INSTITUTO POLITÉCNICO NACIONAL

Escuela Superior de Física y Matemáticas

Apuntes de Estadística y Probabilidad
para el Bachillerato Tecnológico

TESIS MEMORIA DE EXPERIENCIA
PROFESIONAL

QUE PARA OBTENER EL TÍTULO DE
LICENCIADO EN FÍSICA Y MATEMÁTICAS

PRESENTA
AMELIA ALBARRÁN MALDONADO

México, D. F. Mayo de 2004

A quien con sus manitas me
impulsó a dar un paso más,
y con sus besos y sonrisas me
dio vida para continuar.
Al ser más importante en mi
vida.
A tí Alma Luz.

Agradecimientos

.

Agradezco a Dios todo lo que
me ha dado;
principalmente a mi hija y
el gran amor de mis padres.

A los integrantes del jurado Lic. Manuel Robles Bernal, Lic. Salvador Q. Flores García, M. en C. Francisco Ramírez Reyes, Lic. Ma. Elizabeth de la Cruz Santiago y al Dr. Pablo Lam Estrada por su participación en la revisión y sugerencias de esta tesis.

Introducción

Es generalizada la problemática que existe en el proceso de enseñanza-aprendizaje de las matemáticas que se imparten en nuestro país, sobre todo a nivel de secundaria y bachillerato en donde se tiene la formación preparatoria dirigida hacia la formación técnica y profesional de los estudiantes. Por una mal planeación en los cursos de matemáticas que han existido desde los cursos de primaria, que va cambiando poco a poco ante las nuevas ideas de enseñanza-aprendizaje, los estudiantes han sentido que las matemáticas son difíciles de aprender; tal es el grado en este sentir, que se considera a un alumno con "cerebro" como aquel que tiene la facilidad en el proceso de aprendizaje de estas áreas. Por lo cual, los métodos tradicionales que se han aplicado en estos niveles de enseñanza han probado ser poco eficaces en el proceso de enseñanza-aprendizaje.

Hoy en día se ha estado aplicando en algunos países el proceso de enseñanza-aprendizaje en el contexto del objeto-sujeto, el cual consiste en plantear el conocimiento de la matemática con la participación del estudiante y del profesor al mismo tiempo, es decir, se da la información del conocimiento a los estudiantes quienes de inmediato lo aplican para que dicho conocimiento lo vayan analizando y aprendiendo, y de esta manera comprendan el objeto final del conocimiento, mientras que el profesor está para motivar, aclarar dudas, dar la reflexiones pertinentes, establecer deducciones y finalizar con conclusiones para que el proceso de enseñanza-aprendizaje sea lo más efectivo posible para todos los participantes.

La materia de Estadistica y Probabilidad que se imparte en el quinto semestre en el Instituto SENDA 2000, incorporada a la Secretaria de Educación Pública, está dirigido a estudiantes del tronco común del bachillerato tecnológico con la finalidad de que adquieran el conocimiento necesario para poderlos aplicar en situaciones concretas y directas en los distintos campos científicos y tecnológicos para la aplicación de técnicas en la organización de datos, obtención de propiedades elementales e interpretación de las mismas, así como en el mecanismo de cálculo e interpretaciones de probabilidades de eventos. Esta materia tiene como antecedente la Aritmética, Álgebra, Geometría, Trigonometría, Geometría Analítica y Cálculo Diferencal.

A través de cerca de once años de labor en SENDA 2000, he tratado de establecer en lo mejor posible el preceso de enseñanza-aprendizaje en el que esté involucrado el objeto-sujeto. Aún cuando, uno se enfrenta siempre con la mala preparación en el conocimiento con la que llegan los estudiantes y al rechazo e indiferencia ante el estudio de las matemáticas, cuyo fin último es el de acreditar la materia como quiera que sea. Sin embargo, creo que si la actitud del maestro es profesional, en estas escuelas la enseñanza cumplirá con los objetivos para los que se ha crado.

Estos apuntes se han desarrollado para dar el conocimiento que permita que el estudiante los maneje y aplique a casos específicos. Por supuesto, la reflexión, la inducción, la deducción y las conclusiones son expuestas para el mejor entendimiento. Por ejemplo, las deducciones han de ser sencillas, sobre todo la idea es que ejerciten la razón de los conocimientos que estudian. El lenguaje está establecido de acuerdo al nivel en que se encuentran los estudiantes; es claro que no es un nivel elevado y que presenta muchas deficiencias, pero obtienen la herramienta necesaria que han de utilizar en el medio en el que se desarrollarán. También, hacemos notar que no presentamos ejercicios sino algunos ejemplos, con el propósito de que los estudiantes deben de ser parte de la creación, posiblemente ficticia, de eventos que a ellos se les ocurra partiendo de situaciones reales a los que se van a enfrentar; éstos serán desarrollados en las actividades que se han diseñado para tal fin. Al final los estudiantes presentan un trabajo de los reaultados que han obtenido. Por supuesto, la participación en equipos es un factor esencial en todo este proceso para que surja la interacción de ideas que permitan, en algunas ocasiones, la discusión de los temas tratados.

El objetivo como profesor es el de motivar, aclarar dudas y guiar al estudiante en el conocimiento previo que se les ha dado para que lo puedan aplicar de manera improvisada y racional, y que puedan observar todo tipo de errores y aciertos a los que se pueden enfrentar. Esperando con esto el interés, la creatividad y la comprensión de un conocimiento que es de utilidad tanto personal como social. La idea es el de preparar a los estudiantes para que puedan hacer un buen papel en la parte laboral dentro de la sociedad.

Los apuntes no son el conocimiento completo de la estadística y probabilidad, pero sí podemos decir que son los conocimientos básicos necesarios que en ese nivel se deben de tener. Los temas correspondientes al programa, y que aquí presentamos, se pueden encontrar en cualquier libro de estadística y probabilidad elemental. Lo importante es de que tanto los maestros como los estudiantes deben de aprender de todos los fenómenos que se presentan en cada ciclo escolar para el mejoramiento de la educación en nuestro país, independientemente del lugar en que se da el conocimiento.

Contenido

Capítulo 1

La Estadística

En este capítulo, iniciaremos el estudio de la estadística presentando su definición y clasificación. Además, de definir las herramientas básicas y los objetos que utiliza la estadística para su desarrollo.

Podemos decir que el alumno ha tenido cierto contacto con la estadística aunque sea a través de los medios de comunicación. Saben intuitivamente que es la referencia, o medición en alguna forma, de ciertos datos que dan información de algún evento, pero también tienen conocimiento de que estos datos pueden ser graficados de distintas maneras. Esto de alguna manera, representa alguna ventaja para la enseñanza de la estadística.

La estadística está relacionada con los métodos científicos en la toma, organización, recopilación, presentación y análisis de datos, tanto para la deducción de conclusiones como para la toma de decisiones razonables de acuerdo con tales análisis obtenidos.

En un sentido más estricto, el término "estadística" es utilizado para denotar los mismos datos o números que se derivan de ellos, como por ejemplo, promedios. Así se habla de estadística de empleo, estadística de accidentes, etc.

En este capítulo, explicaremos las herramientas que utiliza la estadística para su estudio en el análisis y la toma de decisiones de los datos obtenidos.

1.1 Estadística Descriptiva e Inductiva

En una colección de datos que atañen a las características de un grupo de individuos u objetos, es a menudo imposible o poco práctico observar la totalidad de los individuos, sobre todo si éstos son muchos. En lugar de examinar al grupo entero llamado **población** o **universo**, se examina una pequeña parte del grupo llamada **muestra**. Así, podemos tener la colección de datos tales como las alturas y pesos de los estudiantes de una escuela en un año escolar o el número de cerrojos defectuosos y no defectuosos producidos por una fábrica en una producción determinada.

Una población puede ser **finita** o **infinita**. Por ejemplo, la población consistente de alturas y pesos de alumnos de una escuela en un año escolar es finita, mientras que la población formada por todos los posibles sucesos (caras o cruces) en tiradas sucesivas de una moneda es infinita.

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

Al no poder estar absolutamente ciertos de la veracidad de tales inferencias, se ha de utilizar con frecuencia en estas conclusiones el término de **probabilidad**.

La parte de la estadística que trata solamente de describir y analizar un grupo dado sin sacar conclusiones o inferencias de un grupo mayor se llama **estadística descriptiva** o **estadística deductiva**. Esta parte de la estadística es el objeto estudio de este trabajo.

1.2 Variables Discretas y Continuas

En esta sección presentaremos algunos conceptos matemáticos importantes.

Una **variable** es un símbolo tal como

$$a, b, c, \ldots, x, y, z, A, B, C, \ldots, X, Y, Z,$$

que pueden tomar un valor cualquiera de un conjunto determinado de ellos, llamado **dominio** de la variable. Si la variable puede tomar solamente un valor, entonces ésta se llama **constante**. Por lo general, las primeras letras del alfabeto son consideradas como constantes.

Una variable que en teoría puede tomar cualquier valor entre dos valores dados se llama **variable continua**, si no es así, se llama **variable discreta**.

Por ejemplo, en una familia el número n de hijos puede tomar cualquiera de los valores $0, 1, 2, \ldots$, pero no puede ser 2.5 o 3.439; es así, una variable discreta. Por otro lado, la altura h de un individuo puede ser 1.68 metros, 1.676 metros o 1.67689 metros, dependiendo de la exactitud de la medida; esta es una variable continua.

Los datos que vienen definidos por una variable discreta o continua se llaman **datos discretos** o **datos continuos**, respectivamente. En general, las *medidas* dan origen a datos continuos, mientras que las *enumeraciones* o *conteos* originan datos discretos.

1.3 Representación de Datos

Si a cada valor de una variable x le hacemos corresponder uno o más valores de otra variable y, entonces decimos que y está en **correspondencia** con x y escribimos $y = f(x)$ para indicar esta correspondencia. La variable x se llama **variable independiente** e y es la **variable dependiente** de la variable x. Si a cada valor de x le corresponde un solo valor de y, entonces decimos que y está en **función** con x.

La dependencia funcional o correspondencia entre dos variables está a menudo representada en una *tabla*. Sin embargo, puede también indicarse por medio de una ecuación que relacione las variables, tal como $y = f(x) = 5x-3$, y en este caso, para los valores $x = -2, -1, 0, 1, 2$ obtenemos los valores $y = -13, -8, -3, 2, 7$ los cuales los podemos representar en cualquiera de las siguientes tablas:

x	-2	-1	0	1	2
y	-13	-8	-3	2	7

x	y
-2	-13
-1	-8
0	-3
1	2
2	7

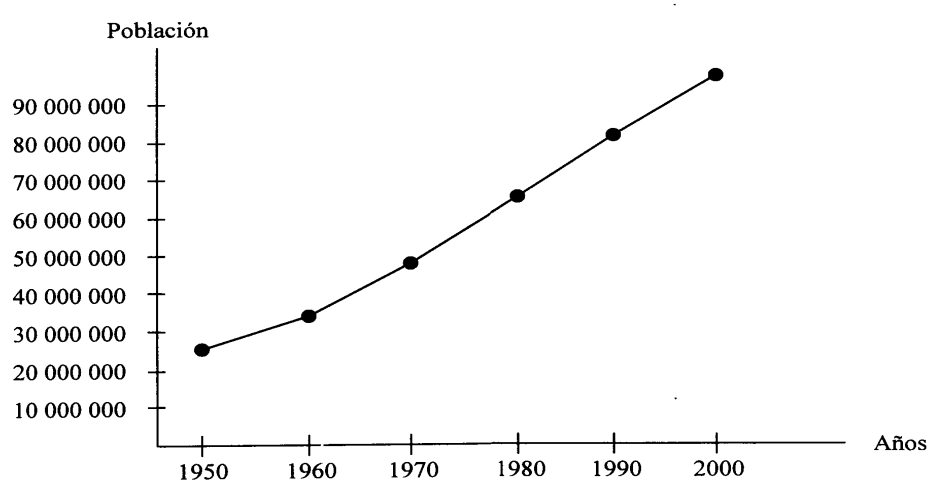
El uso de las tablas nos permite expresar en síntesis los resultados obtenidos. Por ejemplo, en la Tabla 1 podemos expresar la población de México del año de 1950 al año 2000.

Año	Estados Unidos Mexicanos (Población)
1950	25 791 017
1960	34 923 129
1970	48 225 238
1980	66 846 833
1990	81 249 645
2000	97 483 412

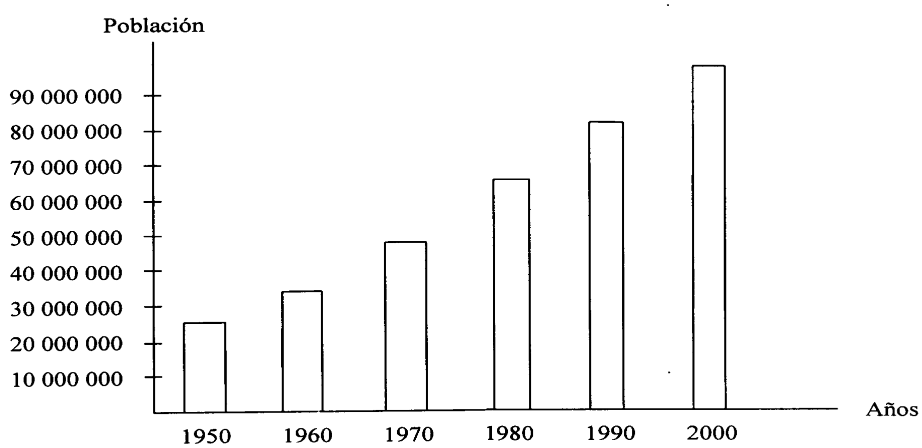
Tabla 1. Población de México, 1950-2000
(Información INEGI)

También, la Tabla 1 puede ser representada por medio de gráficas a través de los siguientes tipos:

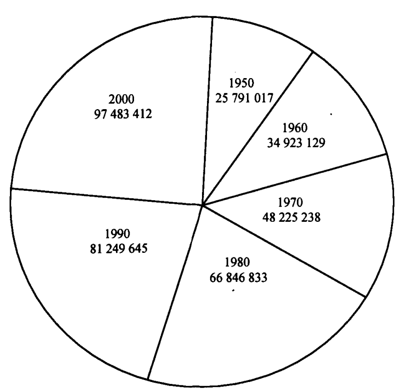
(a) **gráfica de línea**



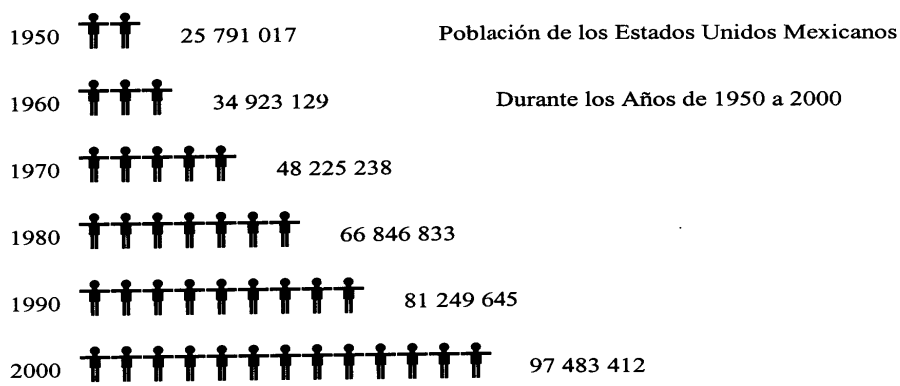
(b) **gráfica de barras**



(c) **gráfico circular**



(d) **pictograma**



1.4 Actividades

Las actividades para este capítulo son las siguientes:

1.- Formar a todos los estudiantes del grupo, en grupos de cinco personas para iniciar las actividades a desarrollar.

2.- Conocer el origen y avances de la estadística, así como su influencia y aplicaciones en la sociedad.

3.- Establecer las divisiones de la estadística y su objeto de estudio.

4.- Comentar y discutir los conceptos de estadística, estadística descriptiva y estadística inductiva, así como todos los conceptos relacionados con éstos (por ejemplo los conceptos de población, muestra, infinito, finito), creando para esto ejemplos que den claridad y comprensión.

5.- Discutir los conceptos de variable continua, variable discreta, datos continuos y datos discretos. Crear ejemplos con respecto a estos conceptos.

6.- En la graficación de datos, identificar todos los conceptos que en este tema aparecen, y desarrollar una serie de datos prácticos de algún evento para que lo presenten en los distintos tipos de representación gráfica que se han tratado.

7.- ¿En qué secretarias del gobierno federal creen que la estadística es aplicable?

8.-¿Conocen algún otro lugar donde la estadística se aplique?

9.- ¿Es posible que en tu casa se aplique la estadística?

Capítulo 2

Distribuciones de Frecuencia

En este capítulo analizaremos la forma en la que los datos obtenidos pueden ser distribuidos de alguna manera. Aquí es donde el alumno empieza a tener ciertos problemas con la comprensión de los conceptos presentados.

2.1 Distribuciones de Frecuencia

La **toma de datos** es la obtención de una colección numérica que no han sido ordenados de alguna manera. Mientras que una **ordenación de datos** es una colección de datos numéricos tomados en un orden creciente o decreciente de magnitud. La diferencia entre el mayor y el menor de los números se llama **recorrido** o **rango** de los datos .

Cuando se disponen de un gran número de datos, es útil distribuirlos en **clases** o **categorías** y determinar el número de individuos que pertenencen a cada clase, que es la **frecuencia de clase**. Una ordenación tabular de los datos en clase, reunidas las clases y con las frecuencias correspondientes a cada una, se conoce como una **distribución de frecuencias** o **tabla de frecuencias**.

La Tabla 2 es una distribución de frecuencias de alturas registradas con aproximación en metros de 100 estudiantes de la Escuela $\mathcal{X}$. Por ejemplo, la primera clase o categoría comprende las alturas de 1.46 a 1.50 metros, y viene indicada con el símbolo 1.46 - 1.50. Puesto que 17 estudiantes tienen una altura perteneciente a esta clase, la correspondiente frecuencia de clase es 5.

Altura (metros)	Número de estudiantes
1.46 - 1.50	17
1.51 - 1.55	31
1.56 - 1.60	36
1.61 - 1.65	10
1.66 - 1.70	5
1.71 - 1.75	1
	Total 100

Tabla 2. Alturas de 100 estudiantes en la Escuela $\mathcal{X}$

Los datos ordenados y resumidos en una distribución de frecuencia se llama **datos agrupados**. Aunque con el proceso de agrupamiento generalmente se pierde parte del detalle original de los datos, pero tiene la ventaja de presentarlos "todos" en una sencilla tabla que facilita el hallazgo de las relaciones que pueda haber entre ellos, puestas así de manifiesto.

Un símbolo que define una clase, tal como 1.46 - 1.50 de la Tabla 2, se conoce como **intervalo de clase**. Los números extremos, 1.46 y 1.50, son los **límites de la clase**; el menor 1.46 es el **límite inferior** de la clase y el mayor 1.50 es el **límite superior** de la clase. Los términos clase e intervalo de clase se utilizan a menudo indistintamente, aunque el intervalo de clase es realmente un símbolo para la clase.

Un intervalo de clase que, al menos en teoría, no tiene límite superior o inferior, se conoce como **intervalo de clase abierto**. Al referirse a la edad de grupos de individuos, el intervalo de clase "mayores de 65 años" es un intervalo de clase abierto.

Si las alturas se registran con aproximación de metros en la Tabla 2, el intervalo de clase 1.46 - 1.50 teóricamente incluye todas las medias desde 1.455 a 1.505 metros. Estos números se conocen como **límites reales de clase** o **límites verdaderos de clase**; el menor de ellos es el **límite real inferior**, y el mayor de ellos es el **límite real superior**. Los límites reales de clase se obtienen sumando al límite superior de un intervalo de clase el límite inferior del intervalo de clase contiguo superior y dividido por 2.

En algunas ocaciones, los límites reales de clase se utilizan para simbolizar las clases. Por ejemplo, las diferentes clases de la primera columna de la Tabla 2 podrían indicarse por 1.455 - 1.505, 1.505 - 1.555, etc. Sin embargo, con tal notación aparece una ambigüedad, pues los límites reales de clase no coincidirían con las observaciones reales. Así, si se tiene una observación de 1.505, no sería posible discernir si pertenece al intervalo de clase 1.455 - 1.505 o al 1.505 - 1.555.

El **tamaño** o **anchura** de un intervalo de clase es la diferencia entre los límites reales de clase que lo forman, y también se conoce como **anchura**, **tamaño** o **longitud de clase**. Si todos los intervalos de la clase de una distribución de frecuencias tienen igual anchura, esta anchura común se representa por c. En tal caso, c es igual a la diferencia entre dos sucesivos límites de clases inferiores o superiores.

La **marca de clase** es el punto medio del intervalo de clase y se obtiene sumando los límites inferior y superior de la clase y se divide por 2. Así, la marca de clase del intervalo 1.46 - 1.50 es 1.475. La marca de clase también se llama **punto medio de la clase**.

Para análisis matemáticos posteriores, todas las observaciones pertenecientes a un intervalo de clase dado se suponen coincidentes con la marca de clase.

2.2 Reglas Generales para Formar las Distribuciones de Frecuencia

Las reglas generales para formar las distribuciones de frecuencia son las siguientes:

- Determinar el mayor y el menor de los datos registrados y así encontrar el rango (diferencia entre el mayor y el menor de los datos).
- Dividir el rango en un número conveniente de intervalos de clase del mismo tamaño. Si esto no es posible, utilizar intervalos de clase de diferente tamaño o intervalos de clase abiertos. El número de intervalos de clase se toma generalmente entre 5 y 20 dependiendo de los datos.

Los intervalos de clase se eligen también de forma que las marcas de clase o puntos medios coincidan con datos realmente observados. Esto tiende a minorar el llamado **error de agrupamiento**, en los análisis posteriores. Sin embargo, los límites reales de clase no coincidirán con los datos observados.

- Determinar el número de observaciones que caen dentro de cada intervalo de clase, es decir, encontrar las frecuencias de clase.

2.3 Histogramas y Polígonos de Frecuencia

Dos representaciones gráficas de las distribuciones de frecuencia son utilizadas.

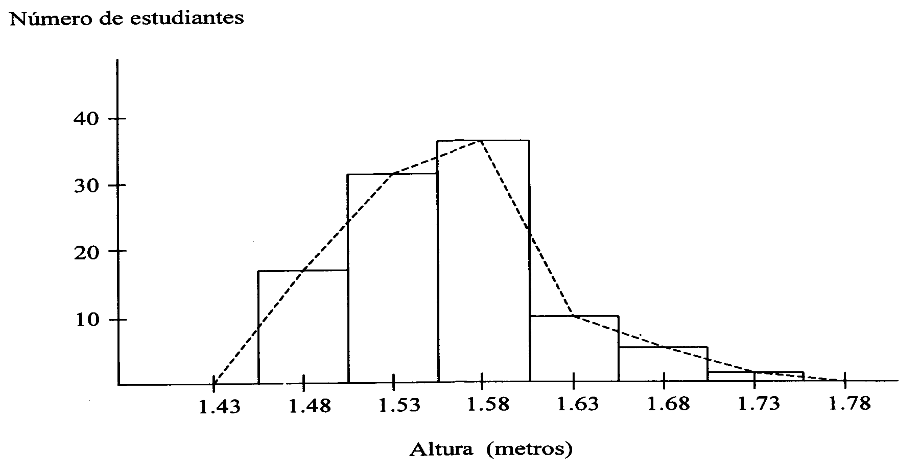
Un **histograma** o **histograma de frecuencia** consiste de una serie de rectángulos que tienen

- Sus bases sobre el eje horizontal (eje de las x's) con centros en las marcas de clase y longitud igual al tamaño de los intervalos de clase.

- Superficies proporcionales a las frecuencias de clase.

  Si los intervalos de clase tienen todos igual tamaño, las alturas de los rectángulos son proporcionales a las frecuencias de clase y se acostumbra en tal caso a tomar las alturas numéricamente iguales a las frecuencias de clase. Si los intervalos de clase no son de igual tamaño, estas alturas deberán ser calculadas.

Un **polígono de frecuencias** es un gráfico de línea trazado sobre las marcas de clase. Puede obtenerse uniendo los puntos medios de los techos de los rectángulos en el histograma como se muestra en la siguiente gráfica:



Como se observa en la gráfica anterior, se acostumbra prolongar el polígono hasta las marcas de clase inferior y superior inmediatas.

2.4 Distribución de Frecuencia Relativa

La **frecuencia relativa** de una clase es la frecuencia de la clase dividida por el total de frecuencias de todas las clases y se expresa generalmente como porcentaje. Por ejemplo, la frecuencia relativa de la clase 1.56 - 1.60 de la Tabla 2 es 36/100=36%. La suma de las frecuencias relativas de todas las clases es evidentemente 1 ó 100%.

Si las frecuencias que aparecen en una tabla se substituyen por las correspondientes frecuencias relativas, la tabla resultante se llama **distribución de frecuencias relativas**, **distribución porcentual** o **tabla de frecuencias relativas**.

Las representaciones gráficas de distribuciones de frecuencia relativas pueden obtenerse del histograma o del polígono de frecuencias, al cambiar la escala vertical de frecuencia a frecuencia relativa, conservándose el mismo diagrama. Los gráficos que resultan se llaman **histogramas de frecuencias relativas** o **histogramas porcentuales** y **polígonos de frecuencias realtivas** o **polígonos porcentuales**, respectivamente.

2.5 Distribución de Frecuencia Acumulada. Ojivas

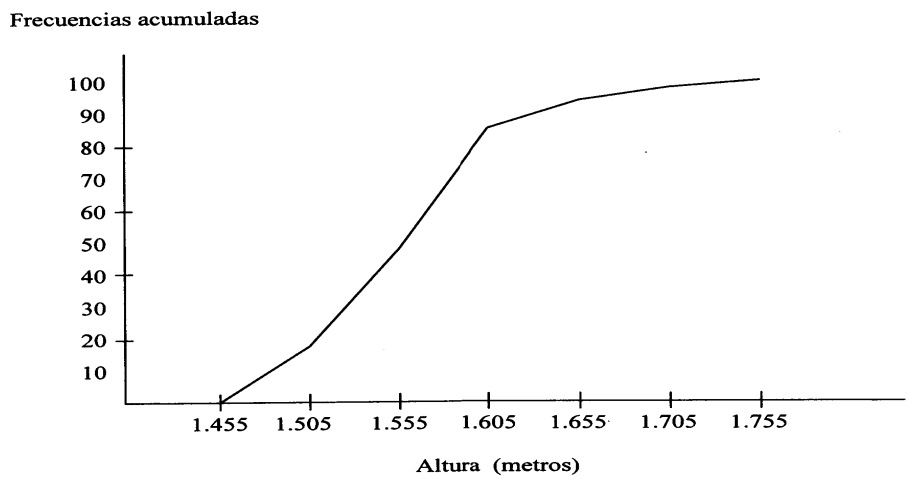
La frecuencia total de todos los valores menores que el límite real superior de clase de un intervalo de clase dado se conoce como **frecuencia acumulada**, inclusive hasta ese intervalo de clase. Por ejemplo, la frecuencia acumulada hasta el intervalo de clase 1.56 - 1.60 inclusive en la Tabla 2 es 84, significando que 84 estudiantes tienen alturas menores que 1.605 metros.

Una tabla que represente la frecuencia acumulada se llama **distribución de frecuencias acumulados**, **tabla de frecuencias acumuladas** o brevemente **distribución acumulada**, y se muestra en la Tabla 3 para la distribución de la altura de los estudiantes.

Altura (metros)	Número de estudiantes
Menor que 1.455	0
Menor que 1.505	17
Menor que 1.555	48
Menor que 1.605	84
Menor que 1.655	94
Menor que 1.705	99
Menor que 1.755	100

Tabla 3. Distribución acumulada de alturas de estudiantes en la Escuela $\mathcal{X}$

Un gráfico que muestre las frecuencias acumuladas menores que cualquier límite real superior de clase trazado sobre los límites reales superiores de clase se llama **polígono de frecuencias acumulados** u **ojiva**, y se muestra en el siguiente gráfico para la distribución de la altura de los estudiantes.



En algunos casos es preferible considerar una distribución de frecuencias acumuladas de todos los valores mayores o iguales al límite real inferior de clase de cada intervalo de clase. En este caso, consideremos las alturas de 1.505 metros o más, 1.555 metros o más, etc., ésta se llama a veces **distribución acumulada "o más"** mientras que la considerada anteriormente es la **distribución acumulada "menor que"**. De una se obtiene fácilmente la otra. Las correspondientes ojivas se llaman **"o más"** o **"menor que"**. Siempre que se refiera uno a distribuciones acumuladas u ojivas sin especificar, se entenderá que son del tipo "menor que".

La **frecuencia relativa acumulada** o **frecuencia porcentual acumulada** es la frecuencia acumulada dividida por la frecuencia total. Por ejemplo, la frecuencia relativa acumulada de alturas menores que 1.605 metros es $84/100 = 84\%$., estableciendo con ello que el 84% de los estudiantes tienen altura menor de 1.605 metros.

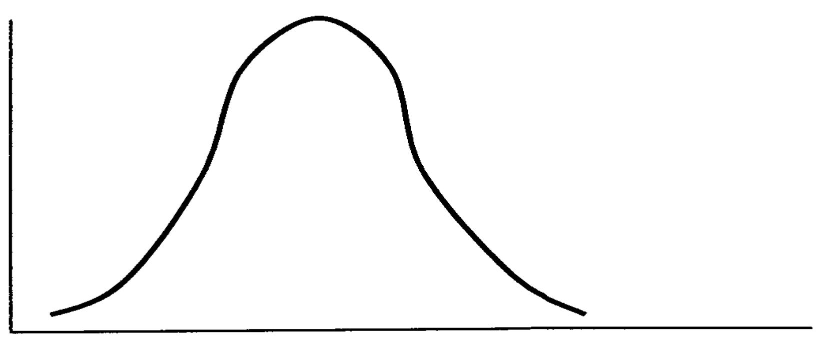
Si se utiliza la Tabla 3 y su gráfico, expuestos anteriormente, las frecuencias relativas acumulas en lugar de las frecuencias acumuladas, entonces los resultados se llaman **distribuciones de frecuencias relativas acumuladas** o **distribuciones porcentuales acumuladas** y **polígonos de frecuencias relativas acumuladas** u **ojivas porcentuales**, respectivamente.

El conjunto de datos puede considerarse normalmente como perteneciente a una muestra extraída de una población grande. A causa de las muchas observaciones que podemos realizar en la población, es posible teóricamente (para datos continuos) elegir los intervalos de clase muy pequeños y todavía tener un número adecuado de observaciones dentro de cada clase. Así se tiene que el polígono de frecuencias o el de frecuencias relativas para una población grande puede estar formado por pequeños segmentos rectos que aproximan el conjunto a una curva; las curvas de este tipo pueden llamarse **curvas de frecuencias** o **curvas de frecuencias relativas**.

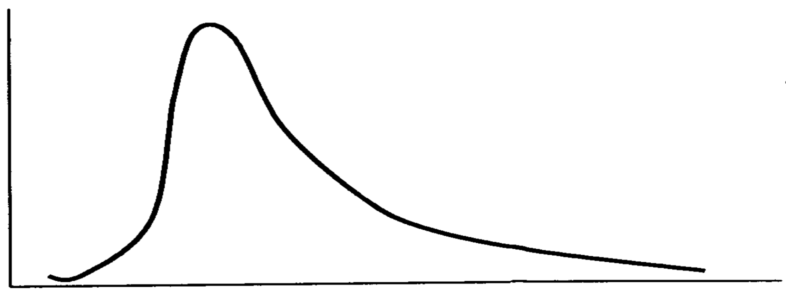
Es de esperar que tales curvas provengan de la suavización de los polígonos de frecuencias o de los polígonos de frecuencias relativas de la muestra; la aproximación es tanto más exacta conforme aumenta el tamaño de la muestra. Por esta razón, una curva de frecuencias se conoce como un **polígono de frecuencias suavizado**.

De manera análoga, las **ojivas suavizadas** provienen de la suavización de los polígonos de frecuencias acumuladas u ojivas. Normalmente es más sencillo suavizar una ojiva que un polígono de frecuencias.

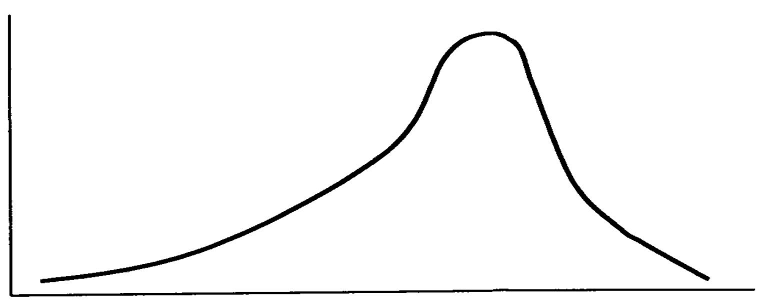
Las curvas de frecuencias presentan determinadas formas características que las distinguen como se indican en las siguientes gráficas:



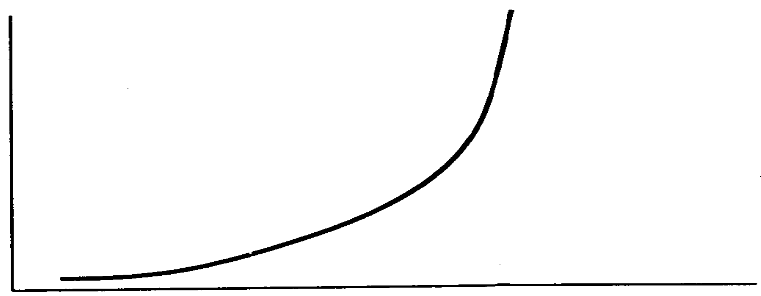
Simétrica o bien formada



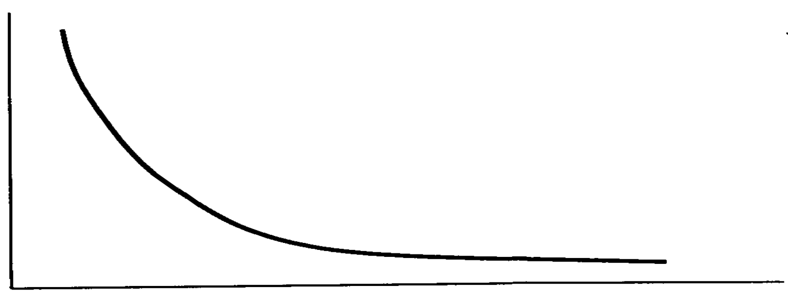
Sesgada a la derecha
(sesgo positivo)



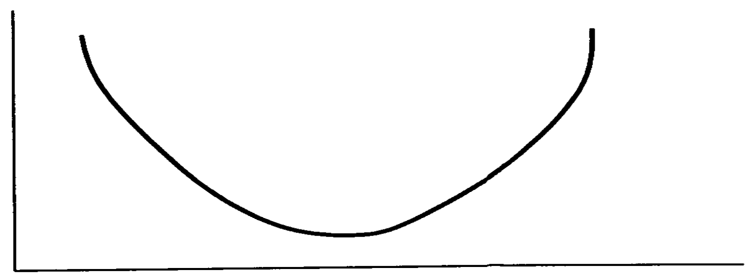
Sesgada a la izquierda
(sesgo negativo)



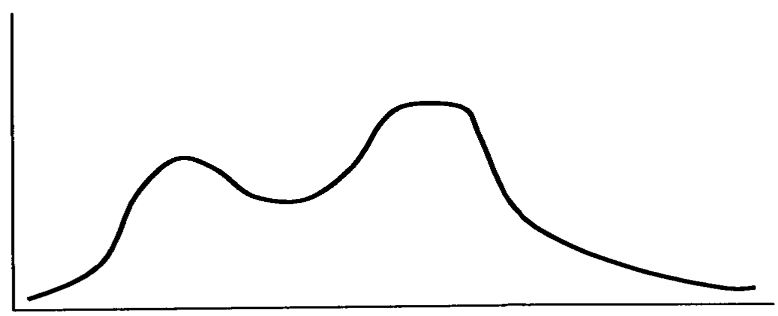
En forma de J



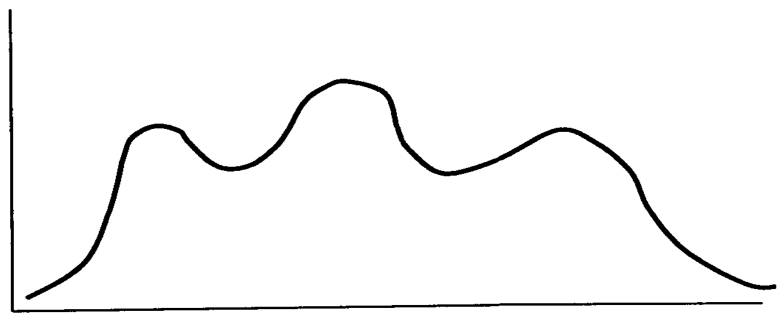
En forma de J invertida



En forma de U



Bimodal



Multimodal

Las **curvas de frecuencia simétrica** o **bien formadas** se caracterizan por el hecho de que las observaciones que equidistan del máximo central tienen la misma frecuencia. Un ejemplo importante es la **curva normal**.

En las **curvas de frecuencia moderadamente asimétricas** o **curvas de frecuencias sesgadas** la cola de la curva a un lado del máximo central es mayor que al otro lado. Si la cola mayor se presenta a la derecha de la curva se dice que ésta está **sesgada a la derecha** o que tiene **sesgo positivo**, mientras que si ocurre lo contrario se dice que la curva está **sesgada a la izquierda** o que tiene **sesgo negativo**.

En las curvas de **forma de** J o de J **invertida**, el máximo se presenta en un extremo.

Las curvas de frecuencias de **forma de** U tienen el maximo en ambos extremos.

Las curvas de frecuencias **bimodal** tienen dos máximos.

Una curva de frecuencias **multimodal** tienen más de dos máximos.

2.6 Actividades

1.- Formar a todos los estudiantes del grupo, en grupos de cinco personas para iniciar las actividades a desarrollar.

2.- ¿Por qué creen que al tener una serie de datos, éstos deben de ordenarse?

3.- Comenten los conceptos de frecuencia de clase y distribución de frecuencias. Además, desarrollar un ejemplo en donde puedan aplicar estos conceptos de tal manera que se puedan establecer los intervalos de clase y punto medio de clase.

4.- Discutir las reglas generales para formar una distribución de frecuencias.

5.- ¿Por qué es factible utilizar un sistema de coordenadas rectangulares para representar a los histogramas? ¿Es necesario que el sistema sea rectangular?

6.- Crear un ejemplo en el que se pueda graficar a través de un histograma. Dibuje el polígono de frecuencias.

7.- Discutir el concepto de distribución de frecuencias relativas. ¿Podrían dar una explicación del por qué es utilizable la palabra "relativa" en este concepto?

8.- Buscar en un diccionario enciclopédico el significado de la palabra "ojiva".

9.- ¿Cuáles son los distintos tipos de curvas? ¿Han visto algún tipo de curva que exprese algún dato? Si la respuesta es afirmativa, decir en dónde y qué tipo de datos eran. Si la respuesta es negativa, ¿puedes construir algunos ejemplos sencillos?

Capítulo 3

Media, Mediana, Moda y otras Medidas de Centralización

Este capítulo es fundamental dentro del programa, en el que el alumno deberá de comprender las distintas medidas de centralización que más son usadas. De incio, el problema de entendimiento es la notación y simbología que aquí se usan. Sin embargo, cuando realizan las actividades a desarrollar se aclaran todas las dudas.

3.1 Notación y Simbología

El símbolo X_i (léase: "X sub i") denota cualquiera de los n valores $X_1, X_2, \ldots, X_n$ que una variable X puede tomar. La letra i en X_i, la cual puede representar cualquiera de los números $1, 2, \ldots, n$, se llama **índice** o **subíndice**. Similarmente, pueden utilizarse como subíndices cualquier otra letra distinta de i, como j, k, l, p, q, s.

El símbolo $\sum_{i=1}^{n} X_i$ se utiliza para indicar la suma de todas la X_i's desde $i = 1$ hasta n $(i = 1, 2, \ldots, n)$, es decir,

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n.$$

Cuando no exista confusión sobre los valores $X_1, X_2, \ldots, X_n$, escribiremos la suma anterior simplemente por $\sum_i X_i$ o $\sum X_i$. El símbolo $\sum$ es la letra griega *sigma*, denotando sumación.

Así tenemos por ejemplo que

$$\sum_{i=1}^{n} X_i Y_i = X_1 Y_1 + X_2 X_2 + \cdots + X_n Y_n,$$

$$\begin{aligned}\sum_{i=1}^{n} aX_i &= aX_1 + aX_2 + \cdots + aX_n \\ &= a(X_1 + X_2 + \cdots + X_n) = a \sum_{i=1}^{n} X_i,\end{aligned}$$

donde a es una constante. Más simplemente, tenemos que

$$\sum aX_i = a \sum X_i.$$

3.2 Promedios y Medidas de Centralización

Un **promedio** es un valor que es típico o representativo de un conjunto de datos. Como tales valores tienden a situarse en el centro del conjunto de datos ordenados según su magnitud, los promedios se conocen también como **medidas de centralización**.

Se pueden definir diferentes tipos de medidas de centralización, las más comunes son la **media aritmética** o brevemente **media**, la **mediana**, la **moda**, la **media geométrica** y la **media armónica**. Cada una de ellas tienen sus ventajas e inconvenientes, dependiendo la aplicación de una u otra de los resultados que se pretendan sacar de los datos.

3.3 Media Aritmética

La **media aritmética** o **media** de un conjunto de n números $X_1, X_2, \ldots, X_n$ se representa por $\bar{X}$ (leáse: "X barra") y de define como

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{3.1}$$

y las **desviaciones** de $X_1, X_2, \ldots, X_n$ de su media aritmética se definen como:

$$X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}.$$

Si los números $X_1, X_2, \ldots, X_k$ se presentan $f_1, f_2, \ldots, f_k$ veces, respectivamente (es decir, se presentan con frecuencias $f_1, f_2, \ldots, f_k$), la media aritmética es

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + \cdots + f_kX_k}{f_1 + f_2 + \cdots + f_k} = \frac{\sum_{i=1}^{k} f_iX_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_iX_i}{n} \tag{3.2}$$

donde $n = \sum_{i=1}^{k} f_i$ es la **frecuencia total**, es decir, el número de casos.

En algunas ocasiones se asocia a los números $X_1, X_2, \ldots, X_k$ ciertos **factores** o **pesos** $w_1, w_2, \ldots, w_k$ que dependen del significado o importancia de cada unos de los números. En este caso,

$$\bar{X} = \frac{w_1X_1 + w_2X_2 + \cdots + w_kX_k}{w_1 + w_2 + \cdots + w_k} = \frac{\sum_{i=1}^{k} w_iX_i}{\sum_{i=1}^{k} w_i} \tag{3.3}$$

se llama **media aritmética ponderada** Nótese la similitud con la ecuación (3.2), que puede considerarse como una media aritmética con los pesos f_1, f_2,..., f_k.

Algunas propiedades de la media artimética son las siguientes:

- La suma algebraica de las desviaciones de un conjunto de números de su media aritmética es cero.

- La suma de los cuadrados de las desviaciones de un conjunto de números X_i's de cualquier número a es mínima solamente si $a = \bar{X}$.

- Si f_1 números tienen de media m_1, f_2 números tienen de media m_2,..., f_k números tienen de media m_k, entonces la media de todos los números es

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + \cdots + f_kX_k}{f_1 + f_2 + \cdots + f_k}$$

  es decir, una media aritmética ponderada de todas las medias.

- Si A es cualquier *supuesta media artimética* (que puede ser cualquier número) y si $d_i = X_i - A$ son las desviaciones de X_i de A, las ecuaciones (3.1) y (3.2) se convierten en

$$\bar{X} = A + \frac{\sum_{i=1}^{n} d_i}{n} \tag{3.4}$$

  y

$$\bar{X} = A + \frac{\sum_{i=1}^{k} f_i d_i}{\sum_{i=1}^{k} f_i} \tag{3.5}$$

  donde $n = \sum_{i=1}^{k} f_i$. Nótese que (3.4) y (3.5) están resumidas en la ecuación $\bar{X} = A + \bar{d}$.

Cuando los datos se presentan mediante una distribución de frecuencia, todos los valores caen dentro de unos intervalos de clase dados que se consideran coincidentes con las marcas de clase o puntos medios de cada intervalo. Las fórmulas (3.2) y (3.5) son válidas para tales datos agrupados si se interpreta X_i como la marca de clase, f_i su correspondiente frecuencia de clase, A una marca de clase cualquiera y $d_i = X_i - A$ las desviaciones de X_i de A.

Los cálculos con las fórmulas (3.2) y (3.5) se llaman a veces **métodos largos** y **cortos**, respectivamente.

Si todos los intervalos de clase tienen igual tamaño c, las desviaciones $d_i = X_i - A$ pueden expresarse como cu_i, donde los u_i's pueden ser números positivos o negativos o cero, es decir, 0, ± 1, ± 2,... y la fórmula (3.5) se convierte en

$$\bar{X} = A + \frac{\sum_{i=1}^{k} f_i u_i}{n} \cdot c \tag{3.6}$$

la cual es equivalente a la ecuación $\bar{X} = A + c\bar{u}$. Este se llama **método clave** para el cálculo de la media. Es un método muy corto y debería emplearse siempre para datos agrupados cuando los intervalos de clase son iguales. Adviértase que en el método clave los valores de la variable X *se transforman* en los valores de la variable u de acuerdo con $X = A + cu$.

3.4 Mediana

La **mediana** de una colección de datos ordenados en orden de magnitud es el valor medio o la media artimética de los valores medios.

Por ejemplo, la mediana de los números 3, 4, 4, 5, 6, 8, 8, 8, 10 es 6. Mientras que la mediana de los números 5, 5, 7, 9, 11, 12, 15, 18 es (9+11)/2=10.

Para datos agrupados, la mediana se obtiene mediante interpolación y viene dada por

$$M = L_1 + \frac{1}{f_M} \cdot \left(\frac{n}{2} - \left(\sum f \right)_1 \right) \cdot c \tag{3.7}$$

donde

L_1	es el límite real inferior de la clase mediana (es decir, la clase que contiene la mediana);
n	es el número total de datos (es decir, frecuencia total);
$(\sum f)_1$	es la suma de las frecuencias de todas las clases por debajo de la clase mediana;
f_M	es la frecuencia de la clase mediana;
c	es el tamaño del intervalo de la clase mediana.

Geométricamente, la mediana es el valor de X (abscisa) que corresponde a la vertical que divide un histograma en dos partes de igual área. Este valor de X se denota por $\tilde{X}$.

3.5 Moda

La **Moda** de una serie de números es aquel valor que se presenta con la mayor frecuencia, es decir, es el valor más común. La moda puede no existir, incluso si existe puede no ser única.

Por ejemplo, en la serie de valores 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 tiene moda 9. No tiene moda la serie de valor 3, 5, 8, 10, 12, 15, 16. En la serie de valor 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9 se tienen dos modas 4 y 7, y se llama **bimodal**. Una distribución que tiene una sola moda se llama **unimodal**.

En el caso de datos agrupados donde se ha construido una curva de frecuencias para ajustar los datos, la moda será el valor (o valor) de X correspondientes al máximo (o máximos) de la curva. Este valor de X se representa a veces por $\hat{X}$. De una distribución de frecuencias o un histograma de moda puede sacarse de la fórmula

$$Mo = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot c \tag{3.8}$$

donde

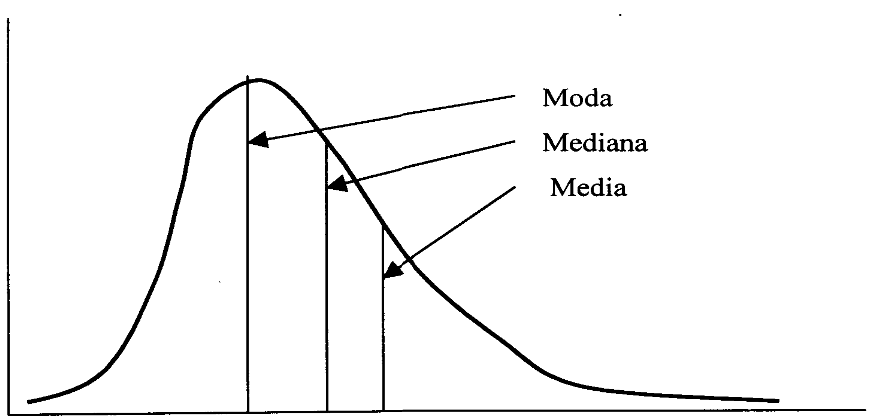
L_1	es el límite real inferior de la clase modal (es decir, la clase que contiene la moda);
Δ_1	es el exceso de la frecuencia modal sobre la frecuencia de la clase contigua inferior
Δ_2	es el exceso de la frecuencia modal sobre la frecuencia de la clase contigua superior
c	es el tamaño del intervalo de la clase modal.

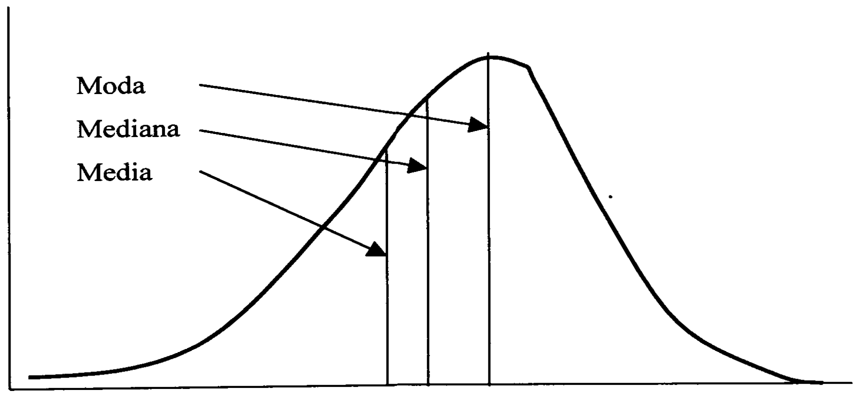
3.6 Relación Empírica entre Media, Mediana y Moda

Para curvas de frecuencias unimodales que sean moderadamente sesgadas (asimétricas), se tiene la relación empírica

$$\text{Media - Moda} = 3(\text{Media - Mediana}).$$

En las siguientes gráficas muestran las posiciones relativas de la media, mediana y moda para curvas de frecuencia que están sesgadas a la derecha y a la izquierda, respectivamente. Para curvas simétricas, la media, moda y me





3.7 Media Geométrica y Media Armónica

La **media geométrica** G de una serie de n números $X_1, X_2, \ldots, X_n$ está dada por

$$G = \sqrt[n]{X_1 X_2 \cdots X_n}. \tag{3.9}$$

Por otro lado, la **media armónica** H de una serie de n números $X_1, X_2, \ldots, X_n$ es la recíproca de la media aritmética de los recíprocos de los números, es decir,

$$H = \frac{1}{\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{X_i}} \tag{3.10}$$

o equivalentemente

$$\frac{1}{H} = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{X_i}. \tag{3.11}$$

La relación entre las medias aritmética, geométrica y armónica para una serie de números positivos $X_1, X_2, \ldots, X_n$ está dada como sigue:

$$H \leq G \leq \tilde{X} \tag{3.12}$$

donde la igualdad se presenta solamente cuando todos los números $X_1, X_2, \ldots, X_n$ son idénticos.

3.8 Raíz Cuadrada del Cuadrado de la Media

La **raíz cuadrada del cuadrado de la media (R.M.S.)** o **media cuadrática** de una serie de números $x_1, x_2, \ldots, x_n$ está dada por

$$R.M.S. = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}. \tag{3.13}$$

Este tipo de promedio se usa frecuentemente en aplicaciones físicas, sobre todo en el análisis estadístico de los experimentos que se realizan en el laboratorio.

3.9 Cuartiles, Deciles y Percentiles

Si una serie de datos se colocan en orden de magnitud, el valor medio (o la media aritmética de los dos valores medios) que divide al conjunto de datos en dos partes iguales es la *mediana*. Por extensión, de esta idea se puede pensar en aquellos valores que dividen a los datos en cuatro partes iguales. Estos valores, representados por Q_1, Q_2 y Q_3 se llaman **primero**,

segundo y **tercer cuartil**, respectivamente. El valor de Q_2 es igual al de la mediana.

Análogamente, los valores que dividen los datos en diez partes iguales se llaman **deciles** y se representan por $D_1, D_2, \ldots, D_9$, mientras que los valores que dividen los datos en cien partes iguales se llaman **percentiles** y se representan por $P_1, P_2, \ldots, P_{99}$. El quinto decil y el quincuagésimo percentil se corresponden con la mediana. Los percentiles P_{25} y P_{75} se corresponden con el primer y tercer cuartil, respectivamente.

En conjunto, cuartiles, deciles, percentiles y otros valores obtenidos por subdivisiones análogas de los datos se llaman **cuantiles**.

3.10 Actividades

1.- Formar a todos los estudiantes del grupo, en grupos de cinco personas para iniciar las actividades a desarrollar.

2.- ¿Por qué es importante la notación y la sombología en muchas áreas del conocimiento? Dar ejemplo de notación y simbología en diferentes áreas del conocimiento humano.

3.- Analizar cada una de las partes que contenga la notación simbólica de la suma de una cantidad finita de términos.

4.- Después de haber discutido las medidas de centralización: media, mediana y moda, obtener algunas diferencias de éstas en base a sus propiedades.

5.- El promedio de vida de un mexicano, ¿qué tipo de medida de centralización es?

6.- ¿Es posible que la moda sea más grande que la mediana, y esta última más grande que la media, o no existe relación alguna? Trate de argumentar sus respuestas. Si es posible, desarrollen algunos ejemplos que les permitan aclarar dichas relaciones.

7.- ¿Cuál es la relación que existe entre la media aritmética y la media armónica?

8.- Desarrolle un ejemplo en el que pueda obtener los valores que dividan los datos en diferentes partes iguales.

Capítulo 4

Desviación Típica y otras Medidas de Dispersión

Al grado en que los datos numéricos tienden a extenderse alrededor de un valor medio se le llama **variación** o **dispersión** de los datos. Se utilizan distintas medidas de dispersión o variación, las más empleadas son el rango, la desviación media, el rango semi-intercuartilítico, el rango entre percentiles 10-90 y la desviación típica.

4.1 Rango

El **rango** de un conjunto de números es la diferencia entre el mayor y el menor de todos ellos.

Por ejemplo, el rango de los números 2, 3, 3, 5, 5, 5, 8, 10, 12 es 10, y en ocaciones este rango se da por la simple anotación de los números mayor y menor: 2 a 12.

4.2 Desviación Media

La **desviación media** de una serie de n números $X_1, X_2, \ldots, X_n$ viene dada por

$$D.M. = \frac{\sum_{i=1}^{n} |X_i - \tilde{X}|}{n} \tag{4.1}$$

donde $\tilde{X}$ es la media aritmética de los números y $|X_i - \tilde{X}|$ es el valor absoluto de las desviaciones de las diferencias X_i de $\tilde{X}$. (El **valor absoluto** de un número real a se define como: $|a| = a$ si $a \geq 0$ y como $|a| = -a$ si $a < 0$; así, $|-4| = 4$ y $|6| = 6$.)

Si $X_1, X_2, \ldots, X_k$ se presentan con frecuencias $f_1, f_2, \ldots, f_k$, respectivamente, la desviación media puede escribirse como

$$D.M. = \frac{\sum_{i=1}^{k} f_i |X_i - \tilde{X}|}{n} \tag{4.2}$$

donde $n = \sum_{i=1}^{k} f_i$. Esta forma es útil para datos agrupados donde las diferentes X_i's representan las marcas de clase y las f_i's las correspondientes frecuencias de clase. Ocasionalmente, la desviación media se define como desviaciones absolutas de la mediana u otro promedio en lugar de la media. Una propiedad interesante de la suma $\sum_{i=1}^{n} |X_i - a|$ es que es mínima cuando a es la mediana, es decir, la desviación media respecto a la mediana es mínima.

Sería más apropiado utilizar el término **desviación media absoluta** que el de desviación media.

4.3 Desviación Cuartílica

La **desviación cuartílica**, también llamada **rango semi-intercuartílico**, de una serie de datos se define por

$$\text{Rango semi-intercuartilítico} = Q = \frac{Q_3 - Q_1}{2}$$

donde Q_1 y Q_3 son el primer y tercer cuartil de los datos. El rango intercuartílico $Q_3 - Q_1$ se emplea a veces, pero el rango semi-intercuartilítico es más utilizado como medida de dispersión.

4.4 Rango entre Percentiles 10-90

El **rango entre percentiles 10-90** de una serie de datos viene definida por

$$\text{Rango percentil 10-90} = P_{90} - P_{10}$$

donde P_{10} y P_{90} son los percentiles décimo y nonagésimo de los datos.

El **rango semipercentil 10-90**, $(P_{90} - P_{10})/2$, puede también emplearse aunque su empleo no es común.

4.5 Desviación Típica

La **desviación típica** de una serie de n números $X_1, X_2, \ldots, X_n$, se representa por s y se define por

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}} \tag{4.3}$$

Así, s es la raíz cuadrada del cuadrado medio de las desviaciones a la media, o como a veces se le llama, **raíz del cuadrado medio de las desviaciones**.

Si $X_1, X_2, \ldots, X_k$ se presentan con frecuencias $f_1, f_2, \ldots, f_k$, respectivamente, la desviación típica puede escribirse como

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i(X_i - \bar{X})^2}{n}} \tag{4.4}$$

donde $n = \sum_{i=1}^{k} f_i$.

Las propiedades de la desviación típica son las siguientes:

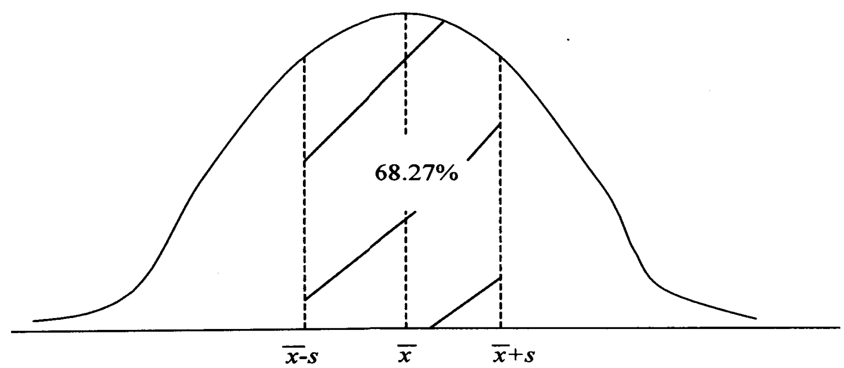
- La desviación típica puede definirse como

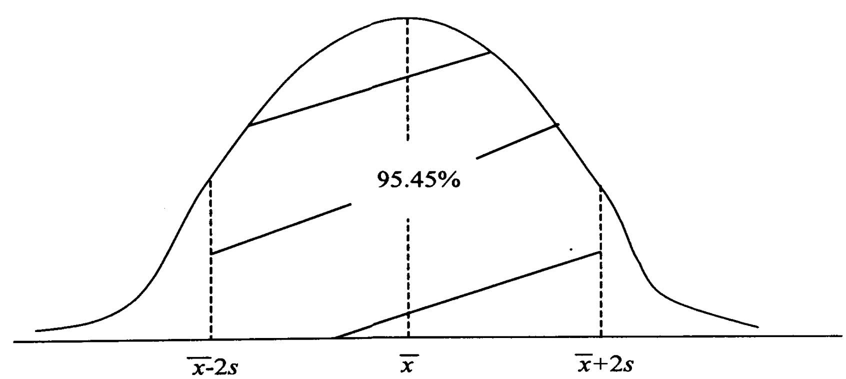
$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - a)^2}{n}}$$

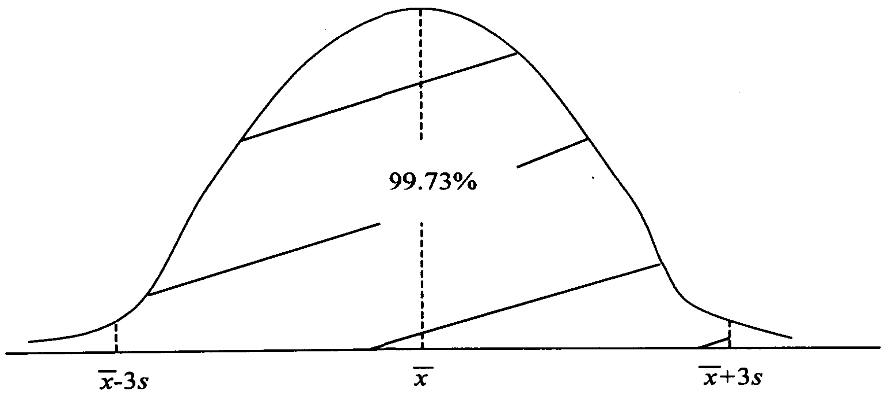
  donde a es un promedio que puede ser distinto de la media aritmética. De todas las desviaciones típicas, la mínima es aquella para la que $a = \bar{X}$. Esta propiedad suministra una razón de peso para definir la desviación típica como se ha definido anteriormente.

- Para distribuciones normales, resulta que el 68.27% de los casos están comprendidos entre $\bar{X} - s$ y $\bar{X} + s$, es decir, el valor de la desviación típica a ambos lados de la media. El 95.45% de los casos están comprendidos entre $\bar{X} - 2s$ y $\bar{X} + 2s$, es decir, el doble del valor de la desviación típica a ambos lados de la media. El 99.73% de los casos están comprendidos entre $\bar{X} - 3s$ y $\bar{X} + 3s$, es decir, el triple del valor de la desviación típica a ambos lados de la media. Obsérvese las gráficas que se presentan en seguida.

  Para distribuciones moderadamente asimétricas, los porcentajes anteriores pueden mantenerse aproximados.


68.27%
$\overline{x}$-s
$\overline{x}$
$\overline{x}$+s


95.45%
$\overline{x}$-2s
$\overline{x}$
$\overline{x}$+2s



En algunas ocasiones, la desviación típica de los datos de una muestra viene definida con $(n-1)$ en lugar de n en los denominadores de las expresiones (4.3) y (4.4), ya que el valor resultante representa un estimador mejor de la desviación típica de una población de la que se ha tomado una muestra. Para valores grandes de n (por ejemplo, $n > 30$), prácticamente no hay diferencia entre las dos definiciones. También, cuando se necesita el estimulador mejor, puede obtenerse siempre multiplicando la desviación típica calculada con la primera definición por $\sqrt{\dfrac{n}{n-1}}$. De aquí que se acustumbre a utilizar la primera definición. Para distribuciones moderadamente asimétricas se tienen las siguientes fórmulas empíricas entre las medidas de dispersión, a saber:

$$\begin{aligned}\text{Desviación media} &= \frac{4}{5}\ (\text{Desviación típica}).\\[1ex] \text{Rango semi-intercuartílico} &= \frac{2}{3}\ (\text{Desviación típica}).\end{aligned}$$

Estas son consecuencias del hecho de que para distribuciones normales se tiene que la desviación media y el rango semi-intercuartílico son, respectivamente, iguales a 0.7979 y 0.6745 veces la desviación típica.

4.6 Varianza

La **varianza** de un conjunto de datos se define como el cuadrado de la desviación típica y viene dada por s^2 en (4.3) y (4.4).

Cuando es necesario distinguir la desviación típica de una población de la desviación típica de una muestra sacada de esta ponblación, se emplea el símbolo s para la última y σ para la primera. Así, s^2 y σ^2 representan la **varianza muestral** y la **varianza poblacional**.

Supóngase dos series de datos de n_1 y n_2 números (o dos distribuciones de frecuencia con frecuencia totales n_1 y n_2) cuyas variaciones vienen dadas por s_1^2 y s_2^2, respectivamente, y que tienen la *misma* media $\bar{X}$. Entonces, la **varianza combinada** para ambas series (o ambas distribuciones de frecuencia) está dada por

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}.$$

Esta es una media aritmética ponderada de las varianzas. Este resultado puede generalizarse a 3 o más series de datos.

El cálculo de la desviación típica tiene algo de error, debido al agrupamiento de los datos en clases (error de agrupamiento). Para ajustarnos a la realidad se utiliza la **varianza corregida**

$$\text{Varianza corregida} = \text{Varianza de datos corregidos} - \frac{c^2}{12}$$

donde c es el tamaño del intervalo de clase. La corrección introducida $\frac{c^2}{12}$ se conoce como la **corrección Sheppard**. Se utiliza en distribuciones continuas donde las "colas" van gradualmente a cero en ambas direcciones.

Los estadísticos difieren en lo que se refiere a *cuándo* y *si* debe aplicarse la corrección de Sheppard. Ciertamente no debe aplicarse sin haber hecho un examen completo de la situación. Esto se debe a que frecuentemente se tiende a *sobre corregir* y así sustituir unos errores por otros.

4.7 Dispersión Absoluta y Relativa. Coeficiente de Variación

La dispersión o variación real determinada por la desviación típica u otra medida de dispersión se llama **dispersión absoluta**. Sin embargo, una dispersión o variación de 5 centímetros en la medida de una distancia de 1 metro tiene un efecto totalmente distinto al que tendría la misma variación de 5 centímetros en una distancia de 10 metros. Una medida de este efecto viene dada por la **dispersión relativa** definida por

$$\text{Dispersión realtiva} = \frac{\text{Dispersión absoluta}}{\text{Promedio}}$$

Si la dispersión absoluta es la desviación típica s y el promedio es la medida $\bar{X}$, la dispersión relativa se conoce por **coeficiente de variación** o **coeficiente de dispersión** dada por

$$\text{Coeficiente de variación} = V = \frac{s}{\bar{X}}$$

y está generalemente expresado como un porcentaje. Adviértase que el coeficiente de variación es independiente de las unidades utilizadas. Por esta razón es útil para comparar distribuciones donde las unidades pueden ser diferentes. Un inconveniente del coeficiente de variación es que deja de ser útil cuando $\bar{X}$ está próximo a cero.

4.8 Variable Normalizada. Referencias Tipificadas

La variable

$$z = \frac{X - \bar{X}}{s}$$

que mide la desviación de la medida en unidades de desviación típica se llama **variable normalizada** y sus cantidades son adimensionales (es decir, independientes de las unidades empleadas).

Si las desviaciones de la medida vienen dadas en unidades de desviación típica, se dice que están expresadas en **unidades tipificadas** o **referencias tipificadas**. Son de gran valor en la comparación de distribuciones.

4.9 Actividades

1.- Formar a todos los estudiantes del grupo, en grupos de cinco personas para iniciar las actividades a desarrollar.

2.- Analizar el concepto de rango a través de diferentes ejemplos.

3.- Calcule el valor absoluto de diferentes números.

4.- Discutir la palabra "mínima" dando ejemplos de esto. Por ejemplo, ¿cuál es la edad mínima de las personas que habitan la casa donde viven?

5.- Defina lo que es la disviación media.

6.- De algunas diferencias que existen entre la desviación media y la desviación típica. Puede considerar las propiedades de la desviación típica para que tenga más claridad en sus respuestas.

7.- Crear ejemplos para calcular la varianza, varianza muestral y varianza poblacional.

8.- Dar argumentos que permitan diferenciar las palabras "relativo" y "absoluto".

9.- ¿Cuál es la relación entre la dispersión relativa y la dispersión absoluta?

10.- Desarrollar un trabajo final de los ejemplos creados o investigados en donde se obtengan todos los datos estadísticos.

Capítulo 5

Conjuntos

En este capítulo daremos la herramienta indispensable que permita establecer el desarrollo de la probabilidad. Para empezar, daremos una introducción sobre las permutaciones, para luego continuar con el estudio del análisis combinatorio y la teoría de conjuntos.

5.1 Permutaciones

Supóngase que tenemos n objetos diferentes y que queremos ordenar k de estos objetos de alguna manera específica. Entonces, tenemos que hay n maneras de elegir el primer objeto. Después de haber elegido el primer objeto, tenemos que habrán $n-1$ maneras de escoger el segundo objeto. Siguiendo con el proceso, finalmente tendremos $n-k+1$ formas de elegir el k-ésimo objeto, y deducimos entonces que el número de **ordenaciones**, o **permutaciones** diferentes estará dado por:

$$n(n-1)(n-2)\cdots(n-k+1)$$

donde notamos que el producto anterior tiene k factores. Así, dicho producto será el número de *permutaciones* de n objetos tomados de k en k. En el caso particular en que tomemos $k=n$, el producto anterior se convierte en:

$$n! = n(n-1)(n-2)\cdots 2\cdot 1$$

el cual es llamado el **factorial** de n. De esta manera, podemos escribir

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

La fracción anterior tendrá sentido para $k = n$ si definimos $0! = 1$.

Ejemplo El número de permutaciones diferentes que consisten de 3 letras cada una y que pueden formarse de las 7 letras A, B, C, D, E, F, G es

$$\frac{7!}{(7-3)!} = \frac{7!}{4!} = \frac{7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{4\cdot 3\cdot 2\cdot 1} = 7\cdot 6\cdot 5 = 210.$$

5.2 Análisis Combinatorio

Notemos que en una permutación estamos interesados en el orden de la *distribución* de los objetos. Así, si *abc* es una permutación diferente de *cba*. Sin embargo, en muchas ocasiones es de interés solamente seleccionar o elegir los objetos sin orden alguno. Este último tipo de selecciones son llamadas **combinaciones**. Por ejemplo, *abc* y *cba* determinan la misma combinación.

El número total de combinaciones de k objetos seleccionados de n dados (también llamadas **combinaciones de** n **objetos de** k **en** k) está dada por el llamado **coeficiente binomial**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}.$$

Una propiedad de los coeficientes binomiales es la siguiente:

$$\binom{n}{k} = \binom{n}{n-k}.$$

Ejemplo El número de maneras en las cuales 3 cartas pueden elegirse de un total de 7 cartas diferentes es

$$\binom{7}{3} = \frac{7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{(3\cdot 2\cdot 1)(4\cdot 3\cdot 2\cdot 1)} = 35.$$

5.3 El Binomio de Newton

Los coeficientes binomiales provienen de la expansión binomial

$$(x+y)^n = \binom{n}{0} x^n + \binom{n}{1} x^{n-1}y + \binom{n}{2} x^{n-2}y^2 + \cdots + \binom{n}{n} y^n$$

llamado el **Binomio de Newton**.

Ejemplo

$$\begin{aligned} (x+y)^4 &= \binom{4}{0} x^4 + \binom{4}{1} x^3y + \binom{4}{2} x^2y^2 + \binom{4}{3} xy^3 + \binom{4}{4} y^4 \\ &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4. \end{aligned}$$

5.4 Conjuntos y sus Operaciones

Un **conjunto** es una colección de objetos, llamados **elementos** del conjunto. Denotaremos a un conjunto por las letras mayúsculas A, B, C, etc., y a los elementos por letras minúsculas a, b, c, etc. Si un elemento a pertenece a un conjunto A escribimos $a \in A$. Si a no es elemento de A escribimos $a \notin A$. Si a y b son elementos de A escribimos $a, b \in A$.

Un conjunto puede definirse haciendo una lista de sus elementos o, si es posible, describiendo una propiedad que tienen todos sus elementos. El primero se denomina **método de extensión** y el segundo **método de comprensión**. Así, por ejemplo, tenemos que el conjunto de todas las vocales definido por el método de extensión lo podemos expresar como $\{a, e, i, o, u\}$, o por el método de comprensión $\{x \mid x \text{ es una vocal}\}$, léase "el conjunto de los elementos x tales que x es una vocal" donde la línea vertical | se lee "tal que". También, el conjunto $\{x \mid x \text{ es un triángulo en un plano}\}$ es el conjunto de todos los triángulos en un plano, el cual no se puede definir por el método de extensión. Finalmente, si lanzamos un par de dados comunes, los "números" o "puntos" posibles que pueden resultar sobre la cara superior de cada dado son los elementos del conjunto $\{1, 2, 3, 4, 5, 6\}$.

Si cada elemento de un conjunto A también es un elemento de un conjunto B, entonces decimos que A es un **subconjunto** de B y escribimos $A \subset B$ o $B \supset A$ y se lee "A está contenido en B" o "B contiene a A". Podemos observar que $A \subset A$, para cualquier conjunto A.

Si A y B tienen exactamente los mismos elementos, entonces escribimos $A = B$; en caso contrario $A \neq B$. Si $A \subset B$ pero $A \neq B$, entonces decimos que A es un subconjunto **propio** de B.

Si $A \subset B$ y $B \subset C$, entonces tenemos claramente que $A \subset C$.

Para nuestros propósitos, especificaremos un conjunto en el cual todos los subconjuntos que se consideran serán subconjuntos de éste, el cual es llamado el **conjunto universo** y es denotado por la letra U cuyos elementos son llamados **puntos**. También es útil considerar un conjunto que no tiene elementos llamado el conjunto **vacío** y denotado por $\emptyset$; éste es subconjunto de cualquier conjunto. Por ejemplo, si lanzamos un dado, el conjunto de todos los resultados posibles es el universo $\{1, 2, 3, 4, 5, 6\}$; el conjunto de resultados que consisten de las caras 7 o 10 sobre un sólo dado es el conjunto vacío.

En algunas ocasiones, el conjunto universo U puede ser representado geométricamente por el conjunto de puntos dentro de un rectángulo. En tal caso, los subconjuntos de U se representan por conjuntos de puntos dentro de círculos. Tales diagramas, bien conocidos, se denominan **diagramas de Venn**, y sirven para darnos una intuición geométrica respecto a las posibles relaciones entre conjuntos.

Sobre los subconjunots de un conjunto universo U podemos realizar diferentes oparaciones entre ellos, a saber:

1. Unión. El conjunto de los elementos que pertenencen a A o a B se llama la **unión** de A y B y se escribe $A \cup B$.

2. Intersección. El conjunto de todos los elementos que pertenecen tanto a A como a B se llama la **intersección** de A y B y se escribe $A \cap B$.

3. Diferencia. Al conjunto de todos los elementos de A que no pertenencen a B se llama la **diferencia** de A y B, denotada por $A - B$.

4. Complemento. Si $B \subset A$, entonces $A - B$ se llama el **complemento de** B **relativo a** A y se escribe B^c_A. Si $A = U$, entonces $U - B$ simplemente diremos que es el **complemento** de B y se escribe B^c. El complemento de $A \cup B$ se escribe $(A \cup B)^c$.

Para cualesquiera A, B, C conjuntos de un universo U se tienen los siguientes teoremas:

Teorema 5.4.1 (Ley conmutativa de las uniones)

$$A \cup B = B \cup A.$$

Teorema 5.4.2 (Ley asociativa de las uniones)

$$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C.$$

Teorema 5.4.3 (Ley conmutativa de las intersecciones)

$$A \cap B = B \cap A.$$

Teorema 5.4.4 (Ley asociativa de las intersecciones)

$$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C.$$

Teorema 5.4.5 (Ley distributiva)

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Teorema 5.4.6 (Ley distributiva)

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Teorema 5.4.7 $A - B = A \cap B^c$.

Teorema 5.4.8 *Si* $A \subset B$*, entonces* $B^c \subset A^c$.

Teorema 5.4.9 $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$.

Teorema 5.4.10 $A \cup U = U$, $A \cap U = A$.

Teorema 5.4.11 (Ley De Morgan)

$$(A \cup B)^c = A^c \cap B^c.$$

Teorema 5.4.12 (Ley De Morgan)

$$(A \cap B)^c = A^c \cup B^c.$$

Teorema 5.4.13 $A = (A \cap B) \cup (A \cap B^c)$.

Si los conjuntos A y B satisfacen que $A \cap B = \emptyset$, entonces decimos que A y B son **disjuntos**.

5.5 Actividades

1.- Formar a todos los estudiantes del grupo, en grupos de cinco personas para iniciar las actividades a desarrollar.

2.- Crear ejemplos en donde se calculen permutaciones de objetos.

3.- Desarrollar expansiones binomiales en ejemplos concretos.

4.- Establecer el concepto de análisis combinatorio.

5.- Dar ejemplos de subconjuntos de un conjunto universo determinado.

6.- Aplicar los diagramas de Venn en casos particulares para obtener la unión, intersección, y diferencia entre conjuntos.

7.- ¿Cómo podría representarse el conjunto vacío mediante los diagramas de Venn?

8.- Verificar los teoremas sobre conjuntos a ejemplos específicos.

9.- Aplicar el método de extensión y el método de comprensión a conjuntos específicos.

Capítulo 6

La Probabilidad

La probabilidad tuvo sus inicios en el siglo XVII con los esfuerzos de los matemáticos como Fermat y Pascal al resolver problemas relacionados con los juegos de azar, y es hasta el siglo XX cuando se desarrolla una teoría matemática rigurosa, es decir, basada sobre axiomas, definiciones y teoremas. La probabilidad tiene aplicaciones en ingeniería, ciencias y matemáticas, así como en la agricultura, la administración de empresas, la medicina y la sicología.

6.1 Experimentos Aleatorios, Espacio Muestral y Eventos

Un principio fundamental de los experimentos es que si se efectúan repetidamente bajo condiciones similares, entonces obtenemos resultados que son esencialmente los mismos. Sin embargo, hay experimentos en los cuales esto no ocurre. Estos últimos son llamados **experimentos aleatorios**. Por ejemplo, si lanzamos una moneda, el resultado del experimento es un "sol", simbolizado por s, o una "águila", simbolizada por a, es decir, será uno de los elementos del conjunto $\{s, a\}$, y si lanzamos la misma moneda dos veces el resultado del experimento será uno de los elementos del conjunto $\{ss, sa, as, aa\}$.

Un conjunto $\mathcal{S}$ que consiste de todos los resultados de un experimento aleatorio se llama un **espacio muestral** y cada uno de los resultados se llama **punto muestral**. Obsérvese que $\mathcal{S}$ corresponde al conjunto universo. Si el espacio muestral tiene un número finito de puntos, entonces decimos que el espacio muestral es **finito**; en caso contrario, decimos que el espacio muestral es **infinito**. Cuando el espacio muestral es infinito y los puntos se pueden ir numerando con el conjunto de los números naturales ($\{1, 2, 3, \ldots\}$), entonces decimos que el espacio muestral es **infinito numerable**, en caso contrario diremos que el espacio muestral es **infinito no numerable**. Un espacio muestral que es finito o infinito numerable se suele llamar **espacio muestral discreto**, en caso contrario, es decir si es infinito no numerable, se suele llamar **espacio muestral no discreto**.

Un **evento** es un subconjunto A del espacio muestral $\mathcal{S}$, es decir, es un conjunto de resultados posibles en un experimento. Si el resultado de un experimento es un elemento de A, decimos que el evento A ha ocurrido. Un evento que consta de un sólo punto de $\mathcal{S}$ frecuentemente se llama un **evento elemental** o **simple**.

Ejemplo. Si lanzamos una moneda dos veces, el evento de que sólo resulte una cara es el subconjunto del espacio muestral que consiste de los puntos sa y as.

Como eventos particulares tenemos el evento en sí mismo, que es el **evento cierto** o **seguro** ya que un elemento de $\mathcal{S}$ debe de ocurrir, y el conjunto $\emptyset$, que se llama el **evento imposible** puesto que un elemento de $\emptyset$ no puede ocurrir.

Observemos que puesto que los eventos son conjuntos, esto se pueden operar a través de las propiedades de los conjuntos. Por ejemplo, $A \cup B$ es el evento "A o B". Si los conjuntos corresponden a eventos excluyentes, es decir $A \cap B = \emptyset$, entonces decimos que los eventos son **mutuamente excluyentes**. Esto quiere decir que pueden ocurrir ambos.

6.2 Probabilidad

Cuando se realiza un experimento aleatorio siempre hay incertidumbre sobre si el evento específico ocurrirá o no. De esto, podemos establecer que como medida de *probabilidad* con la que podemos esperar que un evento ocurra es conveniente asignar un número entre 0 y 1. Si se está seguro que el evento ocurrirá decimos que su probabilidad será del 100% ó 1, pero si estamos seguros de que el evento no ocurrirá decimos que su probabilidad será cero. Por ejemplo, la probabilidad será de 1/4 cuando existe el 25% de oportunidad de que ocurra y un 75% de que no ocurra; equivale a decir que la *probabilidad* contra su ocurrencia es del 75% al 25% o 3 a 1.

Existen dos procedimientos importantes por medio de los cuales podemos obtener estimaciones para la probabilidad de un evento. El primero es el llamado **enfoque clásico** o **a priori**, en el cual se tiene un evento que puede ocurrir de m maneras diferentes de un número total de n maneras posibles, todos igualmente factibles, entonces la probabilidad del evento es m/n. Por ejemplo, supóngase que se tiene la probabilidad de que resulte una cara en un sólo lanzamiento de una moneda. Puesto que hay dos maneras igualmente factibles del resultado de la moneda, simplmente "sol" o "águila" (suponiendo que la moneda no se pierda ni se caiga verticalmente), y de estas dos maneras una cara puede aparecer en una sóla manera, entonces tenemos que la probabilidad requerida es 1/2. Al llegar a este resultado, suponemos que la moneda no está alterada, es decir, no está cargada.

El segundo procedimiento es el llamado **enfoque de frecuencia relativa** o **a posteriori** el cual consiste en lo siguiente: si después de n repeticiones de un experimento, donde n es muy grande, un evento ocurre m veces, entonces la probabilidad del evento es m/n. Este procedimiento también se llama **probabilidad empírica** del evento. Por ejemplo, si lanzamos una moneda 1000 veces y nos encontramos que 574 veces resultan caras, estimamos que la probabilidad de dicha cara es 574/1000=0.574.

Ambos procedimientos mencionados, el enfoque clásico y el de frecuencia, presentan ciertas dificultades, el primero debido a la vaguedad de las "igualmente factibles" y el segundo debido a la vaguedad determinada por las palabras "número muy grande". Así que los matemáticos se dirigieron a la cuestión axiomática, utilizando conjuntos, para resolver estas dificultades.

6.3 Axiomas y Teoremas de la Probabilidad

Supóngase que tenemos un espacio muestral $\mathcal{S}$. Si $\mathcal{S}$ es discreto, entonces todos los subconjuntos corresponden a eventos y recíprocamente, pero si $\mathcal{S}$ es continuo, solamente subconjuntos especiales (llamados **medibles**) corresponderán a eventos. Así, a cada evento A en la clase $\mathcal{C}$ de eventos le asociamos un número real $P(A)$, es decir, P es una función de valor real definida en $\mathcal{C}$, llamada **función de probabilidad**, y diremos que $P(A)$ es la **probabilidad del evento** A, si se satisfacen los siguientes axiomas:

Axioma 1. Para cada evento A en la clase $\mathcal{C}$

$$P(A) \geq 0.$$

Axioma 2. Para el evento cierto o seguro $\mathcal{S}$ en la clase C

$$P(\mathcal{S}) = 1.$$

Axioma 3. Para cualesquier número de eventos mutuamente excluyentes $A_1, A_2, \ldots$ en la clase $\mathcal{C}$

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

Notemos, en particular, que del Axioma 3, si A y B son dos eventos excluyentes en la clase $\mathcal{C}$, entonces

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

A través de los axiomas anteriores para una función de probabilidad P, se tienen los siguientes teoremas:

Teorema 6.3.1 *Si* $A_1 \subset A_2$, *entonces* $P(A_1) \leq P(A_2)$ *y* $P(A_2 - A_1) = P(A_2) - P(A_1)$.

Teorema 6.3.2 *Para cada evento* A*, su probabilidad está entre* 0 *y* 1*, es decir,*

$$0 \leq P(A) \leq 1.$$

Teorema 6.3.3 *El evento imposible tiene probabilidad cero, es decir,* $P(\emptyset) = 0$.

Teorema 6.3.4 *Si* A^c *es el complemento de* A*, entonces* $P(A^c) = 1 - P(A)$.

Teorema 6.3.5 *Si* $A = A_1 \cup A_2 \cup \cdots \cup A_n$*, donde los eventos* $A_1, A_2, \ldots, A_n$ *son mutamente excluyentes, entonces*

$$P(A) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

En particular, si $A = \mathcal{S}$*, el espacio muestral, se tiene que*

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1.$$

Teorema 6.3.6 *Si* A *y* B *son dos eventos cualesquiera, entonces*

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

Generalizando cuando A_1, A_2, A_3 *son tres eventos cualesquiera, entonces*

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Teorema 6.3.7 *Para dos evento* A *y* B *se tiene que*

$$P(A) = P(A \cap B) + P(A \cap B^c).$$

Teorema 6.3.8 *Si un evento A debe de resultar en uno de los eventos mutuamente excluyentes $A_1, A_2, \ldots, A_n$, entonces*

$$P(A) = P(A \cap A_1) + P(A \cap A_2) + \cdots + P(A \cap A_n).$$

Ahora bien, construyamos algunas asignaciones de probabilidades. Por ejemplo, supóngase que tenemos un espacio muestral $\mathcal{S}$ el cual consiste únicamente de los eventos elementales $A_1, A_2, \ldots, A_n$ entonces, por el Teorema 6.3.5, tenemos que

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1.$$

Se excluye que podemos elegir arbitrariamente cualquier número no negativo para las probabilidades de estos eventos simples y cuando se satisfaga el Teorema 6.3.1. En particular, si suponemos *probabilidades iguales* para todos los eventos simples, entonces

$$P(A_k) = \frac{1}{n} \text{ para } k = 1, 2, \ldots, n$$

y si A es un evento compuesto por m eventos simples tenemos que

$$P(A) = \frac{m}{n}.$$

Esto equivale al enfoque clásico de la probabilidad establecido en la sección anterior. También podríamos emplear otros procedimientos para asignar probabilidades, como el de la frecuencia relativa establecida en la sección anterior.

La asignación de probabilidades provee un *modelo matemático* y su éxito debe de probarse experimentalmente tal y como se realiza en física o en cualquier otra ciencia.

Ejemplo. Se lanza sólo un dado. Se desea encontrar la probabilidad de que resulte 2 ó 5. Para esto, tenemos que el espacio muestral es $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. Luego, si asignamos probabilidades iguales a los puntos muestrales, o sea, si suponemos que el dado no está alterado, entonces

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}.$$

Así, el evento que resulte 2 ó 5, indicado por $2 \cup 5$, estará dado por

$$P(2 \cup 5) = P(2) + P(5) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

6.4 Probabilidad Condicional

Consideraremos ahora dos eventos A y B de los cuales $P(A) > 0$. Denotaremos por $P(B|A)$ la probabilidad de B cuando A ha ocurrido. Puesto que A ha ocurrido, A se convierte en el nuevo espacio muestral reemplazando el original $\mathcal{S}$. De aquí que la definición de $P(B|A)$ es:

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

equivalentemente

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

Así, podemos decir, de la definición anterior, que la probabilidad de A ocurra tantas veces la probabilidad de que B ocurra ya que A ha ocurrido. Decimos entoncces que $P(B|A)$ es la **probabilidad condicional** de B dada A, es decir, la probabilidad de B cuando A ha ocurrido. La probabilidad condicional satisface los axiomas establecidos en la Sección 6.3.

Ejemplo. Hallar la probabilidad de que en un sólo lanzamiento de un dado resulte un número menor que 4 cuando

(a) No se da ninguna información.

(b) El lanzamiento resultó un número impar.

Para rsolver (a), tenemos que si B es el evento menor que 4, entonces $B = \{1, 2, 3\}$. Luego, suponiendo probabilidades iguales para los puntos muestrales, tenemos que

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

Por otro lado, para obtener (b), tenemos que si A es el evento número impar, entonces $P(A) = 3/6 = 1/2$. También, $P(A \cap B) = 2/6 = 1/3$. Luego, entonces

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

Por lo tanto, el saber que el resultado del lanzamiento es un número impar aumenta la probabilidad de $1/2$ a $2/3$.

Veamos dos teoremas sobre la probabilidad condicional.

Teorema 6.4.1 *Para tres eventos cualesquiera A_1, A_2, A_3 se tiene*

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_3).$$

El teorema anterior establece que la probabilidad de que A_1, A_2 y A_3 ocurran es igual a la probabilidad de que A_1 ocurra tantas veces la probabilidad de que A_2 ocurra cuando A_1 ha ocurrido tantas veces la probabilidad de que A_3 ocurra dado que A_1 y A_2 han ocurrido. Este resultado puede ser generalizado a n eventos.

Teorema 6.4.2 *Si un evento A debe de resultar en uno de los eventos mutuamente excluyentes $A_1, A_2, \ldots, A_n$, entonces*

$$P(A) = P(A_1)P(A|A_1) + P(A_2)P(A|A_2) + \cdots + P(A_n)P(A|A_n).$$

Si $P(B|A) = P(B)$, es decir, la probabilidad de que B ocurra no está afectada por la ocurrencia o no ocurrencia de A, entonces decimos que A y B son **eventos independientes**. Esto es equivalentes a decir que

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

Si A_1, A_2, A_3 son independientes, entonces deben de ser independientes por parejas, es decir,

$$P(A_i \cap A_k) = P(A_i)P(A_k)$$

si $i \neq k$ donde $i, k = 1, 2, 3$. Pero también debemos de tener que

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3).$$

Teorema 6.4.3 (Regla de Bayes) *Supóngase que $A_1, A_2, \ldots, A_n$ son eventos mutuamente excluyentes cuya unión es el espacio muestral $\mathcal{S}$, es decir, uno de los eventos debe de ocurrir. Entonces, si A es cualquier evento, se cumple que*

$$P(A_k|A) = \frac{P(A_k)P(A|A_k)}{\displaystyle\sum_{i=1}^{n} P(A_k)P(A|A_i)}.$$

Con el teorema anterior, podemos hallar las probabilidades de los diferentes eventos $A_1, A_2, \ldots, A_n$ que pueden causar la ocurrencia de A. Por tal razón, con frecuencia se hace referencia al Teorema de Bayes como el Teorema sobre la *Probabilidad de Causas*.

6.5 Actividades

1.- Formar a todos los estudiantes del grupo, en grupos de cinco personas para iniciar las actividades a desarrollar.

2.- Establecer la importancia del estudio de la probabilidad.

3.- Dar ejemplos concretos de experimentos aleatorios.

4.- Discutir los conceptos de espacio muestral, punto muestral y el de evento.

5.- ¿Cuándo se establece que dos eventos son mutuamente excluyentes? Dar ejemplos.

6.- Discutir el concepto de probabilidad.

7.- Establecer las diferencias que hay entre el enfoque clásico o a priori y el enfoque como frecuencia relativa o a posteriori.

8.- Dar ejemplos en el que se apliquen los axiomas y teoremas de la probabilidad.

9.- De un ejemplo en el que se puedan aplicar diferentes probabilidades.

10.- Discutir el concepto de probabilidad condicional.

Conclusiones

Se debe de hacer un gran esfuerzo para dar un conocimiento que sea motivante a los estudiantes. No es fácil aplicar estas nuevas ideas, pero en ciertos casos ha dados buenos resultados entre los estudiantes. Por ejemplo, se da disposición de trabajo en forma colectiva que permite la discusión de los conocimientos que se están adquiriendo, de tal manera que la clase es el primer contacto con éstos. Se desarrolla la habilidad de análisis y comprensión de los conocimientos de manera racional para que los estudiantes no sientan que las matemáticas son difíciles y aburridas. Comprenden cómo es que las matemáticas son útiles en nuestra vida cotidiana al crear ejemplos donde la estadística y la probabilidad son aplicables, por lo que deben de estar concientes del mundo real en que viven para así poder obtener con éxito lo que se les ha pedido y puedan aclarar las dudas que van surgiendo. Con todo esto, se espera que el estudiante se sienta como parte de la actividad del saber y motivados para el trabajo de un conocimiento posterior.

Bibliografía

[1] Hoel, P. G., *Estadística Elemental*, Compañía Editorial Continental, S. A. de C. V., 1979.

[2] Mendenhall, W., *Introducción a la Probabilidad y la Estadística*, Wadsworth Internacional/Iberoamérica, 1982.

[3] Spiegel, M. R., *Probabilidad y Estadística*, Mc Graw-Hill/Interamericana de México, S. A. de C. V., 1991-1975.

Índice