuencia absoluta	Frecuencia acumulada
A	50.5	55.5	53	1538	1538
B	55.5	60.5	58	2367	3905
C	60.5	65.5	63	0986	4891
D	65.5	70.5	68	0754	5645
E	70.5	75.5	73	1736	7481

El cálculo para obtener el 3° cuartil se hace de la siguiente forma:

$$\text{Posición de } Q_3 = \frac{3\times(7481+1)}{4} = \frac{3\times 7482}{4} = \frac{22446}{4} = 5611.5$$

El valor está contenido en la clase D, que comprende los valores del 4891 al 5645.

En donde:

L_i =Límite inferior de la clase que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =65.5
F_{i-1} =Frecuencia acumulada de la clase anterior a la que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado = 4891
n = Tamaño de la muestra =7481
f_i = Frecuencia de la clase que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =754
w = ancho del intervalo que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =5

$$Q_3 = L_i + \frac{\frac{3n}{4} - F_{i-1}}{f_i} \times w = 65.5 + \frac{(\frac{3\times 7481}{4}) - 4891}{754} \times 5 = 65.5 + \frac{5611.25 - 4891}{754} \times 5 = 65.5 + 4.776 = 70.276$$

El valor del 3° cuartil es 70.276

El cálculo para obtener el 4° decil se hace de la siguiente forma:

$$\text{Posición de } D_4 = \frac{4 \times (7481+1)}{10} = \frac{4 \times 7482}{10} = \frac{29928}{10} = 2992.8$$

El valor está contenido en la clase B, que comprende los valores del 1538 al 3905.

En donde:

L_i =Límite inferior de la clase que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =55.5
F_{i-1} =Frecuencia acumulada de la clase anterior a la que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =1538
n= Tamaño de la muestra =7481
f_i = Frecuencia de la clase que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =2367
w= ancho del intervalo que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =5

$$D_4 = L_i + \frac{\frac{4n}{10} - F_{i-1}}{f_i} \times w = 55.5 + \frac{(\frac{4 \times 7481}{10}) - 1538}{2367} \times 5 = 55.5 + \frac{2992.9 - 1538}{2367} \times 5 = 55.5 + 0.615 = 56.115$$

El valor del cuarto decil es 56.115

El cálculo para obtener el 24° percentil se hace de la siguiente forma:

$$\text{Posición de } P_{24} = \frac{24 \times (7481+1)}{100} = \frac{24 \times 7482}{100} = \frac{179568}{100} = 1795.68$$

El valor está contenido en la clase B, que comprende los valores del 1538 al 3905.

En donde:

L_i =Límite inferior de la clase que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =55.5
F_{i-1} =Frecuencia acumulada de la clase anterior a la que , según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =1538
n= Tamaño de la muestra =7481
f_i = Frecuencia de la clase que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =2367
w= ancho del intervalo que, según la frecuencia absoluta acumulada, contiene a la posición del fractil deseado =5

$$P_{24} = L_i + \frac{\frac{24n}{100} - F_{i-1}}{f_i} \times w = 55.5 + \frac{(\frac{24 \times 7481}{100}) - 4891}{2367} \times 5 = 55.5 + \frac{1795.94 - 4891}{2367} \times 5 = 55.5 + (-6.54) = 48.96$$

El valor del 24° percentil es 48.96

Ejercicios Propuestos

De los siguientes datos:

67	35	26	39	95
68	97	26	75	35
46	67	16	64	57
67	46	18	36	35
26	35	19	85	37

Escribe le letra que corresponda a la respuesta correcta.

( )	Media aritmética	a	77.5
( )	mediana	b	51.5
( )	moda	c	22.5
( )	Tercer cuartil	d	48.68
( )	Rango medio	e	96.25
( )	6° decil	f	69.2
( )	87° porcentil	g	39
		h	40.5
		i	67
		j	35
		k	87.25

A todos los estudiantes de un grupo de sexto se le aplico un examen de matemáticas, que consistía de 25 aciertos. El número de aciertos por alumno se da en la siguiente lista:

18	22	13	24	23	21	22	23	07	18
22	12	08	21	22	16	25	24	06	12
06	11	07	12	11	13	14	16	08	01
15	10	06	13	10	11	02	19	17	04
23	09	05	09	05	04	00	18	15	

.
Con esos datos "sin agrupar" obtén

a) la media
b) la mediana
c) la moda
d) el primer de decil.
e) El tercer cuartil
f) El 38° porcentil

Ahora con los datos agrupados calcula la:

a) la media
b) la mediana
c) la moda

Calcula la media geométrica de: 3.16, 4.08, 5.78, 3.49, 6.54, 7.89, 9.87.

Calcula la media armónica de 2, 4, 6, 8, 12 y 24.

III. MEDIDAS DE DISPERSIÓN O VARIABILIDAD

Las medidas de dispersión contribuyen a tener una información más completa del conjunto de datos y pueden ayudar a analizar el comportamiento de la población en estudio.

Las medidas de tendencia central, en muchas ocasiones, pueden dar un cuadro incompleto o erróneo de la información del conjunto que se analiza.

Por ejemplo:

Un profesor de matemáticas tiene a su cargo dos grupos, cada uno de 15 alumnos, que obtuvieron las siguientes calificaciones finales.

Grupo A	Grupo B
10,10,10,9,9,9,9,8,4,3,2,4,1,2,0	8,6,6,7,6,6,7,6,6,2,4,7,6,6,7

El profesor obtiene el promedio de cada grupo, es decir, obtiene la media aritmética de cada grupo de datos.

Promedio del grupo A =

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10+10+10+9+9+9+9+8+4+3+2+4+1+2+0}{15} = \frac{90}{15} = 6$$

Promedio del grupo B =

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{8+6+6+7+6+6+7+6+6+2+4+7+6+6+7}{15} = \frac{90}{15} = 6$$

El maestro se da cuenta que los dos grupos tienen el mismo promedio, pero en el grupo A existen alumnos muy buenos y muy malos, en donde reprobaron 7 alumnos. En cambio el grupo B no cuenta con alumnos brillantes, pero todos ellos son constantes y únicamente reprobaron 2 alumnos. Si el maestro se basara en elegir a un grupo con base en su promedio, le daría igual escoger al grupo A o B. Pero, si supiera las condiciones del grupo seguramente el profesor elegiría al grupo B, porque no tendría tantos problemas de reprobación.

En estas características radica la importancia de obtener otro tipo de mediciones de los datos, llamadas medidas de dispersión o variabilidad, que son:

Rango o Amplitud

Rango percentil

Rango intercuartil

Desviación media

Desviación estándar

Varianza

Coeficiente de variación.

Porcentaje de variación.

III.1. Medidas de dispersión para datos individuales.

Las medidas de dispersión, al igual que las de tendencia central, pueden calcularse para datos individuales como para datos agrupados.

III.1.1.Rango de una muestra

El rango R de un grupo de datos, conocido también como amplitud, es la diferencia entre el límite superior de la muestra menos el límite inferior de la misma. Esta medida es rápida de calcular pero aproximada, y a la vez es su mayor desventaja, porque con sólo un valor muy grande o muy pequeño puede variar drásticamente.

Rango = Limite inferior de la muestra - limite superior de la muestra = U-L

U = Límite superior de la muestra
L = Límite inferior de la muestra

III.1.2. Rango percentil

En el capítulo anterior se analizaron los fractiles con sus diferentes denominaciones, cuartiles (dividían a la muestra en 4 grupos con el mismo número de elementos), deciles (dividían a la muestra ordenada en 10 partes con el mismo número de datos) y los percentiles (dividían a la muestra en 100 partes cada una de ellas con el mismo número de elementos).

El rango percentil es muy semejante al fractil y con frecuencia se confunden, porque el fractil ubica el dato que se encuentra en un "x" porcentaje de los datos, en cambio, el rango percentil se emplea para determinar el tanto por ciento de los datos que se encuentran por debajo de un cierto valor.

Para mostrar lo anterior, se utilizará un ejemplo del capítulo anterior[12].

En una comunidad de profesionistas, se les pregunta el salario quincenal a 25 de ellos y las respuestas obtenidas son las siguientes:

$2,100	$9,000	$3,900	$ 900	$5,800
$4,500	$3,500	$4,850	$3,300	$7,100
$1,000	$2,500	$8,500	$2,900	$6,500
$1,350	$3,250	$2,300	$3,100	$3,400
$8,000	$4,800	$1,500	$2,400	$2,500

[12] vid. infra. , p. 78

Generando una tabla de frecuencias de estos datos se tiene que:

Salario	**Frecuencia absoluta**	**Frecuencia absoluta acumulada**	**Frecuencia relativa acumulada Porcentual**
$900	1	1	4%
$1,000	1	2	8%
$1,350	1	3	12%
$1,500	1	4	16%
$2,100	1	5	20%
$2,300	1	6	24%
$2,400	1	7	28%
$2,500	2	9	36%
$2,900	1	10	40%
$3,100	1	11	44%
$3,250	1	12	48%
$3,300	1	13	52%
$3,400	1	14	56%
$3,500	1	15	60%
$3,900	1	16	64%
$4,500	1	17	68%
$4,800	1	18	72%
$4,850	1	19	76%
$5,800	1	20	80%
$6,500	1	21	84%
$7,100	1	22	88%
$8,000	1	23	92%
$8,500	1	24	96%
$9,000	1	25	100%

Mediante la observación de esta tabla se puede determinar la distribución de los diferentes niveles de ingreso, pero no se obtienen los valores precisos de los fractiles y sí quisiéramos saber ¿qué porcentaje (rango percentil) de las personas ganan por lo menos $5,800?, entonces la tabla nos muestra que es el 80%.

III.1.3. Rango Intercuartil

En estadística se suele analizar la situación que se encuentra en medio de una distribución de frecuencias y se refiere al **rango intercuartil,** que es la diferencia entre el tercer cuartil (**Q_3**) y el primer cuartil (**Q_1**), el rango intercuartil se define por la letra **Q** y es:

Rango intercuartil = Q = Q_3-Q_1

Otra medida, es el rango semicuartílico o desviación cuartil, que es la mitad del rango intercuartílico. Designándolo por **Q_D** se tiene:

$$Q_D = \frac{Q_3 - Q_1}{2}$$

A pesar de que el rango intercuartil y la desviación cuartil, como medidas de la variabilidad de las observaciones son más adecuadas que el rango de una muestra, presentan varios inconvenientes que demeritan su uso. Estos son:

a) No toma en consideración todos los valores de la distribución de frecuencias y puede suceder que valores menores que Q_1 y valores mayores que Q_3, estén muy compactados o muy dispersos, pero el valor de Q es el mismo, porque no los toma en cuenta.
b) No es posible, conociendo el valor del rango intercuartil (Q), hacer la ubicación precisa de un dato dentro de la distribución de frecuencias.

Para entender a fondo la aplicación del rango intercuartil trabajemos el ejemplo anterior:

En una comunidad de profesionistas, se les pregunta el salario quincenal a 25 de ellos y las respuestas obtenidas fueron:

$2,100	$9,000	$3,900	$ 900	$5,800
$4,500	$3,500	$4,850	$3,300	$7,100
$1,000	$2,500	$8,500	$2,900	$6,500
$1,350	$3,250	$2,300	$3,100	$3,400
$8,000	$4,800	$1,500	$2,400	$2,500

Ordenando la muestra de menor a mayor.

$ 900	$2,300	$3,100	$3,900	$6,500
$1,000	$2,400	$3,250	$4,500	$7,100
$1,350	$2,500	$3,300	$4,800	$8,000
$1,500	$2,500	$3,400	$4,850	$8,500
$2,100	$2,900	$3,500	$5,800	$9,000

Calculando el primer cuartil se tiene que:

$$Posición\ de\ Q_1 = \frac{1(n+1)}{4} = \frac{25+1}{4} = \frac{26}{4} = 6.5$$

El valor que corresponde a la posición 6.5 (posición que está entre la 6° y la 7° posición, a 5 décimos de la distancia de la 6° posición (2,300) y a 5 décimos de la 7° posición (2,400), debe sumársele a la 6° posición (2,300), 5 décimas de la diferencia entre la 6° y la 7° posición (2,400-2,300=100), es decir, 2,300+(0.5*100)=2,300+50=2,350.5, por lo tanto, el valor del $1^{er.}$ cuartil es 2,350 (**Q_1=2,350**)

Calculando el $3^{er.}$ cuartil se tiene que:

$$Posición\ de\ Q_3 = \frac{3(n+1)}{4} = \frac{3\times(25+1)}{4} = \frac{3\times 26}{4} = \frac{78}{4} = 19.5$$

El valor que corresponde a la posición 19.5 (posición que está entre la 19° y la 20° posición, a 5 décimos de la distancia de la 19° posición (4,850) y a 5 décimos de la 20° posición (5,800), debe sumársele a la 19° posición (4,850), 5 décimas de la diferencia entre la 19| y la 20° posición.(5,800-4,850=950), es decir 4,850+(0.5*950)=4,850+475=5,325.5, por lo tanto, el valor del $3^{er.}$ cuartil es 5,325 (**Q_3=5,325**)

El rango intercuartil es:

$Q = Q_3 - Q_1$ = 5,325-2,350=2,975

De lo que se deriva que más o menos la mitad de las personas tengan salarios comprendidos entre $2,350 y $5,325.

Y la desviación cuartil es:

$$Q_D = \frac{Q_3 - Q_1}{2} = \frac{5{,}325 - 2{,}350}{2} = \frac{2{,}975}{2} = 1{,}487.5$$

Esta indica que tienen una desviación con respecto a la mediana de más-menos $1,487.

Obsérvese que si se cambian los datos anteriores al primer cuartil y después del tercer cuartil, el rango intercuartil no cambia.

$ 50	$2,300	$3,100	$3,900	$35,000
$130	$2,400	$3,250	$4,500	$43,500
$360	$2,500	$3,300	$4,800	$54,000
$1,000	$2,500	$3,400	$4,850	$62,300
$2,100	$2,900	$3,500	$5,800	$90,000

El cálculo del 1^{er} cuartil y del $3^{er.}$ cuartil sigue siendo el mismo debido a que los datos que se utilizaron al obtener el 1^{er} y 3^{e} cuartil no cambiaron. Por lo tanto, el rango intercuartil se conservó igual.

Es por ello, que el rango intercuartil no considera valores muy compactados o muy dispersos anteriores al 1^{er} cuartil o superiores al 3^{e} cuartil.

Esta medida no puede indicar cuántas personas ganan $3,300. Es decir, no puede hacer una ubicación precisa de un dato dentro de la distribución de salarios, porque sólo se puede obtener mediante una tabla de frecuencias.

III.1.4. Desviación Media

La desviación media es una medida que establece el promedio de la distancia de cada dato con respecto a la media aritmética.

Para calcularla se tiene que obtener la distancia entre cada dato y la media de la muestra completa, después se calcula el promedio de estás distancias.

La fórmula de la desviación media es:

$$DM = \frac{\sum_{i=1}^{n} \left| x_i - \overline{x} \right|}{n}$$

$\overline{x}$ = media aritmética de la muestra

n = Tamaño de la muestra

Es importante destacar que se habla de distancia, lo que implica el cálculo de valores absolutos (si esto no fuera así la suma sería cero, debido a que, este resultado es una característica de la media que se mencionó en el capítulo II)

Obtengamos la desviación media de las siguientes distribuciones:

Distribución (a)		Distribución (b)		Distribución ©	
x_i	$\lvert x_i - \bar{x} \rvert$	x_i	$\lvert x_i - \bar{x} \rvert$	x_i	$\lvert x_i - \bar{x} \rvert$
15	0	18	3	26	5
15	0	16	1	21	0
15	0	14	1	17	4
15	0	19	4	36	15
15	0	17	2	24	3
15	0	10	5	22	1
15	0	11	4	12	9
15	0	15	0	10	11
$\bar{X} = \sum_{i=1}^{8} \frac{x_i}{n} = 15$	$\sum_{i=1}^{8} \lvert \bar{X} - X \rvert = 0$	$\bar{X} = \sum_{i=1}^{8} \frac{x_i}{n} = 15$	$\sum_{i=1}^{8} \lvert x_i - \bar{x} \rvert = 20$	$\bar{X} = \sum_{i=1}^{8} \frac{x_i}{n} = 21$	$\sum_{i=1}^{8} \lvert \bar{X} - X \rvert = 48$
$DM = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{n} = \frac{0}{15} = 0$		$DM = \frac{\sum_{i=1}^{n} \lvert \bar{x} - x_i \rvert}{n} = \frac{20}{15} = 1.33$		$DM = \frac{\sum_{i=1}^{n} \lvert \bar{x} - x_i \rvert}{n} = \frac{48}{15} = 3.3$	
$Rango = 15 - 15 = 0$		$Rango = 19 - 10 = 9$		$Rango = 36 - 10 = 26$	

III.1.5. Varianza y Desviación Estándar.

Otra medida de dispersión es la que eleva cada una de las desviaciones al cuadrado (en vez de valores absolutos) $(x_i - \mu)^2$ y después se suman $\left[\sum_{i=1}^{n} (x_i - \mu)^2 \right]$, y por último el resultado se divide entre el tamaño de la población. A este número se le conoce como ***varianza poblacional***.[13]

$$\left[\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{n} \right]$$

[13] La varianza muestral tiene una pequeña diferencia ante la varianza poblacional, ya que la fórmula cambia en su divisor al considerar el tamaño de la muestra menos uno, en vez del tamaño de la muestra.

$$\text{Varianza muestral} = \sum_{i=1}^{n} \frac{(X - \mu)^2}{n - 1}$$

Por ejemplo:

Datos x_i	$\lvert x_i - \bar{x} \rvert$	$\left(x_i - \bar{x}\right)^2$
22	0.4	0.16
13	-8.6	73.96
33	11.4	129.96
16	-5.6	31.36
19	-2.6	6.76
25	3.4	11.56
16	-5.6	31.36
23	1.4	1.96
29	7.4	54.76
35	13.4	179.56
17	-4.6	21.16
19	-2.6	6.76
20	-1.6	2.56
23	1.4	1.96
14	-7.6	57.76
$\sum_{i=1}^{15} x_i = 324$		$\sum_{i=1}^{15} \left(x_i - \bar{x}\right)^2 = 611.6$
$X = \sum_{i=1}^{15} \frac{x_i}{15} = 21.6$		$\sigma^2 = \sum_{i=1}^{15} \frac{\left(x_i - \bar{x}\right)^2}{n} = \frac{611.6}{15} = 40.77333$

La varianza es: 40.77333

El concepto de varianza puede ser útil cuando los datos corresponden a mediciones de un experimento. Si estas mediciones difieren poco entre sí, la media también sería un valor muy cercano a los datos y esto se vería reflejado en su varianza. Por el contrario, si la varianza es un número muy grande reflejaría que los datos están muy alejados entre sí. La operación de elevar al cuadrado la varianza no está expresada en las unidades originales, sino en una unidad elevada al cuadrado. Debido a esto, es necesario restaurar la unidad original, obteniendo la raíz cuadrada. La medida así obtenida se conoce como ***desviación estándar.***

La desviación estándar es la raíz cuadrada de la varianza.[14]

$$\sigma = \text{Desviación estándar poblacional} = \sqrt{Varianza \text{ poblacional}} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}}$$

Del ejercicio anterior obtenemos la desviación estándar:

Distribución		
x_i	$\lvert x_i - \bar{x} \rvert$	$(x_i - \bar{x})^2$
18	3	9
16	1	1
14	1	1
19	4	16
17	2	4
10	5	25
11	4	16
15	0	0
$\bar{X} = \sum_{i=1}^{8} \frac{x_i}{n} = 15$	$\sum_{i=1}^{8} \lvert x_i - \bar{x} \rvert = 20$	$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 72$
$DM = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{n} = \frac{20}{15} = 1..33$		*Varianza* $\sigma^2 = \frac{\sum_{i=1}^{8} (x_i - \bar{x})^2}{n} = \frac{72}{15} = 4.8$
		Desviación Estándar $\sigma = \sqrt{\sigma^2} = \sqrt{4.8} = 2.19$

14

$$\sigma = \text{Desviación estándar muestral} = \sqrt{Varianza \text{ muestral}} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}$$

Obsérvese que, una desviación indica que tan distantes están los valores de la media. Si se considera el intervalo de la forma:

$$\left[\bar{x} - DM, \bar{x} + DM\right] = \left[15 - 1.33, 15 + 1.33\right] = \left[13.67, 16.33\right]$$

Obsérvese que, este intervalo contiene a tres valores de la muestra que son: 14, 15 y 16. En otras palabras, este intervalo abarcaría al 37.5% del total de la muestra.

En cambio, si generamos el mismo intervalo, considerando ahora la desviación estándar en vez de la desviación media, se tendría:

$$\left[x - \sigma,\ x + \sigma\right] = \left[15 - 2.19,\ 15 + 2.19\right] = \left[12.81,\ 17.19\right]$$

Este intervalo contiene a los valores 14, 15,16 y 17 que representan un 50% del total de la muestra.

Esta propiedad de la desviación estándar se presenta en general y hace que sea una medida más utilizada que la desviación media.

III.1.5.1. Desviación Estándar: Fórmula clásica y simplificada.

Anteriormente, se obtuvo la desviación estándar por la fórmula clásica, pero existe otra manera de calcularla por medio de la fórmula simplificada, que además, es la fórmula más utilizada:

$$\sigma = \text{Desviación estándar } = \frac{\sum_{i=1}^{n}(x_i)^2 - n(\bar{x})^2}{n}$$

En donde :

$\sum_{i=1}^{n} x_i$ = Suma de cada valor elevado al cuadrado

n = Tamaño de la muestra

$\bar{x}^2$ = *Media* al cuadrado

Se demostrará que:

$$\sigma^2 = \sum_{i=1}^{n}\frac{(x_i - \bar{x})^2}{n} = \sum_{i=1}^{n}\frac{(x_i)^2 - n(\bar{x})^2}{n}$$

Se tiene que:

$$\sigma^2 = \sum_{i=1}^{n}\frac{(x_i - \bar{x})^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Esto es debido, a las propiedades de la sumatoria (capitulo II), donde:

$$\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$$

c = constante.

Como se observa en la fórmula, aparece un binomio al cuadrado (recuérdese el desarrollo del binomio al cuadrado como: el cuadrado del primer término, más el doble producto del primer por el segundo término, más el cuadrado del segundo término, es decir, $(a + b)^2 = a^2 + 2ab + b^2$).

Continuando con la demostración, tenemos que:

$$\sigma^2=\sum_{i=1}^{n}\frac{\left(x_i-\bar{x}\right)^2}{n}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i\right)^2-2x_i\bar{x}+\left(\bar{x}\right)^2$$

Debido a la propiedad de la sumatoria que dice: $\sum_{i=1}^{n} x_i + y_i = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

Se tiene que:

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i\right)^2-2x_i\bar{x}+\left(\bar{x}\right)^2=\frac{1}{n}\left[\sum_{i=1}^{n}\left(x_i\right)^2-2\bar{x}\sum_{i=1}^{n}x_i+\bar{x}^2\sum_{i=1}^{n}1\right]$$

Dado que la media se define como: $\bar{x}=\dfrac{\sum_{i=1}^{n} x_i}{n}$

Sustituyendo queda:

$$\sigma^2=\frac{1}{n}\left[\sum_{i=1}^{n}(X_i)^2-2\frac{\sum_{i=1}^{n}X_i}{n}\sum_{i=1}^{n}X_i+n\frac{\left[\sum_{i=1}^{n}X_i\right]^2}{n^2}\right]=$$

$$\frac{1}{n}\left[\sum_{i=1}^{n}(X_i)^2-2\frac{\left[\sum_{i=1}^{n}X_i\right]^2}{n}+n\frac{\left[\sum_{i=1}^{n}X_i\right]^2}{n^2}\right]$$

En la última igualdad se observa una suma de fracciones, esto es:

$$-2\frac{\left[\sum_{i=1}^{n} x_i\right]^2}{n} + n\frac{\left[\sum_{i=1}^{n} x_i\right]^2}{n^2} = \frac{-2n\left[\sum_{i=1}^{n} x_i\right]^2 + n\left[\sum_{i=1}^{n} \bar{x}_i\right]^2}{n^2}$$

$$= \frac{\left[\sum_{i=1}^{n} x_i\right]^2 \times (-2n+n)}{n^2} = \frac{n\left[\sum_{i=1}^{n} x_i\right]^2}{n^2} = n\bar{x}^2$$

Debido a esto se tiene que:

$$\sigma^2 = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i)^2 - 2\frac{\left[\sum_{i=1}^{n} x_i\right]^2}{n} + n\frac{\left[\sum_{i=1}^{n} x_i\right]^2}{n^2}\right] = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i)^2 - n\overline{(x)^2}\right]$$

Por lo tanto, queda demostrado que:

$$\sigma^2 = \sum_{i=1}^{n}\frac{(x_i - \bar{x})^2}{n} = \sum_{i=1}^{n}\frac{(x_i)^2 - n(\bar{x})^2}{n}$$

Una vez demostrado se puede obtener la desviación estándar por dos fórmulas equivalentes:

$$\text{Desviación estándar} = \sigma = \sqrt{\sum_{i=1}^{n}\frac{(x_i - \bar{x})^2}{n}} = \sqrt{\sum_{i=1}^{n}\frac{(x_i)^2 - n(\bar{x})^2}{n}}$$

III.2. Medidas de Dispersión para datos agrupados

Las medidas de dispersión, al igual que las de tendencia central para datos agrupados, son sólo una aproximación.

III.2.1. Rango, varianza y desviación estándar

El rango, la varianza y la desviación estándar, para datos agrupados, no cambian en la estructura de su definición, lo que cambia es la manera de obtenerse, por ejemplo:

Clase	Límite inferior	Límite superior	Marca de Clase	Frecuencia absoluta	Frecuencia Relativa	Frecuencia absoluta acumulada	Frecuencia complementaria
1	70	95	82.5	3	0.15	3	17
2	100	125	112.5	5	0.25	8	12
3	130	155	142.5	6	0.3	14	6
4	160	185	172.5	6	0.3	20	0

El **rango** se obtiene de la siguiente forma:

$$\begin{aligned} &Rango \\ &\text{Límite Superior de la última clase - Límite inferior de la primera clase} \\ &= 185 - 70 = 115 \end{aligned}$$

La **varianza** se calcula de la siguiente manera:

$$Varianza = \frac{Suma\left[(Marca - Media)^2 \times Frecuencia\ de\ clase\right]}{Suma\ de\ las\ frecuencias\ absolutas}$$

$$Varianza = \sigma^2 = \frac{\sum_{i=1}^{n} \left(M_i \times \bar{x}\right)^2 \times f\left(x_i\right)}{\sum_{i=1}^{n} f\left(x_i\right)}$$

$$\sigma = Desviación \quad estándar = \sqrt{\operatorname{var} ianza} = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{n} \left(M_i - \bar{x}\right)^2 \times f\left(x_i\right)}{\sum_{i=1}^{n} f\left(x_i\right)}}$$

Recuérdese que la fórmula de la media es:

$$Media = \frac{Suma\ (\ Marce\ de\ clase \times frecuencia\ de\ cada\ clase)}{Suma\ de\ las\ \ frecuencias\ absolutas}$$

$$Media = \overline{X} = \frac{\sum_{i=1}^{n} (\mathrm{M}_{\mathrm{i}} \times f\,(x_i))}{\sum_{i=1}^{n} f\,(x_i)}$$

La solución al ejercicio es:

$$\overline{X} = \frac{(82.5 \times 3) + (112.5 \times 5) + (142.5 \times 6) + (172.5 \times 6)}{20} = \frac{2700}{20} = 135$$

La media es 135.

La varianza es:

$$\sigma^2 = \frac{\left[(82.5 - 135)^2 \times 3\right] + \left[(112.5 - 135)^2 \times 5\right] + \left[(142.5 - 135)^2 \times 6\right] + \left[(172.5 - 135)^2 \times 6\right]}{20}$$

$$\sigma^2 = \frac{8268.75 + 2531.25 + 337.5 + 8437.5}{20} = \frac{19575}{20} = 978.75$$

Dado que la desviación estándar es la raíz cuadrada de la varianza:

$$Desviación\ \ estándar = \sigma = \sqrt{var\ ianza} = \sqrt{\sigma^2} = \sqrt{978.5} = 31.28$$

Al igual que en los datos individuales la varianza tiene otra forma de calcularse como:

$$Varianza = \frac{Suma\ [Marca\ de\ clase \times frecuencia\ de\ la\ clase] - suma\ de\ las\ frecuencias\ absolutas \times media\ al\ cuadrado}{Suma\ de\ las\ frecuencias\ absolutas}$$

$$Varianza = \sigma^2 = \frac{\sum_{i=1}^{n}(M_i^2 \times f_i) - \sum f_i \times \overline{X}^2}{\sum_{i=1}^{n} f_i}$$

$$\sigma^2 = \frac{\left(82.5^2 \times 3 + 112.5^2 \times 5 + 142.5^2 \times 6 + 172.5 \times 6\right) - \left(20 \times 135^2\right)}{20}$$

$$\sigma^2 = \frac{20418.75 + 63281.25 + 121837.5 + 178537.5 - 364500}{20} = \frac{19575}{20} = 978.75$$

La desviación estándar es 31.28.

III.2.2. Significado y uso de la desviación estándar.

Junto con la media aritmética, la desviación estándar es el parámetro más usado en estadística. La desviación estándar es una forma de medir el grado de dispersión de los datos alrededor de la media.

Veamos el siguiente ejemplo para comprender la utilidad de la desviación estándar:

A un grupo de 265 personas se les aplicó un examen de 30 preguntas obteniendo los siguientes resultados:

Número de preguntas contestadas correctamente	No. de personas. (frecuencia)
0	2
1	3
2	4
3	5
4	6
5	7
6	6
7	8
8	9
9	10
10	11
11	9
12	12
13	13
14	15
15	13
16	12
17	14
18	12
19	11
20	10
21	13
22	10
23	9
24	8
25	8
26	7
27	6
28	5
29	3
30	4

Si se quisiera saber cuál fue el promedio de preguntas correctas por persona, y se tuviera el número de aciertos por cada una de ellas, el problema es muy sencillo, se resuelve de la manera tradicional. Pero, si los datos se presentan como el número de personas que tienen cierta cantidad de aciertos, al hacer la suma de las frecuencias entre el número de datos, se puede incurrir en un error (de hecho este valor siempre será el mismo sin importar cómo estén distribuidas las frecuencias).

Ahora, calculando la media aritmética para estos datos se tiene que:

Número de preguntas contestadas correctamente	No. de personas. (frecuencia)
0	2
1	3
2	4
3	5
4	6
5	7
6	6
7	8
8	9
9	10
10	11
11	9
12	12
13	13
14	15
15	13
16	12
17	14
18	12
19	11
20	10
21	13
22	10
23	9
24	8
25	8
26	7
27	6
28	5
29	3
30	4
n=31	$\sum_{i=0}^{31} f(x_i) = 265$
$\bar{x} = \frac{\sum f(x_i)}{n} = \frac{265}{31} = 8.54$	

Cambiemos las frecuencias

Número de preguntas contestadas correctamente	No. de personas. (frecuencia)
0	15
1	12
2	10
3	5
4	6
5	7
6	10
7	8
8	9
9	4
10	11
11	9
12	12
13	13
14	2
15	13
16	3
17	4
18	7
19	11
20	6
21	5
22	8
23	9
24	8
25	10
26	12
27	6
28	13
29	3
30	14
n=31	$\sum_{i=0}^{31} f(x_i) = 265$
$\bar{x} = \frac{\sum f(x_i)}{n} = \frac{265}{31} = 8.54$	

Para encontrar la respuesta correcta a nuestra pregunta y obtener el número de aciertos, simplemente tendría que sumar el número de aciertos de cada persona entre el total de personas.

La primera tabla indica: 2 personas obtuvieron cero aciertos; tres personas obtuvieron tres aciertos; dos personas obtuvieron cuatro aciertos; y así sucesivamente.

Si se desea saber el número total de aciertos habría que sumar

$$\underbrace{(0+0)+(1+1+1)+(2+2+2+2)+\ldots+(29+29+29)+(30+30+30+30)}_{265 \text{ sumandos}}$$

Esta suma tiene 265 sumandos, los mismos sumandos que tendría la suma de los aciertos de cada una de las personas.

Ahora, se puede ver que:

$(0+0)+(1+1+1)+(2+2+2+2)+\ldots+(29+29+29)+(30+30+30+30)$

$=2(0)+3(1)+4(2)+\ldots+3(29)+4(30)$ y esto es equivalente a multiplicar el número de aciertos por el número de personas que lo obtuvieron, como se desarrolla en el siguiente cuadro:

Número de preguntas contestadas correctamente x_i	No. de personas. (frecuencia) $f(x_i)$	$f(x_i)\times x_i$	x_i^2	$x_i^2 f(x_i)$
0	2	0	0	0
1	3	3	1	3
2	4	8	4	16
3	5	15	9	45
4	6	24	16	96
5	7	35	25	175
6	6	36	36	216
7	8	56	49	392
8	9	72	64	576
9	10	90	81	810
10	11	110	100	1100
11	9	99	121	1089
12	12	144	144	1728
13	13	169	169	2197
14	15	210	196	2940
15	13	195	225	2925
16	12	192	256	3072
17	14	238	289	4046
18	12	216	324	3888
19	11	209	361	3971
20	10	100	400	4000
21	13	273	441	5733
22	10	220	484	4840
23	9	207	529	4761
24	8	192	576	4608
25	8	200	625	5000
26	7	182	676	4732
27	6	162	729	4374
28	5	140	784	3920
29	3	87	841	2523
30	4	120	900	3600
	$\sum f(x_i) = 265$	$\sum x_i \times f(x_i) = 4104$		$\sum x_i^2 f(x_i) = 7736$
		$\bar{x} = \frac{\sum x_i \times f(x_i)}{\sum f(x_i)} = \frac{4104}{265}$ $\bar{x} = 15.487$		
$\sigma^2 = \frac{\sum x_i^2 (f(x_i)) - n\bar{x}^2}{n} = \frac{7736 - [265 \times (15.487)^2]}{265} = 18.079$				
$\sigma = \sqrt{18.079}2595 = 4.252$				

Obsérvese que, en este procedimiento sí es importante la distribución de las frecuencias. Si obtenemos el mismo cálculo para la segunda tabla el resultado es diferente:

Número de preguntas contestadas correctamente x_i	No. de personas. (frecuencia) $f(x_i)$	$f(x_i)\times x_i$	x_i^2	$x_i^2 f(x_i)$
0	15	0	0	0
1	12	12	1	12
2	10	20	4	40
3	5	15	9	45
4	6	24	16	96
5	7	35	25	175
6	10	60	36	360
7	8	56	49	392
8	9	72	64	576
9	4	36	81	324
10	11	110	100	1100
11	9	99	121	1089
12	12	144	144	1728
13	13	169	169	2197
14	2	28	196	392
15	13	195	225	2925
16	3	48	256	768
17	4	68	289	1156
18	7	126	324	2268
19	11	209	361	3971
20	6	120	400	2400
21	5	105	441	2205
22	8	176	484	3872
23	9	207	529	4761
24	8	192	576	4608
25	10	250	625	6250
26	12	312	676	8112
27	6	162	729	4374
28	13	364	784	10192
29	3	87	841	2523
30	14	420	900	12600
	$\sum f(x_i) = 265$	$\sum x_i \times f(x_i) = 3921$		$\sum x_i^2 f(x_i) = 81511$
		$\bar{x} = \frac{\sum x_i \times f(x_i)}{\sum f(x_i)} = \frac{3921}{265}$ $\bar{x} = 14.79622642$		

$$\sigma^2 = \frac{\sum x_i^2 (f(x_i)) - n\bar{x}^2}{n} = \frac{81511 - \left[265 \times (14.796)^2\right]}{265} = \frac{23496.77176}{265} = 88.66706325$$

$$\sigma = \sqrt{88.66706325} = 9.416318986$$

Ahora nos podemos preguntar: ¿Cuántas desviaciones estándar (atrás de la media y delante de ella) se necesitan para generar un intervalo que contenga el 75% de los datos?

El porcentaje de los datos se encuentra en un intervalo de:

$$\bar{x} - 1.5\sigma = 15.487 - 2(4.252) = 6.93$$
$$\bar{x} + 1.5\sigma = 15.487 + 2(4.252) = 23.991$$

Dado que, no existe ningún dato de 6.93, lo podríamos aproximar a 7 sin riesgo de cometer error, al igual que con 23.991 se aproxima a 24.

Se generó el intervalo de los datos (7,24), para poder obtener que porcentaje de los datos se encuentran en este intervalo, sumaremos las frecuencias:

Número de preguntas contestadas correctamente	No. de personas (frecuencia)
7	8
8	9
9	10
10	11
11	9
12	12
13	13
14	15
15	13
16	12
17	14
18	12
19	11
20	10
21	13
22	10
23	9
24	8
	$\sum f(x_i) = 199$

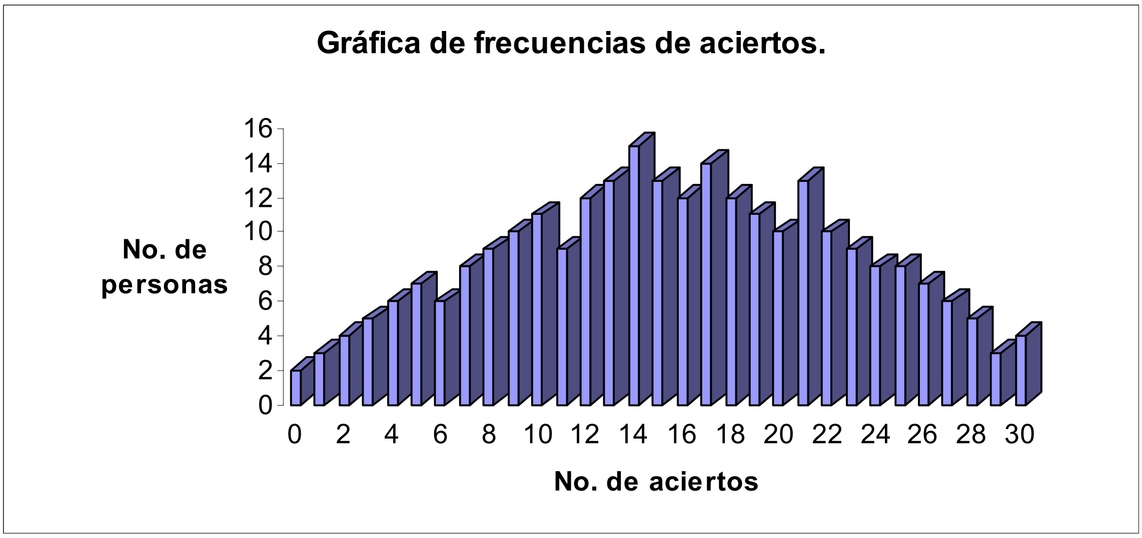
Obteniendo el porcentaje por medio de una regla de tres, tenemos:

$$\frac{265}{199} = \frac{100\%}{x\%} = \frac{199 \times 100\%}{265} = \frac{19900\%}{265} = 75.09\%$$

La desviación estándar puede proporcionar información interesante sobre todo tipo de poblaciones. Ésta característica, que acabamos de obtener, es un resultado general que fue descubierto por el matemático ruso Pafnuty Lvovich Chevyshev, y lo enunció en el siguiente teorema:

Teorema de Chevyshev: Para cualquier conjunto de datos que tengan media μ y desviación estándar σ, podemos esperar que en un intervalo de la forma $(\mu - k\sigma, \mu + k\sigma)$ el porcentaje de datos que están contenidos en este intervalo es aproximadamente $1 - \frac{1}{k^2}$ donde k>1.

Observemos el gráfico que dibuja la siguiente distribución.



Aplicando el teorema de Chevyshev, se espera que la dispersión de los datos con respecto a la media, en términos de las desviaciones estándar sean las siguientes:

Por lo menos el 68% de los datos bajo la curva caerán en el intervalo (μ-1.5σ,μ+1.5σ)

a) Por lo menos el 75% de los datos bajo la curva caerán en el intervalo (μ-2σ,μ+2σ)

b) Por lo menos el 89% de los datos bajo la curva caerán en el intervalo (μ-3σ,μ+3σ)

c) Por lo menos el 94% de los datos bajo la curva caerán en el intervalo (μ-4σ,μ+4σ)

La utilidad del teorema de Chevyshev proviene del hecho que requiere pocos conocimientos acerca del conjunto de datos. Por ejemplo:

El promedio de calificaciones en el curso de Estadística es de 6.6 con un desviación estándar de 1.4. Por medio del teorema de Chevyshev obtengamos:

a) Un intervalo que tenga al menos el 75% de las calificaciones.

b) El porcentaje de alumnos que obtienen calificaciones en un intervalo de 6 a 9.

En el primer caso se espera que los datos se dispersen alrededor de la media en un 75% ,es decir:

$$1 - \frac{1}{k^2} = 75\% \quad es\ decir\ \ 1 - \frac{1}{k^2} = 0.75$$

Despejando a K se obtiene lo siguiente:

$$\frac{1}{k^2} = 1 - 0.75 \text{ esto es } \frac{1}{k^2} = 0.25, \text{ entonces } \frac{1}{k^2} = 0.25 \text{ esto implica que}$$

$$k^2 = \frac{1}{0.25} = 4 \text{ lo que significa que } k^2 = 4$$

$$\textit{por} \text{ lo que, } k = \pm\sqrt{4} \text{ entonces k puede obtener los valores de - 2 y + 2}$$

El intervalo de calificaciones se determina:

$$\mu \pm k\sigma = 6.6 \pm 2(1.4) = 6.6 \pm 2.8$$

El límite inferior de nuestro intervalo es igual a 6.6 –2.8=3.6

Y el límite superior es igual a 6.6 +2.8= 9.4

Lo que nos muestra que el 75% de los alumnos obtienen calificaciones de 3.6 a 9.4.

Para el segundo caso, se conoce que:

$$\mu - k\sigma = 6 \qquad (1)$$
$$\mu + k\sigma = 9 \qquad (2)$$

Resolviendo el siguiente sistema de ecuaciones por el método de suma y resta; primero se multiplica por negativo uno, la ecuación (1) queda:

$$-\mu + k\sigma = -6 \qquad (1)$$
$$\mu + k\sigma = 9 \qquad (2)$$

Después, sumando la ecuación (1)a la ecuación (2) queda:

$$2k\sigma = 3 \text{ pero } \sigma = 1.4 \text{ entonces } 2(1.4)\text{k} = 3 \quad \text{lo que resulta } 2.8\text{k} = 3 \text{ por lo que}$$

$$\text{k} = \frac{3}{2.8} = 1.0714$$

Utilizando el teorema de Chevyshev:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{1.0714^2} = 1 - \frac{1}{1.1479} = 1 - 0.8712 = 0.1288$$

Lo que nos indica que el porcentaje de alumnos que tienen calificaciones entre 6 y 9 es del 12.88% y este se obtuvo de multiplicar el resultado del teorema de Chevyshev por 100.

III.3. Coeficiente y porcentaje de variación

El coeficiente de variación mide el promedio de las distancias de los datos entre sí (el grado de dispersión entre los datos o que tan separados están los datos entre sí). Este coeficiente no representa unidades (kilogramos, metros, etc.)

Esta medida se obtiene de dividir la desviación estándar entre la media.

$$Coeficiente\ de\ variación = \frac{Desviación\ estándar}{Media} = \frac{\sigma}{\bar{x}}$$

El **porcentaje de variación** sólo se obtiene de multiplicar el coeficiente de variación por 100. Esta medida indica el porcentaje que representa la desviación estándar conforme a la media.

Ejemplo: Determinemos el coeficiente y el porcentaje de variación de los siguientes datos:12, 25, 35,19, 27, 23, 26, 28, 39, 12.

Datos x_i	$\lvert x_i - \bar{x} \rvert$	$(\lvert x_i - \bar{x} \rvert)^2$
12	12.6	158.6
25	0.4	0.16
35	10.4	108.16
19	5.6	31.36
27	2.4	5.76
23	1.6	2.56
26	1.4	1.96
28	3.4	11.56
39	14.4	207.36
12	12.6	158.76
$\sum_{i=1}^{10} x_i = 246$		$\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert^2 = 686.4$
$\bar{x} = \sum_{i=1}^{10} \frac{x_i}{10} = \frac{246}{10} = 24.6$		$\sigma = \sqrt{\sum_{i=1}^{10} \frac{(\lvert x_i - \bar{x} \rvert)^2}{10}} = \sqrt{6864}$ $= 8.285$

La media es igual a 24.6 y la desviación estándar es 8.285.

El coeficiente de variación es:

$$\text{Coeficiente de variación} = \frac{\text{Media}}{\textit{Desviación estándar}}$$

$$= \frac{\sigma}{\bar{x}} = \frac{8.285}{24.6} = 0.3368$$

$$\text{Porcentaje de variación} = \text{Coeficient e de variación} \times 100$$
$$= 0.3368 \times 100 = 33.68\%$$

En el siguiente ejercicio se presentan dos muestras con la misma media pero con datos diferentes.

Muestra A

5,7,9,11,13 en esta muestra n=5

Muestra B

1,5,9,13,17al igual que la muestra anterior n=5

Se calcularán la media y la desviación estándar de estás dos muestras para obtener el coeficiente de variación.

Muestra A		
Datos x_i	$x_i - \bar{x}$	$\left(x_i - \bar{x}\right)^2$
5	-4	16
7	-2	4
9	0	0
11	2	4
13	4	16
$\sum_{i=1}^{5} x_i = 45$		$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 40$
$\bar{x} = \sum_{i=1}^{5} \frac{x_i}{5} = \frac{45}{5} = 9$		$\sigma = \sqrt{\sum_{i=1}^{5} \frac{(x_i - \bar{x})^2}{5}} = \sqrt{4.444} = 2.108$
Coeficiente de variación $\frac{\sigma}{\bar{x}} = \frac{2.108}{9} = 0.2343$		
Porcentaje de variación Coeficiente de variación $\times 100$ =0.2343 $\times$100=23.43%		

Muestra B		
Datos x_i	$x_i - \bar{x}$	$\left(x_i - \bar{x}\right)^2$
1	-8	64
5	-4	36
9	0	0
11	4	36
17	8	64
$\sum_{i=1}^{5} x_i = 45$		$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 160$
$\bar{x} = \sum_{i=1}^{5} \frac{x_i}{5} = \frac{45}{5} = 9$		$\sigma = \sqrt{\sum_{i=1}^{5} \frac{(x_i - \bar{x})^2}{5}} = \sqrt{17.778} = 8.889$
Coeficiente de variación $\frac{\sigma}{\bar{x}} = \frac{8.889}{9} = 0.9877$		
Porcentaje de variación Coeficiente de variación $\times 100$ =0.8889 $\times$100=88.89%		

Observemos que aún cuando ambos datos tienen la misma media, el coeficiente de variación de la muestra B es mayor que el coeficiente de variación de la muestra A, debido a que, los datos de la muestra B guardan mayor distancia que los datos de la muestra A. Por ello, el coeficiente de variación determina que tan distantes están los datos uno de otro.

3.4. Estadística y Azar.

Los datos estadísticos también pueden ser útiles para analizar eventos que dependen del azar, y en muchas ocasiones, la ocurrencia de estos eventos obedecen a ciertas leyes.

Cuando se elaboran dos o más experimentos se puede observar que las frecuencias relativas de cada dato en un experimento y en otro son muy parecidas, la gran mayoría de los experimentos aleatorios (que dependen del azar) que tienen importancia práctica, exhiben cierta estabilidad. Debido a esto, se tratan de desarrollar modelos matemáticos del comportamiento de estos datos.

Para ello, se requiere la teoría de probabilidades, básica en la estadística matemática.

Por ejemplo, de acuerdo a las calificaciones que se obtienen en el bachillerato, un profesor de probabilidad desea obtener una estadística del rendimiento del grupo; lo que en realidad quiere obtener no es el número de alumnos en particular de la muestra, ni el orden en que los evaluó, sino el número de alumnos que obtuvieron cierta calificación, es decir, lo único que importa es el número de alumnos que obtuvieron 5, 6, y así sucesivamente en escala de 5 a 10. El evento que depende del azar es el número de alumnos que tienen una calificación “X”, que más adelante, en probabilidad, se le conocerá como variable aleatoria.

El profesor desconoce cuántos alumnos van a obtener una calificación “X”, pero, al obtener los resultados del número de alumnos que obtuvieron cierta calificación (frecuencia con que se obtuvo la calificación “x”), observó que estos resultados tenían un comportamiento regular en distintos grupos de alumnos.

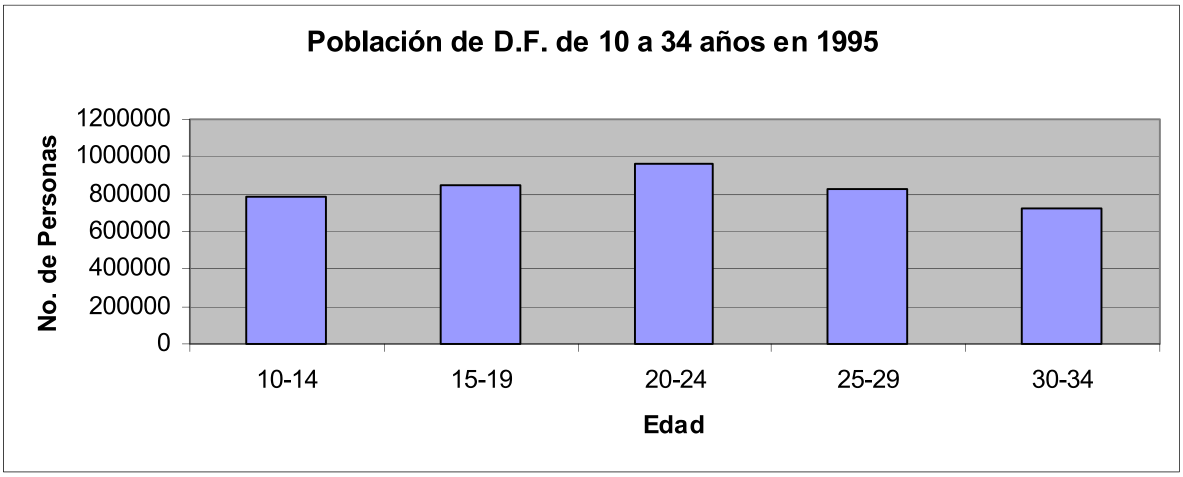
El comportamiento regular más conocido es la **distribución normal de frecuencias.**

Para comprender la distribución normal de frecuencias, en el siguiente ejemplo, graficaremos las frecuencias de edad entre los 10 a 39 años de un grupo representativo de los habitantes del Distrito Federal en 1995.

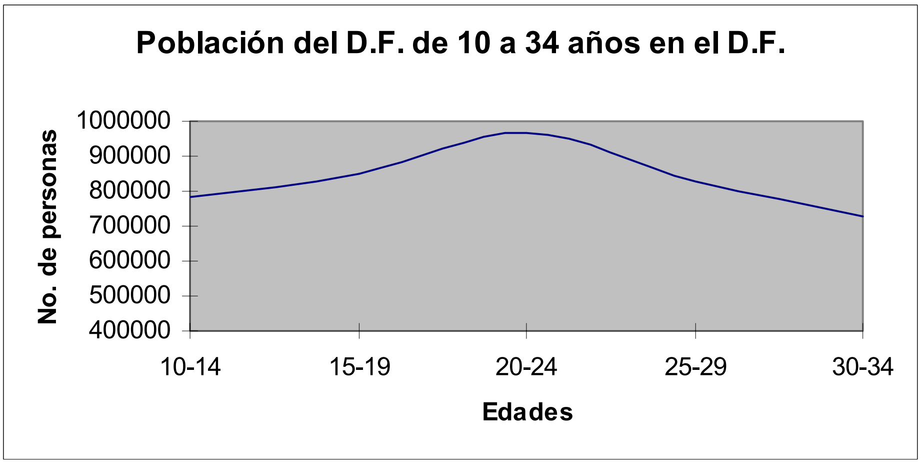
Los datos del grupo representativo son los siguientes:

Edades	Frecuencia
10-14	781 808
15-19	848 466
20-24	964 075
25-29	826 960
30-34	725 263
35-39	643 428

Graficando se tiene:



Censo de población y vivienda para el D.F. (INEGI, 1996)



Observemos que, estas gráficas parecieran tener la forma de una campana., donde su pico coincide con ser la media de la distribución. (Cabe hacer notar que la media se obtiene para datos agrupados).

Corroboremos que la media coincide con el pico de la campana de la siguiente manera:

Edades	Frecuencia	Marca de Clase	Frecuencia por Marca de clase.
10-14	781,808	12	9381696
15-19	848 466	17	14423072
20-24	964 075	22	21209650
25-29	826,960	27	22327920
30-34	725 263	32	23208416
	$\sum_{i=10}^{34} x_i = 4146522$		$\sum_{i=10}^{34} x_i f_i = 90550754$
$\mu = \frac{\sum_{i=10}^{34} x_i f_i}{\sum_{i=10}^{34} x_i} = \frac{90550754}{4146522} = 21.8377$			

La media se encuentra en la barra con la frecuencia más alta o con el pico de la campana.

La distribución normal se reconoce por su grafica de frecuencias que tiene las siguientes características:

- La parte más alta de la gráfica coincide con la media de los datos.
- La media =mediana =moda.
- Es simétrica alrededor de la media.
- Las frecuencias de los datos disminuyen conforme se consideran datos que están mas alejados de la media, y esto pasa tanto en datos menores o mayores que la media. Por esta razón, el gráfico de frecuencias tiene una forma de campana.

La distribución normal de frecuencias es un modelo matemático del comportamiento de los datos, pero existen muchas más distribuciones que modelan el comportamiento de éstos.

Para conocer más acerca de las diferentes distribuciones de datos es necesario conocer la teoría de probabilidad, que nos ayuda, a través de las distribuciones, a modelar los diversos comportamientos de los datos, para predecir la ocurrencia de un evento.

Ejercicios Propuestos

1. Calcule el rango, la desviación media, la varianza y la desviación estándar de los siguiente valores correspondiente a la velocidad promedio (Kilómetros por hora) que llevan los coches en el Periférico a las 22:00 hrs, a su paso por un centro de monitoreo.

95	88	90
87	75	110
78	97	120
90	84	97
105	110	100

2. Los siguientes valores corresponden al peso en kilogramos de una muestra de mujeres de 16 años. Determina la desviación estándar de este grupo de personas.

59	45	42
56	50	48
55	52	52
53	51	54
48	54	49

Calcula el rango, varianza y la desviación estándar de los siguientes datos que se presentan en una tabla de frecuencias para datos agrupados.

3. Esta tabla representa la estatura de un grupo de 1000 personas mayores de 18 años, que fueron a consulta a la clínica 32 del IMSS, la semana pasada.

Clase	Limite inferior (centímetros)	Limite superior (centímetros)	Frecuencia
1	142	151	145
2	152	161	236
3	162	171	258
4	172	181	235
5	182	192	126

4. Esta tabla representa el número de presos que se encuentra en un penal en el estado de Campeche. El grupo está segmentado en edades para clasificar que área del penal que corresponde a cada individuo.

Clase	Limite inferior (años)	Limite superior (años)	Frecuencia
A	18	25	134
B	26	35	670
C	36	45	437
D	46	55	356
E	55	75	128

5. Los registros de una aerolínea demuestran que sus vuelos entre dos ciudades llegan en promedio 4.5 minutos tarde con una desviación estándar de 1.4 minutos. ¿Qué porcentaje de vuelos llegan entre

   a) 2.6 y 6.4 minutos de retraso.

   b) 1.6 y 12.4 minutos de retraso.

6. El gerente de personal de cierta empresa tiene registros diarios del número de empleados que asisten a su trabajo. El promedio de ausentes es de 5.5 y la desviación estándar de 2.5. Como hay muchos días con cero, uno o dos faltantes y solamente pocos días con más de diez, la distribución de frecuencias está muy sesgada. El gerente desea determinar el intervalo en el cual se localizan al 75% de los datos. Ayúdalo a determinarlo.

7. En dieciséis días, un restaurante tuvo los siguientes números de órdenes de pollo y bistec:

Pollo	46	55	43	48	54	65	36	40
	51	53	64	32	41	46	47	47
Bistec	39	41	25	30	46	36	37	23
	30	33	50	44	41	28	35	37

Calcule los dos coeficientes de variación para determinar el artículo para el que el número de órdenes es relativamente más variable.

CONCLUSIONES

La estadística y el cálculo de probabilidades, hoy en día, constituyen una rama independiente de la matemática, con aplicaciones en casi todas las actividades humanas, incluyendo a fenómenos no medibles como la lingüística y la literatura.

De esta manera, el estudio de la Estadística y la Probabilidad es relevante en la formación integral del estudiante del Nivel de Enseñanza Media Superior y como materia dentro del Sistema Educativo de la Escuela Nacional Preparatoria de la UNAM, en particular, y como parte del Sistema Educativo Mexicano en general.

La asignatura de Estadística y Probabilidad en el Nivel de Enseñanza Media Superior proporciona las bases teórico-metodológicas necesarias para que pueda continuar su formación profesional o como etapa terminal de sus estudios, proporcionándole la posibilidad de tomar decisiones frente a sus problemas cotidianos.

A partir de esta investigación, se puede concluir que la enseñanza de la estadística conlleva a nuevos paradigmas dentro del proceso enseñanza-aprendizaje, basándose en los conocimientos mínimos necesarios que se deben tener en este nivel de enseñanza, así como en la forma de transmitirlos al estudiante.

Considerando la secuencia de los contenidos del programa de la Escuela Nacional Preparatoria, se propone que se debe introducir al alumno de una manera sencilla a la Estadística, por medio de un lenguaje cotidiano, sin la formación estricta del lenguaje matemático, utilizando ejemplos e ilustraciones aunados a dar más de un método para la solución de problemas. Esto le permitirá al profesor y alumno relacionar la seriación temática de un curso de Estadística, así como de su utilidad.

Por lo anterior, la enseñanza del manejo de datos en Estadística, se presenta aquí a través del uso de tablas de frecuencias y gráficas a partir de problemas concretos de la vida cotidiana, lo anterior nos permite mostrar parte del lenguaje utilizado en Estadística Descriptiva , para que así el alumno pueda identificar, interpretar y comprender la importancia de cualquier tabla o gráfica que requiera el uso de datos estadísticos, así como, a partir de una colección de datos, desarrolle la habilidad de generar una tabla de frecuencias y bajo ella elegir la grafica que mejor represente el objetivo por alcanzar al iniciar este trabajo estadístico, y finalmente, poder hacer una inferencia sobre esos datos.

Por otra parte, cuando necesitamos encontrar un valor representativo de un conjunto de datos se hace uso de las medidas de tendencia central, que a su vez permiten ubicar a los datos de una muestra, con referencia a un valor central de ellos. Este tema se desarrolla en el capítulo II, en donde no sólo se presenta la fórmula que permite calcular a cada de estas medidas sino que, a través de ejemplos dirigidos, se va construyendo el concepto hasta deducir la fórmula que la identifica, planteamiento que rompe con el esquema que tradicionalmente se ha utilizado en la enseñanza de la Estadística en el nivel de Enseñanza Media Superior.

En muchas ocasiones, las medidas de tendencia central pueden generar un cuadro incompleto o erróneo de la información del conjunto que se analiza, debido a que éstas reducen el conjunto de datos a un valor representativo sin tomar en cuenta el comportamiento del conjunto general de datos. Por lo tanto, la enseñanza de las medidas de dispersión (Capitulo III) contribuye a tener una información más completa del conjunto de datos que se analiza, y pueden ayudar al alumno a reflexionar acerca del comportamiento de la población en estudio.

Es importante destacar el uso y significado de la desviación estándar a través del Teorema de Chevyshev (descrito en el capitulo III) porque nos ayuda a llegar a inferencias muy claras del comportamiento de la población o muestra.

Finalmente, en la última sección de este trabajo se dan las ideas básicas que permiten analizar eventos que dependen del azar.

Es importante destacar que la construcción del conocimiento de modelos matemáticos para representar el comportamiento de los datos, es parte de un curso de Probabilidad necesario en el nivel de Enseñanza Media Superior, en general, y en la Escuela Nacional Preparatoria en particular.

Este trabajo de investigación tiene la finalidad de que cualquier persona, con un conocimiento básico de la matemática, pueda entender y asimilar el conocimiento básico de la Estadística Descriptiva, así mismo pretende facilitar y clarificar la enseñanza en esta materia, sirviendo como material de soporte a los profesores que imparten esta materia o bien, como material de consulta para estudiantes que no hayan cursado la materia pero que requieren adquirir de manera autodidacta estos conocimientos básicos.

Es importante resaltar que la construcción de los conceptos básicos y sus aplicaciones en los ejercicios presentados en este trabajo, han sido probados a lo largo de mi experiencia docente, por lo que se puede afirmar que corresponden a las habilidades adquiridas por los alumnos de este nivel del bachillerato y permite que los alumnos adquieran las bases mínimas para su posterior desarrollo educativo en niveles superiores.

BIBLIOGRAFIA

1. Mendenhall, William, et al. *Estadística Matemática con Aplicaciones*. México, Grupo Editorial Iberoamerica, 1986. p.p. 750

2. Perelman, Y.I. Álgebra Recreativa. México, Ediciones de Cultura Popular,1985.p.p.252.

3. Perelman, Y.I. Matemáticas Recreativas. México, Mir Moscú, 1982.p.119-125

4. *Plan de desarrollo de la Escuela Nacional Preparatoria* 2002-2006 p. 11

5. *Programa de estudios de la Asignatura de Estadística y Probabilidad de la ENP. De la U.N.A.M. 2003-2004*

6. Sánchez Corona, Octavio*. Probabilidad y Estadística*. México, Mc Graw Hill, 2004 .p.p.303

7. Téllez Rojo, Martha María y Zamora Muñoz, Salvador. Año 2 "Estadística herramienta confiable". En: *¿Cómo Ves?* Año 2, Vol. (16), p.22-24

8. Perelló, Marcelino. Año 2000. "Eran pizarroncitos". En: *Ciencia y Desarrollo*. XXVI(150); p.90.

BIBLIOGRAFIA SUGERIDA

Para poder ampliar en el tema se sugiere la siguiente bibliografía.

1. Canavos, George G*. Probabilidad y Estadística, aplicaciones y métodos*. México, Mc. Graw Hill,1988.p.p 651

2. Chao, Lincoln C. *Introducción a la Estadística.* México, Continental, 1985. p.p. 66-103

3. Downie, N.M., et al., *Métodos estadísticos aplicados*. México, Harla, 1988.p.p.380

4. Fuenlabrada de la Vega Trucíos, Samuel, Probabilidad y Estadística, México, Mc. Graw Hill, 2002,p.p.251

5. Infante, G. Said, et al. *Métodos Estadísticos, un enfoque interdisciplinario*. México, Trillas, 1991.

6. Johnson, Robert. *Estadística Elemental*. México, Trillas, 1976.p.p.482

7. Kreyszing,Edwin, *Introducción a la Estadística. Principios y Métodos*. México, Limusa, 1985.p.p. 505

8. Levin, Jack, *Fundamentos de Estadística en la Investigación Social*. México, Harla, 1981. p.p.305.

9. Meyer, Paul L. *Probabilidad y Aplicaciones Estadísticas*. México, Addison-Wesley Iberoamericana, 1992. p.p.480.

10. Spiegel, Murria R. *Estadística*. México, Mc. Graw Hill, 1991.p.p.556

11. Willoughby, Stephen S. Probabilidad y Estadística. México, Cultural, 1981.p.p.212

A P É N D I C E

Un trabajo con datos reales

Este apéndice tiene como finalidad realizar un trabajo estadístico aplicando los conocimientos de los capítulos anteriores a través de datos reales.

Utilicemos los datos correspondientes a la Delegación Azcapotzalco, obtenidos de la muestra del Censo Nacional de Población 2000 donde se utilizarán la variable de ingreso por hogar es decir cuánto es el ingreso promedio mensual de un hogar que se encuentra en esta delegación y la variable teléfono donde se pregunta si cada hogar cuenta o no con línea telefónica.

Aquí se exponen las medidas tendencia central y de dispersión que se puedan derivar de estos datos y su explicación.

Es importante destacar que este trabajo se realizo con datos de una muestra poblacional, y para fines ilustrativos, se utilizaron estos datos como datos censales para así, poder utilizar parámetros poblacionales, que son los expuestos en este texto, por ejemplo:

$$\frac{\sum_{i=1}^{N} y_i}{N}$$

$$\begin{aligned}
&donde\\
&y_i = ingreso \quad del \quad hogar \quad i\\
&N = número \quad de \quad hogares \quad de \quad la \quad delegación \quad Azcapotzal \quad co
\end{aligned}$$

La manera correcta de trabajar los datos de una muestra es por medio de los estimadores de los parámetros poblacionales es decir el estimador para el parámetro anteriormente citado es:

$$\frac{\sum_{i=1}^{N} f_i y_i}{\sum_{i=1}^{N} f_i}$$

$$
\begin{aligned}
&donde:\\
&y_i = Ingreso\ del\ hogar\ i\\
&f_i = Factor\ de\ \exp ansión = \frac{1}{\pi_i}\\
&\pi_i = P(hogar\ i\ sea\ seleccionado\ en\ la\ muestra)
\end{aligned}
$$

Este factor f_i se encuentra grabado en la base de datos que se ha de bajar de la página del INEGI como se indica a continuación y se localiza con la palabra FACTOR.

La variable teléfono.

Los datos se consultaron a través de la página de Internet. De la siguiente forma:

Se ingresa a la página www.inegi.gob.mx. Y se transporta a características del censo, se selecciona la entidad Distrito Federal y el municipio elegido es el 02.

Se selecciona guardar y se crea un archivo dentro del directorio en el que se está trabajando y de ahí se descomprimen 3 archivos de los cuales se selecciona el archivo VHO_f01.dbf (hogares) y se importar desde Excel para ser trabajado.

Se utiliza la variable teléfono y se visualizan los valores 5, 6 y 9 donde el 5 reporta que ese hogar dispone de al menos una línea telefónica, si aparece un 6 quiere decir que este hogar no cuenta con línea telefónica y si es un 9 indica que se desconoce si el hogar cuenta o no con línea.

Para poder llevar a cabo el conteo de hogares con teléfono y sin este servicio se requiere reemplazar el valor 5 por el valor 1, el valor 6 por el valor 0 y en los valores 9 se reemplaza por la palabra Ignorar valor.

Ahora, necesitamos conocer cuántos hogares disponen o no de línea telefónica y en cuantos hogares se desconocía el tener o no línea telefónica. Para esto, se va a crear una tabla dinámica que aparece en el comando datos de Excel.

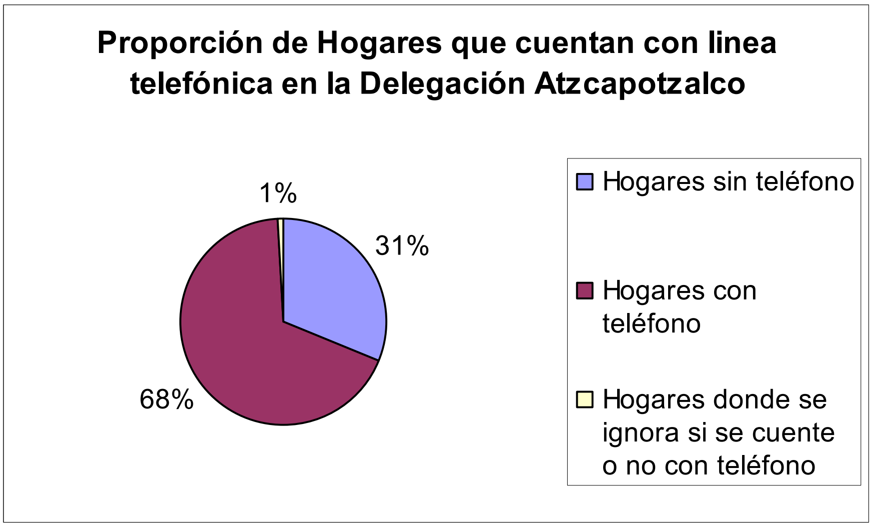
La tabla dinámica aparece de la siguiente manera:

Suma de cuenta	
TELEFONO	Total
0	3079
1	6755
Ignorar valor	69
Total general	9903

Es decir:

TELEFONO	Total
Hogares sin teléfono	3079
Hogares con teléfono	6755
Hogares donde se ignora o no con teléfono	69
Total general	9903

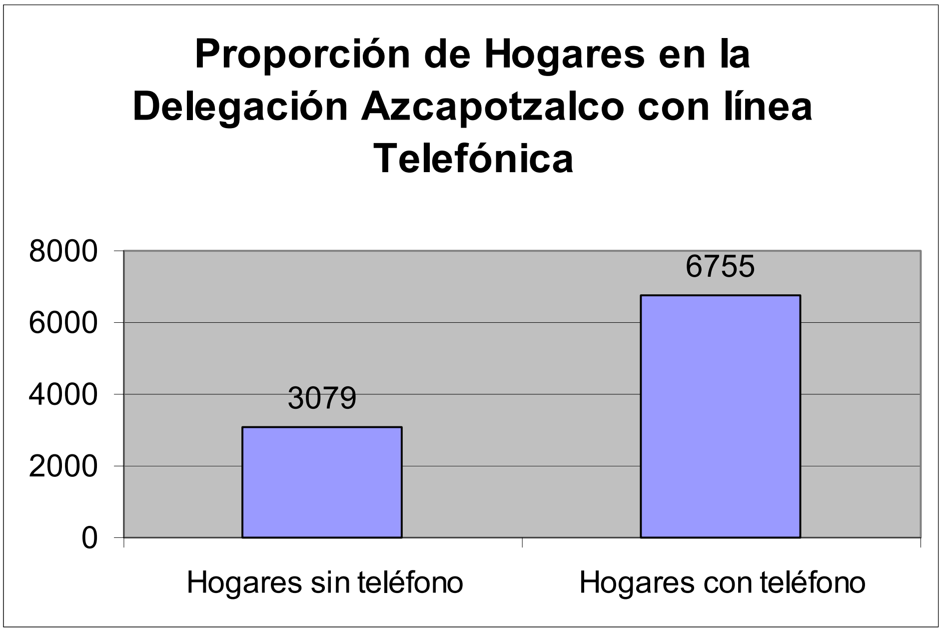
Si se elabora un gráfico de sectores con estos valores se tiene:



Cabe hacer notar que la variable teléfono es del tipo binomial, es decir, sólo tiene dos posibles respuestas, se cuenta o no con este servicio. Debido a ello, sólo se puede calcular la proporción de las personas que cuentan o no con línea telefónica y se debe anular la repuesta de ignorar valor ya que de alguna manera este hogar dispone o no de línea telefónica con línea telefónica. La tabla correspondiente a la variable teléfono ignorando los hogares que ignoran la disponibilidad de este servicio es:

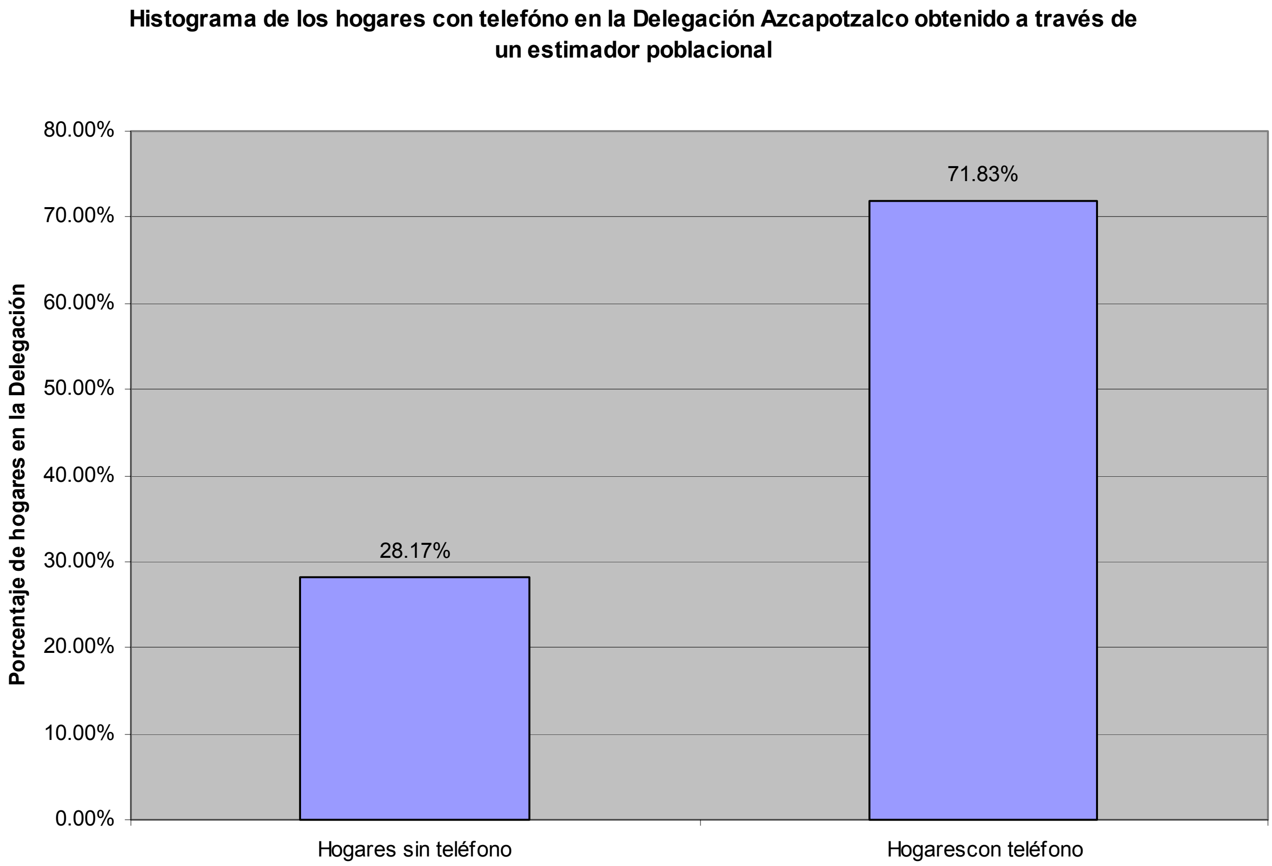
TELEFONO	**Total**	**Porcentaje**
Hogares sin teléfono	3079	31.3%
Hogares con teléfono	6755	68.7%
Total general	9834	100.0%

El histograma de esta variable es:



Para hacer el uso real de este parámetro, se va a obtener como un estimador Poblacional.

TELEFONO	Total (factor×ingreso) ($f_i \times y_i$)	Porcentaje
Hogares sin teléfono	31 111	28.17%
Hogares con teléfono	79 312	71.83%
Total general	110 423	100.0%



La variable ingreso por hogar:

Para trabajar con esta variable sigamos estos pasos:

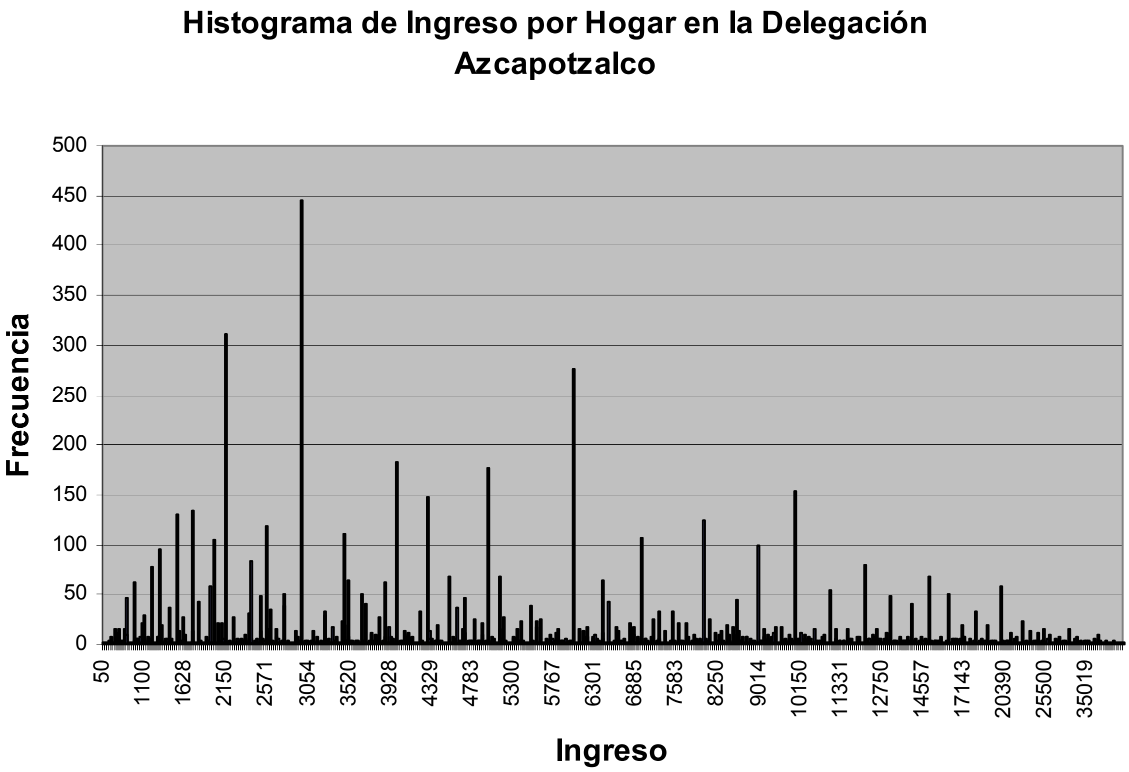
Crear una tabla dinámica con estos datos.

El resultado de esta operación va a ser los datos del ingreso por hogar ordenados de menor a mayor y el número de veces que aparece cada dato en la población.

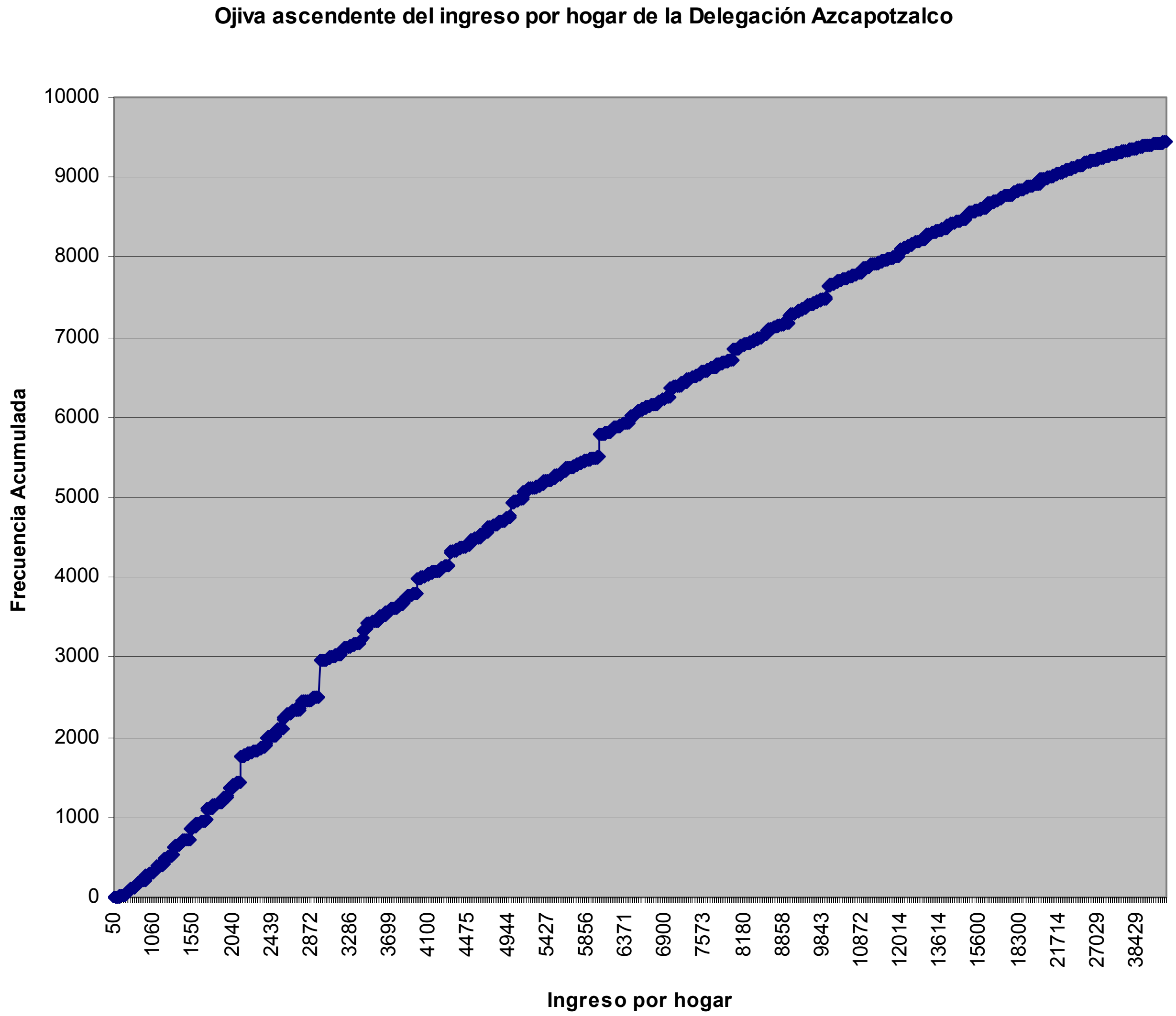
Con estos datos, ya ordenados, eliminemos los hogares con un ingreso “0”, ya que ningún hogar se puede solventar con este salario, esto podría indicar que estos hogares viven de la limosna o del ingreso de otros familiares al hogar y el ingreso de 999,999 ya que este se considera un dato no valido en la percepción del ingreso por hogar.

Una vez eliminados estos datos tracemos un histograma de frecuencias por medio de Excel, sombreando todos los datos y apretando el icono de gráficos.

La gráfica resultante es la que se muestra a continuación:



Se puede observar que está grafica tiene un comportamiento muy disperso, es decir es asimétrica, y que las medidas de tendencia central van a ser muy diferentes entre sí. También se observa que el último valor en el eje "X" es 35,019 pero el ingreso más alto registrado es de 999,998 pero después de este valor sólo existen 96 datos, pero estos no son considerados debido a que se distorsiona la gráfica.



Al igual que la gráfica anterior, se distingue que el último dato en el eje "x" es 53,142 debido que si se considerarán los siguientes datos, la gráfica se distorsionaría.

Calculemos cada una de estás medidas.

El rango medio

Sabemos que el rango medio se calcula como:

$$Rango\ medio = \frac{Lím.\,\mathrm{sup}.\ de\ la\ población - lím.\,\mathrm{inf}.\ \ de\ la\ pobación}{2} = \frac{999\,998 - 50}{2} = 499\,974$$

Es conocido, que esta medida es muy fácil de obtener, pero en este caso no índica un buen promedio de la población, ya que es conocido por todos que el promedio de ingreso en la Delegación Azcapotzalco no es de 499,974 esto se debe a que el límite superior e inferior tiene un rango de valores muy amplio.

La moda

La moda se define como el valor más frecuente dentro de un conjunto de datos y está se obtiene por una simple inspección de los datos. En el caso de los hogares de la Delegación Azcapotzalco la moda es de $3,000 con una frecuencia de 445, y en el histograma se percibe como la barra más alta. Recordemos que está medida indica que la mayoría de la población percibe un ingreso mensual de $3 000.

La media aritmética

Está se calcula multiplicando cada valor del ingreso por su frecuencia, para realizar todas las multiplicaciones se procede en Excel de la siguiente forma:

Se multiplica cada dato por su frecuencia. Es decir que, se multiplica $x_i \times (f(x_i))$

Para conocer el tamaño de la población que juega en esta variable se suman todas las frecuencias de los datos, ya que conocemos que la suma de las frecuencias absolutas es igual al tamaño de la muestra. En este caso el tamaño de la población fue de 9440 hogares.

Después, se tiene que sumar todos los valores de la columna $x_i \times (f(x_i))$. Esta suma da por resultado 69,592,954.

La media aritmética se calcula en este caso como:

$$\bar{x} = \frac{\sum_{i=1}^{9440} x_i \times f(x_i)}{n} = \frac{69592954}{9440} = 7372.13$$

Según está medida, podría indicar que en promedio los hogares de la Delegación Azcapotzalco, perciben $7,312, pero recordemos que la media aritmética se ve seriamente afectada si existen valores extremos, es decir muy distantes de la mayoría de los datos. En este caso, si existen valores extremos pues en la muestra se observan 8 valores que están en el rango de 100,000 hasta 999,998 pesos mensuales. Por ello, este valor no índica un buen promedio de ingreso de esta población.

Ahora el cálculo como un estimador poblacional:

$$\bar{x} = \frac{\sum_{i=1}^{9440} f_i \times y_i}{\sum_{i=1}^{9440} f_i} = \frac{1839707326}{249653} = 7369$$

La mediana

Para poder obtener la mediana correctamente, se tiene que observar si el tamaño de la muestra es un número par o impar. En este caso, el tamaño de la muestra fue 9,440 que es un número par.

Para calcular la posición de la mediana, previamente que nuestros datos se encuentran ordenados de menor a mayor se realiza lo siguiente:

Llamemos 2k=9,440, por consiguiente $k = \frac{9,440}{2} = 44720$

Localicemos los valores que se encuentran en la posición k y k+1, es decir en las posiciones 4720 y 4721.

Los valores que corresponden a estas posición 4,914 y 4,928, respectivamente.

Finalmente, la mediana se obtiene como el promedio de estos dos valores, esto es: $mediana = \frac{k+(k+1)}{2} = \frac{4,914+4,928}{2} = 4921$

La mediana es $ 4,921 es decir, el 50% de la población gana menos de $4,921. Cabe aclarar que este valor no se ve afectado por la presencia de valores extremos sino simplemente por el tamaño de la muestra.

Fractiles, (cuartiles y deciles)

Para dar una idea más precisa de los ingresos de la población realicemos el cálculo de los cuartiles y los deciles.

Los cuartiles:

Primer cuartil. (Q_1)

$$Posición\ \ de\ Q_1 = \frac{n+1}{4} = \frac{9440+1}{4} = 2360.25$$

Este valor se encuentra ubicado entre la posición 2360 y 2361, y estos valores corresponden en la población a 2,786 y 2,786 respectivamente, por lo que la posición del primer cuartil el 2,786. Esto nos indica que el 25% de los hogares de la Delegación Azcapotzalco tienen un ingreso menor o igual a $2,786 mensuales.

Segundo Cuartil

El segundo cuartil corresponde a la posición de la mediana y este tenia un valor de $4,921. Lo cual significa que el 50% de los hogares de esta Delegación perciben un ingreso menor o igual a $4,921 mensuales.

Tercer cuartil. (Q_3)

$$Posición\ \ de\ Q_3 = \frac{3(n+1)}{4} = \frac{3(9440+1)}{4} = \frac{28323}{4} = 7080.75$$

Este valor se encuentra ubicado entre la posición 7080 y 7081, y estos valores corresponden en la población a 8,572 y 8,572 respectivamente, por lo que la posición del tercer cuartil el 8,572. Esto nos indica que el 75% de los hogares de la Delegación Azcapotzalco tienen un ingreso menor o igual a $8,572 mensuales.

Los deciles:

Primer decil.

$$Posición\ \ de\ D_1 = \frac{n+1}{10} = \frac{9440+1}{10} = 944.1$$

Este valor se encuentra ubicado entre la posición 944 y 945, y estos valores corresponden en la población a 1,671 y 1,676 respectivamente, por lo que la posición del primer decil se encuentra a 1 décimo de la distancia entre 1676-.1671=5 y 5(.1)= 0.5 y este valor debe de sumarse a 1,671+0.5 =1,671.5. Esto implica que el 10% de los hogares de la Delegación Azcapotzalco perciben un ingreso menor o igual a 1,671.5.

Segundo decil.

$$Posición\ \ de\ D_2 = \frac{2(n+1)}{10} = \frac{2(9440+1)}{10} = \frac{18882}{10} = 1888.2$$

Este valor se encuentra ubicado entre la posición 1888 y 1889 y estos valores corresponden en la población a 2,357 y 2,357 respectivamente, por lo que la posición del segundo decil es 2,357. Lo cual indica que el 20% de los hogares de la Delegación Azcapotzalco perciben un ingreso menor o igual a $2,357.

Tercer decil.

$$Posición\ de\ D_3 = \frac{3(n+1)}{10} = \frac{3(9440+1)}{10} = \frac{28323}{10} = 2832.3$$

Este valor se encuentra ubicado entre la posición 2832 y 2833 y estos valores corresponden en la población a 2,743 y 2,743 respectivamente, por lo que la posición del tercer decil es 2,743. Lo cual indica que el 30% de los hogares de la Delegación Azcapotzalco perciben un ingreso menor o igual a $2,743.

Cuarto decil.

$$Posición\ de\ D_4 = \frac{4(n+1)}{10} = \frac{4(9440+1)}{10} = \frac{37764}{10} = 3776.4$$

Este valor se encuentra ubicado entre la posición 3776 y 3777 y estos valores corresponden en la población a 3,929 y 3,929 respectivamente, por lo que la posición del cuarto decil es 3,929. Lo cual indica que el 40% de los hogares de la Delegación Azcapotzalco perciben un ingreso menor o igual a $3,929.

Quinto decil.

El quinto decil al igual que el segundo cuartil corresponde al lugar de la mediana y esto es $4,921. Lo cual significa que el 50% de los hogares de esta Delegación perciben un ingreso menor o igual a $4,921 mensuales.

Sexto decil.

$$Posición\ de\ D_6 = \frac{6(n+1)}{10} = \frac{6(9440+1)}{10} = \frac{56646}{10} = 5664.6$$

Este valor se encuentra ubicado entre la posición 5664 y 5665 y estos valores corresponden en la población a 6,000 y 6,000 respectivamente, por lo que la posición del sexto decil es 6,000. Lo cual indica que el 60% de los hogares de la Delegación Azcapotzalco perciben un ingreso menor o igual a $6,000.

Séptimo decil.

$$Posición\ \ de\ D_7 = \frac{7(n+1)}{10} = \frac{7(9440+1)}{10} = \frac{66087}{10} = 6608.7$$

Este valor se encuentra ubicado entre la posición 6608 y 6609 y estos valores corresponden en la población a 7,629 y 7,629 respectivamente, por lo que la posición del séptimo decil es 7,629. Lo cual indica que el 70% de los hogares de la Delegación Azcapotzalco perciben un ingreso menor o igual a $7,629.

Octavo decil.

$$Posición\ \ de\ D_8 = \frac{8(n+1)}{10} = \frac{8(9440+1)}{10} = \frac{75528}{10} = 7552.8$$

Este valor se encuentra ubicado entre la posición 7552 y 7553 y estos valores corresponden en la población a 10,000 y 10,000 respectivamente, por lo que la posición del octavo decil es 10,000. Lo cual indica que el 80% de los hogares de la Delegación Azcapotzalco perciben un ingreso menor o igual a $10,000.

Noveno decil.

$$Posición\ \ de\ D_9 = \frac{9(n+1)}{10} = \frac{9(9440+1)}{10} = \frac{84969}{10} = 8496.9$$

Este valor se encuentra ubicado entre la posición 8496 y 8497 y estos valores corresponden en la población a 15,000 y 15,000 respectivamente, por lo que la posición del noveno decil es 15,000. Lo cual indica que el 90% de los hogares de la Delegación Azcapotzalco perciben un ingreso menor o igual a $15,000.

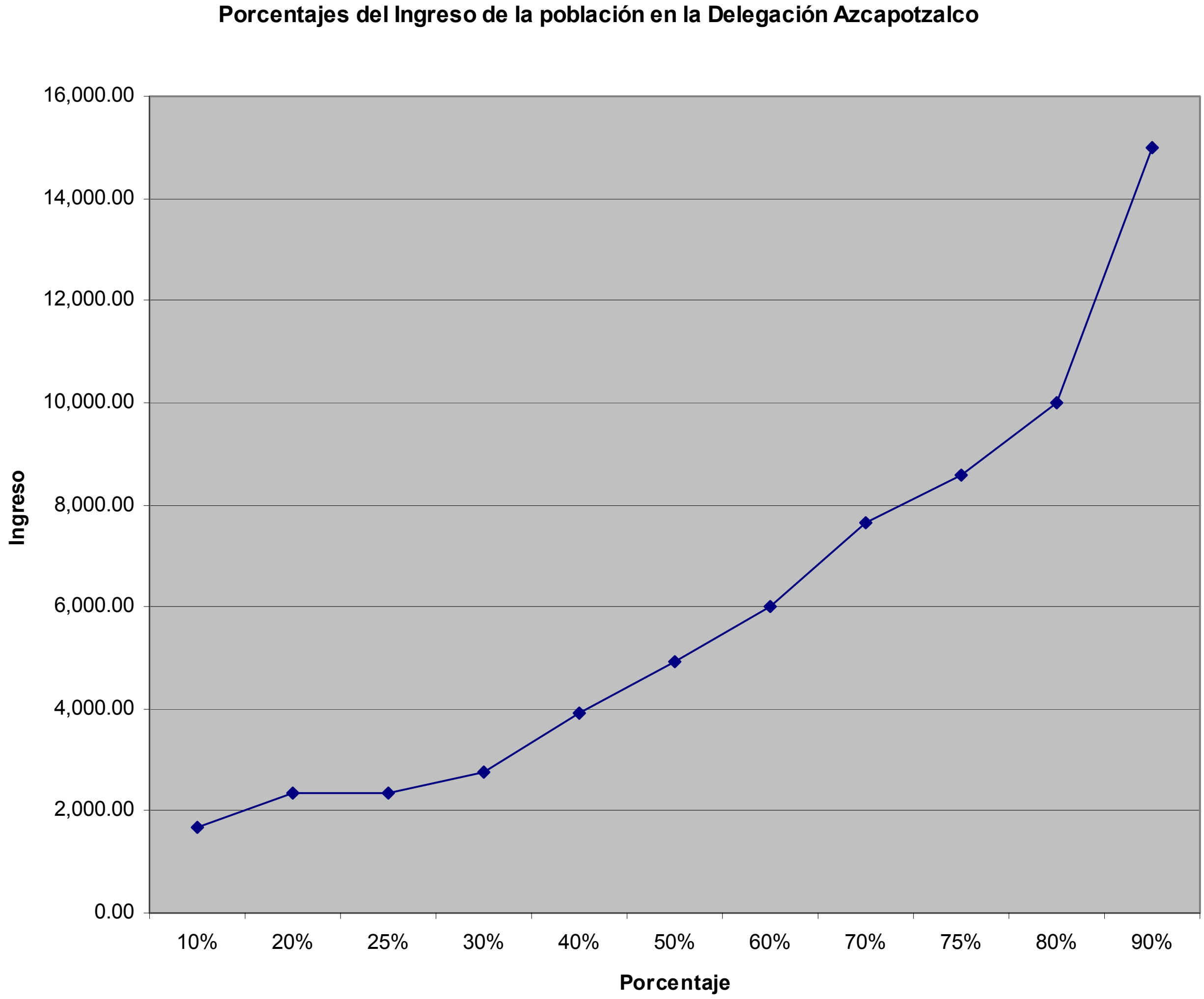
En resumen, se mostrarán los valores de las medidas de tendencia central en el siguiente cuadro:

Medidas de tendencia central	**Valor**	**Indica**
Rango medio	**499,974**	Es el punto medio del límite superior y el límite inferior de la población.
Moda	**3,000.00**	Indica que la mayoría de la población tiene un ingreso de $3,000.
Media Aritmética	**7,312.13**	Es el promedio aritmético de los ingresos en la Delegación Azcapotzalco
Mediana	**4,921.00**	Está indica que el 50% de los hogares(es decir 4720 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $4,921
1° Cuartil	**2360.25**	Este indica que el 25% de los hogares(es decir 2360 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $2360.25
3° Cuartil	**8572.00**	Este indica que el 75% de los hogares(es decir 7080 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $8752.00
1° Decil	**1,671.5.**	Este indica que el 10% de los hogares (es decir 994 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $1,671.5
2° Decil	**2,357.00**	Este indica que el 20% de los hogares(es decir 1988 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $2,357.00

3° Decil	**2,743.00**	Este indica que el 30% de los hogares (es decir 2832 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $2,743.00
4° Decil	**3,929.00**	Este indica que el 40% de los hogares (es decir 3776 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $3,929.00
5° Decil	**4,921.00**	Este indica que el 50% de los hogares(es decir 4720 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $4,921.00
6° Decil	**6,000.00**	Este indica que el 60% de los hogares(es decir 5664 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $6,000.00
7° Decil	**7,629.00**	Este indica que el 70% de los hogares (es decir 6608 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $7,629.00
8° Decil	**10,000.00**	Este indica que el 80% de los hogares (es decir 7552 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $10,000.00
9° Decil	**15,000.00**	Este indica que el 90% de los hogares (es decir 8496 hogares) en la delegación Azcapotzalco tiene un ingreso menor o igual a $15,000.00

Es importante destacar que en este caso, los fractiles es la herramienta estadística que muestra un panorama más claro de la realidad que se presenta en esta delegación.

A continuación mostraremos un grafico con los porcentajes de ingreso de la población.



Es claro con está gráfica no marca el último 10%, debido a que si aparecieran estos últimos valores distorsionarían la gráfica. También es claro que solo el 10% de los hogares de la Delegación Azcapotzalco perciben arriba de $15,000 mensuales. También se distingue que las medidas de tendencia central no están muy próximas y por ello el histograma de ingreso presentaba esas características.

Otro tipo de análisis es el comparar el ingreso correspondiente al 10%, 20%, 25%, 30%...90% de los hogares con respecto al total de los ingresos de la Delegación Azcapotzalco. Este los podemos representar en la siguiente tabla:

Fractil	**Valor del Fractil**	**Ingreso total del fractil**	**Porcentaje del ingreso total del fractil con respecto al ingreso total**	**Porcentaje de ingreso con respecto a la población restante del fractil en cuestión.**
1° Decil	**1,671.5.**	1089500	1.57%	El 90% restante del los hogares de esta Delegación percibe el 98.43%
2° Decil	**2,357.00**	3232575	4.64%	El 80% restante del los hogares de esta Delegación percibe el 95.36%
1° Cuartil	**2360.25**	4189892	6.02%	El 75% restante del los hogares de esta Delegación percibe el 93.98%
3° Decil	**2,743.00**	6033225	8.67%	El 70% restante del los hogares de esta Delegación percibe el 91.33%
4° Decil	**3,929.00**	8824969	12.68%	El 60% restante del los hogares de esta Delegación percibe el 87.32%
5° Decil	**4,921.00**	12917849	18.56%	El 50% restante del los hogares de esta Delegación percibe el 81.44%
6° Decil	**6,000.00**	18051591	25.94%	El 40% restante del los hogares de esta Delegación percibe el 74.06%
7° Decil	**7,629.00**	24387122	35.04%	El 30% restante del los hogares de esta Delegación percibe el 64.96%

3° Cuartil	**8572.00**	28221018	40.55%	El 25% restante del los hogares de esta Delegación percibe el 59.45%
8° Decil	**10,000.00**	32605681	46.85%	El 20% restante del los hogares de esta Delegación percibe el 53.15%
9° Decil	**15,000.00**	43908414	63.09%	El 10% restante del los hogares de esta Delegación percibe el 36.91%
Total	**999,998**	69592954		

Por último calculemos las medidas de dispersión de estos datos.

El rango

Rango = Límite inferior de la muestra – Límite superior de la muestra =999,998-50=999,948

Se observa que el rango de los ingresos por hogar en la Delegación Azcapotzalco es muy amplio, y eso habla de la mala distribución de los salarios en esta delegación, debido a que el 90% tiene un ingreso menor o igual a $15,000.

El rango intercuartil

El rango intercuartil es la diferencia entre el tercer cuartil (Q_3) y el primer cuartil (Q_1), el rango intercuartil se define por la letra Q y es:

$$\text{Rango intercuartil} = Q_3 - Q_1 = 8572-2360.25=6211.75$$

De aquí podemos deducir que más o menos la mitad de los hogares de la delegación Azcapotzalco perciben $6,211.25

Desviación cuartil.

La desviación cuartil es:

$$Q_D = \frac{Q_3 - Q_1}{2} = \frac{8{,}572 - 2360.25}{2} = \frac{6{,}211.75}{2} = 3{,}105.875$$

Este dato indica que los datos tienen una desviación con respecto a la mediana de $3,105.875

La varianza

Para obtener la varianza por medio de Excel se requiere elevar al cuadrado cada dato del ingreso, sumar todos los datos del ingreso elevados al cuadrado y restarle el producto del tamaño de la muestra por la media aritmética elevada al cuadrado. Por último se suman todos los datos elevados al cuadrado señalando la siguiente celda de la columna en blanco y sombreando todos los datos elevados al cuadrado. De aquí se obtiene $\sum_{i=1}^{9440} x_i^{\;2} = 3305879341769$, después se eleva la media aritmética al cuadrado $\bar{x}^2 = 54348373.83$ y se multiplica por el tamaño de la muestra(n = 9440) n$\bar{x}^2 = 54348373.83$(9,440)= 5.13049E+11. Por último se sigue la fórmula de la varianza:

$$\sigma^2 = \frac{\sum_{i=1}^{9440} x_i^2 - n\bar{x}^2}{n} = \frac{3305879341769 \text{ - } 5.13049\text{E} + 11}{9440} = \frac{2792830692781}{9440} = 378836077.9$$

Por lo que la varianza es 378836077.9 y la desviación estándar como la raíz de la varianza es:

$\sigma = \sqrt{\sigma^2} = \sqrt{378836077.9} = 19463.71182$.

Lo cual quiere decir que la desviación de los datos con respecto a la media es de $19,463.71182.

ANEXO
CONTENIDO DEL PROGRAMA DE ESTADISTICA Y PROBABILIDAD

a)Nombre de la unidad: ESTADÍSTICA DESCRIPTIVA.
b) Propósitos de la unidad I.
Que el alumno sea capaz de diferenciar, organizar, representar gráficamente el significado que un conjunto de datos tiene en relación con su fenómeno relativo a su entorno social, para vincular la estadística a su realidad.

HORAS	CONTENIDO	DESCRIPCIÓN DEL CONTENIDO	ESTRATEGIAS DIDACTICAS (actividades de aprendizaje)	BIBLIOGRAFIA
40		En esta unidad:	El profesor, a partir de determinados problemas de la realidad y de otras disciplinas, discutirá con el grupo la utilidad y la importancia de la Estadística Descriptiva. El alumno en forma individual o por equipos, con la asesoría del profesor y en el aula:	Básica: 1, 2, 3, 4, Complementaria: 5, 6, 7, 8, 9, 10.
	Introducción.	Se abordará una reseña de lo que es la Estadística, para que sirve y cómo manejar y presentar la información. Se establecerá el papel que juega en el entorno social y en las diversas disciplinas en la toma de decisiones.	Investigará en que actividades recientes de su vida, ha estado presente la estadística.	
		Se definirán las variables en categóricas y numéricas. Las categóricas en nominales y ordinales; las numéricas en discretas y continuas.	Distinguirá entre variables categóricas y numéricas y sus respectivas clasificaciones. Dará ejemplos de datos numéricos.	
	Datos.	Se abordará el concepto de dato numérico u ordinal.	Organizará datos de su entorno e identificará la naturaleza de éstos. Distinguirá en ellos aspectos de calidad y cantidad; los representará gráficamente interpretando su organización.	
	Clasificación y construcción de	Se clasificarán, de manera general, las diferentes categorías de datos. Se describirán		

	bloques estadísticos.	las condiciones que caracterizan la cantidad y la calidad de ellos. Ëstos se clasificarán en discretos y continuos y se definirán las diversas escalas de medición usadas en la Estadística, a saber: nominal, ordinal y de intervalos.		
	Organización de los datos por medio de tablas.	Se establecerán los métodos para organizar datos usando tablas de frecuencias: sin agrupar, agrupadazas, relativas, acumuladas, relativas acumuladas y tablas de dos variables.	Organizará datos de su entorno a partir de tablas de frecuencias como lo indica el contenido.	
	Tipos de gráficas.	Se describirán diferentes formas de representar datos por medio de gráficas: de barras, circulares, histogramas e histogramas de frecuencias relativas. Además se considerarán gráficas lineales, y polígonos de frecuencias, así como, ojivas; ojivas e histogramas. Se enfatizará el uso de las representaciones gráfi de datos para la configuración de poblaciones.	Tomará problemas de su entorno y los representará mediante cada una de las gráficas que aparecen en el contenido. Planteará, resolverá e interpretará problemas cuya solución requiera elaborar las tablas de frecuencia propuestas en el contenido. Tomará problemas de su entorno y los representará mediante cada una de las gráficas que aparecen en el contenido.	
	Introducción a la sumatoria.	Se establecerán las propiedades de la sumatoria.	Operará con sumatorias	
	Análisis de datos de una variable: Medidas de tendencia central y de localización.	Para organizar los datos se definirán y analizarán las diversas medidas de tendencia central y de localización: media, mediana, moda, media aritmética, media armónica, rango medio y medidas de localización.	Usando datos y gráficas tratará estadísticamente la información recopilada. A través de mediciones numéricas calculará e interpretará las medidas de tendencia central referidas en el contenido.	
		Se definirán y analizarán las diversas medidas de dispersión o variabilidad: rango, rango intercuartil,	A través de mediciones numéricas calculará e interpretará las medidas de	

	Medidas de dispersión o variabilidad. Análisis descriptivo de datos bivariados: Correlación	marca de desviación, suma de cuadrados, varianza. Para datos agrupados en tabla de frecuencia se definirán: desviación estándar y estimación de la desviación estándar. Se definirán y analizarán los diversos métodos que simplifican el cálculo del coeficiente de correlación: coeficiente de correlación de Pearson, suma de productos cruzados y fórmula computada para la correlación.	dispersión o variabilidad referidas en el contenido Calculará coeficientes de correlación. Aplicará los métodos de correlación en la solución de problemas de su entorno.	

Bibliografía básica:

1. Infante, G. Said et al., *Métodos Estadísticos, un enfoque interdisciplinario*. México, Trillas, 1991.
2. Freund, John E. et al., *Estadística Elemental*. México, Prentice Hall, 1992.
3. Colección Sistema de Educación a Distancia, Introducción a los métodos estadísticos 1. México, Universidad Pedagógica Nacional, SEP, 1988.
4. Willoughby, Stephen S., Probabilidad y Estadística, Cultural S.A., 1981

Bibliografía complementaria:

5. Kreysing, Edwin, *Introducción a la Estadística Matemática*. México, Limusa, 1981.
6. Downie, N.M. et al., *Métodos estadísticos aplicados*. México, Harla, 198.
7. Johnson, Robert*, Estadística Elemental*. México, Trillas, 1976.
8. Shao, Stephen P., *Estadística para economistas y administradores de empresas*. México, Herrero, 1970.
9. Hoel, Paul G. *Estadística elemental*. México, C.E.C.S.A., 1976.
10. Spiegel, Murria R*. Estadística*. México, Mc. Graw Hill, 1991

a)Nombre de la unidad: CONJUNTOS
b)Propósitos de la unidad II.

Que el alumno reafirme los conocimientos sobre conjuntos y sus operaciones básicas, previamente adquiridos, para que los aplique en problemas de análisis combinatorio y probabilidad.

HORAS	CONTENIDO	DESCRIPCIÓN DEL CONTENIDO	ESTRATEGIAS DIDACTICAS (actividades de aprendizaje)	BIBLIOGRAFIA
5		En esta unidad:	El profesor, a partir de determinados problemas de la realidad y de otras disciplinas, discutirá con el grupo la utilidad de los conjuntos. El alumno en forma individual o en equipos, con la asesoría del profesor y en el aula:	Básica: 1, 2, 3, 4, Complementaria: 5, 6, 7, 8, 9.
	Conjuntos: Idea intuitiva (por extensión y por comprensión) Conceptos básicos y simbología. Subconjuntos Conjunto Universal. Conjunto Vacío.	Se revisarán los conocimientos adquiridos, en cursos anteriores, sobre conjuntos para que al reafirmarlos se acceda fácilmente a la teoría de probabilidad.	Resolverá problemas de la vida cotidiana en los que se aplique las operaciones con conjuntos.	
	Operaciones con conjuntos.	Se revisarán las operaciones de: unión, intersección, diferencia y complemento. Se representará gráficamente mediante diagramas de Venn- Euler.	Por medio de diagramas de Venn-Euler representará y resolverá problemas en los que se apliquen las operaciones entre conjuntos.	
	Cardinalidad de la unión, de la intersección y del complemento.	Se revisará el concepto de cardinalidad de un conjunto definiendo la cardinalidad de la unión, la intersección y el complemento.	Resolverá problemas en los que determine la cardinalidad de los conceptos descritos en el contenido.	

Bibliografía básica:

1. Cardenas, Humberto et al., *Álgebra Superior*. México, Trillas, 1973.
2. Lovaglia, Florence et al., *Álgebra*. México, Harla, 1969.
3. Swokoski, Earl W., *Álgebra universitaria*. México, CECSA, 1970
4. Grimaldi, Ralph P., Matemáticas *discretas y combinatoria*. México, Addison Wesley Iberoamericana, 1989.

Bibliografía complementaria:

5. Lipchutz, Seymour, *Matemática discreta*.México, Serie Shaum, Mc Graw Hill, 1990.
6. Schaaf, Peters, *Álgebra*, un enfoque moderno. México, Reverté, 1972.
7. Fuller, Gordon, *Álgebra elemental*. México, CECSA, 1994.
8. Meserve, Bruce et al*., Introducción a las Matemáticas*. México, Reverté, 1967.
9. Barnett, Raymond A., *Álgebra*. México, Mc. Graw Hill, 1990.

Nombre de la unidad: PROBABILIDAD.

b) Propósitos de la unidad III.

Que el alumno sea capaz de identificar a la probabilidad como un instrumento confiable en la inferencia y toma de decisiones.

HORAS	CONTENIDO	DESCRIPCIÓN DEL CONTENIDO	ESTRATEGIAS DIDACTICAS (actividades de aprendizaje)	BIBLIOGRAFIA
45		En esta unidad:	El profesor, a partir de determinados problemas de la realidad y de otras disciplinas, discutirá con el grupo la utilidad e importancia de la probabilidad. El alumno individualmente o por equipos, con la asesoría de su profesor y en el aula:	
	Espacio muestral.	Se definirá espacio muestral y su notación. Se establecerá la diferencia entre población y espacio muestral.	Desarrollará ejemplos en los que identifique el espacio muestral de una población.	
	Experimentos y eventos.	Se analizarán y definirán los conceptos de experimentos y eventos probabilísticos.	Desarrollará e interpretará experimentos en los que se incluyan diferentes tipos de eventos vinculados con su entorno.	
	Principio fundamental del conteo.	Se enunciará el principio fundamental del conteo para identificar las diferentes maneras en que puede ocurrir un evento.	Desarrollará e interpretará diversos ejemplos de experimentos aleatorios con y sin repetición, utilizando un diagrama de árbol.	
	Análisis combinatorio.	Se definirán los conceptos de factorial, permutación, combinación y propiedades de la combinatoria.	Desarrollará e interpretará diferentes ejemplos de ordenaciones, permutaciones y combinaciones con y sin repetición de experimentos aleatorios.	

	Concepto de Probabilidad. Eventos. Teoremas de probabilidad.	Se definirá formalmente el concepto de probabilidad asignada a un evento y su notación. Se enfatizará en el uso de las tablas de frecuencia y su representación gráfica, en la obtención de la probabilidad Se definirán evento: elemental, seguro y nulo: eventos independientes, mutuamente excluyentes y no excluyentes; se establecerán las propiedades: la probabilidad de un evento es igual a la suma de las probabilidades de los eventos elementales que los componen. El espacio muestral de una de las variables es un evento seguro y la probabilidad de que ocurra es 1. Se establecerán y aplicarán los teoremas: Para eventos mutuamente excluyentes: P(E U F)= P(E) +P(F) Para eventos complementarios P(E')=1-P (E) Para eventos condicionales: P(E/F)= $\frac{P(E \cap F)}{P(F)}$ Para eventos independientes:	Se resolverán problemas empleando tablas de frecuencia. Desarrollará y resolverá prácticas e las que se determine el espacio muestral, el significado de un evento nulo o exitoso; el significado de una variable aleatoria discreta o continua, en problemas que se vinculen con situaciones de su entorno. En el salón de clase y, bajo la supervisión de su profesor, planteará, resolverá e interpretará problemas que relacionan la teoría de conjuntos con la teoría de la probabilidad. Resolverá e interpretará problemas en los que calcule el valor de todas y cada una de las variables descritas en el contenido. Por ejemplo: Si la probabilidad de que ocurra un evento L es0.1.¿Cuál es la probabilidad de que no ocurra, es decir cual es la probabilidad del evento complementario L'? En una caja se tienen 20 sobres idénticos 10 de los cuales están vacíos, 8 tienen un billete de 10	

	Variables aleatorias: discretas y continuas. Funciones de distribución	P(E U F) = P(E)+P(F)- P(E∩F) P(E∩F)= P(E)P(F) Se definirá y analizará el concepto de variables aleatorias clasificándolas en discretas y continuas. Se mencionará que existen distribuciones de probabilidad de variables aleatorias discretas o continuas. Se describirán someramente, como ejemplos, la función de distribución binomial (discreta) y la normal (continua).	pesos, y 2 uno de 100 pesos. Si se sacan 2 sobres al azar, calcule la probabilidad de que ambos tengan un billete de $100.A) Si el primero se regresa a la caja antes de sacar el segundo. B) Si el primero no se regresa. Resolverá ejercicios sencillos en los que se aplique una distribución binomial.	

Bibliografía básica:

1. Infante, G. Said et al., *Métodos Estadísticos, un enfoque interdisciplinario*. México, Trillas, 1991.
2. Freund, John E. et al., *Estadística Elemental*. México, Prentice Hall, 1992.
3. Colección Sistema de Educación a Distancia, Introducción a los métodos estadísticos 1. México, Universidad Pedagógica Nacional, SEP, 1988.
4. Willoughby, Stephen S., Probabilidad y Estadística, Cultural S.A., 1981

Bibliografía complementaria:

5. Kreysing, Edwin, *Introducción a la Estadística Matemática*. México, Limusa, 1981.
6. Downie, N.M. et al., *Métodos estadísticos aplicados*. México, Harla, 198.
7. Johnson, Robert, *Estadística Elemental*. México, Trillas, 1976.
8. Shao, Stephen P., *Estadística para economistas y administradores de empresas*. México, Herrero, 1970.
9. Hoel, Paul G. *Estadística elemental*. México, C.E.C.S.A., 1976.
10. Spiegel, Murria R. *Estadística*. México, Mc. Graw Hill, 1991.
11. Fuller, Gordon et al., *Álgebra universitaria*, México, CECSA, 1992,