



UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO

24  
2EJ

FACULTAD DE CIENCIAS

USO DE LA CALCULADORA EN LOS CURSOS DE  
ESTADISTICA DEL COLEGIO DE CIENCIAS Y  
HUMANIDADES, NIVEL BACHILLERATO  
(PROPUESTA DIDACTICA)

T E S I S  
QUE PARA OBTENER EL TITULO DE:  
A C T U A R I O  
P R E S E N T A:

MAXIMINO DIAZ GUTIERREZ



FACULTAD DE CIENCIAS  
SECCION ESCOLAR



MEXICO, D. F.

1995

FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

M. en C. Virginia Abrin Bacule  
Jefe de la División de Estudios Profesionales de la  
Facultad de Ciencias  
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis: "USO DE LA CALCULADORA EN  
LOS CURSOS DE ESTADÍSTICA DEL COLEGIO DE CIENCIAS Y HUMANIDADES, NIVEL BACHILLERATO"

realizado por MAXIMINO DIAZ GUTIERREZ

con número de cuenta 6110406-2 , pasante de la carrera de ACTUARIA

Dicho trabajo cuenta con nuestro voto aprobatorio.

**Asesores**

Director de Tesis Propietario	ACT. ISABEL PATRICIA CAFAGGI FELIX
Propietario	M. en C. AGUSTIN ONTIVEROS PINEDA
Propietario	ACT. HUMBERTO SANTILLANA LOYO
Suplente	M. en C. MANUEL JESUS FALCONI MAGANA
Suplente	M. en C. ALEJANDRO BRAVO MOJICA

*[Handwritten signatures and initials]*

*[Handwritten signature]*  
MAT. CESAR GUEVARA BRAVO



A MI ESPOSA.

MARIA TERESA

A MIS HIJOS.

MAX ARTURO

TANIA Y

JOSE ANTONIO

**C A P I T U L O 1**  
**LA ESTADISTICA Y EL METODO CIENTIFICO**  
**CONCEPTOS BASICOS**

1.1 INTRODUCCION .....	2
1.2 BOSQUEJO HISTORICO .....	6
1.3 ESTADISTICA ESCRIPTIVA, ESTADISTICA, INFERENCIAL, POBLACION Y MUESTRA. ....	8
1.4 ALCANCES DE LA ESTADISTICA, METODO CIENTIFICO ..	12
1.5 ESTADISTICA Y METODO CIENTIFICO. ....	15
1.6 MEDICION DE FENOMENOS, ESCALAS DE MEDICION. ...	17
1.7 EJERCICIOS DE REPASO. ....	24
1.8 FENOMENOS ALEATORIOS Y FENOMENOS DETERMINISTAS .	26
1.9 VARIABLES, VARIABLES CONTINUAS, VARIABLES DISCRETAS O DISCONTINUAS, VARIABLES DEPENDIENTES Y VARIABLES INDEPENDIENTES. ....	30

**CAPITULO 2**  
**REPRESENTACION GRAFICA DE DATOS**

2.1 UTILIDAD DE LAS GRAFICAS. ....	34
2.2 GRAFICA LINEAL. ....	35
2.3 GRAFICA DE BARRAS. ....	37
2.4 GRAFICA CIRCULAR. ....	41
2.5 GRAFICA PICTORICA. ....	45
2.6 EJERCICIOS DE REPASO. ....	46

### CAPITULO 3

#### DISTRIBUCION DE FRECUENCIAS E HISTOGRAMAS

3.1 ORGANIZACION, AGRUPACION Y REPRESENTACION DE DATOS. ....	50
3.2 TABLA DE FRECUENCIAS. ....	51
3.3 INTERVALO DE CLASE Y LIMITES DE CLASE. ....	55
3.4 LIMITES REALES DE CLASE. ....	56
3.5 TAMAÑO O ANCHURA DE UN INTERVALO DE CLASE. ....	58
3.6 MARCA DE CLASE. ....	58
3.7 REPRESENTACION GRAFICA DE DATOS. ....	59
3.8 HISTOGRAMA, POLIGONO DE FRECUENCIAS Y LA OJIVA. ....	60
3.9 DISTRIBUCION DE FRECUENCIAS RELATIVAS. ....	63
3.10 DISTRIBUCION DE FRECUENCIAS ACUMULADAS, TABLAS DE FRECUENCIA ACUMULADAS Y OJIVAS. POLIGONOS DE FRECUENCIAS ACUMULADAS, OJIVAS "MENOR QUE" Y "MAYOR QUE". ....	64
3.11 DISTRIBUCION DE FRECUENCIAS RELATIVAS ACUMULADAS, OJIVAS PORCENTUALES. ....	68
3.12 EJERCICIOS DE REPASO. ....	74

**CAPITULO 4**  
**MEDIDAS DE TENDENCIA CENTRAL**  
**MEDIA ARITMETICA, MEDIANA, MODA Y CUANTILES**

4.1	LA NOTACION SIGMA . . . . .	79
4.2	PROMEDIOS Y MEDIDAS DE CENTRALIZACION. . . . .	84
4.3	LA MEDIA ARITMETICA Y SU CALCULO PARA DATOS NO AGRUPADOS. . . . .	85
4.4	CALCULO DE LA MEDIA ARITMETICA PARA DATOS NO AGRUPADOS, USANDO LA CALCULADORA. . . . .	89
4.5	PROPIEDADES DE LA MEDIA ARITMETICA. . . . .	94
4.6	VENTAJAS Y DESVENTAJAS DE LA MEDIA ARITMETICA. . . . .	101
4.7	LA MEDIANA.SU CALCULO PARA DATOS NO AGRUPADOS. . . . .	102
4.8	VENTAJAS Y DESVENTAJAS DE LA MEDIANA. . . . .	104
4.9	LA MODA Y SU CALCULO PARA DATOS NO AGRUPADOS. . . . .	104
4.10	EJERCICIOS DE REPASO. . . . .	105
4.11	CALCULO DE LA MEDIA ARITMETICA PARA DATOS AGRUPADOS. . . . .	110
4.12	CALCULO DE LA MEDIA ARITMETICA PARA DATOS AGRUPADOS, USANDO LA CALCULADORA. . . . .	112
4.13	CALCULO DE LA MEDIANA PARA DATOS AGRUPADOS. . . . .	114
4.14	CALCULO DE LA MEDIANA PARA DATOS AGRUPADOS USANDO CALCULADORA. . . . .	117
4.15	CALCULO DE LA MODA PARA DATOS AGRUPADOS. . . . .	124
4.16	CALCULO DE LA MODA PARA DATOS AGRUPADOS USANDO CALCULADORA. . . . .	126
4.17	CUARTILES, DECILES Y PERCENTILES (CUANTILES). . . . .	128
4.18	CALCULO DE LOS CUANTILES PARA DATOS AGRUPADOS. . . . .	130
4.19	CALCULO DE LOS CUANTILES PARA DATOS AGRUPADOS CON CALCULADORA. . . . .	134
4.20	EJERCICIOS DE REPASO. . . . .	137

**C A P I T U L O 5**  
**MEDIDAS DE DISPERSION, SESGO Y CURTOSIS**

5.1	MEDIDAS DE DISPERSION. ....	142
5.2	RANGO O RECORRIDO. ....	142
5.3	LA VARIANZA Y LA DESVIACION ESTANDAR O TIPICA Y SU CALCULO PARA DATOS NO AGRUPADOS. ....	144
5.4	CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS NO AGRUPADOS, POR FORMULA SIMPLIFICADA. ....	153
5.5	CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS NO AGRUPADOS CON CALCULADORA. ....	155
5.6	EL COEFICIENTE DE VARIACION. ....	157
5.7	EJERCICIOS DE REPASO. ....	159
5.8	CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS AGRUPADOS. ....	162
5.9	CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS AGRUPADOS POR FORMULA SIMPLIFICADA... ..	163
5.10	CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS AGRUPADOS CON CALCULADORA.....	164
5.11	INTERPRETACION DE LA DESVIACION ESTANDAR. LA REGLA EMPIRICA. ....	165
5.12	SESGO. ....	172
5.13	CALCULO DEL COEFICIENTE DE SESGO CON CALCULADORA. ....	174
5.14	CURTOSIS, COEFICIENTE DE CURTOSIS PERCENTILICO. ....	176
5.15	CALCULO DEL COEFICIENTE DE CURTOSIS PERCENTILICO CON CALCULADORA. ....	179
5.16	EJERCICIOS DE REPASO. ....	180

## CAPITULO 6

### MEDIDAS DE CORRELACION

6.1	INTRODUCCION. ....	185
6.2	CORRELACION LINEAL. ....	186
6.3	COEFICIENTE DE CORRELACION. ....	188
6.4	RECTA DE REGRESION. ....	195
6.5	ERROR ESTANDAR DE ESTIMACION. ....	200
6.6	CALCULO DEL COEFICIENTE DE CORRELACION Y DE LA RECTA DE REGRESION CON LA CALCULADORA. ....	203
6.7	EJERCICIOS DE REPASO. ....	208

## CAPITULO 7

### PROBABILIDAD

7.1 BOSQUEJO HISTORICO. ....	214
7.2 TECNICAS DEL CONTEO, FACTORIAL DE UN NUMERO NATURAL. ....	221
7.3 EL PRINCIPIO FUNDAMENTAL DEL CONTEO Y LAS COMBINACIONES. ....	223
7.4 CALCULO DE LAS COMBINACIONES USANDO LA CALCULADORA. ....	225
7.5 EJERCICIOS DE REPASO. ....	226
7.6 ESPACIO MUESTRAL Y EVENTOS. ....	227
7.7 DOS ENFOQUES DE PROBABILIDAD. ....	231
7.8 AXIOMAS DE PROBABILIDAD. ....	241
7.9 TEOREMAS DE PROBABILIDAD. ....	243
7.10 ESPACIOS FINITOS DE PROBABILIDAD. ....	224
7.11 EJERCICIO DE REPASO. ....	250
7.12 ESPACIOS FINITOS EQUIPROBABLES. ....	254
7.13 PROBABILIDAD CONDICIONAL. ....	258
7.14 TEOREMA DE LA MULTIPLICACION. ....	263
7.15 EVENTOS INDEPENDIENTES. ....	266
7.16 EJERCICIOS DE REPASO. ....	269

**CAPITULO 8**

**DISTRIBUCIONES DE PROBABILIDAD**

**\* DISTRIBUCION BINOMIAL.**

**\* DISTRIBUCION NORMAL.**

8.1	DISTRIBUCIONES DE PROBABILIDAD. ....	274
8.2	VARIABLES ALEATORIAS. ....	274
8.3	DISTRIBUCIONES DE PROBABILIDAD DE UNA VARIABLE DISCRETA. ....	276
8.4	VALOR ESPERADO, VARIANZA Y DESVIACION ESTANDAR DE UNA DISTRIBUCION DE PROBABILIDAD DISCRETA. ....	277
8.5	LA DISTRIBUCION BINOMIAL. ....	285
8.6	CALCULO DE LA BINOMIAL USANDO LA CALCULADORA. ....	287
8.7	LA DISTRIBUCION BINOMIAL COMO UNA DISTRIBUCION DE PROBABILIDAD DISCRETA. ....	289
8.8	PROPIEDADES DE LA DISTRIBUCION BINOMIAL. ...	290
8.9	LA DISTRIBUCION NORMAL. ....	291
8.10	LA DISTRIBUCION NORMAL UNITARIA O NORMAL ESTANDAR. ....	299
8.11	CALCULO DE AREAS DE LA DISTRIBUCION NORMAL UNITARIA USANDO LA CALCULADORA. ....	307
8.12	EJERCICIOS DE REPASO. ....	314

**APENDICE**

SOLUCION A LOS EJERCICIOS DE REPASO. ....	319
REFERENCIAS BIBLIOGRAFICAS. ....	326

# **C A P I T U L O 1**

## **LA ESTADISTICA Y EL METODO CIENTIFICO**

### **CONCEPTOS BASICOS**

- 1.1 INTRODUCCION
- 1.2 BOSQUEJO HISTORICO
- 1.3 ESTADISTICA DESCRIPTIVA, ESTADISTICA INFERENCIAL,  
POBLACION Y MUESTRA.
- 1.4 ALCANCES DE LA ESTADISTICA, METODO CIENTIFICO.
- 1.5 ESTADISTICA Y METODO CIENTIFICO.
- 1.6 MEDICION DE FENOMENOS, ESCALAS DE MEDICION.
- 1.7 EJERCICIOS DE REPASO.
- 1.8 FENOMENOS ALEATORIOS Y FENOMENOS DETERMINISTAS
- 1.9 VARIABLES, VARIABLES CONTINUAS, VARIABLES DISCRETAS  
O DISCONTINUAS, VARIABLES DEPENDIENTES Y VARIABLES  
INDEPENDIENTES.

1.1 INTRODUCCION: La asignatura de estadística está ubicada en el quinto y sexto semestre del plan de estudios del Colegio de Ciencias y Humanidades de la Unidad Académica del Ciclo de Bachillerato de la UNAM.

La estadística es una de las herramientas más ampliamente utilizadas en la investigación científica. Se emplea en instituciones gubernamentales y educativas, en los negocios y en la industria, y en otras organizaciones. El empleo juicioso de las técnicas estadísticas permite obtener conclusiones útiles a partir de un conjunto de datos numéricos.

Los objetivos de este trabajo son:

- 1.- Enseñar al estudiante a organizar y resumir datos.
- 2.- Enseñarle la manera de tomar decisiones cuando tiene una gran cantidad de datos, examinando sólo una pequeña parte de ellos.

Los conceptos y métodos necesarios para lograr el primer objetivo se presentan en la estadística descriptiva, y el segundo objetivo se alcanza a través del estudio de la estadística inferencial.

En este trabajo, se estudia la estadística descriptiva en los primeros capítulos y la estadística inferencial en los últimos capítulos.

## MATEMATICAS QUE SE NECESITAN

A través de la experiencia de los profesores que han impartido la materia de estadística durante años, se sabe que muchos estudiantes abordan los cursos de estadística con bastante desgano, y es típico de esos estudiantes sus limitadas bases matemáticas. Han hecho cursos previos de matemáticas con experiencias desalentadoras y escaso éxito, y llegan al curso de estadística convencidos de que tendrán otra mala experiencia.

El desarrollo de éste trabajo no es para un curso de matemáticas, claro que debido a la naturaleza misma de la estadística habrá que hacer algunas operaciones aritméticas para las que el nivel de matemáticas que se necesita es mínimo. Se podría decir que con conocimientos de aritmética y de álgebra elemental es suficiente para la comprensión de los capítulos que se presentan en este trabajo.

Quiero señalar que este trabajo está diseñado para hacer uso de las calculadoras electrónicas, en especial la calculadora CASIO fx-5000f que es la que presento a través de los capítulos, debido a su precio accesible, al alcance de los estudiantes, a sus 128 fórmulas incorporadas, a las 12 memorias de cálculo, y a la

posibilidad de agregar algunas fórmulas que se necesiten y no se encuentren en la memoria de la calculadora.

Las funciones estadísticas de la calculadora simplifican una amplia variedad de operaciones estadísticas comúnmente usadas, con un considerable ahorro de tiempo, de esfuerzo y de reducción de errores de cálculo. Su empleo permite abreviar tiempo y por lo mismo avanzar más rápido en el aprendizaje de la materia, ya que no será necesario calcular tablas laboriosas, ni raíces cuadradas, etc.. Así el esfuerzo dedicado a los cálculos suplidos con la calculadora, se pueden dedicar a la atención e interpretación de problemas planteados en cada capítulo.

A los profesores que imparten la materia les recomiendo no obligar a los estudiantes a memorizar fórmulas. Las evaluaciones deberán efectuarse con un formulario y permitiéndoles el uso de la calculadora tal y como lo recomiendan en el proyecto para la revisión del plan y los programas de estudio del área de matemáticas del C.C.H.

La calculadora es empleada en este trabajo para calcular los siguientes conceptos de estadística:

- La media aritmética, la mediana y la moda.

- Los cuartiles, los deciles y los percentiles.
- La desviación estándar para una población y para una muestra.
- Coeficiente de sesgo.
- Coeficiente de curtosis percentilico.
- Coeficiente de correlación y la recta de regresión.
- Combinaciones de n elementos.
- La distribución binomial.
- La distribución normal.

La bibliografía consultada para éste trabajo y que la presento al final del mismo, está dividida de acuerdo a su importancia, en bibliografía básica y de consulta:

-La bibliografía básica es la compuesta por obras que me sirvieron de apoyo para el desarrollo del cuerpo general del trabajo y que estuve consultando continuamente.

-La bibliografía de consulta es la conformada por los libros que considero como auxiliares para complementar el estudio por parte de los alumnos de los diversos temas tratados al estudiar éste libro.

## 1.2 BOSQUEJO HISTORICO

Desde hace muchos años la Estadística ha sido utilizada como instrumento de investigación, quizá sin saber que se hacía uso de ella, de una manera natural se les fue dando forma a los métodos Estadísticos.

Es imposible imaginar una cultura u organización social sin datos históricos, normas que la rigen, cultura, religión, etc. estos son los hechos que constituyen las manifestaciones esenciales de su existencia, y no es posible también dejar de pensar en la serie de datos que deben ser recopilados para ir modificando y perfeccionando su manera de vida.

Muchos años antes de Cristo, ya se tienen noticias de compilaciones Estadísticas efectuadas en Cerdeña, Egipto, China, Persia, Grecia y Roma. En el Pentateuco de Moisés se habla ya de un censo y Confucio narra como el Rey Yao hizo una Estadística Agrícola e Industrial del País. En general sus fines eran tributarios y con el objeto de determinar los recursos existentes para la guerra.

A medida que los grupos y los conjuntos sociales se perfeccionaban, hasta dar origen al Estado como entidad jurídica, política y administrativa, las enumeraciones,

las recopilaciones de datos estadísticos se hicieron no sólo más frecuentes, sino también más exactas y completas. Por lo mismo, se mejoraban los métodos a seguir para su obtención, los análisis que de los mismos se hacían, arrojaban resultados más precisos y variados, y se multiplicaban sus usos. Decaen con la caída del Imperio Romano para florecer más tarde con Carlomagno, con Guillermo el Conquistador y en España en el siglo XIV. En Inglaterra, en los antiguos registros parroquiales se registraban datos relativos a nacimientos, defunciones y matrimonios en las llamadas "Cuentas de Mortalidad", datos que la Iglesia introduce con carácter obligatorio después del Concilio de Trento.

Al aumentar las aplicaciones de la estadística, aumenta su importancia, y a fines del siglo XVI y principios del XVII, tiene su origen la estadística científica, siendo Sartorio, profesor de Padua quien introduce los métodos cuantitativos de las ciencias naturales en el organismo humano. En este siglo se le considera como disciplina autónoma en la descripción de cosas del Estado y, se consolida como ciencia universitaria en el siglo XVIII.

Se descubren las leyes demográficas y se hacen relaciones entre las defunciones y otros factores y se inicia el cálculo de probabilidades, fundándose la estadística moderna con base en los otros trabajos de

Adolfo Quetelet y A. Agustín Cournot, estableciéndose tres divisiones: estadística metodológica, social o demográfica y descriptiva.

### 1.3 ESTADISTICA DESCRIPTIVA, ESTADISTICA INFERENCIAL, POBLACION Y MUESTRA.

Los métodos estadísticos pueden ser aplicados por igual, tanto en las ciencias sociales como en las ciencias físicas.

Un problema cualquiera involucra una investigación estadística a realizar.

Con el siguiente ejemplo nos daremos cuenta de ello: Supongamos que nos interesa conocer el tipo de películas que más agrada a las personas que viven en el Sur del D.F.

Lo primero que debemos determinar es la "población" con la que vamos a trabajar, esto es, debemos dejar bien claro qué colonias abarca el Sur del D.F.

Después de definir la población, se tiene que decidir sobre la forma en que se entrevistará a cada individuo,

en grupos, por encuestas, por preguntas directas, quiénes harán este trabajo, etc.

Muy frecuentemente resulta imposible obtener los datos que se requieren de toda una población, por lo cual es necesario seleccionar sólo algunos elementos de la población; esto es, se debe obtener una muestra representativa de la población.

Estudiando la muestra obtenemos resultados que podemos inferir para toda la población, esto quiere decir que si sabemos la preferencia de películas de los individuos que participaron en la muestra, se puede suponer que esta preferencia también es de la población en general.

Pero para estudiar la muestra necesitamos recopilar datos haciendo encuestas adecuadas, ordenar la información, presentarla gráficamente y obtener medidas representativas para poder tener resultados óptimos y éstos proyectarlos en términos probabilísticos a la población.

En la explicación de las etapas a seguir en este ejemplo tan sencillo hemos desarrollado las etapas que se siguen en el Método Estadístico, estas las definiremos a continuación de manera más formal.

## ESTADISTICA DESCRIPTIVA

Cuando los datos sobre un hecho se han recopilado, ordenado, presentado en gráficas y se han calculado numéricamente sus características, se dice que se han cubierto las etapas que constituyen el propósito de la Estadística Descriptiva, de la cual damos su definición.

**ESTADISTICA DESCRIPTIVA:** Es la parte de la estadística que recopila, ordena y analiza datos sin sacar conclusiones.

## ESTADISTICA INFERENCIAL

El propósito de la Estadística Inferencial es obtener conclusiones para una población a partir del estudio de una muestra representativa de ella, por lo que:

**LA ESTADISTICA INFERENCIAL** es la parte de la estadística que obtiene conclusiones de la descriptiva, y trata las condiciones bajo las cuales esos resultados son válidos.

## ESTADISTICA

La Estadística Descriptiva y la Estadística Inferencial constituyen una sola disciplina: la Estadística.

**POBLACIÓN:** Es un conjunto de objetos, individuos con propiedades y que poseen características comunes observables de las cuales se desea información, tal como las alturas y pesos de los estudiantes de un colegio, el número de televisores defectuosos y no defectuosos producidos por una fábrica en un día determinado.

Una población puede ser finita o infinita dependiendo de que se tenga un número limitado o ilimitado de datos. Por ejemplo todos los televisores producidos por una fábrica en un día determinado es una población finita, mientras que la población formada por todos los posibles sucesos (águilas y soles) en tiradas sucesivas de una moneda es infinita.

**MUESTRA:** Es un subconjunto de una población. Una muestra se considera representativa de una población si cada uno de los datos que la integran son escogidos independientemente unos de los otros para que todas las unidades de la población tengan la misma posibilidad de ser incluidas en ella.

#### 1.4 ALCANCES DE LA ESTADISTICA, METODO CIENTIFICO.

El fin de la ciencia es lograr la descripción y explicación de los fenómenos naturales, para ello se encarga de buscar las leyes que la rigen.

La mayoría de los científicos siguen una metodología común llamada Método Científico para buscar dichas leyes.

El Método Científico cambia de acuerdo a la ciencia que se aplica, sin embargo el proceso de investigación en cualquier ciencia se resume en cuatro etapas fundamentales que ocurren sucesivamente en un proceso en espiral.

##### PRIMERA ETAPA

Se coleccionan hechos por observación o experimentación. La observación consiste en el estudio de los fenómenos tal como se presentan en la naturaleza. Mientras que la experimentación realizada generalmente en un laboratorio, es el estudio de los fenómenos bajo condiciones de control determinadas por el científico.

## SEGUNDA ETAPA

Se formulan las hipótesis o teorías que expliquen los hechos en términos de relaciones de causa y efecto.

En esta etapa se utiliza principalmente el Método Inductivo, para formular una hipótesis o teoría que sea aplicable a eventos o fenómenos distintos de aquellos a partir de los cuales fue elaborada.

## TERCERA ETAPA

Se efectúan deducciones para determinar qué resultados se obtienen bajo determinadas condiciones.

Aquí se utiliza el método deductivo para obtener proposiciones comprobables mediante la observación y la experimentación.

## CUARTA ETAPA

Se verifican las deducciones para determinar qué resultados se obtienen bajo determinadas condiciones.

Si en la cuarta etapa no hay contradicción entre las deducciones de la hipótesis y los nuevos hechos, no se infiere la validez de la hipótesis, sino que se efectúan nuevas deducciones bajo otras condiciones, y de nuevo

éstas se verifican por observación o experimentación. Esto se hace sucesivamente hasta que se encuentren condiciones en las que la hipótesis requiera modificaciones. Si hay una hipótesis que no ha entrado en contradicción con una extensa gama de hechos observados en condiciones variadas, se le denomina Ley Natural. Sin embargo, aún así no hay certeza de que la ley natural sea válida estrictamente, puede ser que algunos hechos no observados aún refuten o modifiquen dicha ley. Por eso se afirma que en la Ciencia no hay ninguna verdad definitiva.

De otra manera, si en la cuarta etapa existe contradicción entre las deducciones de la hipótesis y los nuevos hechos, la hipótesis se rechaza. Los nuevos hechos se agregan a los de la primera etapa y con ellos el ciclo comienza de nuevo. Se puede decir que el método científico sirve para rechazar hipótesis, pero no para aceptarlas. Así, las leyes naturales son hipótesis que no han sido rechazadas en varios ciclos del método científico.

Las etapas primera y segunda no están diferenciadas claramente, ya que no es posible observar todo en relación a algún fenómeno.

El científico posee siempre algún conocimiento vago de tipo conceptual o intuitivo sobre la hipótesis y esto es lo que lo guía para saber qué observar.

Algunos ejemplos del método científico son: Las leyes de la herencia de Mendel, La teoría de la evolución de las especies de Darwin, etc.

#### 1.5 ESTADISTICA Y METODO CIENTIFICO

La estadística en su funcionamiento interno hace uso del método científico, de manera que para un buen uso de la estadística se requiere conocer el método científico.

Por otro lado cuando el método científico se aplica a fenómenos aleatorios (fenómenos en donde interviene el azar), es la estadística la que auxilia al método científico.

#### AYUDA QUE LE PRESTA LA ESTADISTICA AL METODO CIENTIFICO

##### PRIMERA ETAPA (Observación o experimentación)

La estadística le presta ayuda al método científico en:

- a) Cómo tomar el mejor tamaño de la muestra.
- b) Cómo elegir la muestra de manera que sea representativa (muestreo, diseño de experimentos).

- c) Cómo resumir aspectos principales de lo observado (Estadística descriptiva).
- d) Cómo determinar el grado de consistencia de las observaciones (rangos permisibles).

#### SEGUNDA ETAPA (Formulación de la hipótesis).

La estadística no ayuda, quizá sólo el resumen o construcción de modelos ayude a sugerir la hipótesis.

#### TERCERA ETAPA (Deducciones de la hipótesis)

No hay ayuda de la estadística. Salvo el caso en que las hipótesis sean probabilísticas, por ejemplo las leyes de Mendel.

#### CUARTA ETAPA (Verificación o contrastación)

La estadística permite verificar si hay concordancia o no entre lo esperado y lo observado. Para esto se postulan modelos probabilísticos que expresen la hipótesis o bien el caso contrario (hipótesis de nulidad de efectos, de diferencias) y se contrastan los hechos observados en la hipótesis estadística.

Ejemplos: La cantidad de deporte en horas que practican los alumnos influyen en las calificaciones, es la hipótesis científica. La hipótesis estadística (de

nulidad) es que no influye, cantidad de deporte y calificaciones son independientes.

El colesterol en la sangre aumenta con la edad. Se prueba la hipótesis de que no están asociados.

#### 1.6 MEDICION DE FENOMENOS, ESCALAS DE MEDICION.

En el método científico los hechos observados casi siempre son difíciles de comprender y evaluar en toda su complejidad.

Los hechos señalan una o varias propiedades de los fenómenos determinando las modalidades que presentan dichas propiedades. Para facilitar la tarea de comprensión y evaluación de los hechos observados, permitiendo además la comunicación de un hombre a otro, es necesaria cierta cuantificación o clasificación de las modalidades que presentan las propiedades relevantes de los fenómenos, aún cuando esto implique el riesgo de perder algunos aspectos de las manifestaciones de los fenómenos. Este es el proceso de medición, el cual presenta cinco alternativas que se reflejan en las denominadas escalas de medición, que son clasificadas

como escala nominal, escala ordinal, escala de intervalo, escala de razón y escala absoluta.

## ESCALAS DE MEDICION

### ESCALA NOMINAL

En este nivel de medición únicamente se le dan nombres a las diferentes modalidades de una propiedad que presenta el fenómeno bajo estudio. Esta escala se utiliza como medida de identificación y sus valores son tales que no se pueden ordenar de una manera única.

En el caso de que se empleen números para distinguir las clasificaciones o subclases, estas tienen un valor simbólico y no numérico, esto constituye sólo etiquetas que permitan diferenciar un conjunto de otro.

Ejemplo: Cuando clasificamos a los miembros de una comunidad en alfabetos y analfabetos, en económicamente activos e inactivos, a los miembros de partidos políticos en PRI, PAN, PRD, las carreras que cursa un conjunto de estudiantes, el estado civil de un grupo de personas, solteros, casados, divorciados, viudos y otros, etc.

### ESCALA ORDINAL

En esta escala de medición además de asignar nombres a las modalidades que presentan las propiedades del fenómeno, también, se determina en forma aproximada el grado de intensidad de la propiedad observada, lo cual se refleja en una relación de orden, por ejemplo cuando clasificamos a los miembros de una comunidad en estratos alto, medio y bajo, se puede establecer que alto es mayor que medio y medio mayor que bajo o a la inversa, bajo es menor que medio y medio es menor que alto, pero no se determina que distancia existe entre un miembro del estrato medio respecto a otro del estrato alto. Por lo cual en el nivel ordinal la distancia entre dos unidades no es conocida, únicamente se conoce su relación en cuanto a rango o jerarquía.

En el nivel ordinal cuando se emplean números que se asignan a las características permiten determinar el orden o la posición jerárquica en una escala, pero no tienen significado en lo referente a ¿cuánto? o a ¿cuántas veces? porque son cualidades no aditivas.

### ESCALA DE INTERVALO

La escala de intervalo tiene además de las propiedades de la escala ordinal, la propiedad de que la distancia entre dos valores es una magnitud conocida, la cual da a esta escala un mayor grado de perfección. En el nivel de

intervalo el punto cero (punto de origen) y la unidad de medición son arbitrarias.

El punto seleccionado como cero no es cero verdadero en el sentido de que no indica ausencia total de la cantidad que se está midiendo, y a diferencia de las escalas nominal y ordinal, los valores numéricos que se utilizan, sí tienen propiedades cuantitativas y admite todas las manipulaciones aritméticas.

Ejemplo: Las escalas empleadas en medición de la temperatura: grados centígrados, Fahrenheit y Kelvin, esta última empleada en la ciencia física.

Nos referiremos a las dos primeras escalas por ser las más comunes.

Temperaturas equivalentes en las dos escalas:

GRADOS CENTIGRADOS	0	10	20	30	40	50	80	100
GRADOS FAHRENHEIT	32	50	68	86	104	122	176	212

En la escala de grados centígrados la congelación del agua a nivel del mar constituye el punto cero, y la ebullición a este mismo nivel el punto cien. En la escala Fahrenheit estos puntos equivalen a 32 y 212 grados. En ambas escalas el punto cero es arbitrario; para la escala de grados centígrados pudo haberse seleccionado otro procedimiento para determinar el cero

y el valor cien, por ejemplo la congelación y ebullición a la altura de 5000 metros. Las unidades de medición - grados - están determinadas por el sistema empleado para fijar los valores de 0 y 100. Sin embargo ambas escalas contienen la misma cantidad y el mismo tipo de información.

En este mismo ejemplo podemos comprobar, mediante los valores numéricos, que la razón de las diferencias entre los intervalos son independientes de las cantidades empleadas y del punto cero, por ejemplo

$$\frac{40-30}{30-20} = 1 \qquad \frac{104-86}{86-68} = 1$$

La razón entre iguales diferencias de temperatura es la misma. Las dos escalas proporcionan la misma cantidad de información numérica, esto es así porque están relacionadas linealmente. Un valor en una escala se puede transformar en valor equivalente en otra mediante la relación afin:

$$F = \frac{9}{5} C + 32 \quad \text{O} \quad C = \frac{5}{9} (F - 32)$$

Por ejemplo 50° transformados a Fahrenheit:

$$F = \frac{9}{5}(50) + 32 = \frac{450}{5} + 32 = 90 + 32 = 122$$

212° grados Fahrenheit transformados a centígrados:

$$C = \frac{5}{9}(212-32) = \frac{5}{9}(180) = \frac{900}{9} = 100$$

En esta escala no es correcto afirmar por ejemplo que 100 es el doble de 50, aunque si es correcto decir que la diferencia entre 50 y 100 centígrados será igual a la diferencia entre 212 y 122 grados Fahrenheit.

Otras escalas de intervalo son las escalas de actitudes, las puntuaciones (IQ) de inteligencia, la escala para medir el tiempo en nuestro calendario, el horario etc.

#### ESCALA DE RAZON

Se caracteriza porque tiene un punto cero verdadero, es decir absoluto, además de poseer las características de la escala de intervalo, esto es las distancias entre los valores de la escala son conocidas en el sentido cuantitativo y su "razón" es independiente de las unidades que se utilizan.

En las escalas empleadas en la medición de longitudes, pesos y masas, el punto cero es absoluto, real; por ejemplo en las unidades metro, yarda, kilogramo, libra, metros cúbicos, el valor cero indica ausencia de medición, esto es inexistencia. En la escala de intervalo no podemos afirmar, por ejemplo, que 0 grados centígrados indique ausencia de temperatura.

En la escala de razón se da una razón igual entre dos valores de escalas diferentes, por ejemplo kilogramos y libras; la razón entre 40 y 25 kilogramos es igual a la razón entre 40 y 25 libras. En estas escalas únicamente se puede cambiar la unidad de medida, pero el origen es fijo.

#### ESCALA ABSOLUTA

Es la que tiene todas las características de las anteriores, pero los valores que se les asigna a las modalidades observadas son únicamente números enteros positivos y se utiliza en conteos por ejemplo, cuando queremos contar el número de alumnos que miden más de 1.70 metros de estatura en un colegio determinado etc.

Por lo expuesto anteriormente referente a las escalas de medición se dice que:

En todo estudio científico es necesario recurrir a alguna escala de medición. Si el conocimiento científico

previo del fenómeno es extenso, lo más probable es que las escalas usadas sean de intervalo o de razón. Si se esta en las etapas iniciales del estudio del fenómeno, es más frecuente el uso de escalas nominales u ordinales. La escala absoluta se puede usar en ambos casos dependiendo de la situación.

#### 1.7 EJERCICIOS DE REPASO

En los siguientes incisos indique el nivel de medición que corresponda:

1.- En un grupo de estadística del CCH sur se les preguntó a los alumnos que carrera de la Universidad habían elegido para su ingreso a la facultad y su distribución fue la siguiente:

Facultad de Ciencias Políticas

Facultad de Economía

Facultad de Filosofía y Letras

Facultad de leyes

2.- El VIII Censo de Población clasifica a las personas mayores de 6 años de acuerdo al número de años de estudio terminados y aprobados.

3.- En un periódico aparece la lista de los mejores seis jugadores del año en ajedrez en orden decreciente respecto a su capacidad.

4.- El censo de 1993 señala el lugar de nacimiento de los habitantes censados.

5.- En una escuela se pregunta el ingreso mensual de los alumnos.

6.- Jerarquice los niveles de medición de acuerdo al orden decreciente de perfección.

7.- Dos empleados tienen sueldo de \$ 7,000 y \$ 3,500 pesos respectivamente. ¿El ingreso del primero es el doble que el del segundo?. Indica con qué escala se miden los valores monetarios.

## 1.8 FENOMENOS ALEATORIOS Y FENOMENOS DETERMINISTAS

En general, las matemáticas juegan un papel muy importante, tanto en el proceso de medición como en la especificación de las hipótesis o leyes que gobiernan a los fenómenos. Esto se logra por la introducción de modelos matemáticos para representar las diferentes propiedades de los fenómenos que estudiamos. Los modelos matemáticos constituyen abstracciones de la realidad, ya que sólo representan las propiedades relevantes para la hipótesis planteada, no tomando en cuenta el resto de las propiedades presentes en los fenómenos. Además, estos modelos representan en forma aproximada a la realidad. La aproximación es satisfactoria desde un punto de vista práctico para ciertos fenómenos denominados determinísticos, como las leyes de atracción de Newton, leyes del péndulo, caída libre de un cuerpo etc.; sin embargo, existen fenómenos donde no es posible establecer modelos matemáticos con una representación satisfactoria de la realidad, dichos fenómenos donde interviene el azar se denominan aleatorios. Algunos modelos que incluyen aspectos aleatorios pueden ser útiles en estos casos.

Los fenómenos aleatorios se caracterizan porque no se puede determinar con suficiente precisión su estado final, sino que se tiene cierto grado de incertidumbre. Los fenómenos aleatorios se presentan cuando no se puede conocer el estado inicial del fenómeno con suficiente precisión para determinar su estado final basándose en leyes naturales conocidas. Esto puede deberse a:

- Una variación muy pequeña en el estado inicial produce un cambio muy grande en el estado final. Es el principio de los juegos de azar.
  
- El estado inicial es muy complicado y es impráctico medirlo, por ejemplo situación microscópica de un gas.
  
- Las leyes son conocidas pero complicadas, por lo que en la práctica no se aplican. Por ejemplo, predecir el resultado del lanzamiento de un dado, es posible teóricamente, pero impráctico.
  
- Las leyes naturales involucradas no se conocen lo suficientemente para hacer la predicción. Este es el caso de la mayoría de los fenómenos biológicos y sociales, por ejemplo, ¿cuál será la carrera que escoja

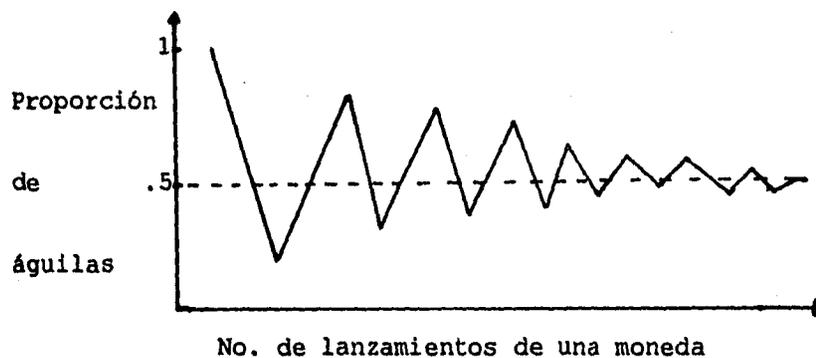
un niño recién nacido?, o ¿ cuál será el tamaño y peso del huevo que pondrá determinada gallina?. Los procesos biológicos y sociales involucrados no son conocidos con exactitud, inclusive no se conoce cuáles son los factores relevantes en éstos fenómenos.

Estrictamente hablando, podemos decir que todo fenómeno es aleatorio; pero si en la práctica se está satisfecho con el grado de aproximación que producen las leyes conocidas y su aplicabilidad es factible a ciertos fenómenos, éstos no se consideran aleatorios. Sin embargo, existe un gran número de fenómenos que desde un punto de vista práctico son aleatorios por una o más causas expuestas. Por ello es importante contar con alguna ayuda metodológica para el estudio de dichos fenómenos. Esta ayuda metodológica es la estadística, que basándose en ciertos postulados probabilísticos, permite hacer descripciones, optimizaciones y predicciones en los fenómenos aleatorios.

Por último sucede que cuando los fenómenos aleatorios son estudiados en gran número de veces en condiciones similares, presentan cierta regularidad llamada estadística. Esta regularidad estadística consiste en que las frecuencias relativas, porcentajes ó proporciones con las cuales ocurren las diferentes

modalidades de una característica del fenómeno aleatorio son muy estables, si desde luego el número de veces que se estudia es grande.

Por ejemplo al estudiar la proporción de "águilas", al ir incrementando el número de veces que se lance una moneda tiende a ser  $\frac{1}{2} = 0.5$ .



Esta regularidad estadística representada por un modelo matemático es la probabilidad que se originó con los juegos del azar, la cual estudiaremos en otro capítulo.

1.9 VARIABLES, VARIABLES CONTINUAS, VARIABLES DISCRETAS  
O DISCONTINUAS, VARIABLES DEPENDIENTES Y VARIABLES  
INDEPENDIENTES.

Al estudiar un fenómeno de la naturaleza es necesario medir las características observables. Esas características que toman diferentes valores son conocidas como variables.

VARIABLE.- Una variable es un símbolo, tal como  $X$ ,  $Y$ ,  $Z$  que puede tomar cualquier valor en un conjunto determinado, por ejemplo los pesos de un grupo de personas, las estaturas de un grupo de alumnos, los salarios que reciben los obreros de una fábrica, etc.

VARIABLE DISCRETA O DISCONTINUA

Una variable se llama discreta, si los valores que toma se pueden contar y sólo pueden expresarse con números enteros bien definidos sin valores intermedios posibles, por ejemplo, el número de alumnos en una clase de estadística del CCH, es una variable discreta porque sólo toma valores como:

1, 2, 3, 4, ...,  $n$  pero no valores intermedios como 2.432 o 18.375. Otros ejemplos: El número de pétalos de una flor, los censos anuales del colegio de profesores, el número de salones en el CCH Sur, el número de billetes

de \$1000 circulando a la vez en la república mexicana, el número de libros de la biblioteca de un colegio, etc.

#### VARIABLE CONTINUA

Una variable es continua si teóricamente puede tomar cualquier valor dentro de un intervalo dado.

Una variable continua y tiene las siguientes características:

- a) El número de valores posibles es infinito.
- b) No tiene los saltos o interrupciones de la variable discreta.
- c) Al hacer una medición se puede hacer con alto grado de precisión, de manera que un metro se puede dividir en centímetros, milímetros, y con instrumentos especializados se pueden tomar mediciones más precisas.
- d) Muy frecuentemente se tratan como si fueran discretas con el fin de facilitar las operaciones. Como en el caso de preguntarle su estatura a una persona, generalmente responde registrando hasta el centímetro más próximo.

Algunos ejemplos de variables continuas son: las temperaturas registradas por un observatorio, el tiempo en horas que dura un examen cualquiera, los salarios de un grupo de profesores, el diámetro de un esfera, la velocidad de un automóvil, etc.

#### VARIABLE DEPENDIENTE E INDEPENDIENTE

Cuando dos variables están relacionadas de tal forma que el valor de la primera queda determinado si se le da un valor a la segunda entonces se dice que la primera es dependiente y la segunda independiente.

A la variable independiente se le pueden asignar valores a voluntad o valores arbitrarios dentro de los límites que el problema lo requiera, mientras que el valor de la variable dependiente queda fijo cuando se asigne un valor a la variable independiente.

Ejemplos: La capacidad de un salón de clases depende del número de sillas; la variable dependiente es la capacidad del salón de clases y la variable independiente es el número de sillas, la producción de toneladas de trigo depende del año (tiempo que se considere), la variable independiente es el tiempo y la variable dependiente es la producción de trigo, la distancia que recorre un automóvil depende del tiempo, etc.

## **CAPITULO 2**

### **REPRESENTACION GRAFICA DE DATOS**

**2.1 UTILIDAD DE LAS GRAFICAS**

**2.2 GRAFICA LINEAL**

**2.3 GRAFICA DE BARRAS**

**2.4 GRAFICA CIRCULAR**

**2.5 GRAFICA PICTORICA**

**2.6 EJERCICIOS DE REPASO**

## 2.1 UTILIDAD DE LAS GRAFICAS

La presentación de los datos estadísticos puede hacerse en tres formas; textual, tabular y gráfica; esto es, puede presentarse en forma escrita, en tablas o en cuadros estadísticos y mediante gráficas.

Las gráficas permiten mostrar, explicar, interpretar y analizar de manera sencilla, clara y efectiva, los datos estadísticos mediante formas geométricas tales como líneas, áreas, volúmenes; además permite la comparación de magnitudes, tendencias y relaciones entre los valores.

Las gráficas son especialmente útiles como medios de divulgación del análisis estadístico ya que las relaciones visuales se captan con facilidad y resulta sencillo recordarlos.

Existen varios tipos de gráficas, siendo las más usuales:

La gráfica lineal.

La gráfica de barras.

La gráfica circular.

La gráfica pictórica.

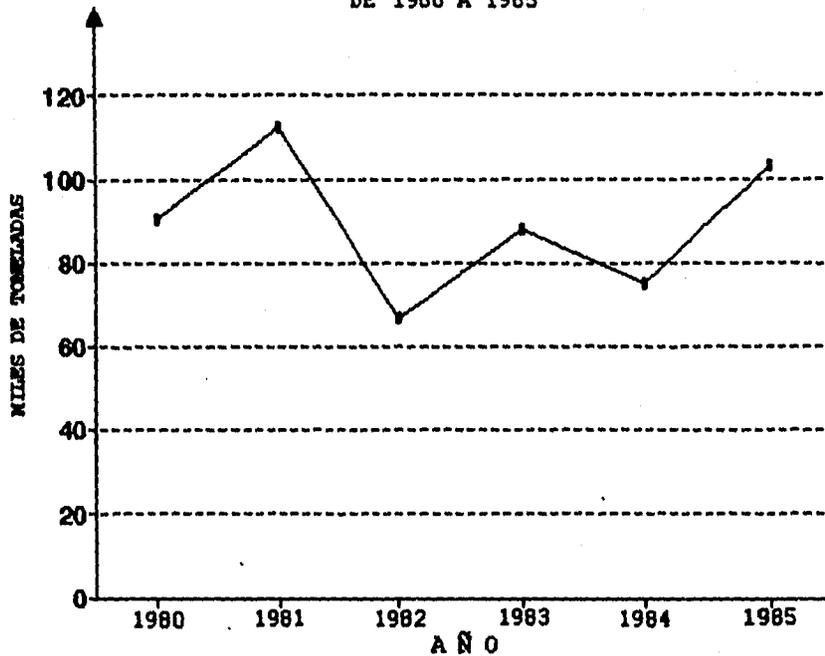
## 2.2 GRAFICA LINEAL

La gráfica lineal es la representación gráfica de la relación numérica que guardan dos conjuntos de datos sobre un plano definido por dos ejes perpendiculares mediante puntos unidos por líneas rectas.

Ejemplo; la siguiente tabla muestra la producción de maíz en México de 1980 a 1985. Representar los datos con una gráfica lineal.

AÑO	PRODUCCION (MILES DE TONELADAS)
1980	90.5
1981	112.6
1982	67.2
1983	88.1
1984	75
1985	103.4

PRODUCCION DE MAIZ EN MEXICO  
DE 1980 A 1985



En este ejemplo la variable dependiente es la producción en miles de toneladas y la variable independiente es el tiempo (años).

Para elaborar este tipo de gráficas se trazan puntos en el sistema de coordenadas tomando los datos de la tabla, por ejemplo (1982, 67.2). Después se unen los puntos por medio de líneas rectas, puesto que no tenemos información de la variación de la población en los años intermedios. Por esta razón este tipo de representación se llama gráfica lineal.

También hay que anotar que el cero tiene generalmente su representación en el eje vertical, pero no en el eje horizontal

### 2.3 GRAFICA DE BARRAS

La gráfica de barras como la gráfica lineal, es la representación de la relación numérica que hay entre dos conjuntos de datos y se construye por medio de rectángulos, donde la altura de cada rectángulo representa el número de miembros de una clase, y el espaciamiento entre ellos, la diferencia entre los datos del otro conjunto.

Las características para elaborar estas gráficas son:

a) La anchura de los rectángulos debe ser igual para cada uno de los rectángulos y se escoge de manera arbitraria.

B) Los rectángulos deben dibujarse separados y el espacio que hay entre ellos debe ser el mismo y se elige arbitrariamente.

C) Los valores de la variable representada en el eje horizontal (variable independiente) deben estar en el centro de las bases de los rectángulos.

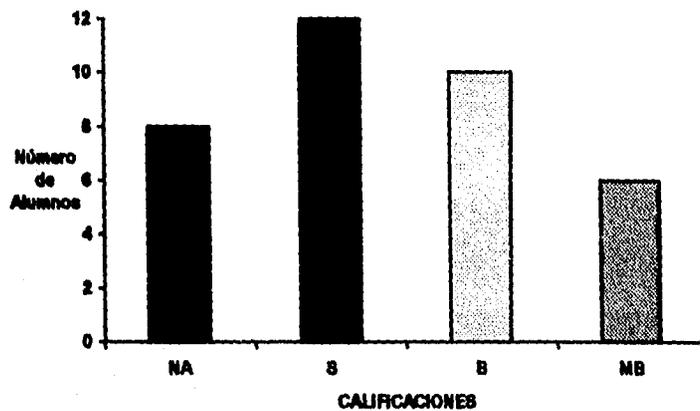
D) Los rectángulos pueden dibujarse en forma horizontal o vertical.

Ejemplo 1.

1.- La siguiente tabla muestra las calificaciones finales del curso de estadística I, del grupo 1516 del C.C.H Sur. Representar los datos utilizando una gráfica de barras.

CALIFICACION	NÚMERO DE ALUMNOS
NA	8
S	12
B	10
MB	6
TOTAL	36

**CALIFICACIONES FINALES DEL CURSO DE ESTADISTICA DEL GRUPO 1618 DEL CCH SUR**

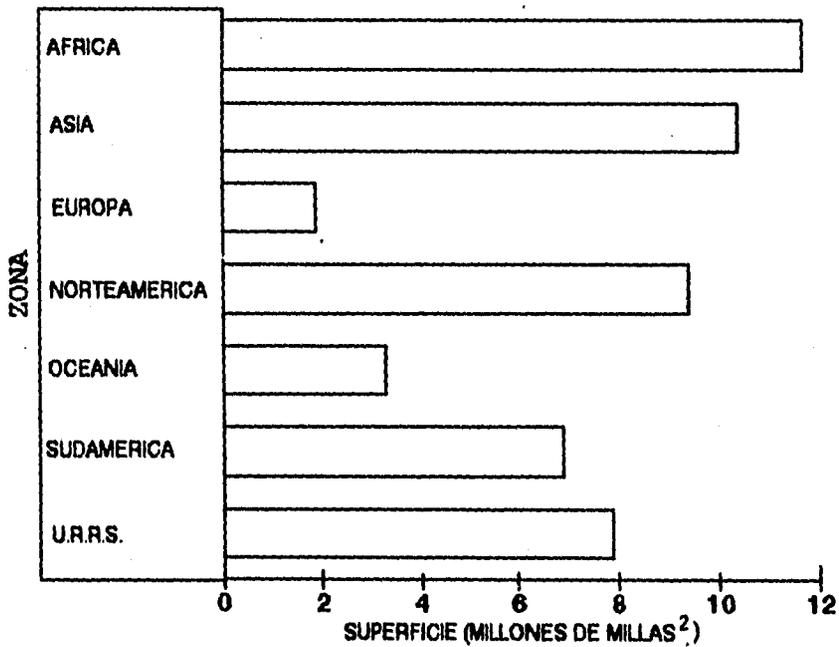


Ejemplo 2

La superficie de distintas zonas del mundo en millones de millas cuadradas aparece en la siguiente tabla.

Representar los datos mediante una gráfica de barras horizontal.

ZONA	SUPERFICIE (Millones de millas cuadradas)
Africa	11.7
Asia	10.4
Europa	1.9
Norteamérica	9.4
Oceania	3.3
Sudamérica	6.9
U.R.R.S	7.9



## 2.4 GRAFICA CIRCULAR

La gráfica circular llamada también gráfica de pastel o gráfica de sectores se emplea para representar la relación existente entre dos conjuntos de datos. Este tipo de gráficas se utiliza de preferencia cuando la variable dependiente o independiente, toma menos de ocho valores.

Para la construcción de estas gráficas, el círculo se divide en pedazos, llamados sectores, con áreas que son proporcionales a las cantidades que han de representarse. Por ejemplo, supóngase que una familia proyecta emplear sus ingresos anuales de la siguiente manera:

DESTINO DEL INGRESO	CANTIDAD (Miles de pesos)
Alimento	\$ 14
Vivienda	10
Vestido	6
Ahorro	2
Otros	8
TOTAL \$ 40	

Para construir la gráfica circular se parte del hecho de que el ingreso anual total es de \$40 mil pesos. Este valor lo hacemos corresponder al número de grados de la circunferencia. \$40 mil - 360°.

Así tomamos el primer dato y calculamos el número de grados que le corresponden mediante una regla de tres simple.

#### ALIMENTOS

\$40 MIL - 360°

14 MIL - X

$$40 x = 14(360^\circ)$$

$$X = \frac{14(360)}{40} = 126^\circ$$

Para los otros datos se sigue el mismo procedimiento.

#### VIVIENDA

\$40 MIL - 360°

10 MIL - X

$$40 X = 10(360^\circ)$$

$$X = \frac{10(360^\circ)}{40} = 90^\circ$$

VESTIDO

\$40 MIL - 360°

6 MIL - X

$$40 X = 6(360^\circ)$$

$$X = \frac{6(360^\circ)}{40} = 54$$

AHORRO

\$40 MIL - 360°

2 MIL - X

$$40 X = 2(360^\circ)$$

$$X = \frac{2(360^\circ)}{40} = 18$$

OTROS

\$40 MIL - 360°

8 MIL - X

$$40 X = 8(360^\circ)$$

$$X = \frac{8(360^\circ)}{40} = 72$$

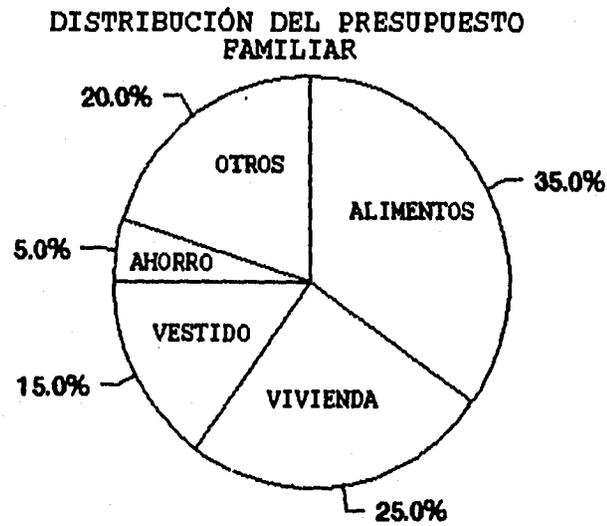
Hay que hacer notar que la suma de los grados correspondientes a los datos debe ser 360°.

Con este procedimiento, haremos la transformación de nuestros valores a porcentajes. Y con los datos obtenidos procederemos a construir la siguiente tabla

PRESUPUESTO FAMILIAR EN MILES DE PESOS  
CON GRADOS Y PORCENTAJES

DESTINO DEL INGRESO	CANTIDAD (MILES DE PESOS)	NÚMERO DE GRADOS	PORCENTAJES
ALIMENTOS	\$ 14	126°	35 %
VIVIENDA	10	90°	25 %
VESTIDO	6	54°	15 %
AHORRO	2	18°	5 %
OTROS	8	72°	20 %
TOTAL	\$ 40 MIL	360°	100 %

Por último haciendo uso del transportador trazamos los sectores resultantes y de esta manera se obtiene la gráfica circular.



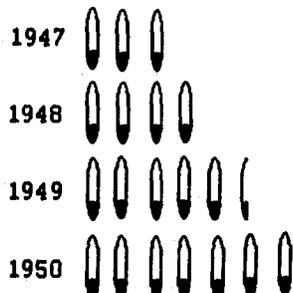
## 2.5 GRAFICA PICTORICA

Las gráficas pictóricas, son muy empleadas por ser atractivas y de fácil comprensión para el público y consisten en ilustraciones, tales como caricaturas, bocetos, croquis, dibujos, etc. Aunque son relativamente fáciles de interpretar, que es su finalidad, por lo general su construcción requiere de especialistas en dibujo.

Ejemplo. La siguiente tabla nos muestra la producción anual de una fábrica de tubos electrónicos, durante cuatro años:

AÑO	NUMERO DE TUBOS
1947	3000
1948	4000
1949	5500
1950	7000

La gráfica pictórica es la siguiente:



PRODUCCIÓN DE TUBOS DE LA FÁBRICA  
(CADA DIBUJO REPRESENTA 1000 TUBOS)

2.6 EJERCICIOS DE REPASO:

1.- El tiempo  $T$  (en segundos) requeridos para una vibración completa de un péndulo simple de longitud  $L$  (en centímetros) es dado por las siguientes observaciones obtenidas en un laboratorio de física.

a) Representar con una gráfica lineal, con  $T$  como función de  $L$ .

b) De la gráfica, estimar  $T$  para el péndulo cuya longitud sea 40 centímetros.

L	10.1	16.2	22.2	33.8	42	53.4	66.7	74.5	86.6	100
T	.64	.81	.95	1.17	1.3	1.47	1.65	1.74	1.87	2.01

2.- La siguiente tabla muestra las velocidades orbitales de los planetas de nuestro sistema solar. Representar los datos utilizando una gráfica de barras.

PLANETA	VELOCIDAD (MILLAS/SEG.)
MERCURIO	29.7
VENUS	21.8
TIERRA	18.5
MARTE	15.0
JUPITER	8.1
SATURNO	6.0
URANO	4.2
NEPTUNO	3.4
PLUTON	3.0

3.- La siguiente tabla muestra la superficie (en millones de millas cuadradas) de los océanos del mundo.

Representar los datos utilizando:

- a) Una gráfica de barras.
- B) Una gráfica circular.

OCEANO	SUPERFICIE (Millones de millas <sup>2</sup> )
PACIFICO	70.8
ATLANTICO	41.2
INDICO	28.5
ANTARTICO	7.6
ARTICO	4.8

4.- La tabla siguiente muestra el estado civil de hombres y mujeres (mayores de 15 años) en México en el año 1960. Representar los datos utilizando unas gráficas circulares del mismo diámetro...

ESTADO CIVIL	HOMBRES (% DEL TOTAL)	MUJERES (% DEL TOTAL)
SOLTERO	24.5	18.8
CASADO	69.8	66.0
VIUDO	3.9	12.8
DIVORCIADO	1.8	2.4

## **CAPITULO 3**

### **DISTRIBUCION DE FRECUENCIAS E HISTOGRAMAS**

- 3.1 ORGANIZACION, AGRUPACION Y REPRESENTACION DE DATOS.
- 3.2 TABLA DE FRECUENCIAS.
- 3.3 INTERVALO DE CLASE Y LIMITES DE CLASE.
- 3.4 LIMITES REALES DE CLASE.
- 3.5 TAMAÑO O ANCHURA DE UN INTERVALO DE CLASE.
- 3.6 MARCA DE CLASE.
- 3.7 REPRESENTACION GRAFICA DE DATOS.
- 3.8 HISTOGRAMA, POLIGONO DE FRECUENCIAS Y LA OJIVA.
- 3.9 DISTRIBUCIONES DE FRECUENCIA RELATIVA.
- 3.10 DISTRIBUCION DE FRECUENCIAS ACUMULADAS, TABLAS DE  
FRECUENCIA ACUMULADAS Y OJIVAS.  
POLIGONOS DE FRECUENCIAS ACUMULADAS, OJIVAS  
"MENOR QUE" Y "MAYOR QUE".
- 3.11 DISTRIBUCIONES DE FRECUENCIAS RELATIVAS ACUMULADAS,  
OJIVAS PORCENTUALES.
- 3.12 EJERCICIOS DE REPASO.

### 3.1 ORGANIZACION, AGRUPACION Y REPRESENTACION DE DATOS.

Supóngase que un profesor, del Colegio de Ciencias y Humanidades desea realizar un estudio para determinar las características físicas de los estudiantes que asisten a sus cursos de Estadística. Una característica de interés es el peso y decide obtener los datos en uno de sus grupos; por medio de la observación directa interroga a cada uno de sus alumnos y obtiene los siguientes datos:

Peso en kilogramos de un grupo de estudiantes de estadística del C.C.H.

66	67	59	62	64
68	61	65	68	67
60	58	68	69	65
70	71	73	71	72
74	77	75	74	63
69	63	68	68	71

Dado que los datos que se obtuvieron en el orden en que fueron compilados son difíciles de interpretar, es necesario ordenarlos en forma creciente o decreciente de acuerdo a su magnitud, siendo la forma creciente la más adecuada.

### DATOS ORDENADOS

58	63	67	69	72
59	64	68	69	73
60	65	68	70	74
61	65	68	71	74
62	66	68	71	75
63	67	68	71	77

También se observa que algunos datos aparecen más de una vez, por lo que al número de veces que se repite un dato le llamaremos frecuencia, representada por  $F$ , y al total de los datos los determinaremos por  $N$ , en este ejemplo  $N=30$ .

### 3.2 TABLA DE FRECUENCIAS

Aunque un conjunto de datos puede hacerse más comprensible y adquirir mayor significado por medio de un ordenamiento; puede lograrse una mayor síntesis agrupando los datos.

Para agrupar a un conjunto de observaciones, se selecciona un conjunto de intervalos contiguos que no se traslapen de modo que cada valor en el conjunto de observaciones pueda colocarse en uno, y sólo en uno de

los intervalos. Estos intervalos se conocen en general como intervalos de clase. Los intervalos de clase estarán referidos a segmentos de la recta numérica de manera tal que los extremos de un intervalo serán números reales, por esta razón a estos extremos se les denomina límites de clase.

Agrupar datos en intervalos de clase genera una simplificación de los datos ordenados denominada distribución de frecuencias, la que estudiaremos más adelante.

Uno de los primeros puntos a considerar, cuando se van a agrupar ciertos datos, es cuántos intervalos van a incluirse. No es conveniente incluir pocos intervalos debido a que hay pérdida de información. Por otra parte si se utilizan demasiados intervalos, no se logra el objetivo de la síntesis. La mejor guía en relación con lo anterior, así como para otras decisiones que deberán tomarse al agrupar los datos, es el conocimiento que se tenga de ellos. Puede ser que se hayan determinado con anterioridad los intervalos de clase, como en el caso de las tabulaciones anuales, cuando se conservan los intervalos de clase de años anteriores con fines comparativos. Al respecto, debe señalarse que no existe una regla universal que deba seguirse, más bien depende, de la precisión que requiera un estudio determinado.

Varios autores con experiencia en estudios estadísticos han establecido algunas reglas empíricas para elegir un

número de intervalos apropiado en relación con el número de datos. A continuación se presenta una tabla con tal característica:

SUGERENCIAS PARA DETERMINAR EL NUMERO  
DE INTERVALOS

NÚMERO DE DATOS DE LA MUESTRA	NÚMERO DE INTERVALOS SUGERIDOS
25	5 - 6
26 - 50	7 - 10
MÁS DE 50	11 - 20

El número de intervalos de clase, especificados por esta tabla debe aumentarse o disminuirse según convenga en beneficio de un buen análisis.

Regresando al caso que estamos estudiando, recordemos que se tienen 30 datos correspondientes al peso en kilogramos de cada alumno. Para construir la distribución de frecuencias, primero se determina el número de intervalos que se considere adecuado. En nuestro ejemplo consideramos que se va a trabajar con 7 intervalos, tomando como base la sugerencia para determinar el número de intervalos expuesta anteriormente. Una vez que se tiene definido el número de intervalos, simbolizado por  $K$ , el tamaño del intervalo de clase (que representamos por la letra  $C$ ) se determina mediante la siguiente expresión:

$$\text{TAMAÑO DEL INTERVALO} = \frac{\text{RANGO}}{\text{NÚMERO DE INTERVALOS}} \quad \text{O SEA } C = \frac{R}{K}$$

Donde el rango o recorrido es la diferencia entre la observación mayor y la observación menor, por lo que:

$$C = \frac{77 - 58}{7} = \frac{19}{7} = 2.71$$

Resulta evidente que es más conveniente utilizar un tamaño  $C=3$  para el intervalo de clase, lo que asimismo, tendrá más significado para el estudiante. Ahora podemos construir los intervalos. Dado que el valor más pequeño en nuestro caso es 58 y el más grande 77, pueden empezarse los intervalos con 57 y terminarse con 77. Esto da los intervalos siguientes.

57 - 59

60 - 62

63 - 65

66 - 68

69 - 71

72 - 74

75 - 77

Como se observa hay 7 intervalos de clase que es lo que queríamos.

queríamos.

Ahora procederemos a construir la tabla de distribución de frecuencias.

Número de intervalo	Intervalo de clase	Datos que se agrupan en cada intervalo	Frecuencia F
1	57 - 59	58, 59	2
2	60 - 62	60, 61, 62	3
3	63 - 65	63, 63, 64, 65, 65	5
4	66 - 68	66, 67, 67, 68, 68, 68, 68, 68	8
5	69 - 71	69, 69, 70, 71, 71, 71	6
6	72 - 74	72, 73, 74, 74	4
7	75 - 77	75, 77	2
			Total N=30

Cuando ya se tiene práctica, el paso anterior se da por obvio y se procede a obtener directamente la forma final de la distribución de frecuencias que es la siguiente:

Tabla de frecuencias del peso en kilogramos de 30 estudiantes de un grupo de estadística del C.C.H.

Número de intervalos	Intervalo de clase	Frecuencia F
1	57 - 59	2
2	60 - 62	3
3	63 - 65	5
4	66 - 68	8
5	69 - 71	6
6	72 - 74	4
7	75 - 77	2
		N=30

### 3.3 INTERVALOS DE CLASE Y LÍMITES DE CLASE

Como se anotó anteriormente, los intervalos de clase

manera que los extremos de cada intervalo son números reales por lo que estos extremos se les denomina límites de clase. Los números 57 y 59 son los límites de clase, el número menor 57 es el límite inferior de la clase y el mayor 59 es el límite superior. Los términos clase e intervalo de clase se utilizan a menudo indefinidamente, aunque el intervalo de clase es realmente un símbolo para la clase.

#### 3.4 LIMITES REALES DE CLASE

Si los pesos de los alumnos se registran con aproximación de kilogramos, el intervalo de clase 57-59 teóricamente incluye todos los pesos desde 56.50 a 59.50 kilogramos. Estos números representados brevemente por los números exactos 56.5 y 59.5 se conocen como límites reales de clase o límites verdaderos de clase, el menor 56.5 es el límite real inferior y el mayor 59.5 es el límite real superior. Estos límites reales de clase se obtienen sumando el límite superior de un intervalo de clase con el límite inferior del intervalo de clase contiguo superior y dividiendo entre dos. Por ejemplo, para encontrar los límites reales de clase del tercer intervalo de nuestra tabla se procede de la siguiente manera:

El límite real inferior del 3er intervalo de clase es igual a la suma del límite superior del 2o intervalo de

clase más el límite inferior del 3er intervalo de clase entre dos.

$$\text{Lím. real inf. del 3er int. de clase} = \frac{62 + 63}{2} = \frac{125}{2} = 62.5$$

Y el límite real superior del 3er intervalo de clase es igual a la suma del límite superior del 3er intervalo de clase más el límite inferior del 4o intervalo de clase entre dos.

$$\text{Lím. real sup. del 3er int. de clase} = \frac{65 + 66}{2} = \frac{131}{2} = 65.5$$

Por lo que los límites reales del tercer intervalo de clase son 62.5 y 65.5 .

Para obtener el límite real inferior del primer intervalo de clase, al límite real superior 59.5 le restamos la anchura del intervalo  $C=3$ , esto es .

$$59.5 - 3 = 56.5.$$

Por lo que el límite real inferior del primer intervalo es 56.5 .

Y para obtener el límite real superior del último intervalo de clase, al límite real inferior del último intervalo de clase le sumamos la anchura del intervalo  $C=3$  de manera que nos resulta  $74.5 + 3 = 77.5$  .

Por lo que el límite real superior del último intervalo de clase es 77.5 .

### 3.5 TAMAÑO O ANCHURA DE UN INTERVALO DE CLASE

El tamaño o anchura de un intervalo de clase es la diferencia entre los límites reales de clase que lo forman y se conoce como anchura de clase, tamaño de clase o longitud de clase. Si todos los intervalos de clase de una distribución de frecuencias tienen igual anchura, esta anchura común es representada por  $C$ . Por ejemplo el tamaño del primer intervalo de clase de la tabla de frecuencia es:

$$C = 59.5 - 56.5 = 3 \text{ por lo que } C = 3 .$$

### 3.6 MARCA DE CLASE

La marca de clase es el punto medio del intervalo de clase, simbolizado por  $X_i$  se obtiene sumando los límites inferior y superior de clase y dividiendo por 2. Así la marca de clase del intervalo 66-68 es  $\frac{66+68}{2} = 67$ .

La marca de clase se llama también punto medio de clase. Para análisis matemáticos posteriores se va a considerar que cada uno de los datos agrupados en un intervalo tiene un valor igual a la marca de clase del intervalo al que pertenecen. De esta manera todos los pesos en el intervalo de clase 66-68 kilogramos se consideran como 67 kilogramos.

Este supuesto es necesario tomando en cuenta que al formar la distribución por frecuencias se pierden los valores individuales de los datos, lo cual se hace considerando que el error que se introduce suele ser más pequeño, y por lo tanto, se desprecia. Así los valores de las  $X_i$  (marcas de clase) se integran en una nueva columna que se adiciona a la tabla de distribución de frecuencias.

Número de intervalo	Intervalos de clase	Fi	Marca de clase
1	57 - 59	2	58
2	60 - 62	3	61
3	63 - 65	5	64
4	66 - 68	8	67
5	69 - 71	6	70
6	72 - 74	4	73
7	75 - 77	2	76
		N=30	

### 3.7 REPRESENTACION GRAFICA DE DATOS

Al clasificar los datos en diversas clases de tamaños y con ello generar una distribución de frecuencias, las características más importantes de la distribución aparecen con mayor claridad, sin embargo, una presentación gráfica las destaca aún más.

En general se utilizan tres tipos de gráficas para presentar datos distribuidos por frecuencias. Estas son el histograma, el polígono de frecuencias y la ojiva.

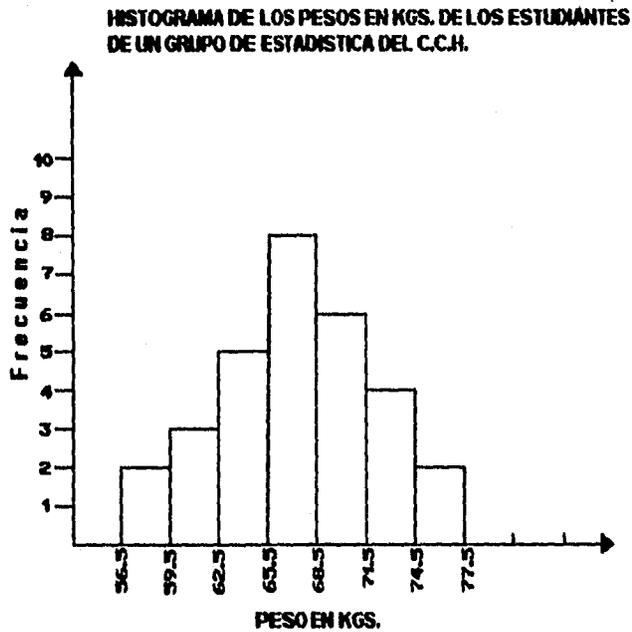
### 3.8 HISTOGRAMA, POLIGONO DE FRECUENCIAS Y LA OJIVA

#### HISTOGRAMA

Histograma o histograma de frecuencias consiste en una serie de rectángulos sin espacios intermedios que se dibuja sobre un par de ejes cartesianos, anotando en las bases (eje horizontal o eje X) los límites reales de los intervalos de clase con centro en las marcas de clase y longitud igual al tamaño de los intervalos. Y como altura de los rectángulos (eje vertical o eje Y) las frecuencias. El eje vertical generalmente empieza con cero, mientras que la escala horizontal puede iniciarse con cualquier número adecuado que convenga como punto de partida para iniciar los intervalos.

En este tipo de gráficas la frecuencia en una clase se representa por la altura del rectángulo correspondiente. En histogramas donde se utiliza densidad de frecuencias (número de datos por unidad de tamaño de intervalo, que se determina dividiendo la frecuencia entre el tamaño del intervalo  $\frac{f_i}{C}$ ) la frecuencia queda representada por el área del rectángulo, por lo que el área del histograma de frecuencias es igual a la suma de todas las frecuencias, es decir, al tamaño de nuestra población.

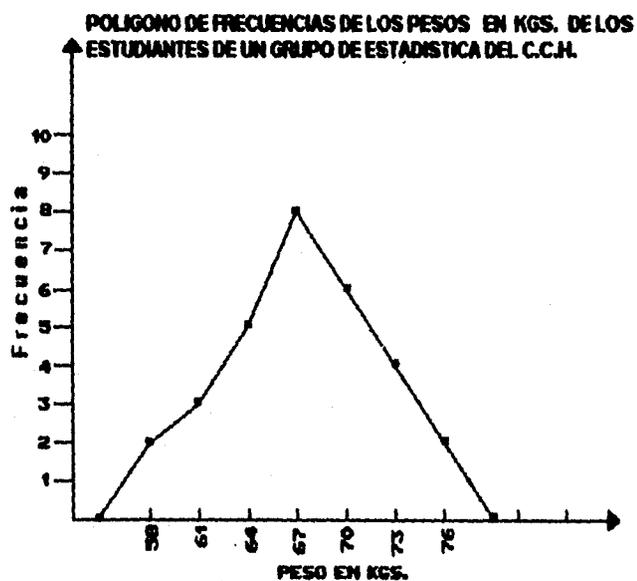
El histograma para el peso de los alumnos es el siguiente:



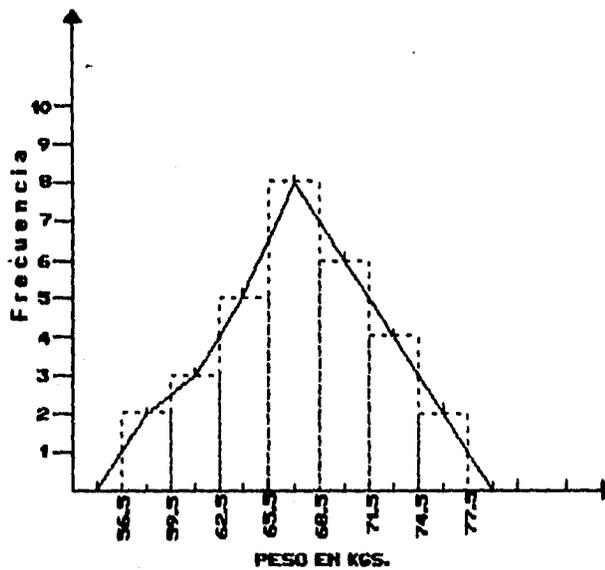
#### POLIGONO DE FRECUENCIAS

El polígono de frecuencias es otra forma de representar gráficamente una distribución de frecuencias. Para trazar un polígono de frecuencias, primero se hace una marca en los puntos medios de la parte superior de cada uno de los rectángulos que representan los intervalos de clase sobre el eje horizontal de la gráfica. Después se

observa que la altura de los rectángulos de un determinado punto corresponde a la frecuencia del intervalo de clase pertinente. Por último se unen los puntos medios con líneas rectas obteniéndose de esta manera el polígono de frecuencias. Hay que hacer notar que el polígono de frecuencias se lleva hasta el eje horizontal en los extremos hasta los puntos, que serán los puntos medios si hubiera un intervalo de clase adicional en cada extremo del histograma, esto permite que el área total quede incluida. Por lo que el área total bajo el polígono de frecuencias equivale al área bajo el histograma.



La siguiente figura muestra el polígono de frecuencias sobrepuesto sobre el histograma. Esta figura permite ver, para el último conjunto de datos, la relación que existe entre las dos formas gráficas.



### 3.9 DISTRIBUCIONES 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 (regla de 3). Por ejemplo, la frecuencia relativa de la clase 66-68 es  $\frac{8}{30} = 0.2666$  o sea 26.66 %. La suma de las

frecuencias relativas de todas clases es evidentemente 1 ó 100%. Si las frecuencias de nuestra tabla de frecuencias se sustituyen por las correspondientes frecuencias relativas, la tabla realmente se llama distribución de frecuencias relativas, distribución porcentual o tabla de frecuencias relativas.

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

### 3.10 DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS

TABLAS DE FRECUENCIAS ACUMULADAS Y OJIVAS

POLÍGONOS DE FRECUENCIAS ACUMULADAS, OJIVAS

"MENOR QUE" Y "MAYOR QUE"

#### DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS

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 hasta ese intervalo de clase inclusive. Por ejemplo la frecuencia

acumulada hasta el intervalo de clase 66-68 inclusive en  $2+3+5+8=18$  significando que 18 estudiantes tienen pesos menores que 68.5 kgs.

una tabla que represente las frecuencias acumuladas se llama distribución de frecuencias acumuladas, tabla de frecuencias acumuladas o brevemente distribución acumulada, se elaboran con los límites reales de clase y pueden ser "menores que" y "mayores que" por ejemplo para el peso de los estudiantes:

TABLA DE FRECUENCIAS ACUMULADAS MENOR QUE :

PESO EN KGS.	NÚMERO DE ESTUDIANTES
MENOR QUE 56.5	0
MENOR QUE 59.5	2
MENOR QUE 62.5	5
MENOR QUE 65.5	10
MENOR QUE 68.5	18
MENOR QUE 71.5	24
MENOR QUE 74.5	28
MENOR QUE 77.5	30

POLIGONO DE FRECUENCIAS ACUMULADAS U OJIVA "MENOR QUE"

Una gráfica que muestre frecuencias acumuladas menores que cualquier límite real superior de clase trazado sobre los límites reales superiores de clase se llama POLÍGONO DE FRECUENCIAS ACUMULADAS U OJIVA y se muestra

en la siguiente figura, para la distribución del peso de los estudiantes.

POLIGONO DE FRECUENCIAS ACUMULADAS (OJIVA) "MENOR QUE" DEL PESO DE LOS ESTUDIANTES DE UN GRUPO DE ESTADISTICA DEL C.C.H.

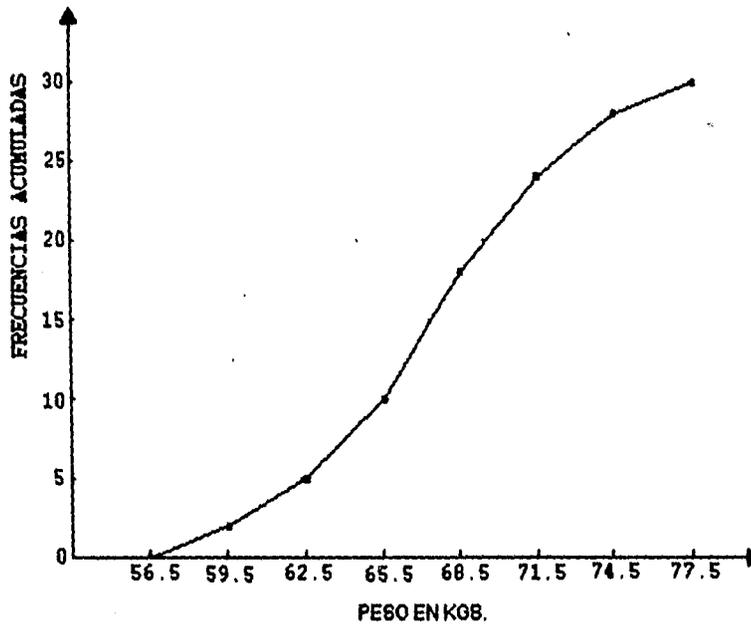


TABLA DE FRECUENCIAS ACUMULADAS "MAYOR QUE"

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 real de clase. En este caso consideramos los pesos de los estudiantes como "mayor

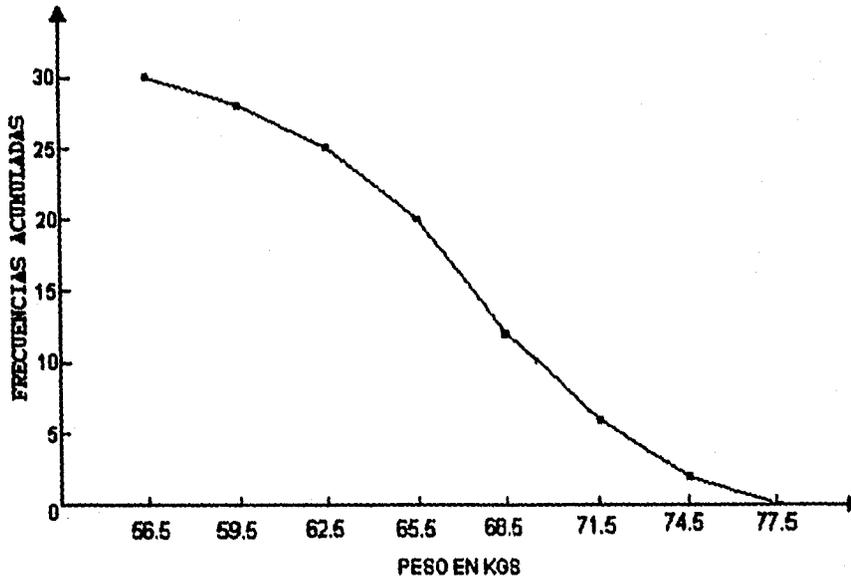
que" o más que 56.5 kgs., "mayor que" o más que 59.5 kgs. ect. esta se llama distribución mayor que o más que, mientras que la considerada anteriormente es la distribución acumulada "menor que". De la una se obtiene fácilmente la otra. Las correspondientes ojivas se llaman "mayor que" y "menor que". La tabla que representa las frecuencias acumuladas "MAYOR QUE" del peso de los estudiantes es:

PESO EN KGS.	NÚMERO DE ESTUDIANTES
MAYOR QUE 56.5	30
MAYOR QUE 59.5	28
MAYOR QUE 62.5	25
MAYOR QUE 65.5	20
MAYOR QUE 68.5	12
MAYOR QUE 71.5	6
MAYOR QUE 74.5	2
MAYOR QUE 77.5	0

#### POLIGONO DE FRECUENCIAS ACUMULADAS U OJIVA "MAYOR QUE"

Una gráfica que muestra las frecuencias acumuladas "MAYOR QUE" es el polígono de frecuencias acumuladas "MAYOR QUE", el cual se muestra a continuación.

POLIGONO DE FRECUENCIAS ACUMULADAS (OJIVA) "MA YOR QUE" PARA EL PESO EN KGS. DE LOS ESTUDIANTES DE UN GRUPO DE ESTADISTICA DEL C.C.H.



### 3.11 DISTRIBUCIONES DE FRECUENCIAS RELATIVAS ACUMULADAS, OJIVAS PORCENTUALES.

La frecuencia relativa acumulada o frecuencia porcentual acumulada es la frecuencia acumulada dividida por la frecuencia total (regla de 3). Por ejemplo, la frecuencia relativa acumulada de los pesos menores que 68.5 kgs. es  $\frac{18}{30} = 0.6$  O 60%, queriendo con ello decir que el 60% de los estudiantes tienen pesos menores que 68.5 kgs.

Si se utilizan las tablas y gráficas anteriores, las frecuencias relativas acumuladas en lugar de las frecuencias acumuladas, los resultados se llaman distribuciones de frecuencias relativas acumuladas y polígonos de frecuencias relativas acumuladas u ojivas porcentuales respectivamente.

Como un resumen de este capítulo elaboraremos una tabla de frecuencias con los conceptos hasta aquí estudiados.

NÚMERO DE INTERVALO	INTERVALOS DE CLASE	Fi	Xi	LÍMITES DE CLASE		LÍMITES REALES DE CLASE	
				INF	SUP	INF	SUP
1	57-59	2	58	57	59	56.5	59.5
2	60-62	3	61	60	62	59.5	62.5
3	63-65	5	64	63	65	62.5	65.5
4	66-68	8	67	66	68	65.5	68.5
5	69-71	6	70	69	71	68.5	71.5
6	72-74	4	73	72	74	71.5	74.5
7	75-77	2	76	75	77	74.5	77.5
		N=30					

FRECUENCIA RELATIVA	FRECUENCIA ACUMULADA OJIVAS	FRECUENCIA RELATIVA ACUMULADA OJIVAS PORCENTUALES
6.66 %	2	6.66 %
10 %	5	16.66 %
16.66 %	10	33.33 %
26.66 %	18	60 %
20 %	24	80 %
13.33 %	28	93.33 %
6.66 %	30	100 %

Ejemplo: La siguiente tabla muestra el precio del kilogramo de cierta marca de arroz en 25 supermercados diferentes.

140	147	130	142	130
175	128	144	137	177
152	170	170	179	162
160	133	158	135	157
166	155	180	174	159

A partir de estos datos elaborar:

I) La distribución de frecuencias de los datos.

II) El histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas (ojiva) menor que.

Solución.

De acuerdo a lo expuesto anteriormente se siguen los siguientes pasos:

a) Ordenamiento de datos.

128	137	152	160	174
130	140	155	162	175
130	142	157	166	177
133	144	158	170	179
135	147	159	170	180

b) Determinar el número de intervalos que deseen utilizarse. Con base al número de datos de la muestra, que en este caso son 25, se pueden escoger de 5 o 6 intervalos, consideremos  $K=6$ .

c) Obtener el tamaño del intervalo.

$$T.I = C = \frac{\text{Rango}}{K} \text{ por lo que } C = \frac{180-128}{6} = \frac{52}{6} = 8.66$$

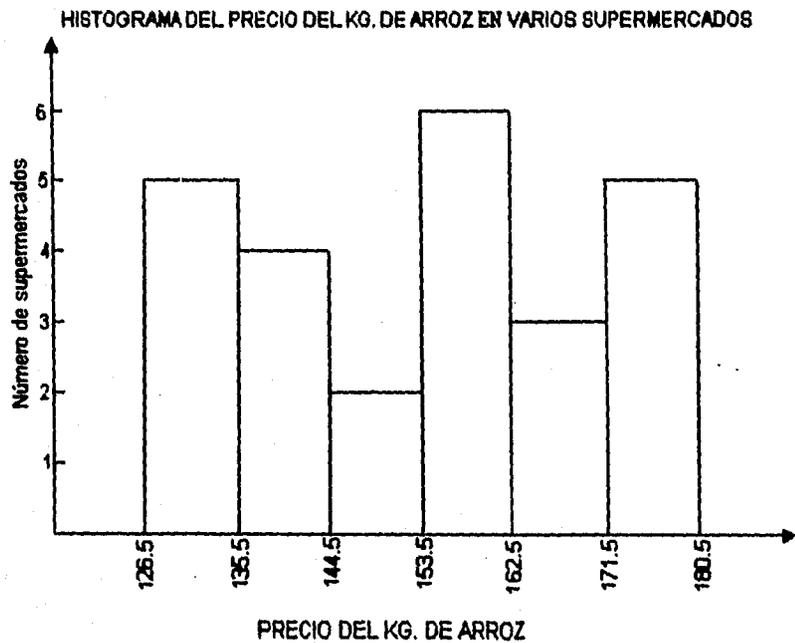
De manera que el tamaño del intervalo resulta  $C=9$ .

d) Una vez que se obtiene el tamaño del intervalo, elaborar la distribución de frecuencias de los datos en varias columnas como en el resumen del capítulo.

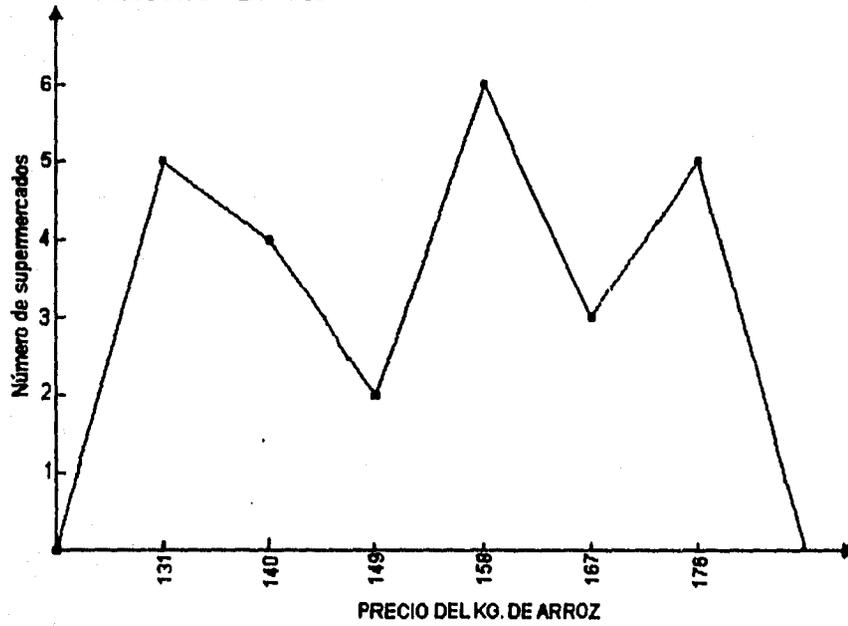
Número de intervalo	Intervalo de clase	Fi	Xi	Límite de clase	
				Inf.	Sup.
1	127 - 135	5	131	127	135
2	136 - 144	4	140	136	144
3	145 - 153	2	149	145	153
4	154 - 162	6	158	154	162
5	163 - 171	3	167	163	171
6	172 - 180	5	176	172	180
		N=25			

Límites reales de clase		Frecuencia relativa	Frecuencia acumulada	Frecuencia relativa acumulada
Inf.	Sup.			
126.5	- 135.5	20 %	5	20 %
135.5	- 144.5	16 %	9	36 %
144.5	- 153.5	8 %	11	44 %
153.5	- 162.5	24 %	17	68 %
162.5	- 171.5	12 %	20	80 %
171.5	- 180.5	20 %	25	100 %

d) Construir el histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas (ojiva) "menor que"



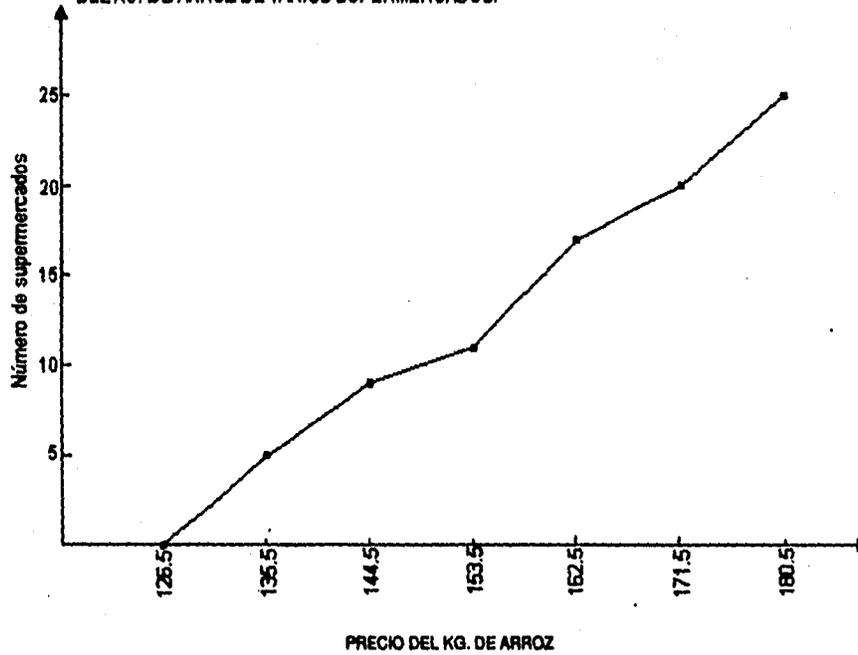
**POLIGONO DE FRECUENCIAS DEL PRECIO DEL KG. DE ARROZ DE VARIOS SUPERMERCADOS**



**Tabla de frecuencias acumuladas "menor que"**

PRECIO DEL KG.	No. DE SUPERMERCADOS
MENOR QUE 126.5	0
MENOR QUE 135.5	5
MENOR QUE 144.5	9
MENOR QUE 153.5	11
MENOR QUE 162.5	17
MENOR QUE 171.5	20
MENOR QUE 180.5	25

POLIGONO DE FRECUENCIAS ACUMULADAS (OJIVA) "MENOR QUE" PARA EL PRECIO DEL KG. DE ARROZ DE VARIOS SUPERMERCADOS.



### 3.12 EJERCICIOS DE REPASO

1.- Los siguientes valores son el número de litros de aceite vendidos en una tienda de abarrotes durante 30 domingos consecutivos.

67	56	64	78	88	57
65	70	66	67	67	61
61	60	72	69	73	64
69	75	68	65	66	80
65	62	71	72	61	68

En base a esos datos.

a) Determinar una distribución de frecuencias.

b) Construir el histograma, el polígono de frecuencias y los polígonos de frecuencias acumuladas (ojivas) "menor que" y "mayor que" .

2.- La siguiente tabla muestra una distribución de frecuencias de la duración en horas de 400 focos con diferentes características, que fabrica una compañía.

Duración (horas)	Número de focos
300 - 399	14
400 - 499	46
500 - 599	58
600 - 699	76
700 - 799	68
800 - 899	62
900 - 999	48
1000 - 1099	22
1100 - 1199	6
N= 400	

Con referencia a esta tabla determinar:

- a) Límite superior de la sexta clase e inferior de la segunda clase.
- b) Marca de clase de la séptima clase.
- c) Límites reales de la última clase.
- d) Tamaño del primer intervalo de clase.

- e) Frecuencia relativa de la cuarta clase.
- f) Porcentaje de focos cuya duración es menor que 800 hrs.
- g) Porcentaje de focos cuya duración es mayor o igual a las 700 hrs.
- h) Porcentaje de focos cuya duración es al menos de 400 hrs. pero menor de 1100 hrs.

3) Para los datos del problema anterior construir:

- a) Un histograma.
- b) Un polígono de frecuencias.
- c) Un histograma de frecuencias relativas.
- d) Un polígono de frecuencias acumuladas (ojiva) "MENOR QUE".

## CAPITULO 4

### MEDIDAS DE TENDENCIA CENTRAL MEDIA ARITMETICA, MEDIANA, MODA Y CUANTILES

- 4.1 LA NOTACION SIGMA  $\Sigma$
- 4.2 PROMEDIOS Y MEDIDAS DE CENTRALIZACION
- 4.3 LA MEDIA ARITMETICA Y SU CALCULO PARA DATOS NO AGRUPADOS
- 4.4 CALCULO DE LA MEDIA ARITMETICA PARA DATOS NO AGRUPADOS, USANDO LA CALCULADORA.
- 4.5 PROPIEDADES DE LA MEDIA ARITMETICA
- 4.6 VENTAJAS Y DESVENTAJAS DE LA MEDIA ARITMETICA.
- 4.7 LA MEDIANA Y SU CALCULO PARA DATOS NO AGRUPADOS.
- 4.8 VENTAJAS Y DESVENTAJAS DE LA MEDIANA.
- 4.9 LA MODA Y SU CALCULO PARA DATOS NO AGRUPADOS.
- 4.10 EJERCICIOS DE REPASO
- 4.11 CALCULO DE LA MEDIA ARITMETICA PARA DATOS AGRUPADOS.
- 4.12 CALCULO DE LA MEDIA ARITMETICA PARA DATOS AGRUPADOS, USANDO LA CALCULADORA
- 4.13 CALCULO DE LA MEDIANA PARA DATOS AGRUPADOS
- 4.14 CALCULO DE LA MEDIANA PARA DATOS AGRUPADOS USANDO CALCULADORA.
- 4.15 CALCULO DE LA MODA PARA DATOS AGRUPADOS.

4.16 CALCULO DE LA MODA PARA DATOS AGRUPADOS USANDO  
CALCULADORA.

4.17 CUARTILES, DECILES Y PERCENTILES (CUANTILES)

4.18 CALCULO DE LOS CUANTILES PARA DATOS AGRUPADOS.

4.19 CALCULO DE LOS CUANTILES PARA DATOS AGRUPADOS CON  
CALCULADORA

4.20 EJERCICIOS DE REPASO.

#### 4.1 LA NOTACION SIGMA ( $\Sigma$ ) :

En matemáticas son utilizados con bastante frecuencia algunos símbolos que nos ayudan a expresar en forma abreviada muchas operaciones, uno de estos símbolos que es utilizado en estadística, es la notación sigma.

Para entender la notación sigma veamos el siguiente ejemplo. Supongamos que tenemos seis valores de una variable X: 1,3,6,7,10 y 13, y que cada uno lo podemos etiquetar de la siguiente manera:

$$x_1 = 1, x_2 = 3, x_3 = 6, x_4 = 7, x_5 = 10, x_6 = 13$$

El símbolo  $x_1$  se lee "x subíndice uno"

El símbolo  $x_2$  , se lee "x subíndice dos"

,

.

.

.

,

El símbolo  $x_6$  se lee "x subíndice seis"

La suma de estos seis valores es:

$$1+3+6+7+10+13$$

para representar esta suma podemos utilizar los símbolos

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

Esta suma la podemos abreviar utilizando la letra mayúscula del alfabeto griego  $\Sigma$  SIGMA que indica que deben sumarse todos los términos colocados delante de ella.

De manera que el símbolo:

$\sum_{i=1}^{i=6} X_i$  indica la suma de todas las  $X_i$  y se lee "X subíndice-i"

desde  $i=1$ , hasta  $i=6$ , esto es:

$$X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = \sum_{i=1}^{i=6} X_i$$

Los números que aparecen arriba y abajo del signo  $\Sigma$  nos indican apartir del cual  $X_i$  vamos a empezar a sumar y en el cual vamos a terminar, respectivamente. Así la expresión anterior se lee "La suma desde que  $i$  es igual a 1 hasta  $i$  igual a 6 de las  $X$  sub-índice  $i$ ". Por lo que la  $i$  toma el valor de 1 después de 2 etc. o sea que va tomando valores enteros, en orden ascendente, empezando con el 1 y terminando con el 6, y que tales valores se van sumando.

Ahora si en lugar de tener 6 valores tenemos  $X_1 + X_2 + X_3 + \dots + X_n$  valores y deseamos hablar de alguno de ellos sin especificar a cual no referimos, podemos utilizar entonces el símbolo  $X_i$  que se lee "X subíndice  $i$ ", lo cual indica que la  $X_i$  puede representar a cualquiera de los 1,2,3,...,  $n$  subíndices. También se puede utilizar como subíndice general otra letra que no sea la  $i$ , por ejemplo  $j, k, q, r, s$ , etc.

$$X_1 + X_2 + X_3 + \dots + X_n = \sum_{i=1}^n X_i .$$

Otro ejemplo de la notación sigma es el siguiente:

Si tenemos una variable simbolizada  $X_i$  que nos representa las edades de 10 niños

$$X_i = X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}$$

Edades: 3, 3, 3, 4, 4, 5, 6, 6, 6, y 7

Entonces  $\sum X_i$  representa la suma de todas las edades de los niños es decir,

$$\begin{aligned} \sum_{i=1}^{10} X_i &= X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} \\ &= 3 + 3 + 3 + 4 + 4 + 5 + 6 + 6 + 6 + 7 = 47 \end{aligned}$$

En algunos casos el símbolo  $\sum$  se emplea para indicar la suma de una parte de la  $X_i$ , es decir de algunos valores simbolizados por  $X_i$ . Si en este ejemplo sólo deseáramos sumar las magnitudes correspondientes a los primeros 5 niños. la suma puede expresarse como:

$$\sum_{i=1}^5 X_i = X_1 + X_2 + X_3 + X_4 + X_5 = 3 + 3 + 3 + 4 + 4 = 17$$

o si la suma corresponde a los valores de los niños simbolizados por  $X_4$  a  $X_7$ ; es decir  $X_4 + X_5 + X_6 + X_7$  deberá expresarse:

$$\sum_{i=4}^7 X_i = 4 + 4 + 5 + 6 = 19$$

ALGUNOS EJEMPLOS MAS DEL SIMBOLO  $\sum$

$$1. \sum_{k=1}^7 k = 1 + 2 + 3 + 4 + 5 + 6 + 7$$

$$2. \sum_{i=1}^{100} i^3 = 1^3 + 3^3 + \dots + 100^3$$

$$3. \sum_{k=1}^{n} X^k = X^1 + X^2 + \dots + X^n$$

$$4. \sum_{i=1}^{2^n} \frac{1}{2^i} = \frac{1}{2^1} + \frac{1}{2^1} + \dots + \frac{1}{2^n}$$

$$5. \sum_{k=2}^{n+1} (X-5)^{k+1} = (X-5)^2 + (X-5)^3 + (X-5)^4 + (X-5)^5$$

Hay que notar en los ejemplos anteriores , que las letras i y k no necesariamente pueden aparecer como subindice sino que puedan ser exponentes, coeficientes, divisores etc.

### Ejercicios

Desarrollar cada una de las siguientes expresiones:

$$1. \sum_{i=1}^n X_i Y_i$$

$$2. \sum_{k=1}^n 5 X_k$$

$$3. \sum_{i=1}^n (X_i Y_i Z_i)^3$$

$$4. \sum_{i=1}^n \frac{f_i X_i}{n}$$

$$5. \sum_{i=1}^n \frac{f_i (X_i - \bar{X})^2}{n}$$

$$6. \sum_{k=1}^n \frac{k+2}{5}$$

Algunas veces cuando se tiene la suma de un conjunto de términos es necesario simbolizarla con  $\Sigma$  para poder desarrollar demostraciones matemáticas.

#### Ejemplo

1)  $1^2 + 2^2 + \dots + n^2$  se puede simbolizar como:

$$1^2 + 2^2 + \dots + n^2 = \sum_{k=1}^{n} k^2$$

2)  $10X_1 + 10X_2 + \dots + 10X_n$  se puede simbolizar como:

$$10X_1 + 10X_2 + \dots + 10X_n = \sum_{k=1}^{n} 10X_k$$

#### Ejercicios

Expresar las siguientes sumas utilizando el signo  $\Sigma$

1.-  $f_1 X_1 Y_1 + f_2 X_2 Y_2 + \dots + f_n X_n Y_n =$

2.-  $ca_1 + ca_2 + \dots + ca_n =$

3.-  $(a_1 + b_1) + (a_2 + b_2) + \dots + (a_n + b_n) =$

4.-  $1^4 + 2^4 + \dots + 100^4 =$

5.-  $f_1 (X_1 - Y_1)^2 + f_2 (X_2 - Y_2)^2 + \dots + f_n (X_n - Y_n)^2 =$

#### 4.2 PROMEDIOS Y MEDIDAS DE CENTRALIZACION

Como se mostró en el capítulo anterior, la presentación gráfica proporciona una descripción visual general de los datos obtenidos. Sin embargo, hay limitaciones para su empleo. Por una parte, no siempre resulta posible presentar visualmente los datos; por otra generalmente la descripción gráfica no se presta al tratamiento matemático necesario para el análisis estadístico. Por estas razones, los expertos en estadística generalmente prefieren utilizar, entre otras medidas, los PROMEDIOS O MEDIDAS NUMERICAS DE TENDENCIA CENTRAL. Algún promedio o valor central que se calcule a partir de los datos de la muestra, proporcionando una imagen mental de los datos y permitiendo al mismo tiempo realizar inferencias acerca de la naturaleza de la población.

PROMEDIO.- Un promedio es un valor 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.

Existen varios tipos de medidas de centralización siendo las más comunes:

LA MEDIDA ARITMETICA

LA MEDIANA

LA MODA

#### 4.3 LA MEDIDA ARITMETICA Y SU CALCULO PARA DATOS NO AGRUPADOS.

Una de las medidas de tendencia central más común y utilizada es el promedio aritmético de un conjunto de observaciones. También se le conoce como la media aritmética, o como la media de un conjunto de observaciones.

La media aritmética de un conjunto de observaciones

$X_1, X_2, \dots, X_n$  es igual a la suma de valores individuales de la muestra dividida entre el número total de observaciones.

Recordamos que nos interesa tanto la muestra como la población, cada una de las cuales posee una media. Para distinguir entre las dos usaremos el símbolo  $\bar{X}$  para la media de la muestra y  $\mu$  (la letra griega mu) para la media de la población.

Ya que las  $n$  observaciones muestrales, se pueden denotar con los símbolos  $X_1, X_2, X_3, \dots, X_n$ . La fórmula para la media muestral es:

$$\bar{X} = \frac{\sum_{i=1}^{i=n} X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad \text{con } i=0,1,2,3,\dots,n$$

y cada uno de los símbolos  $\bar{X}$ ,  $X_i$  y  $n$  nos representan:

$\bar{X}$  = La media aritmética de la muestra y se lee " X barra"

$X_i$  = Datos expresados como valores individuales.

$n$  = Números de elementos de la muestra.

La fórmula para calcular la media aritmética de la población queda expresada por la fórmula.

$$\mu = \bar{X} = \frac{\sum_{i=1}^{i=n} X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad \text{con } i=0,1,2,3,\dots,n$$

donde los símbolos  $\mu$ ,  $X_i$  y  $n$  nos representan:

$\mu$  = media aritmética de la población.

$X_i$  = Datos expresados como valores individuales.

$N$  = Números de elementos de la población (Tamaño de la población).

Ejemplo Obtener la media aritmética para la siguiente muestra de 11 datos. 1, 1, 2, 2, 2, 3, 4, 5, 5, 9, 10

utilizando la fórmula  $\bar{X} = \frac{\sum X_i}{n}$ , se calcula la media de

la siguiente forma:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{1+1+2+2+2+3+4+5+5+9+10}{11} = \frac{44}{11} = 4 \quad \bar{X}=4$$

Hay que observar que aun cuando hay 11 datos para  $x_i$  solamente hay 7 datos diferentes.

El dato 1 aparece dos veces.

El dato 2 aparece tres veces.

El dato 5 aparece dos veces.

Y los otros datos aparecen solamente una vez. Puede calcularse la media, multiplicando el número de veces que ha ocurrido el dato por cada dato diferente y sumando después los productos así obtenidos.

$$\bar{X} = \frac{2(1)+3(2)+1(3)+1(4)+2(5)+1(9)+1(10)}{11} = \frac{44}{11} = 4$$

por lo que  $\bar{X}=4$

Los valores que no están encerrados en paréntesis representan el número de veces que cada dato distinto se repite y los valores dentro del paréntesis son los distintos datos. Entonces, si se emplea  $f_i$  para designar el número de veces que cada dato se repite o frecuencia

del dato, y  $x_i$  para designar los distintos datos, la media muestral  $\bar{X}$  puede expresarse como:

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

Por último el cálculo de la media puede hacerse en forma horizontal o vertical indistintamente, y se calculará únicamente para muestras, a menos que se especifique de otra manera.

Para la forma vertical en este ejemplo su cálculo es:

$x_i$	$f_i$	$f_i x_i$
1	2	2
2	3	6
3	1	3
4	1	4
5	2	10
9	1	9
10	1	10
	11	44

$$\bar{X} = \frac{\sum_{i=1}^n f_i x_i}{n} = \frac{44}{11} = 4$$

$$\therefore \bar{X} = 4$$

#### 4.4 CALCULO DE LA MEDIA ARITMETICA PARA DATOS NO AGRUPADOS, USANDO LA CALCULADORA

Utilizando la calculadora fx casio 5000f para este ejemplo se procede de la manera siguiente:

##### INGRESO DE DATOS

- 1.- PRESIONE MODE X para el modo SD DE ESTADISTICA  
(Aparece en la pantalla SD)
  
- 2.- PRESIONE SHIFT SCL EXE PARA BORRAR LAS MEMORIAS  
ESTADISTICAS (NOTA SCL=AC)
  
- 3.- INGRESE LOS ELEMENTOS DE DATOS UNO POR UNO  
PRESIONANDO LA TECLA DT LUEGO DE CADA INGRESO.

Si tiene un número de elementos de datos que son idénticos, puede ingresarlos como un grupo, usando uno de los dos procedimientos diferentes que a continuación se mencionan.

PRIMER PROCEDIMIENTO

xi	fi	INGRESO
1	2	1 DT DT
2	3	2 DT DT DT
3	1	3 DT
4	1	4 DT
5	2	5 DT DT
9	1	9 DT
10	1	10 DT

(NOTA: DT =  $\sqrt{x}$ )

Cada vez que se presiona DT sin ingresar un nuevo elemento de dato. Se ingresa nuevamente el elemento de dato ingresado previamente.

SEGUNDO PROCEDIMIENTO

Xi	Fi	INGRESO
1	2	1 SHIFT [;] 2 DT
2	3	2 SHIFT [;] 3 DT
3	1	3 DT
4	1	4 DT
5	2	5 SHIFT [;] 2 DT
9	1	9 DT
10	1	10 DT

(NOTA [;]=)

El valor que se ingresa luego de (;) le indica a la calculadora cuántas veces repetir el elemento del dato ingresado previamente.

Una vez que se ha utilizado cualquiera de los dos procedimientos señalados se procede a conocer el resultado de la media aritmética, así como la suma y el número, de los datos operados.

OPERACION DE TECLA	RESULTADO	SIMBOLO
SHIFT X EXE	MEDIA	$\bar{X}=4$
SHIFT IK2 EXE	SUMA DE DATOS	$\sum f_i x_i = 44$
SHIFT IK3 EXE	NUMERO DE DATOS	N=11

(NOTA:  $\bar{X}=1$ , 1k=prog.)

#### BORRADO DE UN ELEMENTO DE DATO INGRESADO

El método usado para borrar los datos ingresados dependen de cuando y cómo los datos fueron ingresados.

##### Ejemplo 1

40DT 20DT 30DT 50DT

PARA BORRAR 50DT, PRESIONE LA TECLA C1 (NOTA C1=X<sup>n</sup>)

##### Ejemplo 2

40DT 20DT 30DT 50DT

PARA BORRAR 20DT, PRESIONE 20 c1

Ejemplo 3

30DT 50DT 120 SHIFT [;]

PARA BORRAR 120 SHIFT [;] PRESIONE AC

Ejemplo 4

30DT 50DT 120 SHIFT [;] 31

PARA BORRAR 120 SHIFT , 31 , PRESIONE AC

Ejemplo 5

30DT 50DT 120 SHIFT [;] 31 DT

PARA BORRAR 120 SHIFT [;] 31 DT, PRESIONE C1

EJEMPLO 6

50 DT 120 SHIFT [;] 31 DT 40 DT 30 DT

PARA BORRAR 120 SHIFT [;] 31 DT PRESIONE 120 SHIFT [;]

31 C1

Es necesario que antes de que utilices la calculadora en fórmulas programadas como la  $\bar{X}$ , etc., desarrolles manualmente las operaciones indicadas en las fórmulas.

Ejemplo Encontrar la medida aritmética del siguiente conjunto de datos 4,5,5,7,7,8,9,10,10,11 y 12.

$X_i$	$f_i$	$f_i X_i$
4	1	4
5	2	10
7	2	14
8	1	8
9	1	9
10	2	20
11	1	11
12	1	12
	11	88

$$\bar{X} = \frac{\sum_{i=1}^n f_i X_i}{n} = \frac{88}{11} = 8$$

Para comprobar este resultado con la calculadora se procede de esta manera

#### INGRESO DE DATOS

- 1.- PRESIONAR MODE X PARA EL MODO SD DE ESTADISTICA
- 2.- PRESIONAR SHIFT SCI EXE PARA LIMPIAR LAS MEMORIAS ESTADISTICAS
- 3.- INGRESAR LOS ELEMENTOS DE LOS DATOS UNO POR UNO PRESIONANDO LA TECLA DT LUEGO DE CADA INGRESO

$X_i$	$f_i$	INGRESO
4	1	4 DT
5	2	5 DT DT o 5 SHIFT [;] 2 DT
7	2	7 DT DT o 7 SHIFT [;] 2 DT
8	1	8 DT
9	1	9 DT
10	2	10 DT DT o 10 SHIFT [;] 2 DT
11	1	11 DT
12	1	12 DT

OPERACION DE LA TECLA	RESULTADO	SIMBOLO
SHIFT X EXE	MEDIA	$\bar{X}=8$
SHIFT IK2 EXE	SUMA DE DATOS	$\sum f_i X_i=88$
SHIFT IK3 EXE	NUMERO DE DATOS	$N=11$

#### 4.5 PROPIEDADES DE LA MEDIA ARITMETICA

Antes de mencionar la primera propiedad es conveniente dar una explicación de lo que se entiende por las desviaciones que hay entre un conjunto de datos y su media aritmética.

Definición.- Las desviaciones de un conjunto de datos respecto a su media aritmética son las diferentes cantidades que resultan de restar el valor de cada uno de los datos y su media aritmética, se representa como  $(X_i - \bar{X})$ .

PRIMERA PROPIEDAD DE LA MEDIDA ARITMETICA

La suma algebraica de las desviaciones de un conjunto de datos  $X_1, X_2, X_3, \dots, X_n$  de su medida aritmética  $\bar{X}$  es igual a cero.

$$\sum_{i=1}^n (X_i - \bar{X}) = 0$$

Ejemplo: tomando los datos del ejercicio anterior

4, 5, 5, 7, 7, 8, 9, 10, 10, 11, y 12 con  $\bar{X} = 8$

$X_i$	$(X_i - \bar{X})$	$X_i$	$(X_i - \bar{X})$
4	4 - 8 = -4	4	-4
5	5 - 8 = -3	5	-3
5	5 - 8 = -3	5	-3
7	7 - 8 = -1	7	-1
7	7 - 8 = -1	7	-1
8	8 - 8 = 0	8	0
9	9 - 8 = 1	9	1
10	10 - 8 = 2	10	2
10	10 - 8 = 2	10	2
11	11 - 8 = 3	11	3
12	12 - 8 = 4	12	4
			0

Utilizando las frecuencias de cada dato

$\sum f_i (X_i - \bar{X}) = 0$  El ejemplo anterior nos queda como:

$X_i$	$f_i$	$(X_i - \bar{X})$	$f_i (X_i - \bar{X})$
4	1	-4	-4
5	2	-3	-6
7	2	-1	-2
8	1	0	0
9	1	1	1
10	2	2	4
11	1	3	3
12	1	4	4
			0

#### SEGUNDA PROPIEDAD DE LA MEDIA

La suma de los cuadrados de las desviaciones de un conjunto de números  $X_i$  respecto a cualquier  $a$  es mínimo solamente si  $a = \bar{X}$ .

#### Ejemplo

Tomamos el siguiente conjunto de datos 2,3,4,5 y 6 el cual tiene como media  $\bar{X}=4$  y  $n=5$ .

Apliquemos esta propiedad de la media para:

$a = \bar{X} = 4$  ,  $a = 2$  ,  $a = 3$  ,  $a = 5$  y  $a = 6$  entonces se tiene:

PARA  $a=\bar{x}=4$

$x_i$	$(x_i - 4)$	$(x_i - 4)^2$
2	-2	4
3	-1	1
4	0	0
5	1	1
6	2	4
		<hr/>
		10

PARA  $a=2$

$x_i$	$(x_i - 2)$	$(x_i - 2)^2$
2	0	0
3	1	1
4	2	4
5	3	9
6	4	16
		<hr/>
		30

PARA  $a=3$

$x_i$	$(x_i - 3)$	$(x_i - 3)^2$
2	1	1
3	0	0
4	1	1
5	2	4
6	3	9
		<hr/>
		15

PARA  $a=5$

$X_i$	$(X_i-5)$	$(X_i-5)^2$
2	-3	9
3	-2	4
4	-1	1
5	0	0
6	1	1
		<hr/>
		15

PARA  $a=6$

$X_i$	$(X_i-6)$	$(X_i-6)^2$
2	-4	16
3	-3	9
4	-2	4
5	-1	1
6	0	0
		<hr/>
		30

Las tablas anteriores muestran los cuadrados y la suma de los cuadrados de las desviaciones, cuando éstas se toman respecto a la media y a otros valores del conjunto de datos. Puede verse que la suma de los cuadrados es más pequeña cuando las desviaciones se toman respecto a la media.

### TERCERA PROPIEDAD DE LA MEDIA

MEDIA ARITMETICA PONDERADA: Si a un conjunto de datos

$X_1, X_2, X_3, \dots, X_n$  se le asignan ciertos factores o pesos  $W_1, W_2, W_3, \dots, W_n$  que dependen de la importancia de cada uno de los datos, entonces.

$$\bar{X} = \frac{W_1 X_1 + W_2 X_2 + \dots + W_n X_n}{W_1 + W_2 + \dots + W_n} = \frac{\sum W_i X_i}{\sum W_i}$$

Se le llama media aritmética ponderada.

Ejemplo: Si un examen final del curso se valora como tres veces los exámenes parciales y un estudiante tiene una calificación de examen final de 85 y calificaciones de exámenes parciales de 70 y 90, su calificación final será:

$$\bar{X} = \frac{1(70) + 1(90) + 3(85)}{1 + 1 + 3} = \frac{415}{5} = 83$$

Por lo que su media aritmética ponderada es  $\bar{X}=83$

#### CUARTA PROPIEDAD DE LA MEDIA

Si  $f_1$  números tienen como media  $m_1$ .

$f_2$  números tienen como media  $m_2$

,

.

.

.

,

$f_n$  números tienen como media  $m_n$

entonces la media de todos los números es:

$$\bar{X} = \frac{f_1 m_1 + f_2 m_2 + \dots + f_n m_n}{f_1 + f_2 + \dots + f_n}$$

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

Ejemplo: Cuatro grupos de estudiantes, formados por 15, 20, 10 y 18 individuos registran una media de pesos de 58, 62, 60 y 67 kilogramos respectivamente. Hallar el peso promedio de todos los estudiantes.

Entonces:

$$f_1 = 15 \quad m_1 = 58$$

$$f_2 = 20 \quad m_2 = 62$$

$$f_3 = 10 \quad m_3 = 60$$

$$f_4 = 18 \quad m_4 = 67$$

$$\bar{X} = \frac{15(58) + 20(62) + 10(60) + 18(67)}{15+20+10+18} = \frac{3916}{63} = 62.158$$

El peso promedio de todos los estudiantes es  $\bar{X}=62.16$  Kgs.

#### 4.6 VENTAJAS Y DESVENTAJAS DE LA MEDIA ARITMETICA

1. UNICIDAD. Para un determinado conjunto de datos, existe una y sólo una media aritmética.
2. SIMPLICIDAD. La media aritmética es fácil de comprender y de calcular.
3. Dado que todos y cada uno de los valores de un conjunto de datos intervienen en el cálculo de la media, ésta es afectada por cada valor. Por lo que los valores extremos (muy bajos o muy altos) influyen en su cálculo, y en algunos casos, pueden alterarla tanto que resulta inconveniente como una medida de tendencia central.

Por ejemplo. Si los honorarios anuales de cinco médicos que ejercen en cierta área son:

\$ 60,000.00,  
 \$ 60,000.00,  
 \$ 60,000.00,  
 \$ 60,000.00  
 \$100,000.00

La media es:

$$\bar{X} = \frac{60,000 + 60,000 + 60,000 + 60,000 + 100,000}{5} = 68,000$$

Evidentemente el único honorario de \$100,000 ha elevado la media en \$8,000. Si la media de los honorarios de los cinco médicos es \$68,000, cabría pensar que dichos honorarios se han distribuido alrededor de \$68,000. Por esta razón, la media de \$ 68,000 no representa adecuadamente en este caso, la distribución de frecuencias de los honorarios.

#### 4.7 LA MEDIANA Y SU CALCULO PARA DATOS NO AGRUPADOS

Otra medida de tendencia central es la mediana. La mediana de un conjunto de observaciones generalmente se define de la siguiente manera.

La mediana de un conjunto finito de valores arreglados en orden de magnitud, es aquel valor que divide al conjunto en dos partes iguales. Si el número de valores es impar, el valor central que es el valor  $(n+1)/2$  ésimo, es la mediana. Si el número de valores es par, no se tiene una sola observación en medio sino dos. En este caso se toma de mediana como la media de estas dos observaciones centrales.

Ejemplo 1. Obtener la mediana del siguiente conjunto de datos:

10,10,11,12,12,13,14,15,16,17,17

En este caso los datos ya están ordenados. Y como el número de datos 11 es impar, la mediana se encuentra en el  $(n+1)/2$  datos o sea  $(11+1)/2 = 12/2 = 6$  es el sexto dato.

10,10,11,12,12, 13 ,14,15,16,17,17

Por lo que Mediana = 13, que es el valor central con cinco datos antes y cinco después.

Ejemplo 2: Calcular la mediana de los siguientes 14 valores.

10,3,10,12,9,7,9,6,7,10,8,6,7,2

Primero arréglese los valores ya sea en orden ascendente o descendente. El arreglo en orden ascendente para estos 14 datos se muestra a continuación.

2,3,6,6,7,7,7,8,9,9,10,10,10,12

Como hay 14 valores (par), la mediana es la media de los valores séptimo y octavo. Esto es:

2,3,6,6,7,7, 7,8, 9,9,10,10,10,12

$$\text{Mediana} = \frac{7+8}{2} = \frac{15}{2} = 7.5$$

#### 4.8 VENTAJAS Y DESVENTAJAS DE LA MEDIANA

1. Unicidad. Como ocurre con la media, sólo existe una mediana para un determinado conjunto de datos.
2. Simplicidad. La mediana es fácil de calcular.
3. No es afectada tan drásticamente por los valores extremos como lo es la media.

#### 4.9 LA MODA Y SU CALCULO PARA DATOS NO AGRUPADOS

La moda de un conjunto de valores es aquel valor que se presenta con mayor frecuencia. Si todos los valores son distintos, no hay moda; por otra parte, un conjunto de valores puede tener más de una moda.

De todas las medidas de tendencia central, la moda es la que más fácilmente se determina, ya que se obtiene por inspección, y no por cómputo.

Ejemplo 1: Encontrar la moda del siguiente conjunto de datos.

10, 10, 11, 12, 13, 13, 13, 14, 15, 16, 16                      Moda=13

Ejemplo 2: Calcular la moda de:

10, 11, 12, 12, 13, 14, 15, 16, 17, 17

Cuando hay dos modas se llama bimodal

Ejemplo 3: Calcular la moda de:

4,4,4,5,5,6,7,7,7,8,8,9,10,11,11,11,12,13,13,14

Cuando hay más de dos modas se llama Multimodal

Ejemplo 4: Encontrar la moda de

4,6,7,8,9,10,11,12

En este caso la moda no existe.

La moda resulta sumamente útil para expresar la tendencia central de observaciones correspondientes a características, cualitativas tales como, estado civil, ocupación, estado de nacimiento etc. por ejemplo, supóngase que los pacientes que se atendieron en una clínica de salud mental durante un determinado año recibieron uno de los siguientes diagnósticos: retraso mental, síndrome cerebral orgánico, psicosis, neurosis y alteración de la personalidad. El diagnóstico que ocurriera con más frecuencia en el grupo de pacientes se llamaría diagnóstico modal.

#### 4.10 EJERCICIOS DE REPASO

1. Las calificaciones de un estudiante en 10 exámenes fueron:

84,91,72,68,78,80,90,87,77,79.

Hallar la media aritmética de las calificaciones

2. Diez medidas del diámetro de un cilindro fueron registradas como, 3.88, 4.09, 3.92, 3.97, 4.02, 3.95, 4.03, 3.92, 3.98 y 4.06 pulgadas.

Hallar la media aritmética de las medidas, utilizando la calculadora.

3. Los salarios anuales de 4 hombres fueron:

\$5,000, \$6,000, \$6,500 y \$30,000.

a) Hallar la media aritmética de sus salarios

b) Decir si este promedio es representativo de los salarios.

4. Comprobar que la suma de las desviaciones de un conjunto de datos de su media aritmética es cero (primera propiedad de la media) de:

4, 5, 6, 6, 6, 7, 8, 8, 8, 9 y 10

nota: primero hay que calcular la media aritmética.

5. Las puntuaciones finales de un estudiante en Matemáticas, Física, Inglés y Biología son, respectivamente, 82, 86, 90 y 70. Si la importancia que se asigna a estas materias son de 5, 3, 2 y 1 respectivamente, determinar el promedio de puntuación adecuado.

6. Tres profesores de estadística registraron una calificación media en sus exámenes de 79, 82 y 84.

Sus clases estaban formadas por 32,25 y 17 estudiantes, respectivamente. Hallar la calificación media para todas las clases.

7. Un grupo de 13 personas que coleccionan estampas, recolectan cada una.

36,28,32,34,36,39,40,50,49,46,44,43 y 46

En base a esos datos calcula:

- a) La media aritmética
- b) La mediana
- c) La moda
- d) La suma de las desviaciones de cada dato con respecto a la media aritmética
- e) ¿Cuál de las tres medidas centrales, escogerías como representativa?

8. Los siguientes datos representan los salarios anuales de 10 ingenieros.

salario	fi
\$ 250,000	1
\$ 110,000	3
\$ 110,000	3
\$ 90,000	2
\$ 80,000	1
	<hr/>
	10

En base a esos datos calcula:

- a) La media aritmética
- b) La mediana
- c) La moda
- e) Decir cuál es la más representativa y por qué.

9. Los siguientes valores son los niveles de glucosa en la sangre extraída a 10 niños en ayunas.

Número	Valor
1	56
2	62
3	63
4	65
5	65
6	65
7	65
8	68
9	70
10	72

En base a esos datos calcule:

- a) La media aritmética
- b) La mediana
- c) La moda

10. Una muestra de once pacientes admitidos para diagnóstico y evaluación en una sala psiquiátrica recientemente abierta en un hospital general, tuvieron las duraciones siguientes de su internación.

Número	Duración del internado (días)
1	29
2	14
3	11
4	24
5	14
6	14
7	28
8	14
9	18
10	22
11	14

Encuentre:

- a) La duración media de la internación para estos pacientes
- b) La mediana
- c) La moda

11. A ocho pacientes en el ala de convalecencia de un hospital general se les permitió que eligieran entre cuatro tipos de carne para la comida. Sus elecciones

fueron las siguientes: pollo, pescados, hígado, pollo, pollo, tortas de carne y pollo. ¿Cuál fué la elección modal?

#### 4.11 CALCULO DE LA MEDIA ARITMETICA PARA DATOS AGRUPADOS

Cuando se agrupan los datos, en intervalos, las observaciones individuales pierdan su identidad. Observando una distribución de frecuencias puede determinarse el número de observaciones que caen dentro de los diferentes intervalos de clase, pero no pueden determinarse los valores reales. Debido a esto, deben plantearse ciertas hipótesis acerca de los valores cuando se calcule una medida descriptiva a partir de datos agrupados. Como consecuencia de estas suposiciones, los resultados obtenidos son sólo aproximaciones de los valores verdaderos.

Al calcular la media a partir de datos agrupados, se supone que todos los valores que caen dentro de un determinado intervalo de clase se localizan en el punto medio del intervalo.

Para encontrar la media, se multiplica cada punto medio por la frecuencia correspondiente, se suman estos productos y se divide entre la suma de las frecuencias.

Si los datos representan una muestra de datos, el cálculo de la media puede mostrarse simbólicamente como:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} \quad \text{o} \quad \bar{x} = \frac{\sum f_i x_i}{n}$$

Donde la  $x_i$  representan el punto medio de cada intervalo de clase.

Para calcular la media a partir de una población de valores agrupados en un número finito de clases, se procede igual que en la ecuación anterior.

Ejemplos.

Ejemplo 1. Calcular la media aritmética para la tabla de distribución de frecuencias del problema del peso en Kgs. de los estudiantes de un grupo de estadística del C.C.H.

NUMERO DE INTERVALO	INTERVALO DE CLASE	$f_i$	$x_i$	$f_i x_i$
1	56.50 - 59.50	2	58	116
2	59.50 - 62.50	3	61	183
3	62.50 - 65.50	5	64	320
4	65.50 - 68.50	8	67	536
5	68.50 - 71.50	6	70	420
6	71.50 - 74.50	4	73	292
7	74.50 - 77.50	2	76	152
		30		2019

$$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{2019}{30} = 67.30$$

$$\therefore \bar{x} = 67.30 \text{kg}$$

4.12 CALCULO DE LA MEDIA ARITMETICA PARA DATOS  
 AGRUPADOS, USANDO LA CALCULADORA

Para calcular la media aritmética de este problema usando la calculadora CASIO 5000 su procede de la siguiente forma.

1. PRESIONE MODE X PARA EL MODO SD DE ESTADISTICA.  
(aparece SD en la pantalla)
2. PRESIONE SHIFT SC1 EXE PARA BORRAR LAS MEMORIAS ESTADISTICAS
3. INGRESE LOS ELEMENTOS DE DATOS UNO POR UNO PRESIONANDO LA TECLA DT LUEGO DE CADA INGRESO

$X_i$	$f_i$	INGRESO
58	2	58 DT DT
61	3	61 DT DT DT
64	5	64 SHIFT [;] 5 DT
67	8	67 SHIFT [;] 8 DT
70	6	70 SHIFT [;] 6 DT
73	4	73 SHIFT [;] 4 DT
76	2	76 DT DT

OPERACION DE TECLA	RESULTADO	SIMBOLO
SHIFT X EXE	MEDIA	$\bar{X}=67.30$
SHIFT IK2 EXE	SUMA DE DATOS	$\sum f_i x_i = 2019$
SHIFT IK3 EXE	NUMEROS DE DATOS	$\sum f_i = N = 30$

Ejemplo 2. La siguiente tabla muestra las calificaciones obtenidas en una prueba de inteligencia por un grupo de niños que participaron en un experimento. Calcular su media.

NUMERO DE INTERVALO	INTERVALO DE CLASE	fi	Xi	fixi
1	10 - 19	5	14.5	72.5
2	20 - 29	19	24.5	465.5
3	30 - 39	10	34.5	345.0
4	40 - 49	13	44.5	578.5
5	50 - 59	4	54.5	218.0
6	60 - 69	4	64.5	258.0
7	70 - 79	2	74.5	149.0
		57		2086.5

$$\Sigma fixi = 2086.5$$

$$X = \frac{2086.5}{57} = 36.60$$

$$\therefore X = 36.60$$

Comprobación de este resultado por medio de la calculadora

1. MODE X  $\Rightarrow$  SD

2. SHIFT SCI EXE

3. INGRESO

```

14.5 SHIFT [;] 5
24.5 SHIFT [;] 19
34.5 SHIFT [;] 10
44.5 SHIFT [;] 13
54.5 SHIFT [;] 4
64.5 SHIFT [;] 4
74.5 DT DT

```

OPERACION DE LA TECLA	RESULTADO	SIMBOLO
SHIFT X EXE	MEDIA	$\bar{X}=36.60$
SHIFT 1K2 EXE	SUMA DE DATOS	$\sum fixi=2086.5$
SHIFT 1K3 EXE	NUMERO DE DATOS	$\sum fi=N=57$

#### 4.13 CALCULO DE LA MEDIANA PARA DATOS AGRUPADOS

Quando se calcula la MEDIA a partir de datos agrupados, se supone que los valores dentro de un intervalo de clase se localizan en el punto medio, sin embargo, al calcular la MEDIANA; se supone que dichos valores estan distribuidos uniformemente en todo el intervalo.

Tomando como base la tabla del peso de los estudiantes de estadística del C.C.H. Procederemos a calcular la mediana como una de forma ilustrar su desarrollo.

NUMERO DE INTERVALO	INTERVALO DE CLASE	FRECUENCIA $f_i$	FRECUENCIA ACUMULADA
1	57 - 59	2	2
2	60 - 62	3	5
3	63 - 65	5	10
4	66 - 68	8	18
5	69 - 71	6	24
6	72 - 74	4	28
7	75 - 77	2	30

El primer paso para calcular la mediana a partir de datos agrupados es localizar el dato en donde se encuentra la mediana utilizando la fórmula  $n/2$ . Para

muestra distribución, tenemos  $n/2=30/2=15$ , por lo que el valor 15 es el dato central donde se encuentra la mediana.

El segundo paso es localizar el intervalo de clase en el que se encuentra el dato 15. Observando la tabla se ve que:

El primer intervalo de clase que tiene 2 frecuencias, se encuentran los datos 1 y 2.

El segundo intervalo de clase que tiene 3 frecuencias, se encuentran los datos 3,4 y 5.

El tercer intervalo de clase que tiene 5 frecuencias, se encuentran los datos 6,7,8,9 y 10.

El cuarto intervalo de clase que tiene 8 frecuencias se encuentran los datos 11,12,13,14,15,16,17 y el 18.

Por lo que el valor de la mediana se encuentra en el cuarto intervalo de clase. Esta en algún punto entre 65.5 y 68.5 que son los límites reales de la clase.

El tercer paso es calcular la mediana mediante un proceso de interpolación, aplicado la siguiente fórmula

$$\text{MEDIANA} = L_1 + \left( \frac{n/2 - f_a}{f_m} \right) C$$

$L_1$  = Limite real inferior del intervalo de clase donde se encuentra la mediana.

$f_a$  = Frecuencia acumulada del intervalo de clase anterior al intervalo de clase en donde se encuentra la mediana.

$C$  = Tamaño del intervalo de clase en donde se encuentra la mediana.

$f_m$  = Frecuencia del intervalo de clase donde se encuentra la mediana.

En nuestro ejemplo se tiene:

$N/2 = 30/2 = 15$  La mediana esta en cuarto intervalo de clase.

$$L_1 = (65+66)/2 = 65.5$$

$$f_a = 2+3+5 = 10$$

$$C = 68.5 - 65.5 = 3$$

$$f_m = 8$$

$$\text{MEDIANA} = 65.5 + \left( \frac{15-10}{8} \right) 3$$

MEDIANA =  $65.5 + 1.875 = 67.375$

MEDIANA = 67.375 Kgs.

#### 4.14 CALCULO DE LA MEDIANA PARA DATOS AGRUPADOS USANDO LA CALCULADORA

La calculadora CASIO 5000 viene incorporada con un total de 12 memorias independientes para el almacenamiento de las fórmulas.

Cuando se almacena una fórmula en una memoria, esta permanece hasta que se le borra. Los contenidos de las memorias del usuario se retienen aún si se presiona la tecla AC o si se apaga la unidad.

Para almacenar una fórmula en la calculadora se siguen los siguientes pasos.

##### 1. PRESIONE MODE 2 PARA CAMBIAR EL MODO WRT

Aparece en la pantalla algo similar a

P	☉	1	2	3	4	5	6	7	8	9	A	B
					wrt						D	
S=											675	

La línea superior muestra las memorias de fórmulas que se encuentran vacías y listas para almacenar. Estas memorias serán referidas como memoria

PO,P1,P2,...,P9,PA y PB. Cada vez que se almacena una fórmula en una de las memorias, el nombre de la memoria en uso se reemplaza en la pantalla con un guión.

El número en la línea inferior de la pantalla muestra el número total de pasos restantes que quedan para la fórmula en la memoria. Este es un punto muy importante. Se puede almacenar lo que desee en cualquiera de las memorias de las fórmulas del usuario, en tanto que el número de pasos de fórmulas ya almacenados no exceda de 675 en las memorias de las fórmulas del usuario.

La presentación usada aquí muestra que las memorias de fórmulas del usuario están todas vacías. Cada vez que se almacena una fórmula en una de las memorias, el número en esta presentación se convierte más pequeño. Cuando se ingresan las fórmulas, la posición ingreso se indica siempre por un cursor parpadeando. Cuando en la memoria solamente existen seis o menos pasos de fórmulas restantes, el cursor cambia a ■ , y el ingreso se convierte imposible cuando se terminan de usar todos los pasos. Si esto ocurre, para lograr más espacios, borre las fórmulas que ya no necesite.

#### ALMACENAMIENTO DE UNA FORMULA

1. PRESIONE MODE 2 PARA CAMBIAR AL WRT

2. USE LAS TECLAS ← o → para seleccionar un nombre de memoria de fórmula.

La memoria de fórmula corrientemente seleccionada se indica por su propio nombre parpadeando en la pantalla. Las teclas ← o → mueven la selección hacia la izquierda o derecha. Por supuesto, este paso puede ser omitido si el nombre de la memoria de fórmula que se encuentra parpadeando es el que se desea.

3. ESPECIFIQUE UN MODO DE CALCULO PARA LA MEMORIA DE FORMULA SELECCIONADA.

La siguiente es una lista de operaciones de teclas y las restricciones para cada modo de cálculo

COMP MODE [+]

BASE n MODE [-]

No puede realizarse cálculos con funciones científicas.

No puede especificarse unidades de medición angular.

No puede recuperarse las fórmulas incorporadas.

SD MODE [X]

No pueden usarse Abs ni  $\sqrt{\quad}$

DT no puede ingresarse consecutivamente LR MODE [/]

No pueden usarse Abs ni  $\sqrt{\quad}$

DT no puede ingresarse consecutivamente

4. PRESIONE LA TECLA EXE

La pantalla debe aparecer como se muestra aquí. El 0 en la parte inferior derecha indica la cantidad de pasos que hay entre la posición presente del cursor y el comienzo de la fórmula. Como todavía no se ha ingresado nada, este número indica cero pasos en la memoria

WTR	D
S=	0

5. INGRESE LA FORMULA

FORMULA DE LA MEDIANA  $M=L_1 + \left( \frac{n/2 - fa}{f_m} \right) C$

FORMULA PARA INGRESAR A LA CALCULADORA

$M= L + ((N/2 - A)/B)C$  DONDE:

$L= L_1$

$n= n$

$A= fa$

$B= f_m$

$C= C$

AC ALPHA M SHIFT= ALPHA L+((ALPHA N/2-ALPHA A)/ ALPHA B)

x ALPHA C EXE

APARECE EN LA PANTALLA

$M=L + ((N/2-A)/B)xC$
L? 0

Observe que cada vez que se ingresa algo, el número que indica el número de pasos se convierte más grande. Si se

utiliza las teclas ← o → para mover el cursor parpadeante en la pantalla, el número cambiará para indicar la cantidad de pasos que existe entre la presente posición y el comienzo de la fórmula.

#### 6. PRESIONE MODE 1 PARA ALMACENAR EL PROGRAMA

EJECUCION DE UNA FORMULA ALMACENADA EN UNA MEMORIA DE LAS FORMULAS DEL USUARIO.

Una vez que se almacena una fórmula, se le puede ejecutar en cualquier momento especificando el nombre de la fórmula del usuario.

1. PRESIONE MODE 1 PARA INGRESAR AL MODO RUN.
2. PRESIONES PROG. SEGUIDO POR EL NOMBRE DE LA MEMORIA DE LA FORMULA.
3. PRESIONE LA TECLA EXE PARA EJECUTAR LA FORMULA SELECCIONADA.

Una vez que se ejecuta una fórmula, la operación siguiente depende del tipo de fórmula que se recupera. Por ejemplo, los indicadores aparecen para indicarle el ingreso de los valores para las variables etc.. Para nuestro ejemplo tenemos:

1. MODE 1

2. PROG O EXE EN LA PANTALLA

$M=L+((N/2-A)/B) \times C$
L ?                      0

$L_1=L=65.5$   
 $n=30$   
 $f_a=A=10$   
 $f_m=B=8$   
 $C=3$

65.5

EN LA PANTALLA

65.5	D
L?	0

EXE

$M=L + ((N/2-A)/B) \times C$	D
N?	0

30

30-	D
N?	0

EXE

$M=L + ((N/2-A)/B) \times C$	D
A?	0

10

10-	D
A?	0

EXE

$M=L + ((N/2-A)/B) \times C$	D
B?	0

8

8-	D
B?	0

EXE

$M=L + ((N/2-A)/B) \times C$	D
C?	0

3

3-	D
C?	O

EXE

$M=L + ((N/2 - A) / B) \times C$	
M=	67.375

MEDIANA=67.375

BORRADO DE LAS FORMULAS DEL USUARIO.

DESPUES DE ALMACENAR LAS FORMULAS PARA LAS MEMORIAS DEL USUARIO, SE PUEDEN BORRAR INDIVIDUALMENTE O TODAS A LA VEZ.

BORRADO DE UNA FORMULA ESPECIFICA DEL USUARIO

1. PRESIONE MODE 3 PARA INGRESAR AL MOD PCL

Se pueden borrar las fórmulas del usuario solamente en el modo PCL.

2. USE LAS TECLAS ← o → PARA SELECCIONAR UN NOMBRE EN LA MEMORIA DE FORMULA.

Las memorias que contienen fórmulas se indican por sus nombres de memoria siendo reemplazadas por un guión.

3. PRESIONE LA TECLA AC PARA BORRAR LA FORMULA DEL USUARIO SELECCIONADA.

Luego de que se presiona la tecla AC, el número de la memoria en donde se ha borrado la fórmula aparecerá nuevamente en la pantalla en lugar del guión.

#### 4. PRESIONE MODE 1 PARA VOLVER AL MODO RUN

#### BORRADO DE TODAS LAS FORMULAS DEL USUARIO

1. PRESIONE MODE 3 PARA INGRESAR AL MODO PCL

2. PRESIONE SHIFT MCL

Esto borra todas las fórmulas del usuario, y ocasiona que todos los nombres de la memoria del usuario aparezcan en la pantalla.

Ejemplo

#### BORRADO DE TODAS LAS MEMORIAS

MODE 3

P	-	1	2	3	4	5	6	7	8	9	A	B
											D	
	S=											658

SHIFT MCL

P	Ø	1	2	3	4	5	6	7	8	9	A	B
											D	
	S											675

#### 4.15 CALCULO DE LA MODA PARA DATOS AGRUPADOS

La moda de un conjunto de datos se definió, como el valor que se repite con mayor frecuencia. Cuando se designa la moda de datos agrupados, se refiere por lo general a la clase modal, donde la clase modal esta

localizada en el intervalo de clase con la frecuencia más alta.

NUMERO DE INTERVALO	INTERVALO DE CLASE	fi	Xi
1	57 - 59	2	58
2	60 - 62	3	61
3	63 - 65	5	64
4	66 - 68	8	67
5	69 - 71	6	70
6	72 - 74	4	73
7	75 - 77	2	76

En nuestro ejemplo la clase modal se encuentra en la cuarta clase 66-68, o bien 65.5-68.5, utilizando los límites reales de clase.

En algunos casos es suficiente decir que la moda está correspondida entre 65.5-68.5 o basta con tomar como moda el punto medio de clase, en este caso 67.

Pero hay ocasiones en que esto no es suficiente y hay que fijar un valor más exacto, por lo que se procede a utilizar la siguiente fórmula:

$$\text{MODA} = L_1 + \frac{D_1}{D_1 + D_2} \cdot C$$

$L_1$  = Límite real inferior de clase, donde se encuentra la moda.

$D_1$  = Diferencia entre la frecuencia modal y la frecuencia del intervalo de clase contiguo inferior.

D<sub>2</sub>= Diferencia entre la frecuencia modal y la frecuencia del intervalo de clase contiguo superior.

C= Tamaño del intervalo de clase donde se encuentra la moda.

En el ejemplo

$$L_1 = (65+66)/2=65.5$$

$$D_1 = 8-5=3$$

$$D_2 = 8-6=2$$

$$C = 68.5 - 65.5 = 3$$

$$\text{MODA} = 65.5 + \frac{3}{3+2} \cdot 3$$

$$\text{MODA} = 65.5 + 1.8$$

$$\text{MODA} = 67.30 \text{ Kgs.}$$

#### 4.16 CALCULO DE LA MODA PARA DATOS AGRUPADOS USANDO CALCULADORA

Como la fórmula anterior no esta incorporada en los 128 fórmulas programadas de la calculadora, procederemos a almacenarla en las 12 memorias independientes, siguiendo los pasos indicados para la formula de la mediana.

1. PRESIONE MODE 2 PARA CAMBIAR AL WRT
2. USE LA TECLA → PARA SELECCIONAR PI NOMBRE DE MEMORIA DE LA FORMULA
3. ESPECIFIQUE UN MODO DE CALCULO PARA LA MEMORIA DE FORMULAS SELECCIONADAS EN ESTE CASO MODE X PARA SD DE

ESTADISTICA

4. PRESIONE LA TECLA EXE

5. INGRESE LA FORMULA

$$D = L + (S / (S+T)) C$$

$$L_1 = L$$

$$D_1 = S$$

$$D_2 = T$$

$$C = C$$

AC ALPHA D SHIFT= ALPHA L+ (ALPHA S/(ALPHA S+ALPHA T))

ALPHA C

6. PRESIONE MODE 1 PARA ALMACENAR EL PROGRAMA.

PARA EJECUTAR LA FORMULA DE LA MODA, ALMACENADA EN LAS

FORMULAS DEL USUARIO SE PROCEDE DE LA SIGUIENTE FORMA:

1. MODE 1

2. PROG. 1 EXE EN LA PANTALLA

D=L + (S/ (S+T)) C
SD D
L? 0

65.5

65.5_ SD D
L? 0

EXE

D=L + (S/ (S+T)) C
SD D
S? 0

3

3_ SD D
S? 0

EXE

D=L + (S/ (S+T)) C
SD D
T? 0

2

2		
-	SD	D
T?		O

EXE

$D=L + (S/(S+T)) C$	C
SD	D
C?	O

3

3		
-	SD	D
C?		O

EXE

$D=L + (S/(S+T)) C$	C
SD	D
D=	67.3

MODA= 67.30

#### 4.17 CUARTILES, DECILES Y PERCENTILES (CUANTILES)

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 el conjunto de datos en dos partes iguales es la mediana. Por extensión de esta idea se puede pensar que aquellos valores que dividen a los datos en cuatro partes iguales, se llaman CUARTILES y están representados por  $Q_1$ ,  $Q_2$  y  $Q_3$ , primero, segundo y tercer cuartil, respectivamente.

De esta manera, los valores que dividen a los datos en 10 partes iguales se llaman DECILES y se representan por  $D_1, D_2, \dots, D_9$ , mientras que los valores que dividen los datos en cien partes iguales se llaman PERCENTILES y se representan por  $P_1, P_2, P_3, \dots, P_{99}$ .

El segundo cuartil  $Q_2$ , el quinto decil  $D_5$  y el quincuagésimo percentil  $P_{50}$  se corresponden con el valor de la mediana.

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

#### CALCULOS DE LOS CUANTILES PARA DATOS NO AGRUPADOS.

El cálculo de cuartiles, deciles y percentiles para datos no agrupados no representa ninguna dificultad, por lo que únicamente daremos un ejemplo utilizando los cuartiles.

Ejemplo: Supongamos que la edad en años cumplidos de 16 niños es la siguiente:

8, 9, 5, 13, 6, 14, 5, 8, 5, 10, 12, 11, 7, 11, 14 y 5

El primer paso es ordenar los datos, generalmente en orden creciente.

5, 5, 5, 5, 6, 7, 8, 8, 9, 10, 11, 11, 12, 13, 14, 14

En seguida dividimos al conjunto de datos en 4 partes iguales y así obtenemos los cuartiles

5, 5, 5, 5    6, 7, 8, 8    9, 10, 11, 11,    12, 13, 14  
                   ∨                   ∨                   ∨  
                   5.5                   8.5                   11.5

$Q_1 = 5.5$  ,  $Q_2 = 8.5$  y  $Q_3 = 11.5$

Que se pueden interpretar de la siguiente manera:

Una cuarta parte (25%) de los niños tienen edades menores que 5.5 años, otra cuarta parte (25%) de 5.5 a menos de 8.5 años, otra cuarta parte edades de 8.5 a menos de 11.5 años, y la otra cuarta parte, de 11.5 años a 14 años.

#### 4.18 CALCULO DE LOS CUANTILES PARA DATOS AGRUPADOS.

NUMERO DE INTERVALO	INTERVALO DE CLASE	FRECUENCIA $f_i$
1	57 - 59	2
2	60 - 62	3
3	63 - 65	5
4	66 - 68	8
5	69 - 71	6
6	72 - 74	4
7	75 - 77	2
		30

Para el cálculo de los cuartiles, deciles y percentiles se utiliza la fórmula de la mediana:

$$\text{Mediana} = L_1 + \left( \frac{n/2 - f_a}{f_m} \right) C$$

$$Q_1 = L_1 + \left( \frac{1/4 n - f_a}{f_{Q_1}} \right) C$$

$L_1$  = Limite real inferior del intervalo de clase donde se encuentra el  $Q_1$

$f_a$  = Frecuencia acumulada del intervalo de clase anterior al intervalo de clase que contiene el  $Q$

$f_{Q_1}$  = Frecuencia del intervalo de clase donde se encuentra el  $Q_1$

$C$  = Tamaño del intervalo de clase donde se encuentra el  $Q_1$

En nuestro ejemplo:

$$1/4 n = 1/4(30) = 7.5$$

El  $Q_1$  se encuentra en el dato 7.5. El dato 7.5 se encuentra en el tercer intervalo de clase.

$$L_1 = (62 + 63)/2 = 62.56$$

$$f_a = 2 + 3 = 5$$

$$f_{Q_1} = 5$$

$$C = 65.5 - 62.5 = 3$$

$$Q_1 = 62.5 + \left( \frac{7.5 - 5}{5} \right) 3$$

$$Q_1 = 62.5 + 1.5 = 64$$

PARA CALCULAR  $Q_3$  :

$$Q_3 = L_1 + \left( \frac{3/4n - f_a}{F_{Q_3}} \right) C$$

$L_i$  = Limite real inferior de la clase que contiene el  $Q_3$

$f_a$  = Frecuencia acumulada del intervalo de clase anterior al intervalo de clase donde se encuentra  $Q_3$  .

$f_o$  = Frecuencia del intervalo de clase que contiene a  $Q_3$  .

$C$  = Tamaño del intervalo de clase que contiene a  $Q_3$  .

POR LO QUE

$$3/4n = 3/4 (30) = 22.5$$

$$L_i = 68.5$$

$$f_a = 18$$

$$f_o = 6$$

$$C = 3$$

$$Q_3 = 68.5 + \left( \frac{22.5 - 18}{6} \right) 3$$

$$Q_3 = 68.5 + 2.25 = 70.75$$

$$Q_3 = 70.75$$

$L_1$  = Limite real inferior de la clase que contiene el  $Q_3$

$f_a$  = Frecuencia acumulada del intervalo de clase anterior al intervalo de clase donde se encuentra  $Q_3$  .

$f_{Q_3}$  = Frecuencia del intervalo de clase que contiene a  $Q_3$  .

$C$  = Tamaño del intervalo de clase que contiene a  $Q_3$  .

POR LO QUE

$$3/4n = 3/4 (30) = 22.5$$

$$L_1 = 68.5$$

$$f_a = 18$$

$$f_{Q_3} = 6$$

$$C = 3$$

$$Q_3 = 68.5 + \left( \frac{22.5 - 18}{6} \right) 3$$

$$Q_3 = 68.5 + 2.25 = 70.75$$

$$Q_3 = 70.75$$

PARA CALCULAR  $P_{10}$

$$P_{10} = L_1 + \left( \frac{10/100 (n) - f_a}{f_{R_0}} \right) C$$

$$10/100 (n) = 1/10 (n) = 1/10 (30) = 3$$

$$L_1 = 59.5$$

$$f_a = 2$$

$$f_P = 3$$

10

$$C = 3$$

$$P_{10} = 59.5 + \left( \frac{3 - 2}{3} \right) 3$$

$$P_{10} = 59.5 + 1 = 60.5$$

$$P_{10} = 60.5$$

PARA CALCULAR  $P_{90}$

$$P_{90} = L_1 + \left( \frac{90/100 (n) - f_a}{f_P} \right) C$$

90

$$90/100 (n) = 90/100 (30) = 2700/100 = 27$$

$$L_1 = 71.5$$

$$f_a = 24$$

$$f_P = 4$$

90

$$C = 3$$

$$P_{90} = 71.5 + \left( \frac{27 - 24}{4} \right) 3$$

$$P_{90} = 71.5 + 2.25 = 73.75$$

$$P_{90} = 73.75$$

4.19 CALCULO DE LOS CUANTILES PARA DATOS AGRUPADOS CON  
CALCULADORA.

PROCEDEREMOS A INCORPORAR LA FORMULA EN LA CALCULADORA  
PARA LOS CUANTILES.

1. PRESIONE MODE 3
2. USAR LA TECLA → PARA SELECCIONAR P2
3. ESPECIFICAR MODE X PARA SD DE ESTADISTICA
4. PRESIONE LA TECLA EXE

5. INGRESE LA FORMULA  $X = L + \left( \frac{MN/P - A}{B} \right) C$

AC ALPHA X SHIFT = ALPHA L + ((ALPHA M ALPHA N/ALPHA P -  
ALPHA A)/ALPHA B)×ALPHA C

6. PRESIONE MODE 1 PARA ALMACENAR EL PROGRAMA.

De los ejemplos anteriores calcular  $Q_1$ ,  $Q_3$ ,  $P_{10}$  y  $P_{90}$

CALCULO DE  $Q_1$

EN LA CALCULADORA  $X = L + ((MN/P-A)/B)C$

1. MODE 1
2. PROG. 2 EXE

$$L_1 = L = 62.5$$

$$M = 1$$

$$N = 30$$

$$P = 4$$

$$A = 5$$

B = 5

C = 3

L ?	62.5	EXE
M ?	1	EXE
N ?	30	EXE
P ?	4	EXE
A ?	5	EXE
B ?	5	EXE
C ?	3	EXE

X= 64    o    Q<sub>1</sub> = 64

CALCULO DE Q<sub>3</sub>

1. MODE 1

2. PROG 2 EXE

L ?	68.5	EXE
M ?	3	EXE
N ?	30	EXE
P ?	4	EXE
A ?	18	EXE
B ?	6	EXE
C ?	3	EXE

X=70.75    o    Q<sub>3</sub> = 70.75

CALCULO DE P<sub>10</sub>

1. MODE 1

2. PROG.    2    EXE

L ? 59.5 EXE

M ? 10 EXE

N ? 30 EXE

P ? 100 EXE

A ? 2 EXE

B ? 3 EXE

C ? 3 EXE

X= 60.5    O    P<sub>10</sub>=60.5

CALCULO DE P<sub>90</sub>

1. MODE 1

2. PROG 2 EXE

L ? 71.5 EXE

M ? 90 EXE

N ? 30 EXE

P ? 100 EXE

A ? 24 EXE

B ? 4 EXE

C ? 3 EXE

X= 73.75    O    P<sub>90</sub>= 73.75

#### 4.20 EJERCICIOS DE REPASO

1. La siguiente tabla muestra la distribución de edades de casos de cierta enfermedad reportada durante un año en un estado particular.

EDAD	NUMERO DE CASOS
5-14	5
15-24	10
25-34	20
35-44	22
45-54	13
55-64	<u>5</u>
	75

EN BASE A ESOS DATOS CALCULAR:

- a) LA MEDIA ARITMETICA
- b) LA MEDIANA
- c) LA MODA
- d) EL CUARTIL UNO  $Q_1$
- e) EL CUARTIL TRES  $Q_3$
- f) EL PERCENTIL DIEZ  $P_{10}$
- g) EL PERCENTIL NOVENTA  $P_{90}$
- h) COMPROBAR LOS RESULTADOS UTILIZANDO LA CALCULADORA.

2. La siguiente tabla muestra una distribución de frecuencias de la duración en horas de 400 focos con diferentes características, que fabrica una compañía.

DURACION (HORAS)	NUMERO DE FOCOS
300-399	14
400-499	46
500-599	58
600-699	76
700-799	68
800-899	62
900-999	48
1000-1099	22
1100-1199	6
	400

CON REFERENCIA A ESTA TABLA CALCULAR:

- a) LA MEDIA ARITMETICA.
- b) LA MEDIANA.
- c) LA MODA.
- d) EL  $Q_1$
- e) EL  $Q_3$
- f) EL  $P_{10}$
- g) EL  $P_{90}$
- h) COMPROBAR CON LA CALCULADORA LOS RESULTADOS

3. Las ganancias de 200 corredores de bolsa estan agrupados en una distribución de frecuencias, como se muestra a continuación.

GANANCIAS DIARIAS	NUMERO DE CORREDORES DE BOLSA
\$ 77.5 - 82.5	5
82.5 - 87.5	12
87.5 - 92.5	13
92.5 - 97.5	22
97.5 - 102.5	30
102.5 - 107.5	35
107.5 - 112.5	32
112.5 - 117.5	20
117.5 - 122.5	15
122.5 - 127.5	10
127.5 - 137.5	6
	<hr/>
	200

**CALCULAR:**

- a) LA MEDIA ARITMÉTICA
- b) LA MEDIANA
- c) LA MODA
- d)  $Q_1$
- e)  $Q_3$
- f)  $P_{10}$
- g)  $P_{90}$
- h) COMPROBAR LOS RESULTADOS CON LA CALCULADORA.

## **C A P I T U L O 5**

### **MEDIDAS DE DISPERSION, SESGO Y CURTOSIS**

- 5.1 MEDIDAS DE DISPERSION
- 5.2 RANGO O RECORRIDO
- 5.3 LA VARIANZA Y LA DESVIACION ESTANDAR O TIPICA Y SU CALCULO PARA DATOS NO AGRUPADOS.
- 5.4 CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS NO AGRUPADOS, POR LA FORMULA SIMPLIFICADA
- 5.5 CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS NO AGRUPADOS CON CALCULADORA.
- 5.6 EL COEFICIENTE DE VARIACION.
- 5.7 EJERCICIOS DE REPASO.
- 5.8 CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS AGRUPADOS.
- 5.9 CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS AGRUPADOS POR LA FORMULA SIMPLIFICADA.
- 5.10 CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS AGRUPADOS CON CALCULADORA.
- 5.11 INTERPRETACION DE LA DESVIACION ESTANDAR. LA REGLA EMPIRICA.
- 5.12 SESGO.
- 5.13 CALCULO DEL COEFICIENTE DE SESGO CON CALCULADORA.
- 5.14 CURTOSIS, COEFICIENTE DE CURTOSIS PERCENTILICO.

5.15 CALCULO DEL COEFICIENTE DE CURTOSIS PERCENTILICO  
CON CALCULADORA.

5.16 EJERCICIOS DE REPASO

## 5.1 MEDIDAS DE DISPERSION

Al grado en que los datos numéricos tienden a extenderse alrededor de un valor medio se le llama variación o dispersión de los datos. Si todos los datos son iguales, no hay dispersión, si son diferentes hay dispersión en los datos. La magnitud de la dispersión puede ser pequeña, cuando los datos, aunque distintos, están próximos entre sí. Si los datos están ampliamente desparramados, la dispersión es mayor.

Existen distintas medidas de dispersión o variación, las más empleadas son:

- a) EL RANGO O RECORRIDO
- b) LA VARIANZA
- c) LA DESVIACION ESTANDAR O TIPICA

## 5.2 RANGO O RECORRIDO

Es la medida de dispersión más sencilla, se determina mediante la diferencia entre el dato mayor y el dato menor de la muestra.

Si se denota el rango por  $R$ , el dato mayor por  $X_A$  y el menor por  $X_B$ , el rango se calcula como sigue:

$$R = X_B - X_A$$

Ejemplo 1. Calcular el rango de la siguiente distribución:

10,12,12,13,15,15,16,17,19,21,23

Rango=23-10=13 por lo que  $R=13$

Ejemplo 2. Obtener el rango para el caso que se estudia, relativo al peso de los estudiantes.

Rango=77-58=19 por lo que  $R=19$

Este valor ya se había obtenido con otros fines; sin embargo, ahora se obtiene como un índice de la variabilidad que indica la amplitud en la escala de los pesos.

La utilidad del rango es limitada. El hecho de que sólo tome en cuenta dos valores, hace que sea una medida pobre en la dispersión. La ventaja principal de utilizar el recorrido es la sencillez y rapidez de su cálculo.

También hay que hacer notar que el rango no es una medida de dispersión de los datos intermedios respecto al valor típico.

### 5.3 LA VARIANZA Y LA DESVIACION ESTANDAR O TIPICA Y SU CALCULO PARA DATOS NO AGRUPADOS

Algunas veces la media aritmética y el rango no permiten diferenciar la forma en que están distribuidos los datos de un conjunto, por lo que es necesario otra forma de evaluación de la distribución, que nos permita saber cuánto varían los datos entre sí. Un ejemplo nos aclarará esto.

Ejemplo. En una escuela secundaria hay cuatro grupos de 2º año. De cada grupo se escogen, por sorteo, cinco alumnos a los que se aplica una prueba de matemáticas. Las calificaciones obtenidas por los cinco alumnos de cada grupo son las siguientes:

Grupo A: 2,3,6,9,10  $x=6$

Grupo B: 6,6,6,6, 6  $x=6$

Grupo C: 4,6,6,6, 8  $x=6$

Grupo D: 4,4,6,8, 8  $x=6$

Cada uno de estos grupos obtuvo una media de 6, sin embargo es claro que en unos grupos las calificaciones varían entre sí más que en otros. En el grupo A, las calificaciones varían desde 2 hasta 10 inclusive, es decir son heterogéneas.

En cambio, las calificaciones del grupo B son totalmente homogéneas, todas son iguales. En cuanto a los grupos C y D, se puede ver que las calificaciones en el C fueron homogéneas, aunque en menor grado que en el grupo B, y

que las calificaciones en el D fueron heterogéneas, aunque en menor grado que en el grupo A.

Esta información, que obtuvimos al observar directamente las calificaciones, nos fue proporcionada por la media. Sin embargo para evaluar la dispersión o variabilidad de los datos procederemos a calcular el RANGO O RECORRIDO.

El rango de las calificaciones de cada uno de los grupos es:

GRUPO A:	$10-2=8$	R=8
GRUPO B:	$6-6=0$	R=0
GRUPO C:	$8-4=4$	R=4
GRUPO D:	$8-4=4$	R=4

Observamos que el valor más grande, 8 corresponde al grupo con mayor dispersión; el más pequeño, 0 al grupo con menor dispersión, y un valor intermedio corresponde a una dispersión intermedia. Sin embargo, aunque se ve que en C hay menor dispersión que en D, el valor del rango es el mismo para ambos. Podemos entonces observar que, aunque el rango se calculó fácilmente, en algunas ocasiones no permite diferenciar la forma en que se distribuyen los datos en cada grupo. Por lo que volvemos a insistir como se había dicho anteriormente, que es necesaria otra forma de evaluación de la dispersión que nos permita distinguir estas diferencias.

De manera natural podemos pensar en medir la dispersión tomando como punto de referencia a la media aritmética. Esto se puede realizar calculando la dispersión a partir de las diferencias entre cada dato y la media simbolizando esto como  $X_i - \bar{x}$ . Es claro que la desviación de los valores  $X_i$  menores que  $\bar{x}$  resulta negativa, y que la desviación de los valores  $X_i$  mayores que  $\bar{x}$  es positiva.

Entonces el siguiente paso es preguntarse si se puede calcular un promedio de desviaciones sumando todas las desviaciones y dividiendo esta suma entre el número total de desviaciones, es decir;

$$\text{promedio de desviaciones} = \frac{\sum(X_i - \bar{x})}{n}$$

Veamos esto en los ejemplos de las calificaciones de los alumnos de los grupos A, B, C y D.

Recuerde que todos los grupos tienen media 6, es decir

$$\bar{X}_A = \bar{X}_B = \bar{X}_C = \bar{X}_D = 6$$

PARA EL GRUPO A CONSIDERANDO LAS FRECUENCIAS:

$X_i$	$f_i$	$(X_i - \bar{x})$	$f_i(X_i - \bar{x})$
2	1	2-6=-4	-4
3	1	3-6=-3	-3
6	1	6-6=0	0
9	1	9-6=3	3
10	1	10-6=4	4
	5		0

Como se observa la suma de las desviaciones respecto de su media es cero, esto es se cumple la propiedad de la media aritmética. Por lo que también el promedio de las desviaciones siempre es cero.

$$\text{PROMEDIO DE DESVIACIONES} = \frac{\sum f_i(X_i - \bar{x})}{n} = \frac{0}{5} = 0$$

$$\text{PROMEDIO DE DESVIACIONES} = 0$$

Calcule de la misma manera el promedio de desviaciones para los grupos B, C y D.

Como se observó en los ejemplos, la suma de las desviaciones siempre es cero, por lo que el promedio de las desviaciones también es cero. Para evitar esto podemos elevar al cuadrado cada desviación, y después sumar los resultados obtenidos, que serán en todos los casos positivos o igual a cero.

Al dividir la suma de los cuadrados de las desviaciones entre  $n-1$ , donde  $n$  es el número de datos de la muestra, obtendremos una medida de dispersión llamada varianza muestral que denotaremos con  $S^2$

LA VARIANZA de una muestra de  $n$  valores  $X_1, X_2, \dots, X_n$ , se define como la suma de los cuadrados de las desviaciones de los datos con respecto a su media  $\bar{x}$  y dividida entre

( $n-1$ ). Y ésta dada por la fórmula: 
$$S^2 = \frac{\sum(X_i - \bar{x})^2}{n-1}$$

También se puede definir como el promedio o media de los cuadrados de las desviaciones de las medidas respecto de su media aritmética.

Considerando lo anterior calculemos la varianza muestral de los datos del grupo A.

VARIANZA DEL GRUPO A:

$X_i$	$f_i$	$(X_i - \bar{x})$	$(X_i - \bar{x})^2$	$f_i(X_i - \bar{x})^2$
2	1	-4	16	16
3	1	-3	9	9
6	1	0	0	0
9	1	3	9	9
10	1	4	16	16
	5			50

$$S^2 = \frac{\sum f_i(X_i - \bar{x})^2}{n-1} = \frac{50}{5-1} = \frac{50}{4} = 12.5 \quad \therefore S^2 = 12.5$$

Calcule la varianza muestral de los datos de los grupos B, C y D, y después compare estas varianzas entre sí y con la varianza del grupo A.

Al comparar las varianzas muestrales se debe observar que la varianza del grupo A es mayor que la del grupo D y ésta es mayor que la del grupo C, que a su vez es mayor que la del grupo B, es decir:

$$s_A^2 > s_D^2 > s_C^2 > s_B^2$$

De esta observación se concluye que, mientras mayor sea la dispersión de los datos de una distribución, mayor será su varianza, por lo que se deduce que la varianza muestral es una buena medida de la dispersión de los datos de una muestra.

Algunas veces cuando se dispone de todos los valores de la población y se requiere calcular la varianza se procede de la siguiente manera.

**DEFINICION:** La varianza de una población de  $N$  valores  $X_1, X_2, X_3, \dots, X_n$ , se define como el promedio del cuadrado de las desviaciones de los valores con respecto a su media  $\mu$ . La varianza de la población se denota por  $\sigma^2$  ( $\sigma$

es la letra griega sigma minúscula) y esta dada por la fórmula:

$$\sigma^2 = \frac{\sum(X_i - \mu)^2}{N}$$

donde  $\mu$  y  $N$  son, respectivamente, la media y el tamaño de la población.

Si observamos las fórmulas para calcular la varianza de la muestra y la varianza de la población, notamos que al calcular  $S^2$ , se dividió entre  $n-1$ , y al calcular  $s^2$  se dividió entre  $N$ , más adelante daremos una justificación formal de este hecho, por el momento podemos afirmar que  $S^2$ , obtenido en la forma indicada, es un buen estimador del parámetro  $\sigma^2$ , es decir que  $S^2$  nos permite efectuar inferencias acerca de  $\sigma^2$  con mayor precisión.

También se puede apreciar que el denominador  $(n-1)$  es prácticamente igual a  $N$  cuando el tamaño de la muestra es muy grande, cosa que sucede generalmente en problemas prácticos.

Conviene hacer la observación de que la varianza algunas veces está representada por unidades cuadradas y, por lo tanto, no es una medida de dispersión apropiada cuando se desea expresar este concepto en términos de las

unidades originales. Así, si las medidas originales fueron en centímetros como en el caso de la estatura de un grupo de personas, la varianza estará representada en centímetros al cuadrado. Por lo que para obtener una medida de dispersión en las unidades originales, simplemente se toma la raíz cuadrada positiva de la varianza obteniendo como resultado lo que se conoce como desviación estándar.

DEFINICION: La desviación estándar de un conjunto de datos, es la raíz cuadrada positiva del promedio de las desviaciones con respecto a la media elevadas al cuadrado y su fórmula es:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(X_i - \bar{x})^2}{n-1}}$$

De la muestra si  $X_1, X_2, X_3, \dots, X_n$  son los  $n$  valores de una muestra.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(X_i - \bar{x})^2}{N}}$$

De la población si  $X_1, X_2, X_3, \dots, X_n$  son los  $N$  valores de una población.

Ejemplo 1. Si tomamos el resultado de la varianza de las calificaciones del grupo A que vimos anteriormente  $S_A^2=12.5$ . Observamos que la varianza quedo expresada en calificaciones al cuadrado, unidades que no tienen sentido en la práctica, por lo que para medir la dispersión con las mismas unidades con que se midió la variable, debemos calcular la desviación estándar:

$$S = \sqrt{\frac{\sum(X_i - \bar{x})^2}{n-1}} = \sqrt{12.5} = 3.53 \quad \therefore \quad S=3.53$$

Ejemplo 2. Encontrar la varianza y la desviación estándar del siguiente conjunto de datos:

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

Como primer paso se procede a calcular su media aritmética.

$X_i$	$f_i$	$f_i X_i$
4	1	4
5	1	5
6	3	18
7	1	7
8	3	24
9	1	9
10	1	10
	11	77

$$\bar{X} = \frac{\sum f_i X_i}{n} = \frac{77}{11} = 7 \quad \therefore \quad \bar{x} = 7$$

Como segundo paso calculamos la varianza:

$X_i$	$f_i$	$(X_i - \bar{x})$	$(X_i - \bar{x})^2$	$f_i(X_i - \bar{x})^2$
4	1	-3	9	9
5	1	-2	4	4
6	3	-1	1	3
7	1	0	0	0
8	3	1	1	3
9	1	2	4	4
10	1	3	9	9
	11			32

$$S^2 = \frac{\sum f_i(X_i - \bar{x})^2}{n-1} = \frac{32}{10} = 3.2 \quad \therefore S = 1.788$$

Como tercer paso calcular la desviación estándar.

$$S = \sqrt{\frac{\sum f_i(X_i - \bar{x})^2}{n-1}} = \sqrt{3.2} = 1.788 \quad \therefore S = 1.788$$

#### 5.4 CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS NO AGRUPADOS POR LA FORMULA SIMPLIFICADA.

El procedimiento anterior para calcular la varianza y la desviación estándar es bastante laborioso debido a que la media debe restarse de cada valor de la muestra. Resulta especialmente tedioso si la muestra es grande o si la media tiene cifras decimales.

Para un cálculo más eficiente, puede utilizarse la siguiente fórmula:

$$s^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1}$$

UTILIZANDO LAS FRECUENCIAS

$$s^2 = \frac{\sum f_i X_i^2 - \frac{(\sum f_i X_i)^2}{n}}{n-1}$$

Esta fórmula se deduce de:

$$s^2 = \frac{\sum f_i (X_i - \bar{x})^2}{n-1} \quad \text{varianza muestral}$$

Desarrollando  $(X_i - \bar{x})^2 = X_i^2 - 2X_i\bar{x} + \bar{x}^2$ , aplicando las propiedades del símbolo  $\sum$  y al mismo tiempo simplificando se obtiene que:

$$s^2 = \frac{\sum f_i (X_i - \bar{x})^2}{n-1} = \frac{\sum f_i X_i^2 - \frac{(\sum f_i X_i)^2}{n}}{n-1}$$

Ejemplo. Calcular la varianza y desviación estándar del problema anterior:

$X_i$	$f_i$	$f_i X_i$	$X_i^2$	$f_i X_i^2$
4	1	4	16	16
5	1	5	25	25
6	3	18	36	108
7	1	7	49	49
8	3	24	64	192
9	1	9	81	81
10	1	10	100	100
	11	77		571

$$s^2 = \frac{\sum fX_i^2 - \frac{(\sum fX_i)^2}{n}}{n-1} = \frac{571 - \frac{(77)^2}{11}}{11-1} = \frac{571 - \frac{5929}{11}}{10}$$

$$s^2 = \frac{571 - 539}{10} = \frac{32}{10} = 3.2 \quad S^2 = 3.2 \text{ VARIANZA}$$

$$s = \sqrt{\frac{\sum fX_i^2 - \frac{(\sum fX_i)^2}{n}}{n-1}} = \sqrt{3.2} = 1.788 \quad S = 1.788 \text{ DESVIACION ESTANDAR}$$

### 5.5 CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS NO AGRUPADOS CON CALCULADORA

Para calcular la varianza y la desviación estándar con la calculadora se siguen los mismos pasos que utilizamos, cuando calculamos la media aritmética.

Ejemplo: Calcular la varianza y la desviación estándar del problema anterior usando la calculadora.

1. Presionar mode X para el modo SD de estadística.  
Aparece SD en la pantalla.
2. PRESIONAR SHIFT SCI EXE PARA BORRAR LAS MEMORIAS ESTADISTICAS.
3. INGRESAR LOS ELEMENTOS DE DATOS UNO POR UNO PRESIONANDO LA TECLA DT, LUEGO DE CADA INGRESO

1. MODE X → SD
2. SHIFT SCL EXE
3. INGRESO
4. DT
5. DT
6. SHIFT [;] 3
7. DT
8. SHIFT [;] 3
9. DT
10. DT

OPERACION DE TECLA		RESULTADO	SIMBOLO
SHIFT	EXE	DESVIACION ESTANDAR MUESTRAL	$S=1.778$
SHIFT	EXE	DESVIACION ESTANDAR POBLACIONAL	$\sigma=1.705$
SHIFT	x EXE	MEDIA ARITMETICA	$\bar{X}=7$
SHIFT	1k1 EXE	SUMA DE DATOS AL CUADRADO	$\sum f_i x_i^2 = 571$
SHIFT	1k2 EXE	SUMA DE DATOS	$\sum f_i x_i = 77$
SHIFT	1k3 EXE	NUMERO DE DATOS	$n=11$

**NOTA:**

Por último para calcular la varianza y la desviación estándar de una población de datos, se siguen los pasos utilizados anteriormente, recordando que la división es entre N, en lugar de n-1.

## 5.6 EL COEFICIENTE DE VARIACION

La desviación estándar es útil como una medida de variación dentro de un determinado conjunto de datos. Sin embargo, cuando se desea comparar la dispersión en dos conjuntos de datos, el comparar las dos desviaciones estándar puede conducir a resultados ilógicos. Puede ser que las dos variables que intervienen se midan en unidades distintas. Además, aún cuando se utilice la misma unidad de medición, las dos medias pueden ser bastante distintas. Por ejemplo, si se compara la desviación estándar de los pesos de los niños de primer año de primaria con la desviación estándar de los pesos de jóvenes de primer año de secundaria, puede encontrarse que la desviación estándar de estos últimos es numéricamente mayor que la de los primeros debido a que los propios pesos son mayores y no porque la dispersión sea mayor.

Lo que necesitamos en casos como este es una medida de variación relativa, más que una variación absoluta. Dicha medida se encuentra en el coeficiente de variación, que expresa la desviación estándar como un porcentaje de la media y se obtiene dividiendo la desviación estándar entre la media y multiplicando el cociente por 100 para expresarlo como porcentaje. Es decir:

Si  $\bar{x}$  y S son la media y la desviación estándar muestrales de un conjunto de datos, entonces el coeficiente de variación de estos datos es:

$$C.V. = \frac{S}{\bar{x}}(100)$$

Ejemplo: Supóngase que dos muestras de personas del sexo masculino proporcionan los resultados siguientes:

	MUESTRA 1	MUESTRA 2
EDAD	25 AÑOS	11 AÑOS
PESO MEDIO	72.5 KGS.	40 KGS.
DESVIACION ESTANDAR	5 KGS.	5 KGS.

Al comparar las desviaciones estándar se podría concluir que las dos muestras poseen igual variabilidad. Sin embargo, si calculamos los coeficientes de variación de las dos muestras obtenemos:

Coefficiente de variación para los individuos de 25 años

$$C.V. = \frac{5}{72.5}(100) = 6.9 \quad C.V. = 6.9\%$$

Coefficiente de variación para los de 11 años.

$$C.V. = \frac{5}{40}(100) = 12.5 \quad C.V. = 12.5\%$$

Observe que el coeficiente de variación carece de unidades, y que al comparar los dos coeficientes de

variación resulta que el peso de los niños de 11 años, tiene una dispersión, en relación al valor de su media, casi dos veces mayor que el de los pesos de los jóvenes de 25 años.

#### 5.7 EJERCICIOS DE REPASO:

1. Dado el siguiente conjunto de datos:

2, 2, 3, 3, 3, 4, 4, 5, 5, 9, calcular:

- a) La media, la mediana y la moda.
- b) El rango o recorrido.
- c) La varianza y la desviación estándar, utilizando la fórmula no simplificada y la fórmula simplificada.
- d) El coeficiente de variación.
- e) Comprobar los resultados de  $\bar{x}$ ,  $S^2$  y  $S$  con la calculadora.

2. Se seleccionan aleatoriamente 11 cerdos y se obtienen sus pesos al nacer como lo siguiente:

14, 26, 25, 25, 18, 18, 23, 21, 20, 24, 16, obtener:

- a) La media, mediana y moda (nota: ordenar los datos)
- b) El rango o recorrido.
- c) La varianza y desviación estándar por la fórmula no simplificada y por la fórmula simplificada.
- d) El coeficiente de variación.
- e) Comprobar los resultados con la calculadora.

3. De los siguientes grupos de datos, decir cuál presenta más dispersión y cuál presenta menos dispersión (calcular las desviaciones estándar):

- a) 1,3,5,7,9,11
- b) 1,2,3,9,10,11
- c) 1,1,2,10,11,11

4. Los promedios de puntos de calificación de 40 estudiantes se han organizado en una distribución de frecuencias.

Utilícense  $\bar{x}=3.0$  para calcular la varianza y la desviación estándar por cualquiera de los dos métodos y comprobarlos con la calculadora.

PROMEDIO DE PUNTOS DE CALIFICACION	NUMERO DE ESTUDIANTES
2.0	2
2.3	3
2.6	7
2.9	11
3.2	7
3.5	7
3.8	3
	40

5. Los salarios semanales obtenidos por 50 estudiantes universitarios se han organizado en una distribución de frecuencias de la siguiente forma:

$x_i$   $f_i$   
 55     2

60	3
65	4
70	5
75	8
80	10
85	8
90	6
95	<u>4</u>
	50

En base a esos datos calcular:

- a) La media aritmética, mediana y moda.
- b) La varianza y la desviación estándar por cualquier método.
- c) El coeficiente de variación.
- d) Comprobar estos resultados con la calculadora.

#### 5.8 CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS AGRUPADOS.

Para calcular la varianza y la desviación estándar a partir de datos agrupados, se supone que todos los valores que caen dentro de un determinado intervalo de clase se localizan en el punto medio del intervalo, tal y como se planteó al calcular la media. Entonces la

varianza de una muestra está representada por la fórmula:

$$s^2 = \frac{\sum f_i(X_i - \bar{x})^2}{n-1}$$

Donde  $X_i$  representa el punto medio de cada intervalo de clase.

Tomemos como ejemplo el peso del grupo de los estudiantes de estadística del C.C.H. recordando que obtuvieron como media  $\bar{x}=67.3$

No. DE INTERVALO	INTERVALO DE CLASE	$f_i$	$X_i$	$(X_i - \bar{x})$	$(X_i - \bar{x})^2$	$f_i(X_i - \bar{x})^2$
1	56.5 - 59.5	2	58	-9.3	86.49	172.98
2	59.5 - 62.5	3	61	-6.3	39.69	119.07
3	62.5 - 65.5	5	64	-3.3	10.89	54.45
4	65.5 - 68.5	8	67	-0.3	0.09	0.72
5	68.5 - 71.5	6	70	2.7	7.29	43.74
6	71.5 - 74.5	4	73	5.7	32.49	129.96
7	74.5 - 77.5	2	76	8.7	75.69	151.38
		30				672.30

$$s^2 = \frac{\sum f_i(X_i - \bar{x})^2}{n-1} = \frac{672.3}{29} = 23.18 \quad S^2 = 23.18 \text{ (varianza)}$$

$$S = \sqrt{23.18} = 4.81 \quad S = 4.81 \text{ (desviacion estandar)}$$

#### 5.9 CALCULO DE LA VARIANZA Y DESVIACION ESTANDAR PARA DATOS AGRUPADOS POR LA FORMULA SIMPLIFICADA

Aplicando la fórmula simplificada en el problema anterior:

$$s^2 = \frac{\sum f_i x_i^2 - \frac{(\sum f_i x_i)^2}{n}}{n-1}$$

NUM. DE INTERVALO	INTERVALO DE CLASE	f <sub>i</sub>	x <sub>i</sub>	f <sub>i</sub> x <sub>i</sub>	x <sub>i</sub> <sup>2</sup>	f <sub>i</sub> x <sub>i</sub> <sup>2</sup>
1	56.5 - 59.5	2	58	116	3364	6728
2	59.5 - 62.5	3	61	183	3721	11163
3	62.5 - 65.5	5	64	320	4096	20480
4	65.5 - 68.5	8	67	536	4489	35912
5	68.5 - 71.5	6	70	420	4900	29400
6	71.5 - 74.5	4	73	292	5329	21316
7	74.5 - 77.5	2	76	152	5776	11552
		30		2019		136,551

$$s^2 = \frac{\sum f_i x_i^2 - \frac{(\sum f_i x_i)^2}{n}}{n-1} = \frac{136551 - \frac{(2019)^2}{30}}{30-1} = \frac{136551 - 135878.7}{29}$$

$$s^2 = \frac{672.3}{29} = 23.18 \quad (\text{varianza})$$

$$s = \sqrt{23.18}$$

$$s = 4.81 \quad (\text{desviación estándar})$$

#### 5.10 CALCULO DE LA VARIANZA Y LA DESVIACION ESTANDAR PARA DATOS AGRUPADOS CON CALCULADORA

Para calcular la varianza y la desviación estándar para datos agrupados con la calculadora se sigue el mismo procedimiento que aplicamos en datos no agrupados.

Pasos a seguir:

1. MODE X SD
2. SHIFT SCI EXE
3. INGRESO

58 SHIFT [;] 2DT  
 61 SHIFT [;] 3DT  
 64 SHIFT [;] 5DT  
 67 SHIFT [;] 8DT  
 70 SHIFT [;] 6DT  
 73 SHIFT [;] 4DT  
 76 SHIFT [;] 2DT

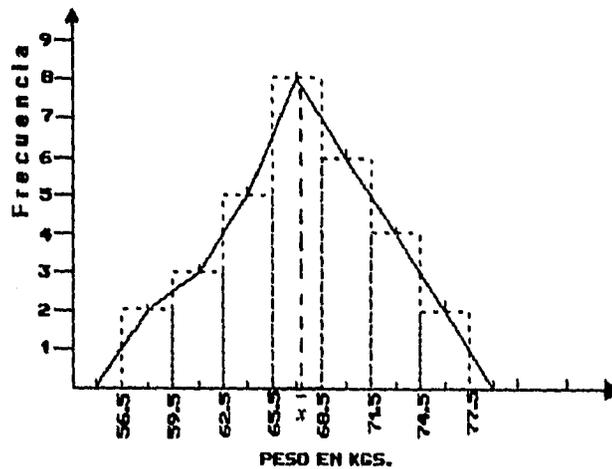
OPERACION DE TECLA		RESULTADO	SIMBOLO
SHIFT X	EXE	DESVIACION ESTANDAR MUESTRAL	$S=4.81$
SHIFT X	EXE	DESVIACION ESTANDAR POBLACIONAL	$\sigma=4.73$
SHIFT X	EXE	MEDIA ARITMETICA	$\bar{x}=67.3$
SHIFT 1k1	EXE	SUMA DE DATOS AL CUADRADO	$\sum f_i x_i^2 = 136,551$
SHIFT 1k2	EXE	SUMA DE DATOS	$\sum f_i x_i = 2019$
SHIFT 1k3	EXE	NUMERO DE DATOS	$n=30$

### 5.11 INTERPRETACION DE LA DESVIACION ESTANDAR

#### LA REGLA EMPIRICA

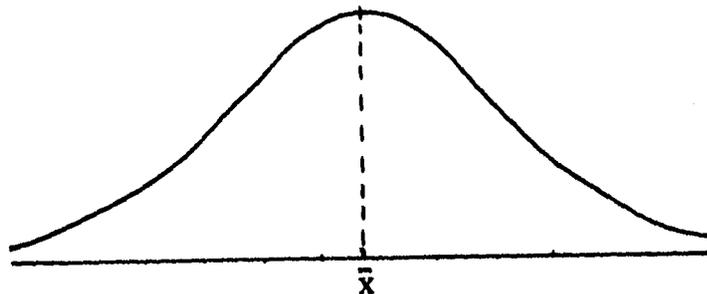
Entre la media y la desviación estándar de un conjunto de datos cuando la distribución de éstos es de un cierto tipo, existe una relación interesante que nos permitirá hacer inferencias importantes en estadística.

Para ello, tomaremos el histograma y el polígono de frecuencias del peso de los estudiantes de estadística del C.C.H. que es el siguiente:



Para esta distribución, se tiene como media  $\bar{x}=67.3$  kgs. y como desviación estándar  $S=4.81$  kgs. Puede observarse que el polígono de frecuencias es aproximadamente simétrico respecto a la recta paralela al eje de las ordenadas que pasa por  $\bar{x}$

Si aumentamos el número de datos y el número de clases, en el histograma disminuirá el ancho de los rectángulos, y el polígono de frecuencias se parecería cada vez más a una curva en forma de campana. A esta forma que toma el polígono de frecuencias se le conoce con el nombre de polígono de frecuencias suavizado y queda de la siguiente forma:



Si lo que acabamos de decir ocurre para cada muestra con un número suficientemente grande de datos, es de esperar que el polígono de frecuencias de la población sea muy parecido a la curva obtenida anteriormente. Esta curva es simétrica respecto de la recta paralela al eje de las ordenadas que pasa por el punto de mayor frecuencia de esta curva. El punto donde dicha recta corta al eje de las abscisas coincide con la media de la población.

Se dice que una población cuyo polígono de frecuencias se parece mucho a la figura anterior, tiene una distribución normal. Este tipo de distribución se estudiará con más detalle en otro capítulo.

Para distribuciones en forma de campana, es decir las distribuciones semejantes a normales, hay una relación

interesante entre la media y la desviación estándar.  
Esta relación se conoce como la regla empírica.

#### REGLA EMPIRICA

Si un conjunto de datos tienen una distribución de frecuencias aproximadamente simétrica, en forma de campana y tiene media  $\bar{x}$  y desviación estándar  $S$ , entonces:

- 1) El intervalo  $\bar{x} \pm S$ ,  $(\bar{x}-S, \bar{x}+S)$  contiene aproximadamente el 68% de los datos.
- 2) El intervalo  $\bar{x} \pm 2S$ ,  $(\bar{x}-2S, \bar{x}+2S)$  contiene aproximadamente el 95% de los datos.
- 3) El intervalo  $\bar{x} \pm 3S$ ,  $(\bar{x}-3S, \bar{x}+3S)$  contiene todos o casi todos los datos.

Verifiquemos la regla empírica con el ejercicio del peso de los estudiantes el cual tiene como media  $\bar{x}=67.3$  y como desviación estándar  $S=4.81$

- 1)  $\bar{x} \pm S, (\bar{x}-S, \bar{x}+S) = (67.3-4.81, 67.3+4.81) = (62.49, 72.11)$
- 2)  $\bar{x} \pm 2S, (\bar{x}-2S, \bar{x}+2S) = (67.3-2(4.81), 67.3+2(4.81))$   
 $= (57.68, 76.92)$

$$3) \bar{x} \pm 3S, (\bar{x} - 3S, \bar{x} + 3S) = (67.3 - 3(4.81), 67.3 + 3(4.81)) \\ = (52.87, 81.73)$$

1) El intervalo (62.49, 72.11) contiene aproximadamente el 68% de los datos. Para ello, presentamos aquí los datos de nuestro ejemplo.

Peso en kilogramos de un grupo de estudiantes de estadística del C.C.H. (Datos ordenados)

58	<u>63</u>	<u>67</u>	<u>69</u>	<u>72</u>
59	<u>64</u>	<u>68</u>	<u>69</u>	73
60	<u>65</u>	<u>68</u>	<u>70</u>	74
61	<u>65</u>	<u>68</u>	<u>71</u>	74
62	<u>66</u>	<u>68</u>	<u>71</u>	75
<u>63</u>	<u>67</u>	<u>68</u>	<u>71</u>	77

Hemos subrayado los datos que pertenecen al intervalo (62.49, 72.11). De los 30 datos, 20 pertenecen al intervalo. Utilizando una regla de tres simple obtenemos:

30 — 100%

20 — X

$$X = \frac{20(100)}{30} = 66.66\%$$

Resulta que el 66.66% de los datos pertenecen al intervalo señalado. Entonces se puede decir que el 66.66% de los estudiantes pesaron entre 62.49 y 72.11 kilogramos.

Observe que este porcentaje es ligeramente menor al que marca la regla empírica, pero recuerde también que esta dice, que el intervalo  $(\bar{X}-S, \bar{X}+S)$  contiene aproximadamente el 68% de los datos.

2) El intervalo (57.68, 76.92) contiene 29 datos de los 30 datos.

<u>58</u>	<u>63</u>	<u>67</u>	<u>69</u>	<u>72</u>
<u>59</u>	<u>64</u>	<u>68</u>	<u>69</u>	<u>73</u>
<u>60</u>	<u>65</u>	<u>68</u>	<u>70</u>	<u>74</u>
<u>61</u>	<u>65</u>	<u>68</u>	<u>71</u>	<u>74</u>
<u>62</u>	<u>66</u>	<u>68</u>	<u>71</u>	<u>75</u>
<u>63</u>	<u>67</u>	<u>68</u>	<u>71</u>	<u>77</u>

por lo que  $\frac{29}{30}(100) = 96.66\%$  de los datos pertenecen al intervalo, entonces se dice que el 96.66% de los estudiantes pesaron entre 57.68 kgs. y 76.92 kgs.

3) El intervalo (52.87, 81.73) contiene todos los datos por lo que se dice que el 100% de los estudiantes pesaron entre 52.87 kgs. y 81.73 kgs.

Esta información puede ser presentada en un cuadro para compararla con los establecidos por la regla empírica.

**PORCENTAJES COMPARADOS CON LA REGLA EMPIRICA.**

INTERVALO	% DE LA REGLA EMPIRICA	% ENCONTRADO
$(\bar{x}-S, \bar{x}+S)$	APROX. 68%	66.66%
$(\bar{x}-2S, \bar{x}+2S)$	APROX. 95%	96.66%
$(\bar{x}-3S, \bar{x}+3S)$	APROX. 100%	100.00%

Estos porcentajes están bastante próximos a los establecidos en la regla empírica. Si combinamos estos resultados con el histograma de los datos, podemos asegurar que los datos muestrales están distribuidos de manera aproximadamente normal.

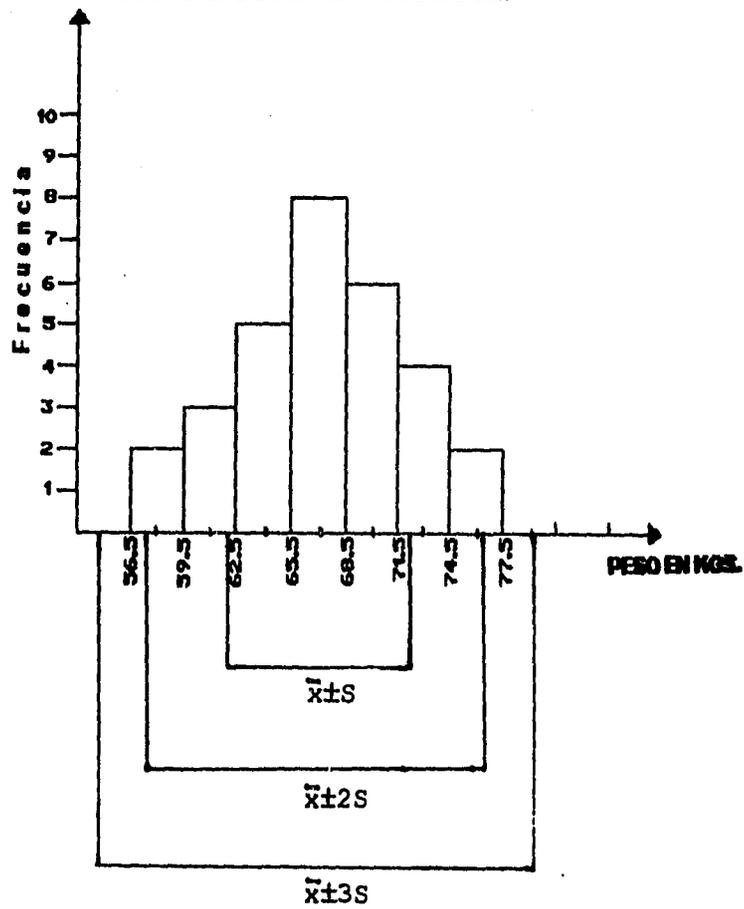
Representemos la regla empírica con el histogramas de nuestro ejemplo.

1)  $\bar{x} \pm s = (62.49, 72.11)$

2)  $\bar{x} \pm 2s = (57.68, 76.92)$

3)  $\bar{x} \pm 3s = (52.87, 81.73)$

HISTOGRAMA DE LOS PESOS EN KGS. DE LOS ESTUDIANTES  
DE UN GRUPO DE ESTADISTICA DEL C.C.H.



La regla empírica puede utilizarse para determinar si aproximadamente un conjunto de datos esta o no distribuido de manera normal, en base a una medida de tendencia central como la media y a una medida de dispersión como la desviación estándar.

#### 5.12 SESGO

El sesgo es el grado de asimetría, o falta de simetría, de una distribución. Si la curva de frecuencia (polígono de frecuencias suavizado) de una distribución tiene una cola más larga a la derecha del punto máximo central que a la izquierda, se dice que esta distribución está sesgada a la derecha o que tiene sesgo positivo. Si es al contrario, se dice que esta sesgada a la izquierda o que tiene sesgo negativo.

En distribuciones sesgadas, la media tiende a situarse con respecto a la moda al mismo lado que la cola más larga. Así una medida de asimetría nos viene dada por la diferencia (Media - moda). Esta medida puede adimensionarse, dividiéndola por una medida de dispersión tal como la desviación estándar.

$$\text{Sesgo} = \frac{\text{Media} - \text{moda}}{\text{desviación estándar}} = \frac{\bar{x} - \text{moda}}{s}$$

Para evitar el empleo de la moda se puede utilizar la relación empírica entre la media, mediana y moda.

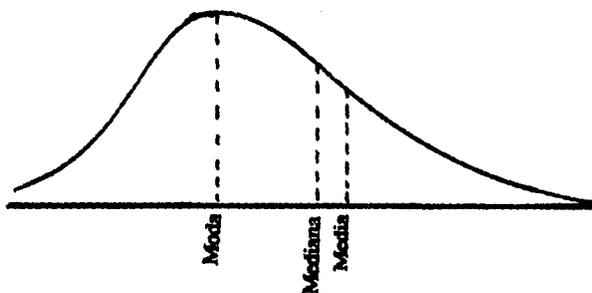
$$\text{Media} - \text{Moda} = 3(\text{Media} - \text{Mediana})$$

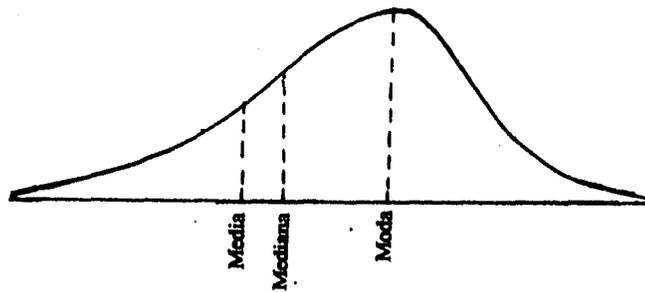
Esta relación empírica se usa para curvas de frecuencias unimodales que sean moderadamente sesgadas (asimétricas).

De manera que la fórmula para encontrar el sesgo queda de la siguiente manera:

$$\text{Sesgo} = \frac{\bar{x} - \text{moda}}{s} = \frac{3(\bar{x} - \text{mediana})}{s}$$

En las siguientes figuras se 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 mediana coinciden.





Ejemplo. Calculemos el sesgo de nuestro ejemplo del peso de los estudiantes, recordando que tiene como medidas centrales y desviación estándar las siguientes:

$$\bar{x}=67.3, \text{Moda}=67.3, \text{Mediana}=67.375 \text{ y } S=4.81$$

Como la media y la moda tienen el mismo valor nos indica que la curva es simétrica.

$$\text{Sesgo} = \frac{\bar{x} - \text{moda}}{S} = \frac{(67.3 - 67.375)}{4.81} = \frac{0}{4.81} = 0$$

tomemos la relación empírica para comprobar el resultado anterior.

$$\text{Sesgo} = \frac{3(\bar{x} - \text{mediana})}{S} = \frac{3(67.3 - 67.375)}{4.81} = \frac{0.225}{4.81} = 0.0467$$

De este resultado podemos concluir que esta curva está ligeramente sesgada a la izquierda, y se puede considerar como simétrica.

### 5.13 CALCULO DEL SESGO CON CALCULADORA

Dado que la fórmula del sesgo no esta incorporada en las 128 fórmulas de la calculadora, procederemos a almacenar en una memoria la fórmula.

1. PRESIONE MODE 2
2. USAR LA TECLA  $\Rightarrow$  PARA SELECCIONAR P3
3. ESPECIFICAR MODE X PARA SD
4. PRESIONE LA TECLA EXE
5. INGRESE LA FORMULA  $S=X-M/D$

AC ALPHA S SHIFT=(ALPHA X - ALPHA M)/ALPHA D

6. PRESIONE MODE 1 PARA ALMACENAR EL PROGRAMA

Sesgo =  $\frac{\bar{x} - moda}{S}$  en la pantalla  $S=(X-M)/D$

Ahora almacenemos la fórmula  $Sesgo = \frac{3(\bar{x} - MEDIANA)}{S}$  que la sustituiremos en la calculadora por  $S=3(X-P)/D$

1. PRESIONAR MODE 2
2. USAR LA TECLA  $\Rightarrow$  PARA SELECCIONAR P4

3. ESPECIFICAR MODE X PARA SD

4. PRESIONAR EXE

5. INGRESAR LA FORMULA  $S = \frac{3(x-P)}{D}$

AC ALPHA S SHIFT=3 (ALPHA-ALPHA P) /ALPHA D

6. PRESIONAR MODE 1 PARA ALMACENAR EL PROGRAMA

Sesgo =  $\frac{3(\bar{x} - \text{MEDIANA})}{S}$  EN LA CALCULADORA  $S=3(X-P)/D$

Ejemplo, calculemos el sesgo del peso de los estudiantes con las fórmulas P3 y P4 incorporadas a la calculadora.

#### PASOS

1. MODE 1

2. PROG 3 EXE

$S = (X-M) / D$

3. x? 67.3 EXE

M? 67.3 EXE

D? 4.81 EXE

SESGO=0 como se tenía

Usando la fórmula  $\text{SESGO} = \frac{3(\bar{x} - \text{MEDIANA})}{S}$

#### PASOS

1. MODE 1

2. PROG 4 EXE

$S = 3(X-P) / D$

3. X? 67.3 EXE

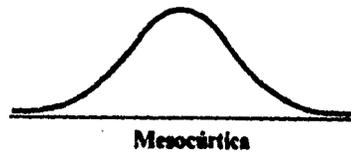
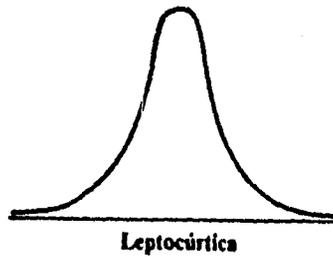
P? 67.375 EXE

D? 4.81 EXE

S=-0.0467

#### 5.14 CURTOSIS, COEFICIENTE DE CURTOSIS PERCENTILICO

Es el grado de apuntamiento de una distribución. Normalmente se toma en relación a la distribución normal. Una distribución que presenta un apuntamiento relativo alto se llama LEPTOCURTICA, mientras que la curva que es más achatada, se llama PLATICURTICA. La distribución normal, que ni es muy apuntada ni achatada, se llama MESOCURTICA.



### COEFICIENTE DE CURTOSIS PERCENTILICO

Una fórmula para medir la picudez de la curva esta basada en los cuartiles y percentiles y se conoce como coeficiente de curtosis percentilico y esta dada por:

$$K = \frac{Q}{P_{90} - P_{10}}$$

Donde  $Q = \frac{Q_3 - Q_1}{2}$  es el rango semi-intercuartilico.

$P_{90} - P_{10}$  es el rango entre los percentiles 10-90.

Para la distribución normal el coeficiente de curtosis percentilico es 0.263, de manera que:

Si  $K > 0.263$  la distribución es platicúrtica.

Si  $K = 0.263$  la distribución es neocúrtica (normal).

Si  $K < 0.263$  la distribución es leptocúrtica.

Calculemos el coeficiente de curtosis percentilico en nuestro ejemplo del peso de los estudiantes recordando que:

$Q_1 = 64$  ,  $Q_3 = 70.75$  ,  $P_{10} = 60.5$  y  $P_{90} = 73.75$

$$Q = \frac{Q_3 - Q_1}{2} = \frac{70.75 - 64}{2} = \frac{6.75}{2} = 3.375 \quad \therefore Q = 3.375$$

$$P_{90} - P_{10} = 73.75 - 60.5 = 13.25$$

$$\text{Por lo que } K = \frac{Q}{P_{90} - P_{10}} = \frac{3.375}{13.25} = 0.254 \quad \therefore K = 0.254$$

Entonces como  $0.254 < 0.263$  se dice que la distribución es ligeramente leptocúrtica y se puede tomar como una distribución normal.

#### 5.15 CALCULO DEL COEFICIENTE DE CURTOSIS PERCENTILICO CON CALCULADORA

Como la fórmula del coeficiente de curtosis no está integrado en las fórmulas de la calculadora, debemos almacenarlo en una memoria.

##### PASOS

1. PRESIONE MODE 2
2. USAR LA TECLA  $\Rightarrow$  PARA SELECCIONAR P5
3. ESPECIFICAR MODE X PARA SD
4. PRESIONAR LA TECLA EXE
5. INGRESAR LA FORMULA  $K = (C - D) / 2 / (M - N)$   
AC ALPHA K SHIFT = ((ALPHA C - ALPHA D) / 2) / (ALPHA M - ALPHA N)
6. PRESIONAR MODE 1 PARA ALMACENAR EL PROGRAMA

Ejemplo. Calcular el coeficiente de curtosis percentilico para nuestro ejemplo, del peso del grupo.

PASOS

1. MODE 1
2. PROG. 5 EXE EN LA PANTALLA  $K = ((C-D)/2)/(M-N)$
3. C? 70.75 EXE  
D? 64. EXE  
M? 73.75 EXE  
N? 70.5 EXE  
K=0.254

5.16 EJERCICIOS DE REPASO

1. Los pesos al nacer (en kilogramos) de 40 becerros nacidos en Toluca en 1990 están agrupados en una distribución de frecuencias como a continuación se muestra:

PESO AL NACER (EN KGS.)	NUMERO DE BECERROS
11.5-14.5	3
14.5-17.5	11
17.5-20.5	15
20.5-23.5	6
23.5-26.5	4
26.5-29.5	1
	40

En base a esos datos calcular:

- a) La media, la mediana y la moda.
- b) La varianza y la desviación estándar por la fórmula no simplificada y por la fórmula simplificada.
- c) El coeficiente de sesgo y decir si es sesgada a la derecha o a la izquierda.
- d) El coeficiente de curtosis percentílico,  $k_e$  interpretarlo.
- e) Comprobar los resultados con la calculadora

2. Para el ejercicio 1 del grupo 4.20 capítulo 4 página 137 calcular:

- a) La varianza y la desviación estándar por cualquier método
- b) El coeficiente de sesgo y decir si es sesgada a la derecha o a la izquierda.
- c) El coeficiente de curtosis percentílico,  $k_e$  interpretarlo.
- d) Comprobar los resultados con la calculadora

3. Para el ejercicio 2 del grupo 4.20 capítulo 4 página

137 calcular:

- a) La varianza y la desviación estándar
- b) El coeficiente de sesgo e interpretarlo.
- c) El coeficiente de curtosis percentílico,  $k$  e interpretarlo.
- d) Comprobar los resultados con la calculadora

4. Para el ejercicio 3 del grupo 4.20 capítulo 4 página

138 calcular utilizando la calculadora:

- a) La desviación estándar
- b) El coeficiente de sesgo e interpretarlo.
- c) El coeficiente de curtosis percentílico,  $k$  e interpretarlo.

5. Para el ejemplo del precio del arroz en kilogramos,

que se vende en 25 supermercados página 70 del capítulo

3, cuya tabla de frecuencias es:

INTERVALO DE CLASE	fi
127-135	5
136-144	4
145-153	2
154-162	6
163-171	3
172-180	5
	25

- a) Calcular la media y la desviación estándar
- b) Calcular los extremos de los intervalos:  $(\bar{x}-S, \bar{x}+S), (\bar{x}-2S, \bar{x}+2S), (\bar{x}-3S, \bar{x}+3S)$  y diga aproximadamente cuántos supermercados tenían sus precios en esos intervalos y calcular sus porcentajes.
- c) Diga si este conjunto de datos está o no distribuido aproximadamente de manera normal.

## **CAPITULO 6**

### **MEDIDAS DE CORRELACION**

- 6.1 INTRODUCCION.
- 6.2 CORRELACION LINEAL.
- 6.3 COEFICIENTE DE CORRELACION.
- 6.4 RECTA DE REGRESION.
- 6.5 ERROR ESTANDAR DE ESTIMACION.
- 6.6 CALCULO DEL COEFICIENTE DE CORRELACION Y DE LA  
RECTA DE REGRESION CON LA CALCULADORA.
- 6.7 EJERCICIOS DE REPASO.

## 6.1 INTRODUCCION

Los técnicos de la Secretaría de Agricultura al analizar la producción de trigo en una de las zonas agrícolas del país, encontraron que la mayor parte de la producción se encontraba en aquellas regiones donde se había utilizado mayor cantidad de un nuevo producto químico para abonar la tierra.

Después de discutir si la concentración de la producción se debía al producto químico utilizado o a algún otro factor, decidieron hacer un análisis de correlación entre la producción de trigo y el nivel de precipitación pluvial, y otro, entre la producción de trigo y la cantidad de abono usado. Se hizo el análisis estadístico y resultó ser mayor el coeficiente de correlación entre la producción del trigo y la cantidad de abono. Por lo anterior, se concluyó que el abono fue el factor que permitió obtener una buena cosecha.

Este tipo de análisis es necesario cuando dos variables en estudio están relacionados en este capítulo estudiaremos este tipo de medidas.

## 6.2 CORRELACION LINEAL

En la vida diaria es común encontrar dos variables que guardan dependencia entre sí, es decir, el valor de una de ellas está sujeto al valor de la otra, o viceversa. Esta interdependencia se conoce como correlación de dos variables.

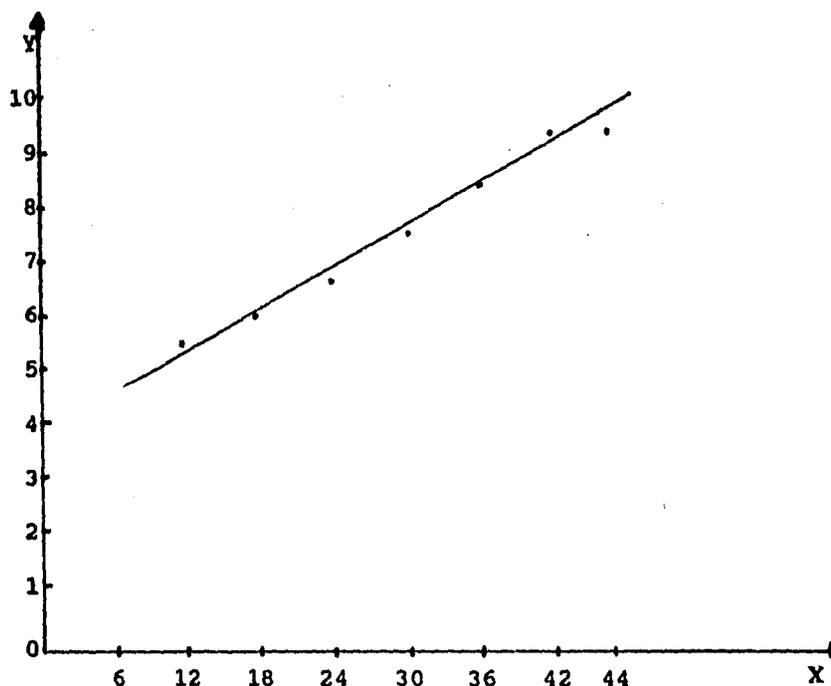
Daremos un ejemplo que nos permita ver el fenómeno de la correlación de dos variables.

En la tabla siguiente se muestra el rendimiento de un cierto cultivo y en función de la cantidad de agua  $x$ . Es decir, suponemos que el rendimiento depende la cantidad de agua

CANTIDAD DE AGUA (X)	RENDIMIENTO DEL CULTIVO (Y)
12	5.27
18	5.68
24	6.25
30	7.2
36	8.02
42	8.71
44	8.42

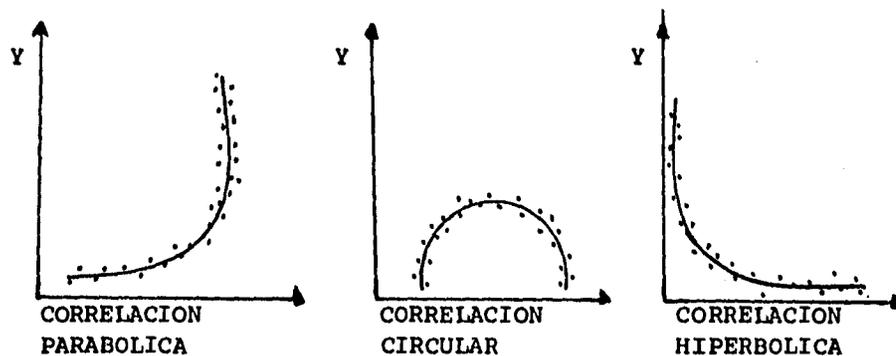
Nuestra pregunta es: ¿será o no cierta esa dependencia?

Al graficar en el plano cartesiano los datos anteriores, se obtiene:



Se puede aventurar decir que las variables están correlacionadas y se agrupan de manera que guarden una correlación lineal, esto es, su gráfica se puede representar por una recta y expresar algebraicamente en una función lineal del tipo:  $Y = mx + b$ .

Es importante hacer notar que puede haber correlaciones no lineales, es decir, que sus gráficas, y por lo tanto sus funciones, no sean lineales, como se observa a continuación.



Nota: en este capítulo sólo mencionaremos la correlación lineal.

### 6.3 COEFICIENTE DE CORRELACION

La correlación lineal tiene por objeto encontrar la función que relaciona linealmente a las dos variables, pero antes es conveniente analizar, por medio del coeficiente de correlación, si la dependencia mencionada es aceptable.

Con este fin definiremos la desviación estándar de la variable  $x$  (cantidad de agua) como:

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

Y para la variable  $Y$  (rendimiento del cultivo), es:

$$S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{N}}$$

Además de la desviación estándar, para cada variable es necesario tener una medida de dispersión de la correlación que existe entre  $X$  y  $Y$ . Esta medida la denominaremos covarianza y la definiremos así:

$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$$

Ahora si podemos definir el coeficiente de correlación o coeficiente  $r$  de Pearson:

$$r = \frac{S_{xy}}{S_x S_y}$$

es decir, que el coeficiente de correlación es el cociente de la división entre la covarianza de las

variables y el producto de las desviaciones estandar de cada variable.

Sustituyendo los valores queda:

$$r = \frac{\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}}{\sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} \sqrt{\frac{\sum (y_i - \bar{y})^2}{N}}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

Para calcular el coeficiente de correlación del problema que nos ocupa, recurrimos a la tabla siguiente:

$x_i$	$y_i$	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
12	5.27	-17.43	-1.81	31.55	303.80	3.28
18	5.68	-11.43	-1.40	16.00	130.64	1.96
24	6.25	-5.43	-0.83	4.51	29.48	0.69
30	7.2	0.57	0.12	0.07	0.32	0.01
36	8.02	6.57	0.94	6.18	43.16	0.88
42	8.71	12.57	1.63	20.49	158.00	2.66
44	8.42	14.57	1.34	19.52	212.28	1.80
206	49.55			98.32	877.68	11.28

$$\bar{x} = \frac{\sum x_i}{n} = \frac{206}{7} = 29.43$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{49.55}{7} = 7.08$$

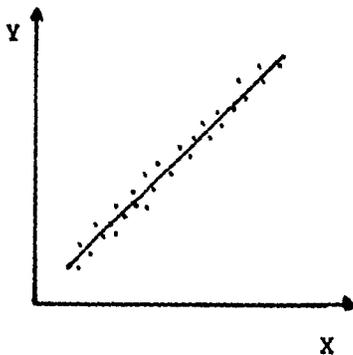
$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} = \sqrt{\frac{877.68}{7}} = 11.20$$

$$S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{N}} = \sqrt{\frac{11.28}{7}} = 1.27$$

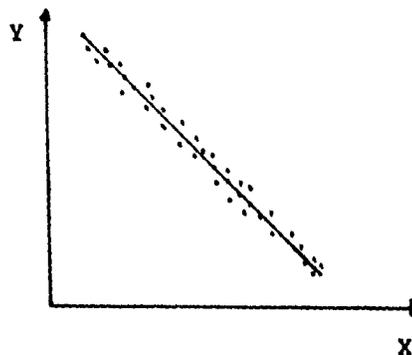
$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N} = \frac{98.32}{7} = 14.05$$

$$r = \frac{S_{xy}}{S_x S_y} = \frac{14.05}{(11.20)(1.27)} = 0.987$$

El coeficiente de correlación puede ser positivo o negativo dependiendo de la covarianza; así tendremos correlación positiva (Y aumenta cuando x aumenta) o correlación negativa (Y disminuye cuando x aumenta), como se muestra:



**CORRELACION POSITIVA**



**CORRELACION NEGATIVA**

Por otro lado, el coeficiente de correlación lineal  $r$  siempre tiene un valor entre  $-1$  y  $+1$ . Un valor igual a  $+1$ , o bien a  $-1$  indica una correlación positiva o una correlación negativa perfecta, respectivamente, es decir, todos los puntos de las observaciones están perfectamente alineados.

- Cuando  $r = 0$  la correlación es nula, es decir, no hay ninguna dependencia entre las variables.

- Cuando  $r$  está entre  $0$  y  $1$ , existe una correlación positiva entre las dos variables  $X$  y  $Y$ . Si estas tienen una correlación positiva, los valores de  $X$  y  $Y$  tienden a moverse en la misma dirección. El valor de una variable tiende a variar directamente con el de la otra.

- Cuando  $r$  está entre  $-1$  y  $0$ , existe una correlación negativa entre  $X$  y  $Y$ . Entonces los valores de  $X$  y  $Y$  tienden a moverse en dirección opuesta; cuando uno aumenta, el otro tiende a disminuir y viceversa.

En nuestro problema la correlación es positiva y muy cercana al valor uno, por lo cual hay una dependencia aceptable entre la cantidad de agua y el rendimiento del cultivo.

Daremos dos ejemplos para analizar el grado de correlación.

Ejemplo 1. Dadas las siguientes parejas de datos, decir si la correlación lineal entre las dos variables es aceptable o no.

$X_i$	$Y_i$
1	5
3	19
-3	-23
6	40
9	61

Procedemos a elaborar la siguiente tabla.

$X_i$	$Y_i$	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
1	5	-2.2	-15.4	33.88	4.84	237.16
3	19	-0.2	-1.4	0.28	0.04	1.96
-3	-23	-6.2	-43.4	269.08	38.44	1883.56
6	40	2.8	19.6	54.88	7.84	384.16
9	61	5.8	40.6	235.48	33.64	1648.36
16	102			593.60	84.80	4155.20

$$\bar{X} = \frac{16}{5} = 3.2$$

$$\bar{Y} = \frac{102}{5} = 20.4$$

$$S_x = \sqrt{\frac{84.80}{5}} = 4.118$$

$$S_y = \sqrt{\frac{4155.20}{5}} = 28.82$$

$$S_x = 4.12$$

$$S_y = 28.82$$

$$S_{xy} = \frac{593.60}{5} = 118.72$$

$$r = \frac{S_{xy}}{S_x S_y} = \frac{118.72}{(4.12)(28.82)} = \frac{118.72}{118.73} = 0.999 \therefore r = 1$$

El coeficiente de correlación es igual a la unidad, por lo que la dependencia es perfecta; es más los puntos están alineados, como se puede verificar al graficarlos.

Ejemplo 2.- Con los siguientes datos apareados, analiza el grado de correlación

$X_i$	$Y_i$
1	1
1	2
1	3
2	2
3	2
3	1
3	3
2	1
2	3

Hagamos la tabla de cálculo.

$X_i$	$Y_i$	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
1	1	-1	-1	1	1	1
1	2	-1	0	0	1	0
1	3	-1	1	-1	1	1
2	2	0	0	0	0	0
3	2	1	0	0	1	0
3	1	1	-1	-1	1	1
3	3	1	1	1	1	1
2	1	0	-1	0	0	1
2	3	0	1	0	0	1
18	18			0	6	6

$$\bar{x} = \frac{18}{9} = 2$$

$$\bar{y} = \frac{18}{9} = 2$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{0}{\sqrt{(6)(6)}} = 0$$

El coeficiente de correlación es nulo, ya que no hay dependencia entre las variables. Lo anterior se puede verificar al hacer la gráfica.

Es conveniente hacer notar que aunque el coeficiente de correlación sea bueno, hay que examinar la dependencia de las variables.

#### 6.4 RECTA DE REGRESION

Una vez que, a través del coeficiente de correlación, sabemos que hay una buena dependencia de las variables, nos interesa conocer la función de relación, es decir, la recta de regresión o la recta de ajuste. Esta función debe cumplir la condición de que la suma de los cuadrados de las distancias verticales de los distintos puntos a la recta de la regresión sea mínima. La forma para lograrlo es a través del METODO DE MINIMOS CUADRADOS.

Este criterio de mínimos cuadrados consiste en definir a la recta de regresión como:

$$y = mx + b$$

la cual tiene dos parámetros:  $m$  y  $b$ , que nos interesa conocer. Al sustituir en ella los valores de  $X_i$ , obtendremos imágenes como los correspondientes de  $Y_i$ , pero estos tendrán un error  $e_i$ . El error  $e_i$  se suma a todos los valores de  $X_i$  como se ve en la tabla siguiente:

$$Y_1 - mx_1 - b = e_1$$

$$Y_2 - mx_2 - b = e_2$$

,

.

.

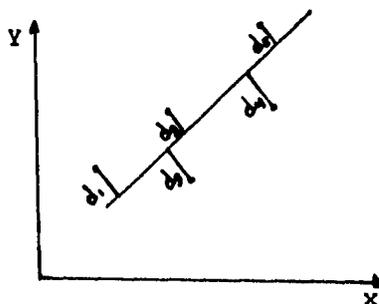
.

,

$$Y_n - mx_n - b = e_n$$

---


$$\Sigma(Y_i - mx_i - b) = \Sigma e_i$$



$$\min \Sigma ( d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 )$$

La suma de  $e_i$  será menor si consideramos su cuadrado así:

$$\sum (Y_i - m X_i - b)^2 = \sum e_i^2$$

Aplicamos ahora el criterio de minimización derivando parcialmente la función  $\sum e_i^2$  respecto a las variables  $m$  y  $b$ .

Para derivar parcialmente respecto a  $m$ , se considera a  $b$  constante. Al derivar parcialmente respecto a  $b$ ,  $m$  es la variable que se considera constante. Por lo anterior podemos escribir:

$$\frac{d}{dm} \sum (Y_i - m x_i - b)^2 = \sum (-2 (Y_i - m x_i - b) X_i) = 0$$

$$\frac{d}{db} \sum (Y_i - m x_i - b)^2 = \sum (-2 (y_i - m x_i - b)) = 0$$

Si resolvemos simultáneamente las dos ecuaciones anteriores, obtenemos los valores de  $m$  y  $b$  así:

$$m = \frac{N \sum x_i y_i - (\sum x_i) (\sum y_i)}{N \sum (x_i^2) - (\sum x_i)^2}$$

$$b = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{N \sum x_i^2 - (\sum x_i)^2}$$

Apliquemos las fórmulas anteriores para encontrar la función lineal que relaciona la cantidad de agua con el rendimiento del cultivo. Para ello requerimos la siguiente tabla:

$X_i$	$Y_i$	$X_i^2$	$X_i Y_i$
12	5.27	144	63.24
18	5.68	324	102.24
24	6.25	576	150.00
30	7.2	900	216.00
36	8.02	1296	288.72
42	8.71	1764	365.82
44	8.42	1936	370.48
206	49.55	6940	1556.50

$$m = \frac{7(1556.50) - (206)(49.55)}{7(6940) - (206)^2} = \frac{688.2}{6144} = 0.11$$

$$b = \frac{(49.55)(6940) - (206)(1556.50)}{7(6940) - (206)^2} = \frac{23.238}{6144} = 3.78$$

Por lo tanto, la recta de regresión se representa por la función lineal siguiente:

$$Y = 0.11 x + 3.78$$

Resumen. De acuerdo con lo que hemos visto en este capítulo, podemos definir dos conceptos:

1) COEFICIENTE DE CORRELACION (r):

Es un número ó índice que nos informa si la correlación es positiva o negativa a través del signo y por lo general se dice que una correlación es buena si se cumple

$$|r| \geq 0.7$$

2) RECTA DE REGRESION:

Es la recta que más se ajusta a las observaciones apareadas y se representa algebraicamente por la expresión:

$$Y = mx + b$$

Las fórmulas de cálculo son las siguientes:

$$r = \frac{\sum (x_i - \bar{X}) (y_i - \bar{Y})}{\sqrt{\sum (x_i - \bar{X})^2 \sum (y_i - \bar{Y})^2}}$$

$$m = \frac{N \sum x_i y_i - (\sum x_i)(\sum y_i)}{N \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{N \sum x_i^2 - (\sum x_i)^2}$$

#### 6.5 ERROR ESTANDAR DE ESTIMACION

Si en la función de la recta de regresión:

$$Y = 0.11X + 3.78$$

sustituimos los valores originales de X, los valores de Y calculados o estimados no serán iguales a los valores de Y originales. Lo anterior se puede verificar.

Resulta fácil comprender que cuanto mayor sea el coeficiente de correlación, menores serán las diferencias entre los valores originales y los calculados en la variable "Y".

Si queremos conocer el grado de dispersión que tienen dichas diferencias, podemos hacerlo a través de la medida denominada error estándar de estimación que se calcula de la siguiente manera:

$$e = \sqrt{\frac{\sum (Y_i - Y_{est})^2}{N}}$$

Para calcular el error estándar de estimación en el problema que nos ocupa, requerimos de la tabla siguiente:

$X_i$	$Y_i$	$Y_{est}$	$(Y_i - Y_{est})$	$(Y_i - Y_{est})^2$
12	5.27	5.10	0.17	0.0289
18	5.68	5.76	-0.08	0.0064
24	6.25	6.42	-0.17	0.0289
30	7.20	7.08	0.12	0.0144
36	8.02	7.74	0.28	0.0784
42	8.71	8.40	0.31	0.0961
44	8.42	8.62	-0.20	0.0400
				0.2931

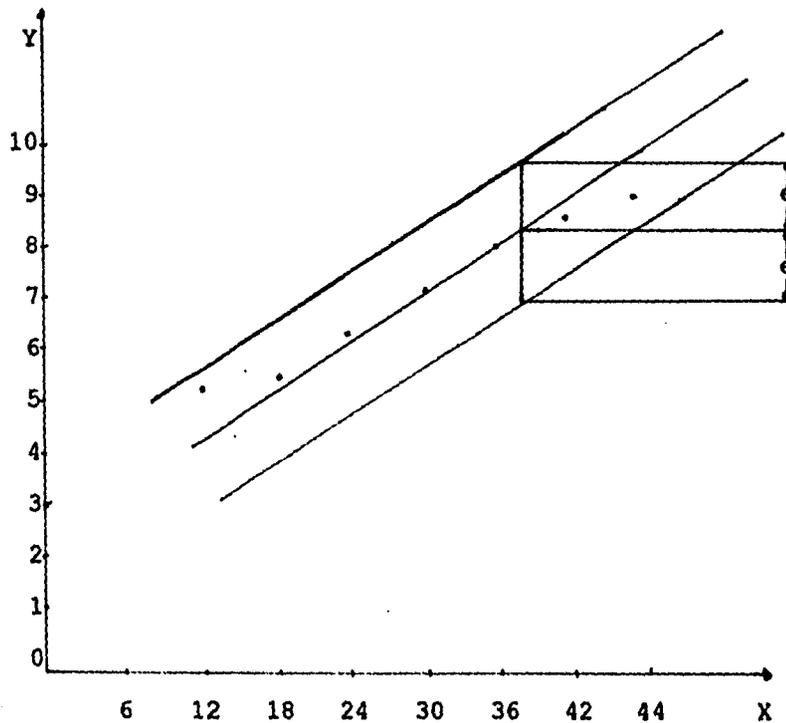
$$Y = 0.11X + 3.78$$

$$e = \sqrt{\frac{\sum (Y_i - Y_{est})^2}{N}} = \sqrt{\frac{0.2931}{7}} = 0.205$$

La interpretación del error estándar de estimación se ve claramente en la gráfica siguiente.

El error estándar de estimación es la medida que se forma verticalmente, arriba y abajo de la recta de

regresión, y permite definir dos rectas paralelas, dentro de las cuales se encuentran el 68% de los puntos de observaciones.



En resumen, diremos que al estar relacionadas dos variables podemos calcular el grado de dependencia entre ellas por medio del COEFICIENTE DE CORRELACION. Si el coeficiente de correlación nos da un resultado aceptable, se pueden calcular los parámetros de la RECTA DE REGRESION.

Al utilizar la RECTA DE REGRESION, los valores estimados no son iguales a los valores originales, por lo cual calculamos el ERROR ESTANDAR DE ESTIMACION. Este valor es como la desviación que tiene los valores estimados tomando la RECTA DE REGRESION como valor medio.

Se puede decir que en una gráfica el punto  $(\bar{x}, \bar{y})$  es el centro de gravedad de los puntos observados.

#### 6.6 CALCULO DEL COEFICIENTE DE CORRELACION Y DE LA RECTA DE REGRESION CON LA CALCULADORA.

Para realizar cálculos de regresión lineal en la calculadora, se debe ingresar el conjunto de datos en el modo LR . [ / ]

Hay que hacer notar que algunos resultados de cálculos se almacenan automáticamente en ciertas memorias de constantes como se indica debajo. Esto significa que cualquiera de los valores previamente almacenados en estas memorias se pierden.

MEMORIAS	RESULTADOS DEL CALCULO
IK 1	SUMA DE LOS CUADRADOS DE DATOS $X_i$
IK 2	SUMA DE LOS DATOS $X_i$
IK 3	SUMA DE LOS ELEMENTOS DE DATOS N
IK 4	SUMA DE LOS CUADRADOS DE DATOS $Y_i$
IK 5	SUMA DE LOS DATOS $Y_i$
IK 6	SUMA DE LOS DATOS $X_i$ e $Y_i$

### REGRESION LINEAL

#### INGRESO DE DATOS

1. PRESIONAR MODE [/] PARA MODO LR
2. PRESIONAR SHIFT SCI EXE PARA BORRAR LAS MEMORIAS ESTADISTICAS.
3. INGRESAR LOS PARES DE DATOS USANDO LA SIGUIENTE OPERACION

(DATO X) SHIFT [,] (DATO Y) DT

Si se tiene un número de pares de datos que son idénticos, se les puede ingresar como un grupo usando uno de los siguientes procedimientos.

### EJEMPLO 1

DATOS: 10/20, 20/30, 20/30, 40/50

INGRESO : 10 SHIFT [,] 20 DT  
          20 SHIFT [,] 30 DT  
          DT  
          40 SHIFT [,] 50 DT

Cada vez que se presiona DT sin ingresar un nuevo par de datos, el par de datos ingresado previamente es ingresado de nuevo (20/30 en ejemplo)

### EJEMPLO 2

DATOS: 10/30, 20/30, 20/30, 20/30, 20/30, 20/30, 40/50

INGRESO: 10 SHIFT [,] 20 DT  
          20 SHIFT [,] 30 SHIFT[;] 5 DT  
          40 SHIFT [,] 50 DT

El valor que se ingresa luego de [;] (5 en el ejemplo) le indica a la calculadora cuántas veces repetir los pares de datos ingresados previamente ( 20/30 en el ejemplo).

### BORRADO DE UN PAR DE DATOS INGRESADOS

El método usado para borrar los datos ingresados dependen de cuándo y cómo los datos fueron ingresados.

**EJEMPLO 1**

10 SHIFT [,] 40 DT  
20 SHIFT [,] 20 DT  
30 SHIFT [,] 30 DT  
40 SHIFT [,] 50

Para borrar 40 SHIFT [,] 50, PRESIONE LA TECLA AC

**EJEMPLO 2**

10 SHIFT [,] 40 DT  
20 SHIFT [,] 20 DT  
30 SHIFT [,] 30 DT  
40 SHIFT [,] 50 DT

Para borrar 40 SHIFT [,], 50 DT, presionar la tecla cl

( Nota  $cl=x^n$  )

**EJEMPLO 3**

10 SHIFT [,] 40 DT  
20 SHIFT [,] 20 DT  
30 SHIFT [,] 30 DT  
40 SHIFT [,] 50 DT

Para borrar 20 SHIFT [,] 20 DT , presione

20 SHIFT [,] 20 CL

El cálculo del coeficiente de correlación  $r$  y de la recta de regresión en el ejemplo estudiado anteriormente se efectua de la manera siguiente:

1. SE PRESIONA MODE [ / ] PARA EL MODO LR (REGRESION LINEAL)

2. SE PRESIONA SHIFT SCI EXE PARA BORRAR LAS MEMORIAS ESTADISTICAS.

3. INGRESO DE LOS DATOS:

12/5.27, 18/5.68, 24/6.25, 30/7.2, 36/8.02,  
42/8.71 y 44/8.42

12	SHIFT	[ / ]	5.27	DT
18	SHIFT	[ / ]	5.68	DT
24	SHIFT	[ / ]	6.25	DT
30	SHIFT	[ / ]	7.2	DT
36	SHIFT	[ / ]	8.02	DT
42	SHIFT	[ / ]	8.71	DT
44	SHIFT	[ / ]	8.42	DT

TERMINO DE CONSTANTE A      SHIFT [A] EXE = 3.78

(Nota: A = 7 ).

COEFICIENTE DE REGRESION B      SHIFT [B] EXE = 0.11

(Nota: B = 8 ).

COEFICIENTE DE CORRELACION r      SHIFT [r] EXE = 0.988

(Nota: r = 9 ).

( NOTA  $\bar{Y} = 4$  y  $\bar{X} = 1$  )

Rendimiento del cultivo Y cuando la cantidad de agua X es igual a 60.

60 SHIFT  $\hat{Y}$  EXE = 10.50 ( Nota:  $\hat{Y} = [ / ]$  )

Cantidad de agua X cuando el rendimiento del cultivo Y es igual a 15

15 SHIFT  $\hat{X}$  EXE = 100.148 ( Nota:  $\hat{X} = [ \times ]$  )

por lo que el coeficiente de correlación  $r = 0.988$

La recta de regresión

$Y = mx + b$  ó  $y = BX + A$  es  $Y = 0.11x + 3.78$

MEMORIA				RESULTADOS DEL CALCULO	
SHIFT	IK	1	EXE	$\sum X_1^2 =$	6940
SHIFT	IK	2	EXE	$\sum X_1 =$	206
SHIFT	IK	3	EXE	N =	7
SHIFT	IK	5	EXE	$\sum Y_1 =$	49.55
SHIFT	IK	6	EXE	$\sum X_1 Y_1 =$	1556.5

## 6.7 EJERCICIOS DE REPASO.

1. Un vitivinicultor observó que las pérdidas de vino durante el añejamiento no eran constantes, sino que variaban de tonel a tonel. Se preguntó si ello ocurría debido a la diferente capacidad de los toneles, o bien había otra causa.

Para satisfacer su curiosidad, obtuvo los datos de los porcentajes de pérdidas en cada tonel según su capacidad, como se reporta a continuación:

Capacidad en $l \times 10^2$	9	2.5	10	16	5	15	7	18
% de pérdida anual	2	4.4	16	15	7.5	7.3	16	11.5

- a) Calcular la covarianza de las variables.
- b) Determinar si existe una buena correlación y en caso afirmativo calcular la función de la recta de regresión.
- c) Comprobar los resultados con la calculadora.

2. Una compañía fabricante de productos farmacéuticos desea conocer si existe correlación entre las ventas de pomada yodada y las ventas totales de la empresa. Para ello disponen de los siguientes datos:

AÑO	VENTAS DE POMADA YODADA \$	VENTAS TOTALES DE LA COMPAÑÍA \$
1968	13,200	34,500
1969	17,400	43,600
1970	19,300	48,900
1971	21,600	54,300
1972	24,000	62,600

- a) Determinar el coeficiente de correlación.
- b) Formular la función de la recta de regresión.
- c) Dibujar las gráficas de la recta y las observaciones.
- d) Comprobar los resultados con la calculadora.

3. El siguiente cuadro nos muestra la altura y el volumen que tienen la principales presas mexicanas\*

NOMBRE	ALTURA (MTS)	VOLUMEN $10^6 \text{ m}^3$
CHICOASEN	240	12.0
INFIERNILLO	148	5.5
LA ANGOSTURA	144	4.4
NETZAHUALCOYOTL	138	8.5
PRESIDENTE ALEMAN	76	8.3
LA AMISTAD	87	11.6
FALCON	50	9.0
MIGUEL HIDALGO	86	10.2
ALVARO OBREGON	90	8.8
LAZARO CARDENAS	97	5.3
ADOLFO LOPEZ MATEOS	107	7.1
VICENTE GUERRERO	63	1.8

a) Calcular el coeficiente de la correlación. Explicar si la correlación es buena o mala, si es positiva o negativa.

b) Encontrar la ecuación de la recta de regresión.

\* Tomado del Comité Nacional de Grandes Presas. Publicado en Revista Ingeniería Civil, septiembre, octubre, 1975  
 Núm. 190, México

c) Graficar las observaciones y verificar visualmente si el punto  $(x,y)$  es el centro de gravedad.

d) Verificar los resultados con la calculadora

4. Considérense diez estudiantes universitarios. A continuación se muestran sus promedios de puntos de calificación en bachillerato y en la universidad.

P.P.C. DEL BACHILLERATO (x)	P.P.C EN LA UNIVERSIDAD (Y)
2.5	2.2
2.6	3.0
2.8	2.4
3.0	2.7
3.2	3.0
3.4	3.0
3.5	3.2
3.6	3.5
3.8	3.4
3.8	3.6

Tomando X como variable independiente y usando la calculadora obtener lo siguiente:

a) El coeficiente de correlación

b) La recta de regresión

## **CAPITULO 7**

### **PROBABILIDAD**

- 7.1 BOSQUEJO HISTORICO
- 7.2 TECNICAS DEL CONTEO, FACTORIAL DE UN NUMERO NATURAL
- 7.3 EL PRINCIPIO FUNDAMENTAL DEL CONTEO Y LAS  
COMBINACIONES
- 7.4 CALCULO DE LAS COMBINACIONES USANDO LA CALCULADORA
- 7.5 EJERCICIOS DE REPASO
- 7.6 ESPACIO MUESTRAL Y EVENTOS
- 7.7 DOS ENFOQUES DE PROBABILIDAD.
- 7.8 AXIOMAS DE PROBABILIDAD
- 7.9 TEOREMAS DE PROBABILIDAD
- 7.10 ESPACIOS FINITOS DE PROBABILIDAD
- 7.11 EJERCICIO DE REPASO
- 7.12 ESPACIOS FINITOS EQUIPROBABLES
- 7.13 PROBABILIDAD CONDICIONAL
- 7.14 TEOREMA DE LA MULTIPLICACION
- 7.15 EVENTOS INDEPENDIENTES
- 7.16 EJERCICIOS DE REPASO

## 7.1 BOSQUEJO HISTORICO

Al servir a la ciencia y a los negocios, las matemáticas de la probabilidad han alcanzado un estado muy superior al de sus orígenes, que estaban ligeramente en el peor de los lados. La teoría de la probabilidad se inspiró en las preguntas de los jugadores que buscaban alguna información interna para ganar en las cartas o en los dados. Tartaglia y Cardano presentaron análisis muy sagaces de los problemas del juego. Pero su trabajo, tal vez demasiado relacionado con el juego para los matemáticos y demasiado matemático para los jugadores, fue olvidado en gran parte. La probabilidad en la forma que la conocemos en la actualidad fue propugnada por un trío de franceses a mediados del siglo XVII: un noble de elevada posición, el caballero De Meré y dos matemáticos esporádicos Blaise Pascal y Pierre de Fermat.

Las principales preocupaciones de Pascal eran la filosofía y la religión. Le entusiasmaba también la geometría "proyectiva", cierta geometría que se refiere a los problemas de perspectiva del dibujo y a las formas de las sombras a que darán lugar las figuras geométricas. Fermat era un jurista de profesión, creó partes de la geometría analítica independientemente de Descartes, pero se le recuerda como uno de los principales teóricos del número de todos los tiempos, una reputación que ganó al quedarse en casa hasta muy

tarde, a la luz de la luna, después de las sesiones en el parlamento local.

En 1651 ó 1652 De Meré y Pascal coincidieron en un viaje a la ciudad de Poitou. Al tratar de encontrar un tópicó de conversación mutuamente interesante con el cual distraerse en el viaje, el mundano De Meré presentó al espiritual Pascal un problema matemático que había producido grandes controversias desde la Edad Media, cómo dividir la polla en un juego de dados que tiene que interrumpirse. Pascal ponderó el problema durante un par de años y finalmente, en 1654, lo comunicó a Fermat para conocer su opinión.

En la célebre correspondencia que siguió al problema puesto por De Meré, Pascal y Fermat empezaron por estar de acuerdo en que, en un juego de dados no terminado, las apuestas en la mesa deberían dividirse según las perspectivas de ganar que tuviese cada jugador. Supongamos por ejemplo, que cada jugador ha apostado 32 doblones (antigua moneda europea de oro) a que el número elegido saldrá tres veces en un dado antes que lo haga el número de otro jugador. Después de que el juego haya seguido durante un rato, el número 6 de De Meré ha salido dos veces, el 4 de su oponente ha salido una sola vez. En este instante De Meré recibe una repentina pregunta; ¿Cómo dividir los 64 doblones de oro de la mesa? El amigo de De Meré podría sostener que dado que

sus posibilidades de conseguir dos tiradas afortunadas son la mitad de probables que las posibilidades de De Meré de obtener una tirada afortunada, tiene derecho a la mitad de lo que corresponde a De Meré, es decir  $21 \frac{1}{3}$  frente a  $42 \frac{2}{3}$  de De Meré. De Meré, por otro lado, podría sostener que en la tirada siguiente del dado lo peor que pudiera sucederle sería perder su ventaja, en cuyo caso, el juego estaría nivelado, y en este supuesto tendría derecho a la división exacta que corresponde a 32 doblones de oro. Si, no obstante, en su próxima tirada tuviera suerte ganaría la apuesta original y recogería la totalidad de los 64 doblones. De Meré arguye, por lo tanto, que incluso antes de la tirada tiene derecho a 32 doblones, más 16 de los que sólo tiene la mitad de la certeza. Y está en lo cierto; Pascal y Fermat lo resolvieron así.

De las investigaciones de Pascal y Fermat en torno a distintas situaciones del juego ha surgido la teoría moderna de la probabilidad: las leyes de la suerte. La idea de que la suerte está regida por leyes puede parecer poco convincente. Pero en verdad las leyes de la probabilidad no impiden la posibilidad de que un individuo tenga una racha de suerte, ni niegan el valor de los presentimientos del juego.

Empiezan a actuar como leyes solo cuando hay muchas repeticiones: al tirar muchas veces los dados, al dar muchas veces las cartas, cuando ocurren muchas colisiones de coches, al considerar las vidas de muchas personas. Este aspecto de la probabilidad se conoce por la ley de los grandes números por las circunstancias especiales que intervienen.

La misma ley da a un individuo sólo una posibilidad remota de tener suerte constantemente, de actuar constantemente mejor de lo que una predicción de probabilidad garantizaría. Por otro lado, una racha de buena suerte no hace disminuir la posibilidad de que un individuo tenga suerte nuevamente en cualquier ocasión determinada; un vendedor que viaja miles de kilómetros al año sin ningún accidente no incurre en un riesgo mayor cada vez que sube a un avión. Las pistas de despegue y el rodar no tienen memoria, y las posibilidades de sobrevivir en un vuelo determinado son tan buenas, la milésima vez, como la primera.

La ley de los grandes números es aplicable hoy a la mayoría de los usos prácticos de la probabilidad. Debido a éstos, la probable exactitud de cualquier provisión aumenta con el número de casos que comprende: el número de moléculas en un recipiente de gas o el número de pólizas de seguro de accidentes suscritos. Esta es una razón por la que las primas son mucho más elevadas en

las pólizas individuales hechas a la medida para cubrir un riesgo particular, que las pólizas ordinarias que puedan extenderse a un número elevado de casos distintos. Por ejemplo, un actuario de seguros puede mirar los archivos referentes al tiempo atmosférico y hallar que, por término medio, en un día del mes de abril en la Ciudad de México, las probabilidades de que llueva son inferiores al 2 por ciento. Pero si un millonario mexicano desea asegurar contra el peligro de lluvia la recepción de la boda de su hija realizada al aire libre, la compañía de seguros no le dará cincuenta casos favorables contra uno, sino que sólo aproximadamente 10 a 1. Es decir, tendría que pagar alrededor de la décima parte del costo de la fiesta. Por otro lado, si un conductor hace la apuesta de 100 dólares con una compañía de seguros de que su coche no dará lugar a 300,000 dólares de daños el año, se le dará un trato mejor. Tendrá que pagar sólo un poco más de lo que corresponde a su participación en todos los daños y perjuicios que hagan colectivamente él y muchos miles de conductores como él.

Las matemáticas de la probabilidad influyen sobre muchos otros factores de la vida moderna. Ayudan al investigador atómico a interpretar las huellas que impresionan en la película las partículas atómicas disparadas desde los ciclotrones. Ayudan al experto en

cohetes a decidir qué factores de seguridad deberían construirse en los costosos sistemas de cohetes dirigidos. Ayudan a valorar a nuestros hijos en las pruebas de inteligencia y hacen posibles las predicciones de votos.

Básicamente, dos leyes son el fundamento de la probabilidad; la ley conjunta para calcular la probabilidad de dos sucesos que se presentan conjuntamente y la ley de exclusividad, para calcular entre dos sucesos la probabilidad de que ocurra uno o el otro.

Regresando a la probabilidad de los juegos, cuando se tira una moneda al aire, ésta asume un estado de impredecibilidad

Nadie puede decir que lado saldrá. A pesar de esto, si se tira aquella moneda un millón de veces, con variaciones cada vez menores, saldrá cara la mitad de veces y cruz la otra mitad. En esencia, ésta es la base de la teoría de la probabilidad, una rama de las matemáticas que se ocupa de las probabilidades, predicciones y de la posibilidad. Enunciadas por primera vez, hace trescientos años, sus aplicaciones más tempranas fueron en el campo del juego, con el que todavía mantiene unos lazos muy fuertes. Pero la

probabilidad (al igual que su asistente, la estadística), se ha convertido en un instrumento moderno indispensable. Pierre Simon de Laplace, prominente en el campo de la probabilidad, la llamó una ciencia que empezó con el juego, pero que evolucionó en "el objeto más importante del conocimiento humano".

Antes de estudiar algunos conceptos de probabilidad estudiaremos algunas técnicas útiles para contar el número de eventos que satisfacen algún conjunto de condiciones.

Estas técnicas son útiles para calcular la probabilidad de un evento, cuando el número total de eventos es grande.

## 7.2 TECNICAS DE CONTEO

### FACTORIAL DE UN NUMERO NATURAL

#### DEFINICION:

$$\begin{aligned}
 0! &= 1 \\
 1! &= 1 \\
 2! &= 2 \times 1 \\
 3! &= 3 \times 2 \times 1 \\
 4! &= 4 \times 3 \times 2 \times 1 \\
 &\vdots \\
 &\vdots \\
 &\vdots \\
 (n-2)! &= (n-2)(n-3) \dots 3 \times 2 \times 1 \\
 (n-1)! &= (n-1)(n-2)(n-3) \dots 3 \times 2 \times 1 \\
 n! &= n(n-1)(n-2) \dots 3 \times 2 \times 1
 \end{aligned}$$

También

$$5! = 5 (4)!$$

$$(n-2)! = (n-2) (n-3)!$$

$$(n-1)! = (n-1) (n-2)!$$

Ejemplo calcular  $n!$  para  $n=6$

$$n! = n(n-1)(n-2)(n-3)(n-4)(n-5)$$

$$6! = 6(6-1)(6-2)(6-3)(6-4)(6-5)$$

$$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

OPERACIONES CON FACTORIALES

$$1) \frac{6!}{4!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1} = 6 \times 5 \quad \circ \quad \frac{6!}{4!} = \frac{6 \times 5 \times 4!}{4!} = 6 \times 5$$

$$2) \frac{15!}{12!} = \frac{15 \times 14 \times 13 \times 12!}{12!} = 15 \times 14 \times 13$$

$$3) \frac{16!}{(16-3)!} = \frac{16 \times 15 \times 14 \times 13!}{13!} = 16 \times 15 \times 14$$

$$4) \frac{14!}{(14-3)! 3!} = \frac{14 \times 13 \times 12 \times 11!}{11! 3!} = \frac{14 \times 13 \times 12}{3!} = \frac{14 \times 13 \times 12}{3 \times 2 \times 1} = 14 \times 13 \times 2$$

$$5) \frac{5!}{(5-5)!} = \frac{5!}{0!} = \frac{5!}{1} = 5! = 5 \times 4 \times 3 \times 2 \times 1$$

$$6) \frac{7!}{(7-0)!} = \frac{7!}{7!} = 1$$

$$7) \frac{10!}{(10-2)! 2!} = \frac{10 \times 9 \times 8!}{8! 2!} = \frac{10 \times 9}{2 \times 1} = \frac{10 \times 9}{2} = 5 \times 9$$

$$8) \frac{(n-1)!}{(n-2)!} = \frac{(n-1)(n-2)!}{(n-2)!} = (n-1)$$

$$9) \frac{n!}{(n-2)!} = \frac{n(n-1)(n-2)!}{(n-2)!} = n(n-1)$$

$$10) \frac{n!}{(n-4)!} = \frac{n(n-1)(n-2)(n-3)(n-4)!}{(n-4)!} = n(n-1)(n-2)(n-3)$$

### 7.3 EL PRINCIPIO FUNDAMENTAL DEL CONTEO Y LAS COMBINACIONES

Si un evento puede realizarse de  $n_1$  maneras diferentes, y si, continuando el procedimiento, un segundo evento puede realizarse de  $n_2$  maneras diferentes, y si, después de efectuados, un tercer evento puede realizarse de  $n_3$  maneras diferentes, entonces el número de maneras en que los eventos puedan realizarse en el orden indicado es el producto  $n_1 \times n_2 \times n_3 \times \dots$

Ejemplo 1. Hallar el número de palabras de tres letras diferentes que se puede formar con las siguientes 6 letras a, b, c, d, e, f.

Representamos las palabras de tres letras por tres cajas:

□ □ □

Ahora la primera letra puede escogerse de 6 formas diferentes; en seguida, la segunda letra se puede escoger de 5 formas diferentes; después de esto, la última letra se puede escoger de 4 formas diferentes, Escribimos cada número en su correspondiente caja como sigue:

6 5 4

Así por el principio fundamental del conteo hay  $6 \times 5 \times 4 = 120$  posibles palabras de 3 letras sin repetición.

Ejemplo 2. ¿De cuántas maneras se puede acomodar una reunión de 4 personas en una fila de 4 sillas?

$$\boxed{4} \boxed{3} \boxed{2} \boxed{1} \quad 4 \times 3 \times 2 \times 1 = 24 \text{ maneras}$$

### COMBINACIONES

Una combinación de  $n$  objetos tomados de  $r$  en  $r$  es una selección de  $r$  objetos de los  $n$  objetos dados, donde el orden no se toma en cuenta, y su fórmula es:

$$C_r^n = \frac{n!}{(n-r)!r!}$$

Para simbolizar las combinaciones se usa  $C_r^n$  o  $\binom{n}{r}$  indistintamente.

Ejemplo 1. Hallar el número de maneras de elegir un comité compuesto de 3 personas de un total de 7 personas

$$C_3^7 = \frac{7!}{(7-3)!3!} = \frac{7 \times 6 \times 5 \times 4!}{4!3!} = \frac{7 \times 6 \times 5}{3 \times 2 \times 1} = 7 \times 5 = 35$$

Se pueden elegir 35 comités diferentes

#### 7.4 CALCULO DE LAS COMBINACIONES USANDO LA CALCULADORA

Para calcular las combinaciones con la calculadora del ejemplo anterior se siguen los siguientes pasos:

1. AC 13 FMLA (en la pantalla)

$$C=n!/(r!x(n-r))$$

$$n=? \quad 0$$

2. 7 EXE (ingrese el valor de n)

$$C=n!/(r!x(n-r))$$

$$r? \quad 0$$

3. 3 EXE (ingrese el valor de r)

$$C= n!/r!x(n-r)$$

$$C= \quad 35$$

Ejemplo 2. De cuántas maneras se puede escoger un comité compuesto de 2 hombres y 2 mujeres de un grupo de 4 hombres y 4 mujeres.

Sin utilizar la fórmula de las combinaciones:

H1,H2,H3 y H4

H1,H2

H1,H3

H1,H4

H2,H3

H2,H4

H3,H4

M1,M2,M3, y M4

M1,M2

M1,M3

M1,M4

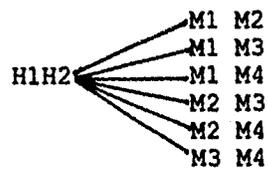
M2,M3

M2,M4

M3,M4

Hay 6 manera de elegir 2 hombres de un grupo de 4 y 6 maneras de elegir 2 mujeres de un grupo de 4. Para formar el comité de 4 personas compuestas de 2 mujeres y

2 hombres, cada una de las seis combinaciones de los hombres se combina con cada una de las seis combinaciones de las mujeres.



Se procede de igual forma para las 5 combinaciones de hombres que faltan, de manera que hay  $6 \times 6 = 36$  comités diferentes que se puedan formar.

Utilizando la fórmula de combinaciones se tiene:

$$C_2^4 \cdot C_2^4 = 6 \times 6 = 36$$

#### 7.5 EJERCICIOS DE REPASO

1. De un grupo de 30 alumnos se desea escoger 3 alumnos  
¿De cuántas maneras diferentes se puede hacer la elección?
2. Un estudiante tiene que contestar 8 de 10 preguntas de un examen.
  - a) ¿Cuántas maneras de escoger tiene?
  - b) ¿Cuántas maneras tiene si las tres primeras son obligatorias?

3. De un lote de 60 artículos, 10 son defectuosos  
¿Cuántas muestras de 6 artículos no defectuosos se pueden obtener?
4. Una granja tiene 10 vacas, 8 puercos, y 4 gallinas.  
¿De cuántas maneras el granjero puede escoger 2 vacas, 3 puercos y 2 gallinas para regalar 7 animales en total?
5. Una clase consta de 7 niños y 4 niñas.  
¿De cuántas maneras un profesor puede elegir un comité donde haya cuatro personas?
- a) Si el comité debe tener dos niñas exactamente
  - b) Un niño exactamente.
  - c) Un niño por lo menos

#### 7.6 ESPACIO MUESTRAL Y EVENTOS

##### ESPACIO MUESTRAL

Brevemente, un espacio muestral es el conjunto universal. Siempre esta relacionado con un experimento.

DEFINICION.- El espacio muestral es el conjunto de todos los resultados posibles de un experimento dado y lo simbolizaremos por S.

### Ejemplos

a) Encontrar el espacio muestral del experimento

"lanzar un dado".

Al lanzar un dado, hay seis posibles resultados

$$S = \{1, 2, 3, 4, 5, 6\}$$

B) Encontrar el espacio muestral del experimento

"lanzar una moneda".

$$S = \{A, S\}$$

**EVENTO:** Evento, es un subconjunto del espacio muestral.

**EVENTO ELEMENTAL:** El evento elemental es un evento que contiene un sólo elemento del espacio muestral.

El conjunto vacío  $\emptyset$  y  $S$  son eventos tal que, el  $\emptyset$  que carece de elementos algunas veces se denomina el evento imposible (o imposibilidad), y  $S$  es el evento cierto o seguro.

Podemos combinar eventos para formar nuevos eventos o eventos compuestos, que consistan en al menos dos eventos.

**EVENTO COMPUESTO:** Un evento compuesto es un subconjunto del espacio muestral, que contiene dos o más eventos. Y

se forman utilizando diferentes operaciones con conjuntos.

a)  $A \cup B$  es el evento que sucede, si A sucede o B sucede o ambos suceden.

b)  $A \cap B$  es el evento que sucede si A y B suceden simultáneamente

c)  $A^C$  (que se lee complemento de A), es el evento que sucede si A no sucede.

EVENTOS MUTUAMENTE EXCLUYENTES: Dos eventos A y B son llamados mutuamente excluyentes, si  $A \cap B = \emptyset$ . En otras palabras, son mutuamente excluyentes si no pueden suceder simultáneamente.

Ejemplo.

Experimento: Láncese un dado y obsérvese el número que aparece en la cara superior. Entonces el espacio muestral es:

$$S = \{1, 2, 3, 4, 5, 6\}$$

Sea A el evento de salir un número par,

B el evento de salir un número impar y

C el evento de salir un número primo

$$A = \{2, 4, 6\}, B = \{1, 3, 5\} \text{ y } C = \{2, 3, 5\}$$

Entonces

$A \cup C = \{2, 3, 4, 5, 6\}$  es el evento de que el número sea par o primo.

$B \cap C = \{3, 5\}$  es el evento de que el número sea impar y primo.

$C^c = \{1, 4, 6\}$  es el evento de que el número no sea primo.

## 7.7 DOS ENFOQUES DE PROBABILIDAD "OBJETIVO Y SUBJETIVO".

### ENFOQUE OBJETIVO DE PROBABILIDAD

Hasta hace muy poco tiempo, la probabilidad era concebida por los estadísticos y matemáticos sólo como un fenómeno objetivo derivado de procesos objetivos.

El concepto de probabilidad objetiva puede caracterizarse aún más, bajo los títulos de:

- 1).- Probabilidad clásica, o a priori.
- 2).- El concepto de probabilidad de frecuencia relativa o a posteriori.

#### 1) PROBABILIDAD CLASICA O A PRIORI.

El estudio de la probabilidad como ya dijimos data del siglo XVII y del trabajo de dos matemáticos, Pascal y Fermat. Gran parte de esta teoría se desarrolló a través

de los intentos para resolver los problemas relacionados con los juegos de azar.

Los ejemplos de los juegos de azar ilustran muy bien los principios que intervienen en la probabilidad clásica. Por ejemplo si se lanza un dado no cargado de seis lados, la probabilidad de que se observe 1 es igual a  $1/6$  y es la misma para las otras cinco caras. Si se elige al azar una carta de una baraja normal bien barajada, la probabilidad de elegir una de corazones es de  $13/52$ . Las probabilidades como éstas se calculan por el proceso de razonamiento abstracto. No es necesario lanzar un dado o tomar una carta para calcular las probabilidades anteriores. En el lanzamiento del dado se dice que es igualmente probable observar cada uno de los seis lados si no existe razón en favor de alguno de ellos. Así mismo, si no existe razón en favor de la elección de una carta en particular de una baraja, se dice que es igualmente probable tomar cada una de las 52 cartas. Puede definirse la probabilidad en el sentido clásico de la siguiente manera:

**Definición:** Si todos los eventos elementales de un fenómeno tienen la misma probabilidad de ocurrir, entonces la probabilidad de ocurrencia de un evento es igual al número de eventos elementales que lo forman

dividido entre el número total de eventos elementales.  
Es decir, para cual quiera que sea el evento E.

$$P(E) = \frac{\text{Número de eventos elementales del evento E}}{\text{Número de eventos elementales del espacio muestral}}$$

Si llamamos a los eventos elementales de un evento E los casos favorables (de E) y al total de eventos los casos posibles, la igualdad anterior se puede escribir en la siguiente forma.

Para cualquier evento E, la probabilidad de ocurrencia de E es:

$$P(E) = \frac{\text{Número de casos favorables}}{\text{Número de casos posibles}}$$

Ejemplo:

Al tirar un dado, encontrar la posibilidad de que salga un 2 o un 5.

El espacio muestral es:  $S = \{1, 2, 3, 4, 5, 6\}$

Si E es el evento de que salga 2 o 5,  $E = \{2, 5\}$

Observamos que el evento E está formado por dos eventos elementales; es decir, el evento E tiene dos casos favorables y dado que hay seis casos posibles, entonces

$$P(E) = \frac{2}{6} = \frac{1}{3} = 0.333$$

Para usar esta definición en el cálculo de probabilidades, se debe cumplir el requisito fundamental que consiste en que todas las probabilidades de los eventos elementales sean iguales; esto equivale, en el ejemplo, a que el dado no este alterado, lo que se cumple si su masa esta repartida homogéneamente en el cubo que lo forma y si este cubo es perfecto.

El uso de la definición clásica para el cálculo de probabilidades presupone que todos los eventos elementales tienen igual probabilidad de ocurrir, lo cual es una suposición teórica. Por lo tanto, las probabilidades así calculadas serán teóricas. Un mal uso de este supuesto nos puede conducir a errores. Por ejemplo, si el dado está alterado, es falso que  $p(E) = 0.333$  puesto que algunas caras caerían hacia arriba más frecuentemente que otras. Pero ¿Cómo saber si el dado que tenemos está alterado?.

La única forma de tener una idea es lanzándolo muchas veces y estimando la probabilidad de ocurrencia de algunos eventos, lo que equivale a utilizar la probabilidad frecuencial

## 2) EL CONCEPTO DE PROBABILIDAD DE FRECUENCIA RELATIVA

### O A POSTERIORI

La aproximación de la frecuencia relativa a la probabilidad depende de la repetición de algún proceso y de la capacidad para contar el número de repeticiones, así como el número de veces que ocurre algún evento de interés. En este contexto, la probabilidad de observar alguna característica, E, de un evento, puede definirse de la manera siguiente.

Definición: Si algún proceso se repite un gran número de veces,  $n$ , y si algún evento resultante con la característica E ocurre  $m$  veces, la frecuencia relativa de ocurrencia de E,  $m/n$ , será aproximadamente igual a la probabilidad de E.

Para expresar esta definición en forma concreta, se escribe:

$$P(E) = \frac{m}{n}$$

Sin embargo, debe tenerse en cuenta que, estrictamente hablando,  $m/n$  es sólo una estimación de  $p(E)$ .

Ejemplo:

Al lanzar una moneda al aire 200 veces y computar los resultados, se obtuvo la siguiente tabla con respecto a las veces que salió águila o sea el evento

$$E = \{\text{salió águila la moneda.}\}$$

La columna encabezada por A indica el número de águilas que aparecen cada 10 tiradas.

La columna  $m = \sum A$  es la suma acumulada de águilas.

La columna  $m/n$  presenta la frecuencia relativa de águilas en  $n$  tiradas. Por ejemplo, en la segunda fila  $m/n = 8/20 = 0.40$ , en la tercera fila  $m/n = 14/30 = 0.47$  que son las frecuencias relativas de águilas en 20 y 30 tiradas.

	1	2	3	4	5	6	7	8	9	10	A	$m = \sum A$	$\frac{m}{n} = \frac{\sum A}{n}$
1	S	A	A	A	S	S	A	S	A	A	6	6	0.60
2	S	S	A	S	S	S	S	A	S	S	2	8	0.40
3	A	A	A	S	S	A	S	A	A	S	6	14	0.47
4	S	A	S	S	A	A	A	A	S	S	5	19	0.48
5	A	S	A	S	S	A	A	A	S	A	6	25	0.50
6	S	A	S	A	A	S	A	S	A	A	6	31	0.52
7	S	A	A	A	A	A	A	S	A	S	7	38	0.54
8	A	A	S	S	A	S	A	A	S	S	5	43	0.54
9	S	S	A	S	A	A	S	S	S	S	3	46	0.51
10	A	S	A	S	A	A	S	A	S	S	5	51	0.51
11	A	S	S	A	S	A	S	A	S	A	5	56	0.51
12	S	A	A	S	A	S	A	A	A	A	7	63	0.53
13	A	S	S	S	A	S	A	S	A	A	5	68	0.52
14	S	S	A	A	S	S	S	A	A	S	4	72	0.51
15	S	A	S	A	S	S	A	S	S	S	3	75	0.50

16	S	S	A	S	S	S	A	S	A	S	3	78	0.49
17	A	S	S	S	A	A	A	S	S	A	5	83	0.49
18	A	S	A	A	S	S	A	A	S	A	6	89	0.48
19	S	A	S	A	S	A	S	A	A	A	6	95	0.50
20	A	S	A	S	S	A	A	S	A	A	6	101	0.505

Observe en este experimento los siguientes puntos:

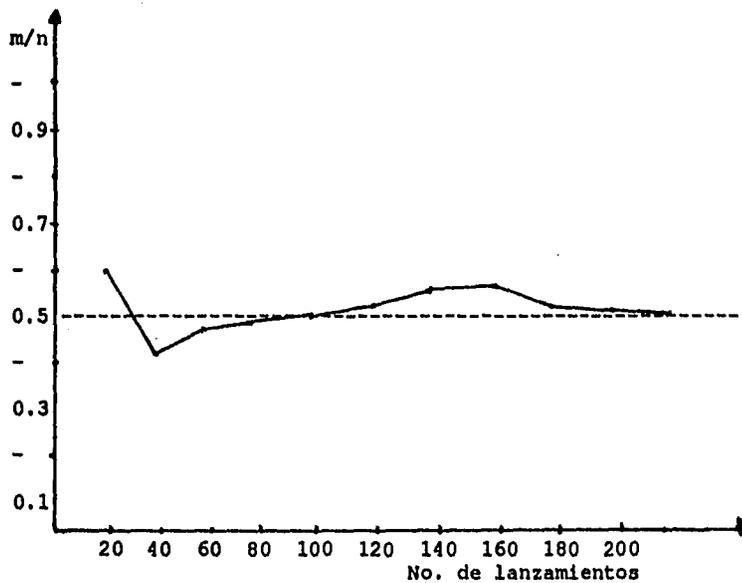
1. Se realiza un experimento (lanzamiento de una moneda) que se repite una y otra vez hasta 200 en condiciones uniformes. La  $n=200$  indica el número de repeticiones del experimento.

2. Hay dos resultados posibles (águila, sol) y se calcula la frecuencia relativa de las águilas aparecidas en diferentes ensayos.

3. Las fluctuaciones de las frecuencias relativas de águilas,  $m/n$  varían considerablemente cuando  $n$  es pequeño, pero cuando  $n$  es grande, la amplitud de las fluctuaciones disminuye. Este fenómeno se expresa diciendo: La frecuencia relativa resulta estable o la frecuencia relativa presenta regularidad estadística a medida que  $n$  crece.

En la siguiente figura se muestra la regularidad estadística mencionada. Gráficamente, en el eje vertical se colocan las frecuencias relativas y en el eje horizontal el número de tiradas  $n$ . En la gráfica se

observa que la amplitud de las fluctuaciones decrece gradualmente cuando  $n$  aumenta: en este ejemplo, tiende a fluctuar alrededor del valor 0.5.



Nótese que se dice tiende a fluctuar alrededor de 0.5 aunque podría hacerlo alrededor de 0.53 o 0.47 o alrededor de cualquier valor. Se ha elegido 0.5 porque se parte del hecho de que cuando no hay fundamentos para preferir uno de los posibles resultados o sucesos a cualquier otro, todos deben considerarse que tienen la misma probabilidad de ocurrir. Así en este caso, en caso de que salga águila o que salga sol tienen la misma probabilidad.

Sin embargo, el principal interés en este punto no está en que la probabilidad sea 0.53, 0.48 o 0.50, sino en ver que cuando el número de experimentos aumenta la frecuencia relativa de las águilas se estabiliza y tiende a un valor determinado. Con esto podemos suponer que cuando el experimento se repite una gran cantidad de veces, la frecuencia relativa podría ser prácticamente igual a un valor  $P$ , con un elevado grado de certeza.

El número  $P$ , que se llama probabilidad del evento  $E$ , se representa por  $P(E)$ . Y  $P(E)=1/2$  en un número tal que la frecuencia  $m/n = 101/200 = 0.505$ , para 200 tiradas, será prácticamente igual a  $P(E)=1/2$  cuando  $n$  es muy grande.

El uso de la definición frecuencial no permite conocer con precisión la posibilidad de ocurrencia de los eventos, sino solamente estimarla mediante muchas observaciones. Por otra parte, aunque la definición clásica permite calcular fácilmente las probabilidades, los resultados obtenidos pueden no concordar con la realidad.

La definición frecuencial de probabilidad considera la experiencia que se tiene de los fenómenos, mientras que la definición clásica los ignora. Es decir, el uso de la definición frecuencial brinda probabilidades a *posteriori*, después de la observación del fenómeno,

mientras que la clásica las brinda a priori, antes de la observación. Esto es, mientras que la clásica es una definición que no considera las condiciones en las cuales se realiza el fenómeno bajo estudio, la frecuencial es una definición que sí considera las condiciones de realización.

#### PROBABILIDAD SUBJETIVA

A principios de la década de 1950, L. J. Savage dió un impulso considerable a lo que conoce como un concepto "personal" de la probabilidad. Este punto de vista sostiene que la probabilidad mide la confianza que tiene determinado individuo en la veracidad de una proposición particular. Este concepto no se basa en la repetición de algún proceso. De hecho, aplicando este concepto de probabilidad, puede evaluarse la probabilidad de un evento que sólo puede ocurrir una vez, por ejemplo, la probabilidad de que se encuentre una cura para el cáncer en los próximos 10 años.

Aunque el punto de vista subjetivo de la probabilidad ha recibido una gran atención desde hace ya varios años, no ha sido completamente aceptado por los estadísticos con orientaciones tradicionales.

## 7.8 AXIOMAS DE PROBABILIDAD

El tratamiento moderno de la teoría de la probabilidad es puramente axiomático. En 1933, el enfoque axiomático de la probabilidad fue formalizado por el matemático ruso, A.N. Kolmogorov. La base de este enfoque está englobado en tres axiomas, a partir de los cuales se construye un sistema completo de la teoría de la probabilidad mediante el uso de la lógica matemática.

A los axiomas también se les llama postulados. Se trata de declaraciones que se aceptan sin demostrar.

Definición: Para cualquier espacio muestral finito  $S = \{a_1, a_2, \dots, a_n\}$ , la probabilidad de un evento E es un número  $P(E)$  tal que cumple los siguientes axiomas de probabilidad.

A1. Para todo evento E,  $0 \leq p(E) \leq 1$

A2.  $P(S) = 1$

A3. Si A y B son eventos mutuamente excluyentes, entonces  $p(A \cup B) = p(A) + p(B)$

El axioma 1 afirma que la probabilidad de un evento puede ser a lo menos cero a lo más uno, o de otra manera una fracción entre cero y uno.

El axioma 2 establece que la probabilidad de un evento seguro (que forzosamente debe ocurrir) es uno. El espacio muestral es un evento seguro ya que contiene todos los resultados posibles de un experimento.

El axioma tres permite determinar la probabilidad de cualquier evento, siempre que se conozcan las probabilidades de los eventos elementales.

#### 7.9 TEOREMAS DE PROBABILIDAD

Los teoremas que veremos se deducen directamente de los axiomas anteriores y aunque no los demostraremos, los ilustraremos mediante ejemplos.

También hay que anotar que los axiomas, así como los teoremas que se estudian en este capítulo, se utilizan para relacionar las probabilidades de distintos eventos entre sí, y son válidos tanto para las estimaciones de probabilidades obtenidas mediante la aplicación de la definición frecuencial, como para las probabilidades calculadas mediante la definición clásica.

Teorema: Si  $\emptyset$  es el conjunto vacío, entonces  $p(\emptyset) = 0$

Teorema: Si  $A^c$  es el complemento del evento  $A$ , entonces  $p(A^c) = 1 - P(A)$

**Teorema:** Si A y B son dos eventos, entonces

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

#### 7.10 ESPACIOS FINITOS DE PROBABILIDAD

**DEFINICION:** Sea S un espacio muestral finitos digamos  $S = \{a_1, a_2, \dots, a_n\}$ . Un espacio finito de probabilidad se obtiene al asignar a cada  $a_i \in S$  un número real  $P_i$  llamado probabilidad de  $a_i$ , que cumple las siguientes propiedades:

- 1) Cada  $P_i$  es no negativa,  $P_i \geq 0$
- 2) La suma de las  $P_i$  es uno,  $\sum P_i = P_1 + P_2 + \dots + P_n = 1$

**EJEMPLOS:**

**Ejemplo 1.-** Al lanzar 3 monedas. Hallar la probabilidad de que:

- a) salga un sol, evento A
- b) Salgan dos soles, evento B
- c) Salgan tres soles, evento C
- d) Salga por lo menos un sol, evento E

El espacio muestral es:

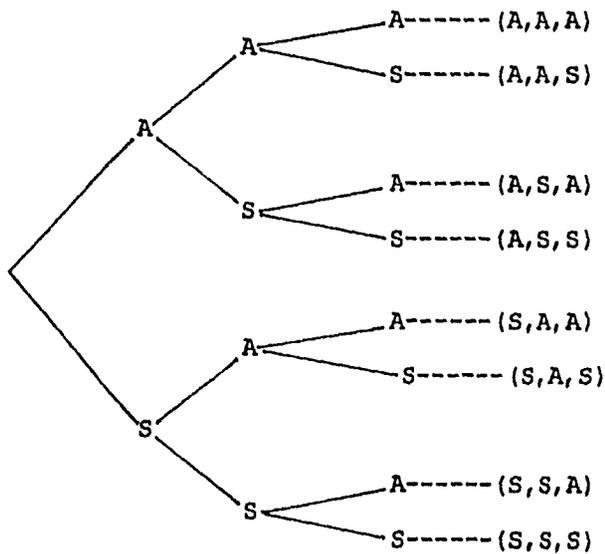
$$S = \{(A, A, A), (A, A, S), (A, S, A), (S, A, A), (S, S, A), (S, A, S), (A, S, S), (S, S, S)\}$$

Si el número de los elementos del espacio muestral de un experimento es  $N(S) = n$ , entonces, el número de eventos de ese experimento es  $2^n$ .

Así en este ejemplo  $2^3 = 8$  nos indica que el espacio muestral consta de 8 resultados.

Otra manera de encontrar el espacio muestral es con la técnica del llamado diagrama de árbol.

DIAGRAMA DE ARBOL



SOLUCION DEL EJEMPLO 1.

a) salga un sol,  $A = \{(A,A,S), (A,S,A), (S,A,A)\}$

entonces  $p(A) = 3/8$

b) salgan dos soles,  $B = \{(A, S, S), (S, A, S), (S, S, A)\}$

entonces  $p(B) = 3/8$

c) salgan 3 soles,  $C = \{(A, A, A)\}$  entonces  $p(C) = 1/8$

d) salga un sol por lo menos,

$E = \{(A, A, S), (A, S, A), (A, S, S), (S, A, A), (S, A, S), (S, S, A),$

$(S, S, S)\}$  entonces  $p(E) = 7/8$

OTRO METODO.

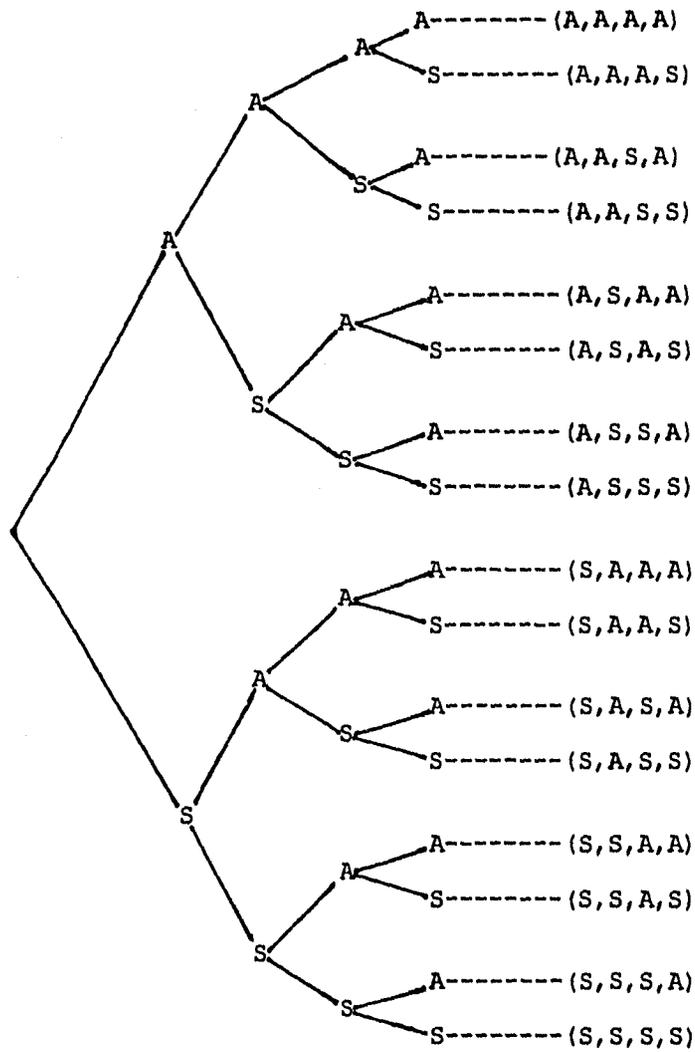
Si  $E^C$  es evento que no salga ningún sol,  $E^C = \{(A, A, A)\}$   
y  $P^C(E) = 1/8$  entonces:

$P(E) = 1 - p(E^C)$ .  $P(E) = 1 - 1/8 = 7/8$

Ejemplo 2. Al lanzar 4 monedas, obtener la probabilidad de que salgan 2 soles.

$2^4 = 16$  nos indica que el espacio muestral consta de 16 resultados.

Utilizando el diagrama de árbol en la siguiente página:



Sea el evento  $H = \{ \text{Que salgan dos soles} \}$ , entonces

$$P(H) = 6/16$$

Ejemplo 3. Al lanzar 5 monedas obtener la probabilidad de que salga por lo menos una águila.

$2^5 = 32$ . El espacio muestral tiene 32 resultados. Ahora si  $H^C$  es el evento que no salga ninguna águila y  $H$  el evento de que salga por lo menos una águila, entonces  $H^C = \{(S,S,S,S,S)\}$  y  $P(H^C) = 1/32$  entonces  $P(H) = 1 - 1/32 = 31/32$

Ejemplo 4. Si se tiran dos dados. Hallar la probabilidad de obtener:

- a) Ocho puntos
- b) De que por lo menos un dado muestre un número par.
- c) El total de puntos mostrados sea divisible por dos o por tres.

El espacio muestral contiene  $6^2 = 36$  resultados

$D_1D_2$	$D_1D_2$	$D_1D_2$	$D_1D_2$	$D_1D_2$	$D_1D_2$
(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

- a) Ocho puntos

$M = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$  entonces  $P(M) = 5/36$

- b) Si  $K = \{\text{Por lo menos un dado muestre un número par}\}$

$P(k) = 27/36$

c) Si  $D = \{\text{El total de puntos mostrados sea divisible por 2 o por 3}\}$   $P(D) = 24/36$

Ejemplo 5.- Supongamos que la probabilidad de que usted llegue a asistir a una universidad es de 0.6; la probabilidad de que trabaje tiempo completo es 0.7 y la probabilidad de que llegue a asistir a una universidad y trabaje tiempo completo es de 0.5 ¿Cuál es la probabilidad de que asista a una universidad o trabaje tiempo completo?

Sea  $P(M) = 0.6$  probabilidad de asistir a una universidad  $P(N) = 0.7$  probabilidad de trabajar tiempo completo  $P(M \cap N) = 0.5$  probabilidad de asistir a una universidad y trabajar tiempo completo.

Entonces la probabilidad de asistir a una universidad o trabajar tiempo completo o  $P(M \cup N)$ , es:

$$P(M \cup N) = P(M) + P(N) - P(M \cap N) = 0.6 + 0.7 - 0.5 = 0.8$$

La probabilidad de que usted ni asista a una universidad ni trabaje tiempo completo es  $p = 1 - 0.8 = 0.20$

### 7.11 EJERCICIOS DE REPASO

1.- Sean A y B eventos con  $P(A) = 3/8$  ,  $P(B) = 1/2$  y

$P(A \cap B) = 1/4$  . Hallar:

a)  $P(A \cup B)$  , b)  $P(A^c)$  y c)  $P(B^c)$

2.- Sean A y B eventos con  $P(A \cup B) = 3/4$  ,  $P(A^c) = 2/3$

y  $P(A \cap B) = 1/4$  . Hallar:

a)  $P(A)$  , b)  $P(B)$

3. Si se tiran tres dados, hallar la probabilidad de que los tres muestren números menores que tres.

4. Tres caballos A, B y C intervienen en una carrera, A tiene el doble de posibilidad de ganar que B, y B el triple de ganar que C. ¿Cuál es la probabilidad de ganar de cada uno de ellos?

5. Una moneda esta cargada (aumentada de peso) de modo que la posibilidad de salir águila sea el triple que la de sol.

Hallar  $p(A)$  y  $P(S)$ .

6.- Tres estudiantes A, B y C intervienen en una prueba de natación, A y B tienen la misma probabilidad de ganar y el doble de la de C. Hallar la probabilidad de que gane B o C.

7.- Hallar la probabilidad de tirar 14 con dos dados de doce caras.

8.- Si se lanzan al aire tres monedas ¿Cuál es la probabilidad de que por lo menos obtengamos dos águilas?

9.- Si se tiran dos dados. Hallar la probabilidad de que:

- a) Por lo menos uno de ellos muestra el 4.
- b) Ambos muestran números menores que 3.

10.- Si la probabilidad de perder cierto juego es 0.4 ¿Cuál es la probabilidad de ganarlo?

11.- Se saca al azar una carta de una baraja americana (52 cartas). Hallar la probabilidad de:

- a) Sacar un rey
- b) Sacar as o reyna

12.- Un psicólogo mete la mano en la jaula y toma una rata al azar, al soltarla en el laberinto. Hallar la probabilidad de que corra.

- a) Hacia la derecha
- b) Hacia la izquierda
- c) Hacia el frente

13.- De 120 estudiantes, 60 estudian francés, 50 estudian español y 20 estudian francés y español. Si se

escoge un estudiante al azar, hallar la probabilidad de que el estudiante.

- a) Estudie sólo francés
- b) Estudie sólo español
- c) Estudie por lo menos un idioma
- d) No estudie ni francés ni español

14.- Suponer que se cruzan dos flores rosadas híbridas (los genes de cada flor contiene un cromosoma rojo y uno blanco). Describir el espacio muestral de la descendencia posible de estas flores y hallar la probabilidad de:

- a) Obtener flores blancas
- b) Obtener flores rojas
- c) Obtener flores rosadas

15.- Un cierto colegio tiene 500 estudiantes y se sabe que:

- 300 leen Francés
- 200 leen Alemán
- 50 leen Ruso
- 20 leen Francés y Ruso
- 30 leen Alemán y Ruso
- 30 leen Alemán y Francés
- 10 leen los tres idiomas

Si se escoge un estudiante al azar de este colegio,

¿Cuál es la probabilidad de que el estudiante:

- i) Lea dos y solamente dos idiomas
- ii) Lea sólo Francés
- iii) Lea cuando menos un idioma
- iv) No lea ningún idioma

16.- Se lanza un dado 50 veces. La tabla siguiente muestra los seis números y la frecuencia con que se repiten.

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

Hallar la frecuencia relativa del evento.

- a) En que aparece un cuatro.
- b) En que aparece un número impar.
- c) En que aparece un número primo.

17.- La probabilidad de que el equipo A gane su primer juego de basquetbol es  $1/2$ , y la probabilidad de que gane su segundo juego es  $1/3$ . ¿Cuál es la probabilidad de que gane por lo menos uno de sus primeros juegos, si la probabilidad de que gane ambos es  $1/6$ ?

18.- Supóngase que una bolsa contiene 10 esferas marcadas 1,2,3...,10. Sea E el evento de extraer una esfera marcada con el número par y F el evento de extraer una esfera marcada con el número 5 o mayor.

- a) ¿Son E y F. mutuamente excluyentes?
- b) Obtenga  $P(E \cup F)$

19.- La probabilidad de que el Sr. Díaz invierta en acciones comunes A es 0.20, en acciones comunes B es 0.30 y en ambas, A y B, 0.10. Hallar la probabilidad de que no invierta ni en A ni en B.

## 7.12 ESPACIOS FINITOS EQUIPROBABLES

Frecuentemente, las características físicas de un experimento sugieren que se asignen iguales probabilidades a los diferentes resultados del espacio muestral. Un espacio finito  $S$  de probabilidad, donde cada punto muestral tiene la misma probabilidad, se llamará espacio equiprobable o uniforme. En particular, si  $S$  contiene  $n$  puntos entonces la probabilidad de cada punto es  $1/n$ . Además si un evento  $A$  contiene  $r$  puntos entonces su probabilidad es  $r \times 1/n = r/n$ . En otras palabras:

$$p(A) = \frac{\text{número de maneras en que el evento } A \text{ puede suceder}}{\text{número de maneras en que el espacio muestral } S \text{ puede suceder.}}$$

Hacemos notar que en la fórmula anterior  $P(A)$  puede utilizarse solamente con respecto a un espacio equiprobable, y no puede usarse en general.

### EJEMPLOS

EJEMPLO 1. Tres tornillos y tres tuercas están en una caja.

Si se escogen 2 piezas al azar. Hallar la probabilidad de sacar un tornillo y una tuerca.

$$S = \begin{matrix} (T01 TU1) & (T02 TU1) & (T03 TU1) & (T01 T02) & (TU1 TU2) \\ (T01 TU2) & (T02 TU2) & (T03 TU2) & (T01 T03) & (TU1 TU3) \\ (T01 TU3) & (T02 TU3) & (T03 TU3) & (T02 T03) & (TU2 TU3) \end{matrix}$$

El espacio muestral S puede ocurrir de 15 maneras.

El evento A puede suceder de 9 maneras

$$A = \begin{matrix} (T01 TU1) & (T02 TU1) & (T03 TU1) \\ (T01 TU2) & (T02 TU2) & (T03 TU2) \\ (T01 TU3) & (T02 TU3) & (T03 TU3) \end{matrix}$$

por lo que  $P(A) = 9/15$

Utilizando combinaciones tenemos:

$$P(A) = \frac{C_1^3 \cdot C_1^3}{C_2^6} = \frac{3 \times 3}{15} = \frac{9}{15}$$

Ejemplo 2. De 10 niños de una clase tres tienen ojos azules. Si se escogen dos niñas al azar. Hallar la probabilidad de que:

- a) Las dos tengan ojos azules
- b) Ninguna tenga ojos azules
- c) Una exactamente tenga ojos azules
- d) Una por lo menos tenga ojos azules

a) Las dos tengan ojos azules

$$P = \frac{C_2^3 \cdot C_0^7}{C_1^{10}} = \frac{3}{45} = \frac{1}{15}$$

b) Ninguna tenga ojos azules.

$$P = \frac{C_0^3 \cdot C_2^7}{C_2^{10}} = \frac{21}{45} = \frac{7}{15}$$

c) Una exactamente tenga ojos azules.

$$P = \frac{C_1^3 C_1^7}{C_2^{10}} = \frac{3 \times 7}{45} = \frac{21}{45} = \frac{7}{15}$$

d) Una por lo menos tenga ojos azules.

$$P = \frac{C_1^3 C_1^7}{C_2^{10}} + \frac{C_2^3 C_0^7}{C_2^{10}} = \frac{21}{45} + \frac{3}{45} = \frac{24}{45} = \frac{8}{15}$$

otra manera de resolver este inciso, con combinaciones:

$$P = 1 - \frac{C_0^3 C_2^7}{C_2^{10}} = 1 - \frac{21}{45} = \frac{45}{45} - \frac{21}{45} = \frac{24}{45}$$

Ejemplo 3. Seis parejas de casados se encuentran en un cuarto. Si se escogen dos personas al azar. Hallar la probabilidad de que:

a) Sean esposos

b) Uno sea hombre y otro mujer

$$a) P = \frac{C_1^6}{C_2^{12}} = \frac{6}{66} = \frac{1}{11}$$

$$b) P = \frac{C_1^6 C_1^6}{C_2^{12}} = \frac{6 \times 6}{66} = \frac{36}{66} = \frac{6}{11}$$

EJEMPLO 4. Una clase tiene 6 niñas y 10 niños. Si se escoge al azar un comité de tres. Hallar la probabilidad de:

a) Seleccionar 3 niños

b) Seleccionar exactamente 2 niños

c) Seleccionar por lo menos un niño.

$$a) \quad P = \frac{C_3^{10} C_0^6}{C_3^{16}} = \frac{120}{560}$$

$$b) \quad P = \frac{C_2^{10} C_1^6}{C_3^{16}} = \frac{45 \times 6}{560} = \frac{270}{560}$$

$$c) \quad P = \frac{C_1^{10} C_2^6}{C_3^{16}} + \frac{C_2^{10} C_1^6}{C_3^{16}} + \frac{C_3^{10} C_0^6}{C_3^{16}} = \frac{150}{560} + \frac{270}{560} + \frac{120}{560} = \frac{540}{560}$$

OTRO PROCEDIMIENTO:

$$p = 1 - \frac{C_0^{10} C_3^6}{C_3^{16}} = 1 - \frac{20}{560} = \frac{560}{560} - \frac{20}{560} = \frac{540}{560}$$

### 7.13 PROBABILIDAD CONDICIONAL

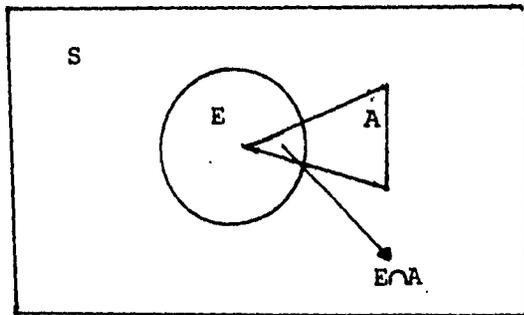
DEFINICION: Sea E un evento arbitrario de un espacio muestral S con  $P(E) > 0$ . La probabilidad de que un evento A suceda una vez que E haya sucedido se define como

$$P(A/E) = \frac{P(A \cap E)}{P(E)}$$

TEOREMA. Sea S un espacio finito equiprobable con eventos A y E, entonces

$$P(A/E) = \frac{\text{Número de maneras en que A y E pueden suceder}}{\text{Número de maneras en que E puede suceder}}$$

El modelo geométrico de la probabilidad condicional se muestra mediante un diagrama de Venn como sigue:



Al definir las probabilidades condicionales, hemos creado en efecto un nuevo espacio muestral  $E$ , el cual es un subconjunto del espacio muestral original  $S$ . Entonces hemos asignado nuevas probabilidades  $P(A/E)$  a los eventos contenidos en  $E$ .

Para justificar esta definición de probabilidad condicional, necesitamos demostrar que los números  $P(A/E)$  en realidad son probabilidades, es decir, satisfacen los tres axiomas de probabilidad

AXIOMA 1  $p(A/E) \geq 0$

$$\text{Si } A \cap E = \emptyset \text{ entonces } P(A/E) = \frac{0}{P(E)} = 0$$

$$\text{Si } A \cap E \neq \emptyset \text{ entonces } P(A/E) = \frac{p(A \cap E)}{p(E)} > 0$$

AXIOMA 2.  $P(E/E) = 1$

$$A=E \Rightarrow A \cap E = E \text{ entonces } P(A/E) = \frac{P(E)}{P(E)} = 1$$

AXIOMA 3. Si  $A \cap E = \emptyset$  (mutuamente excluyentes) entonces  
 $P((A \cup E)/C) = P(A/C) + P(E/C)$

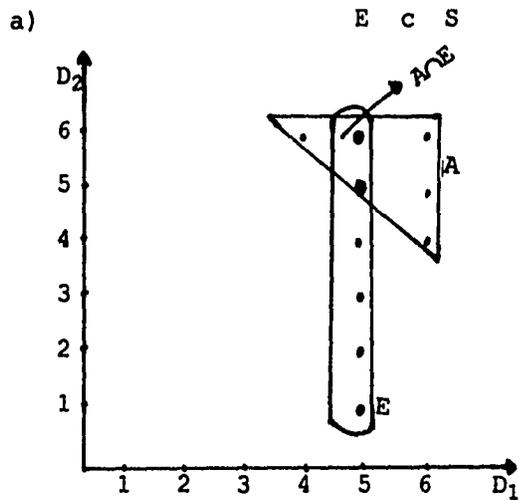
Como  $A \cap E = \emptyset$ , se tiene que  $(A \cap C)$  y  $(E \cap C)$  son mutuamente excluyentes, entonces

$$\begin{aligned} P((A \cup E)/C) &= \frac{P((A \cup E) \cap C)}{P(C)} \\ &= \frac{P((A \cap C) \cup (E \cap C))}{P(C)} \\ &= \frac{P(A \cap C) + P(E \cap C)}{P(C)} \\ &= \frac{P(A \cap C)}{P(C)} + \frac{P(E \cap C)}{P(C)} \\ &= P(A/C) + P(E/C) \end{aligned}$$

### EJEMPLOS

EJEMPLO 1 Si se lanzan dos dados. Hallar la probabilidad de que la suma de sus números sea 10 o mayor si:

- a) Aparece un 5 en el primer dado
- b) Aparece un 5 en uno de los dados



$E = \{\text{aparece un cinco en el primer dado}\}$

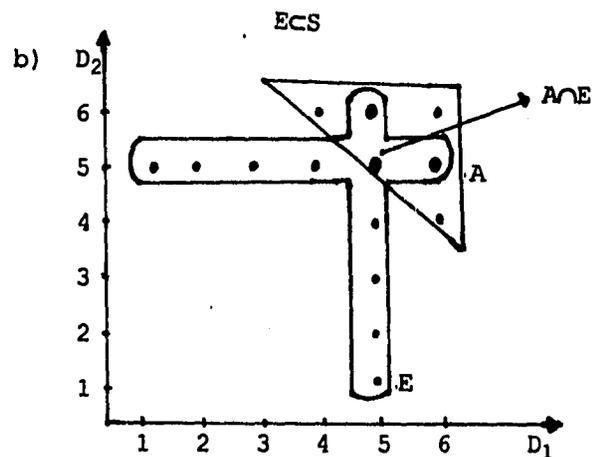
$E = \{(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)\}$

$A = \{\text{La suma de sus números sea 10 o mayor}\}$

$A = \{(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\}$

$A \cap E = \{(5,5), (5,6)\}$

$$P(A/E) = \frac{2}{6}$$



$E = \{\text{Aparece con 5 en uno de los dados por lo menos}\}$

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

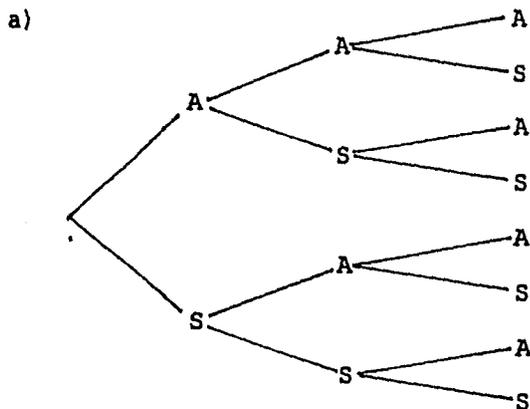
$A = \{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\}$

$A \cap E = \{(5, 5), (5, 6), (6, 5)\}$

$$P(A/E) = \frac{3}{11}$$

EJEMPLO 2. Si se lanzan 3 monedas. Hallar la probabilidad de que todas sean águilas si:

- a) La primera de las monedas es águila
- b) Una de las monedas es águila



$S = \{(A, A, A), (A, A, S), (A, S, A), (A, S, S), (S, A, A), (S, A, S), (S, S, A), (S, S, S)\}$

$E = \{\text{La primera de las monedas es águila}\}$

$E = \{(A, A, A), (A, A, S), (A, S, A), (A, S, S)\}$

$A = \{\text{Todas son águilas}\}$

$A = \{(A, A, A)\}$

$$A \cap E = \{(A, A, A)\}$$

$$P(A/E) = \frac{1}{4}$$

b) E = {Una de las monedas es águila}

$$E = \{(A, A, A), (A, A, S), (A, S, A), (A, S, S), (S, A, A), (S, A, S), (S, S, A)\}$$

A = {Todas las monedas sean águilas}

$$A = \{(A, A, A)\}$$

$$A \cap B = \{(A, A, A)\}$$

$$P(A/E) = \frac{1}{7}$$

EJEMPLO 3. En cierta facultad 25% de los estudiantes reprobaron matemáticas, 15% química y 10% las dos.

a) Si reprobó química. Hallar la probabilidad de que repruebe matemáticas.

b) Si reprobó matemáticas. Hallar la probabilidad de que repruebe química.

$$a) P(M/Q) = \frac{P(M \cap Q)}{P(Q)} = \frac{0.10}{0.15}$$

$$b) P(Q/M) = \frac{P(Q \cap M)}{P(M)} = \frac{0.10}{0.25}$$

## 7.14 TEOREMA DE LA MULTIPLICACION PARA LA PROBABILIDAD

### CONDICIONAL.

Teorema  $P(E \cap A) = P(E) P(A/E)$

Demostración. Por la definición de probabilidad condicional

$$P(A/E) = \frac{P(A \cap E)}{p(E)} \quad \text{con } P(E) > 0$$

Se tiene  $P(E) P(A/E) = P(A \cap E)$  si hacemos  $A \cap E = E \cap A$  entonces  $P(E) p(A/E) = P(E \cap A)$

por lo que  $P(E \cap A) = P(E) P(A/E)$

Por extensión de este teorema tenemos que:

Para los eventos  $A_1, A_2, \dots, A_n$

$$P(A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n) = P(A_1) P(A_2/A_1) P(A_3/A_1 \cap A_2), \dots, \\ P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

### EJEMPLOS

EJEMPLO 1.- Un lote de 10 artículos tiene 3 defectuosos. Si se toman al azar tres artículos del lote uno tras otro. Hallar la probabilidad de que los tres sean no defectuosos.

La probabilidad de que el primer artículo sea no defectuoso es  $7/10$  ya que 7 de los 10 son no defectuosos.

Si el primero es no defectuoso, entonces la probabilidad de que el segundo sea no defectuoso es  $6/9$  puesto que 6 de los 9 sobrantes son no defectuosos.

Si los dos primeros artículos son no defectuosos, entonces la probabilidad de que el tercero sea no defectuoso es  $5/8$  debido a que solamente 5 entre los 8 que quedan son no defectuosos.

Así por el teorema de la multiplicación tenemos

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2/A_1) P(A_3/A_1 \cap A_2) = \frac{7}{10} \times \frac{6}{9} \times \frac{5}{8}$$

EJEMPLO 2. Una urna contiene 7 bolas rojas y 3 bolas blancas. Se sacan 3 bolas de la urna una tras otra. Hallar la probabilidad  $p$  de que las 2 primeras sean blancas y la tercera roja.

$$P = \frac{3}{10} \times \frac{2}{9} \times \frac{7}{8}$$

EJEMPLO 3. Una clase tiene 9 niños y 5 niñas. Si se escogen 3 estudiantes al azar uno tras otro. Hallar la probabilidad de que:

- a) Todos sean niños
- b) El primero y el tercero sean niños y el segundo niña
- c) El primero y el segundo sean niños y el tercero niña

$$\text{a) } P = \frac{9}{14} \times \frac{8}{13} \times \frac{7}{12}$$

$$\text{b) } P = \frac{9}{14} \times \frac{5}{13} \times \frac{8}{12}$$

$$\text{c) } P = \frac{9}{14} \times \frac{8}{13} \times \frac{5}{12}$$

## 7.15 EVENTOS INDEPENDIENTES

Se dice que un evento B es independiente de un evento A si la probabilidad de que B suceda no está influenciada porque A haya o no sucedido.

DEFINICION: Dos eventos A y B son independientes, si, y sólo si:

$P(A \cap B) = P(A) P(B)$  de otro modo son dependientes.

Esta definición es un resultado del teorema de la multiplicación para la probabilidad condicional, ya que  $P(A \cap B) = P(A) P(B/A)$  pero  $P(B/A) = P(B)$  debido a que A y B son independientes

Por lo que  $P(A \cap B) = P(A) P(B)$

De la definición anterior se desprenden los siguientes teoremas.

TEOREMA. Si A y B son eventos independientes con probabilidades  $p(A)$  y  $p(B)$ , entonces:

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

TEOREMA.- Si A y B son eventos independientes, con probabilidades  $P(A)$  y  $p(B)$ , entonces los eventos complementarios  $A^C$  y  $B^C$ , también lo son, y

$$P(A^C \cap B^C) = P(A^C) P(B^C)$$

TEOREMA.- Si A y B son eventos independientes, entonces  $A^C$  y B son independientes y  $P(A^C \cap B) = P(A^C) P(B)$ .

EJEMPLOS

EJEMPLO 1.- La probabilidad de que A dé en el blanco es  $1/3$  y la probabilidad de M es  $1/4$ . Si A y M disparan. Hallar la probabilidad de que se pegue en el blanco.

Si  $P(A) = 1/3$  y  $P(M) = 1/4$ , se busca  $P(A \cup M)$ . Además, la probabilidad de que A o M den en el blanco no depende de que el otro le dé, de manera que el evento A dé en el blanco es independiente de que el evento M dé en el blanco, por lo tanto  $P(A \cap M) = P(A) P(M)$ .

$$\text{Así } P(A \cup M) = P(A) + P(M) - P(A) P(M)$$

$$P(A \cup M) = \frac{1}{3} + \frac{1}{4} - \frac{1}{3} \times \frac{1}{4} = \frac{6}{12} + \frac{3}{12} - \frac{1}{12} = \frac{8}{12} = \frac{2}{3}$$

EJEMPLO 2. - La probabilidad de que un cliente de un restaurante ordene una hamburguesa es 0.65 y la probabilidad de que ordene una cerveza es 0.35. Suponiendo independencia entre los dos eventos, ¿cuál es la probabilidad de que el cliente ordene una hamburguesa y una cerveza?

$$P(H) = 0.65, P(C) = 0.35 \quad P(H \cap C) = 0.65 \times 0.35$$

EJEMPLO 3.- La probabilidad de que un hombre viva 10 años más es  $1/3$ , y la probabilidad de que su esposa viva 10 años más es  $1/4$

Hallar la probabilidad de que:

- Ambos estén vivos dentro de 10 años
- Ninguno este vivo dentro de 10 años
- Al menos uno este vivo dentro de 10 años
- Solamente la esposa este viva dentro de 10 años

Sea  $H = \{ \text{El hombre viva dentro de 10 años} \}$

$M = \{ \text{La esposa viva dentro de 10 años} \}$ , entonces

$$P(H) = 1/3 \quad \text{y} \quad P(M) = 1/4$$

- Calculamos  $P(H \cap M)$ . Puesto que  $H$  y  $M$  son independientes,

$$P(H \cap M) = P(H) \cdot P(M) = \frac{1}{3} \times \frac{1}{4} = \frac{1}{12}$$

- Calculamos  $P(H^C \cap M^C)$

$$\text{como } P(H^C) = 1 - P(H) = 1 - 1/3 = 2/3$$

$$P(M^C) = 1 - P(M) = 1 - 1/4 = 3/4$$

Además  $H^C$  y  $M^C$  son independientes, entonces

$$P(H^C \cap M^C) = P(H^C) \cdot P(M^C) = 2/3 \times 3/4 = 6/12 = 1/2$$

- Calcular  $P(H \cup M) = P(H) + P(M) - P(H \cap M)$  pero como  $H$  y  $M$  son independientes entonces

$$P ( H \cup M ) = P ( H ) + P ( M ) - P ( H ) P ( M )$$

$$P ( H \cup M ) = 1/3 + 1/4 - 1/3 \times 1/4 = 6/12 = 1/2$$

d) Calculamos  $P ( H^C \cap M )$ .

Como  $P ( H^C ) = 1 - p(H) = 1 - 1/3 = 2/3$  y  $H^C$  y  $M$  son independientes, entonces

$$P ( H^C \cap M ) = P ( H^C ) P ( M ) = 2/3 \times 1/4 = 2/12 = 1/6$$

#### 7.16 EJERCICIOS DE REPASO

##### ESPACIOS EQUIPROBABLES

1. De 12 niñas de una clase, 4 tienen pelo negro. Si se escogen tres niñas al azar. Hallar la probabilidad de que:

- a) Los tres tengan pelo negro
- b) Ninguna tenga pelo negro
- c) Una por lo menos tenga pelo negro

2. Tres cucharas y tres tenedores estan en una caja. Si se escogen dos piezas al azar, hallar la probabilidad de sacar una cuchara y un tenedor.

3. Una clase consta de 5 niñas y 12 niños. Si se escoge al azar un comité de 4 hallar la probabilidad de:

- a) Seleccionar cuatro niños
- b) Seleccionar exactamente dos niños

- c) Seleccionar exactamente tres niñas
- d) Seleccionar por lo menos un niño

4. Una caja contiene 20 fusibles buenos y 5 defectuosos. Si seleccionamos al azar 5 fusibles, ¿cuál es la probabilidad de que todos sean buenos?

#### PROBABILIDAD CONDICIONAL

5. Se escogen al azar dos dígitos desde 1 hasta 9. Si la suma es par, hallar la probabilidad de que ambos números sean impares.

6. Sean los eventos A y B con  $P(A) = 1/2$  ,  $P(B) = 1/3$  y  $P(A \cap B) = 1/4$ . Hallar

a)  $P(A/B)$  , b)  $P(B/A)$  , c)  $P(A \cup B)$

7. Se lanza un dado. Si el número es impar, ¿cuál es la probabilidad de que sea primo?

8. Se lanza un par de dados. Si los números que resultan son diferentes, hallar la probabilidad de que su suma sea par.

9. En una escuela el 20% de los alumnos tiene vista defectuosa, el 8% tiene oído defectuoso y el 4% tiene vista y oído defectuoso.

- a) ¿Son los dos eventos (vista y oído defectuosos) eventos independientes?
- b) ¿Cuál es la probabilidad de que un niño tenga oído defectuoso si sabemos que tiene vista defectuosa?
- c) ¿Cuál es la probabilidad de que un niño tenga vista defectuosa, si sabemos que tiene oído defectuoso?
- d) ¿Cuál es la probabilidad de que un niño no tenga oído defectuoso?
- e) ¿Cuál es la probabilidad de que un niño no tenga vista defectuosa?

10. Una clase tiene 12 niños y 6 niñas. Si se escogen tres estudiantes de la clase al azar, uno tras otro. Hallar la probabilidad de que:

- a) Los dos primeros sean niñas y el tercero niño
- b) El primero y el tercero sean del mismo sexo y el segundo del sexo opuesto

## EVENTOS INDEPENDIENTES

11. La probabilidad de que A dé en un blanco es  $1/4$  y la probabilidad de que B dé en el Blanco es  $1/3$ . Si cada uno dispara una vez. Hallar la probabilidad de que sólo uno de ellos le dé al blanco.

12. Una urna A contiene 5 bolas rojas y 3 blancas, y una urna B contiene 2 rojas y 6 blancas. Si se saca una bola de cada urna, ¿cuál es la probabilidad de que las dos sean del mismo color?

13. La caja A contiene 8 artículos de los cuales 3 son defectuosos, y la caja B contiene 5 artículos de los cuales 2 son defectuosos. Se saca al azar un artículo de cada caja, hallar la probabilidad de que:

- a) Ambos artículos sean no defectuosos
- b) De que un artículo sea defectuoso y otro no.

## **CAPITULO 8**

### **DISTRIBUCIONES DE PROBABILIDAD**

**\* DISTRIBUCION BINOMIAL.**

**\* DISTRIBUCION NORMAL.**

**8.1 DISTRIBUCIONES DE PROBABILIDAD**

**8.2 VARIABLES ALEATORIAS**

**8.3 DISTRIBUCIONES DE PROBABILIDAD DE UNA VARIABLE  
DISCRETA**

**8.4 VALOR ESPERADO, VARIANZA Y DESVIACION ESTANDAR DE  
UNA DISTRIBUCION DE PROBABILIDAD DISCRETA**

**8.5 LA DISTRIBUCION BINOMIAL**

**8.6 CALCULO DE LA BINOMIAL USANDO LA CALCULADORA**

**8.7 LA DISTRIBUCION BINOMIAL COMO UNA DISTRIBUCION DE  
PROBABILIDAD DISCRETA**

**8.8 PROPIEDADES DE LA DISTRIBUCION BINOMIAL**

**8.9 LA DISTRIBUCION NORMAL**

**8.10 LA DISTRIBUCION NORMAL UNITARIA O NORMAL ESTANDAR**

**8.11 CALCULO DE AREAS DE LA DISTRIBUCION NORMAL UNITARIA  
USANDO LA CALCULADORA**

**8.12 EJERCICIOS DE REPASO**

## 8.1 DISTRIBUCIONES DE PROBABILIDAD

En este tema relacionaremos los conceptos de probabilidad con lo estudiado en el tema de presentación gráfica de datos.

Trataremos las funciones aleatorias que tienen asociada una sola variable numérica, por lo que los eventos cuya probabilidad de ocurrencia que nos interesarán serán conjuntos numéricos.

## 8.2 VARIABLES ALEATORIAS

Si a cada evento en un experimento probabilístico se le asigna un valor numérico, entonces conforme se observa su resultado se considera una variable aleatoria o fortuita. Este valor numérico será el valor de la variable aleatoria que se estudiará.

**DEFINICION.-** Una variable aleatoria es una variable que toma un valor numérico único para cada uno de los resultados en el espacio muestral de un experimento probabilístico.

En otras palabras, una variable aleatoria se utiliza para indicar el resultado de un experimento probabilístico.

Puede tomar cualquier valor numérico que pertenezca al conjunto de todos los resultados posibles del experimento.

(Se llama "aleatoria" porque el valor que toma es el resultado de un evento fortuito o sujeto al azar).

Cada evento en un experimento probabilístico también debe ser definido de manera que le sea asignado sólo un valor de variable aleatoria, y todo evento debe tener asignado un valor. Típicamente, la variable aleatoria discreta es una cuenta de algo.

El significado de variable aleatoria se aclara con los ejemplos siguientes:

EJEMPLO 1.- Se lanza 8 monedas observándose el número de águilas. La variable aleatoria  $X$  es el número de águilas que caen y pueden tomar valores enteros entre 0 y 8.

EJEMPLO 2.- Se lanzan dos dados, observándose la suma de sus puntos. Una variable asociada a este experimento puede ser el número total de puntos. En este caso, la variable aleatoria  $X$  puede tomar valores ente 2 y 12.

### 8.3 DISTRIBUCIONES DE PROBABILIDAD DE UNA VARIABLE DISCRETA.

Para empezar en el estudio de las distribuciones de probabilidad, se considerará la distribución de probabilidad de una variable aleatoria discreta, que puede definirse como sigue:

DEFINICION.- La distribución de probabilidad de una variable aleatoria discreta es una tabla, gráfica, fórmula o cualquier otro medio que se utilice para especificar todos los valores posibles de una variable aleatoria discreta junto con sus probabilidades respectivas.

#### 8.4 VALOR ESPERADO, VARIANZA Y DESVIACION ESTANDAR DE UNA DISTRIBUCION DE PROBABILIDAD DISCRETA

Recordando que la media aritmética de una distribución de frecuencias es:

$$\bar{X} = \frac{\sum f_i X_i}{n}$$

expresión que puede ser escrita como

$$\bar{X} = \sum X_i \frac{f_i}{n}$$

En otro capítulo se estudió que la probabilidad de un evento es la frecuencia relativa esperada de una ocurrencia. De modo que si sustituimos

$$\frac{f_i}{n} \quad \text{por} \quad p(X_i)$$

puede hallarse la media de una distribución de probabilidad teórica por:

$$\text{valor medio de } X = \sum X_i p(X_i)$$

Recuérdese que la "frecuencia relativa esperada" representa lo que ocurrirá a largo plazo. En consecuencia, la media de  $X$  es la media total de la

población de resultados experimentales. El símbolo para la media de X (en una distribución de probabilidad) es  $\mu$  y la fórmula puede escribirse como:

$$\mu = \sum X_i p(X_i)$$

La media aritmética, el valor esperado o esperanza matemática de una variable aleatoria discreta tienen la misma fórmula, de manera que

$$\mu = E(x) = \sum X_i p(X_i) = X_1 p(X_1) + X_2 p(X_2) + \dots + X_n p(X_n)$$

#### VARIANZA DE DISTRIBUCIONES DISCRETAS

La varianza de distribuciones de probabilidad discretas se define de manera semejante a la varianza de datos muestrales.

$$\sigma^2 = \sum (X_i - \mu)^2 p(X_i)$$

sustituyendo  $\frac{f_i}{n}$  por  $p(X_i)$

como en el caso de la media aritmética  $\mu$ .

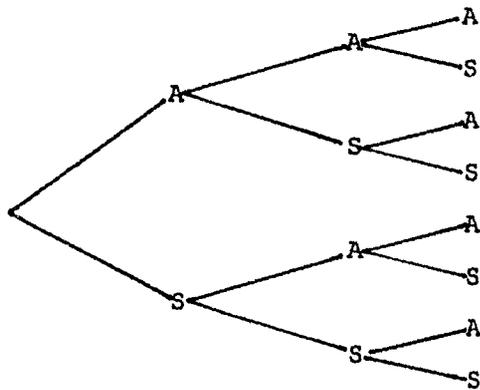
## DESVIACION ESTANDAR DE DISTRIBUCIONES DISCRETAS

La desviación estándar es la raíz cuadrada de la varianza

$$\sigma = \sqrt{\sum p(X_i) (X_i - \mu)^2}$$

### EJEMPLOS

EJEMPLO 1.- Sea  $X$  la variable aleatoria que denota el número de soles obtenidos al lanzar 3 monedas. Hallar su distribución, su histograma, su media aritmética, su varianza y su desviación estándar.



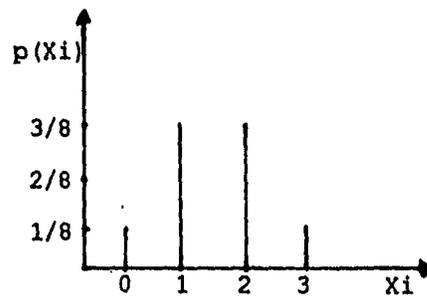
El espacio muestral es

$$S = \{ (A,A,A), (A,A,S), (A,S,A), (A,S,S), (S,A,A), (S,A,S), (S,S,A), (S,S,S) \}$$

DISTRIBUCION DE PROBABILIDAD:

$X_i$	NUMERO DE EVENTOS SIMPLES	$P(X_i)$
0	(A, A, A)	1/8
1	(A, A, S), (A, S, A), (S, A, A)	3/8
2	(A, S, S), (S, A, S), (S, S, A)	3/8
3	(S, S, S)	1/8

HISTOGRAMA DE LA DISTRIBUCION DE  $X_i$



MEDIA ARITMETICA:

$X_i$	$P(X_i)$	$X_i p(X_i)$
0	1/8	0
1	3/8	3/8
2	3/8	6/8
3	1/8	3/8
		12/8

$$\mu = \sum X_i p(X_i) = 12/8 = 3/2 = 1.5$$

LA VARIANZA

Xi	p(Xi)	(Xi-μ)	(Xi-μ) <sup>2</sup>	p(Xi) (Xi-μ) <sup>2</sup>
0	1/8	-3/2	9/4	9/32
1	3/8	-1/2	1/4	3/32
2	3/8	1/2	1/4	3/32
3	1/8	3/2	9/4	9/32
				24/32

$$\sigma^2 = \sum p(Xi) (Xi-\mu)^2 = 24/32 = 3/4 = 0.75$$

LA DESVIACION ESTANDAR :

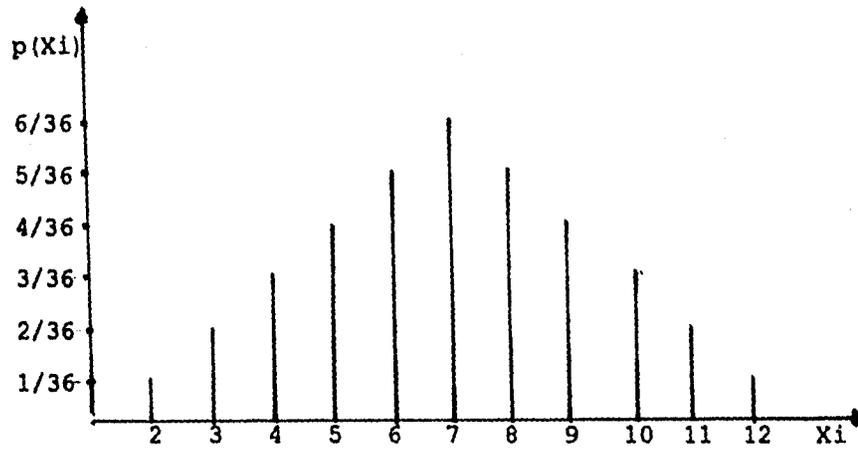
$$\sigma = \sqrt{\sum p(Xi) (Xi-\mu)^2} = \sqrt{0.75} = 0.8660$$

EJEMPLO 2. Al lanzar dos dados, sea x la variable aleatoria que denota la suma de sus puntos mostrados. Hallar su distribución, su media aritmética, su histograma, su varianza y su desviación estandar.

DISTRIBUCION DE PROBABILIDAD:

Xi	NUMERO DE EVENTOS SIMPLES	p(Xi)
2	(1, 1)	1/36
3	(1, 2), (2, 1)	2/36
4	(1, 3), (2, 2), (3, 1)	3/36
5	(1, 4), (2, 3), (3, 2), (4, 1)	4/36
6	(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)	5/36
7	(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)	6/36
8	(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)	5/36
9	(3, 6), (4, 5), (5, 4), (6, 3)	4/36
10	(4, 6), (5, 5), (6, 4)	3/36
11	(5, 6), (6, 5)	2/36
12	(6, 6)	1/36

### HISTOGRAMA DE DISTRIBUCION DE $X_i$



### LA MEDIA ARITMETICA

$X_i$	$p(X_i)$	$X_i P(X_i)$
2	$1/36$	$2/36$
3	$2/36$	$6/36$
4	$3/36$	$12/36$
5	$4/36$	$20/36$
6	$5/36$	$30/36$
7	$6/36$	$42/36$
8	$5/36$	$40/36$
9	$4/36$	$36/36$
10	$3/36$	$30/36$
11	$2/36$	$22/36$
12	$1/36$	$12/36$
		$252/36$

$$\mu = \sum X_i p(X_i) = 252/36 = 7$$

LA VARIANZA

$X_i$	$p(X_i)$	$(X_i - \mu)$	$(X_i - \mu)^2$	$p(X_i) (X_i - \mu)^2$
2	1/36	-5	25	25/36
3	2/36	-4	16	32/36
4	3/36	-3	9	27/36
5	4/36	-2	4	16/36
6	5/36	-1	1	5/36
7	6/36	0	0	0
8	5/36	1	1	5/36
9	4/36	2	4	16/36
10	3/36	3	9	27/36
11	2/36	4	16	32/36
12	1/36	5	25	25/36
				210/36

$$\sigma^2 = \sum p(X_i) (X_i - \mu)^2 = 210/36 = 35/6 = 5.83$$

LA DESVIACION ESTANDAR

$$\sigma = \sqrt{\sum p(X_i) (X_i - \mu)^2} = \sqrt{5.38} = 2.41$$

EJEMPLO 3. Se selecciona al azar una muestra de tres artículos de una caja que contiene 12 de los cuales 3 son defectuosos. Hallar el valor esperado de los artículos defectuosos.

Solución.-

El espacio muestral  $S$  consta de:

$C_3^{12} = 220$  muestras diferentes igualmente posibles de tamaño 3.

$$\text{Cuando } X_1 = 0, P(X_1) = \frac{C_2^3 C_1^3}{C_3^{12}} = \frac{84}{220}$$

$$X_2 = 1, P(X_2) = \frac{C_1^3 C_2^3}{C_3^{12}} = \frac{108}{220}$$

$$X_3 = 2, P(X_3) = \frac{C_1^2 C_2^4}{C_3^{12}} = \frac{27}{220}$$

$$X_4 = 3, P(X_4) = \frac{C_1^1 C_2^5}{C_3^{12}} = \frac{1}{220}$$

Por lo que la probabilidad de que salgan 0, 1, 2 y 3 artículos defectuosos son respectivamente 84/220, 108/220, 27/220 y 1/220

Su distribución de probabilidad es:

$X_i$	$P(X_i)$	$X_i P(X_i)$
0	84/220	0
1	108/220	108/220
2	27/220	54/220
3	1/220	3/220
		165/220

Su valor esperado es  $E(X) = \sum X_i P(X_i) = 165/220 = 0.75$

La distribución binomial es una de las distribuciones de probabilidad que se encuentran con más frecuencia en la estadística aplicada. La distribución se obtiene a partir de un proceso conocido, como el ensayo de Bernoulli en honor al matemático Suizo James Bernoulli (1654-1705), quien hizo contribuciones importantes en el campo de la probabilidad. Cuando en un ensayo de algún proceso o experimento puede conducir solo a uno de dos resultados mutuamente excluyentes, tales como muerto o vivo, enfermo o sano, masculino o femenino etc., se conoce como ensayo de Bernoulli.

Una secuencia de ensayos de Bernoulli forma un proceso de Bernoulli en las siguientes condiciones.

1. Cada ensayo conduce a uno de los dos resultados posibles, mutuamente excluyentes, uno de los resultados posibles se denota (arbitrariamente) como éxito y el otro como fracaso.
- 2, La probabilidad de éxito, denotada por  $p$ , permanece constante de ensayo a ensayo. La probabilidad de fracaso  $1 - p$  se denota por  $q$ , por lo que  $p + q = 1$
3. Los ensayos son independientes, es decir, el resultado de cualquier ensayo particular no es afectado por el resultado de cualquier otro ensayo.

3. Los ensayos son independientes, es decir, el resultado de cualquier ensayo particular no es afectado por el resultado de cualquier otro ensayo.

4. Si estamos interesados en el número de éxitos y no en el orden en que sucedan, entonces aplicamos el siguiente teorema.

TEOREMA. La probabilidad de que K éxitos exactamente en n pruebas repetidas se expresa por.

$$p(K) = b(k, n, p) = C_r^n p^k q^{n-k}$$

k simboliza el número de éxitos.

n simboliza el número de veces que se efectua el experimento.

p simboliza la probabilidad del éxito.

q simboliza la probabilidad del fracaso.

EJEMPLOS:

EJEMPLO 1. Hallar la probabilidad de obtener 2 soles al lanzar 3 monedas.

SOLUCION.

En este caso  $K=2$ ,  $n=3$ ,  $p=1/2$ , y  $q=1-p=1-1/2=1/2$

por lo que  $p(2) = b(2, 3, 1/2) = C_2^3 (1/2)^2 (1/2)^{3-2}$

$$p(2) = 3(1/4)(1/2) = 3/8 = 0.375$$

$$\therefore p(2) = 0.375$$

### 8.6 CALCULO DE LAS BINOMIAL USANDO LA CALCULADORA

Para calcular la binomial en la calculadora se siguen los siguientes pasos. (Tomando los datos del problema anterior).

1. AC 103 FMLA  
(recuperación de la formula de la binomial)

$Px = nCxxPX^{yx}(1$	→
n?	0

2. 3 EXE  
(ingreso del valor n)

$Px = nCxxPX^{yx}(1$	→
x?	0

3. 2 EXE  
(ingreso del valor X)

$Px = nCxxPX^{yx}(1$	→
P?	0

4. 1/2 EXE  
(ingrese el valor de p)

$Px = nCxxPX^{yx}(1$	→
$Px = 0.375$	

$$p(x) = 0.375 \quad \text{o} \quad p(2) = 0.375$$

**EJEMPLO 2.** Al lanzar 6 monedas. Hallar la probabilidad de que:

- a) Salgan 2 águilas exactamente.
- b) Conseguir por lo menos 4 águilas.
- c) No salga ninguna águila.
- d) Salga una águila por lo menos.

Solución.

a) Salgan dos águilas exactamente

$$K=2, n=6, p=1/2 \text{ y } q=1/2$$

$$p(2) = b(2,6,1/2) = C_2^6 (1/2)^2 (1/2)^4 = 15(1/4)(1/16)$$

$$p(2) = 15(1/64) = 15/64 = 0.2343$$

Resultado que se puede comprobar con la calculadora.

b) Conseguir por lo menos cuatro águilas

$$K=4,5,6, n=6, p=1/2 \text{ y } q=1/2$$

$$p(4)+p(5)+p(6) = b(4,6,1/2)+b(5,6,1/2)+b(6,6,1/2)$$

$$p(4) = b(4,6,1/2) = C_4^6 (1/2)^4 (1/2)^2 = 15(1/16)(1/4)$$

$$p(4) = 15/64 = 0.2343$$

$$p(5) = b(5,6,1/2) = C_5^6 (1/2)^5 (1/2) = 6(1/32)(1/2)$$

$$p(5) = 6/64 = 0.0937$$

$$p(6) = b(6,6,1/2) = C_6^6 (1/2)^6 (1/2)^0 = (1)(1/64)(1)$$

$$p(6) = 1/64 = 0.0156$$

$$\begin{aligned} \therefore p(4) + p(5) + p(6) &= 15/64 + 6/64 + 1/64 = 22/64 \\ &= 11/32 = 0.3437 \end{aligned}$$

c) No salga ningún águila.

$$K=0, n=6, p=q=1/2$$

$$p(0) = b(0,6,1/2) = C_0^6 (1/2)^0 (1/2)^6 = 1/64 = 0.015625$$

Todos fracasos equivalen a  $p(0) = q^6 = (1/2)^6 = 1/64$

d) Salga un águila por lo menos.

$k = 1, 2, 3, 4, 5, 6$ ,  $n = 6$ ,  $p = q = 1/2$

$$\begin{aligned} p(1) + p(2) + p(3) + p(4) + p(5) + p(6) &= 1 - p(0) = \\ &= 1 - b(0, 6, 1/2) = 1 - q^6 = 1 - 1/64 = 64/64 - 1/64 = \\ &= 63/64 \end{aligned}$$

Por lo que la probabilidad de por lo menos una águila es igual a uno menos la probabilidad de ninguna águila.

EJEMPLO 3 Supóngase que el 40% de los individuos de una población tiene el grupo sanguíneo B. para una muestra de tamaño 8 extraída de esta población. Encontrar la probabilidad de que se encuentren:

a) Exactamente 3 personas.

b) Menos de 3.

Solución.

a) Exactamente tres personas.

$k = 3$ ,  $n = 8$ ,  $p = 0.4$  y  $q = 0.6$

$$p(3) = b(3, 8, 0.4) = C_3^8 (0.4)^3 (0.6)^5 = 56(0.064)(0.07776) = 0.2787$$

b) Menos de tres.

$K = 0, 1, 2$ ,  $n = 8$ ,  $p = 0.4$  y  $q = 0.6$

$$p(0) + p(1) + p(2) = b(0, 8, 0.4) + b(1, 8, 0.4) + b(2, 8, 0.4)$$

$$P(0) = b(0, 8, 0.4) = q^8 = (0.6)^8 = 0.01679$$

$$P(1) = b(1, 8, 0.4) = C_1^8 (0.4)(0.6)^7 = 8(0.4)(0.02799) = 0.0895$$

$$p(2) = b(2, 8, 0.4) = C_2^8 (0.4)^2 (0.6)^6 = 28 (0.16) (0.0466) = 0.209$$

$$\dots P(0) + P(1) + p(2) = 0.01679 + 0.0895 + 0.209 = 0.3152$$

### 8.7 LA BINOMIAL COMO UNA DISTRIBUCION DE PROBABILIDAD

#### DISCRETA

Ahora si consideramos  $n$  y  $p$  como constantes, entonces la fórmula  $p(k) = b(k, n, p) = C_k^n p^k q^{n-k}$  es una distribución de probabilidad discreta:

K	P(K)
0	$C_0^n q^{n-0} p^0 = q^n$
1	$C_1^n q^{n-1} p^1$
2	$C_2^n q^{n-2} p^2$
3	$C_3^n q^{n-3} p^3$
.	
.	
.	
/	
n	$C_n^n q^{n-n} p^n = p^n$

A esta distribución se le llama distribución binomial, ya que para  $K = 0, 1, 2, 3, \dots, n$  le corresponden los términos sucesivos del desarrollo binomial.

$$(q+p)^n = q^n + C_1^n q^{n-1} p + C_2^n q^{n-2} p^2 + \dots + p^n$$

## 8.8 PROPIEDADES DE LA DISTRIBUCION BINOMIAL

$$\text{Media} = np$$

$$\text{Varianza } \sigma^2 = npq$$

$$\text{Desviación estandar } \sigma = \sqrt{npq}$$

### EJEMPLOS

EJEMPLO 1. La probabilidad de que un artículo producido por una fábrica sea defectuoso es 0.02. Un cargamento de 10,000 artículos se envía a sus almacenes. Hallar el número esperado de artículos defectuosos y la desviación estándar.

$$E(x) = np, \quad E(x) = (10,000)(0.02) = 200$$

$$\sigma = \sqrt{npq}, \quad \sigma = \sqrt{(10,000)(0.02)(0.98)} = \sqrt{196} = 14$$

EJEMPLO 2. El 5% de los tornillos producidos por una fábrica son defectuosos. En una remesa de 400 tornillos. Hallar el valor esperado de tornillos defectuosos.

$$E(x) = np, \quad E(x) = (400)(0.05) = 20$$

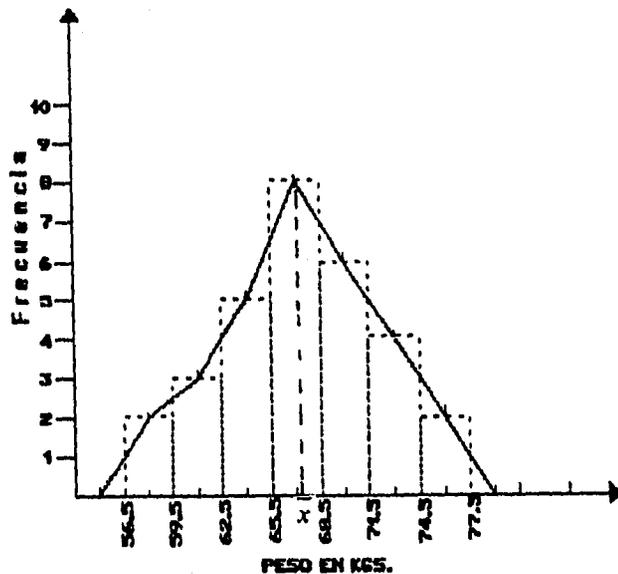
## 8.9 LA DISTRIBUCION NORMAL

### DISTRIBUCIONES CONTINUAS DE PROBABILIDAD

La distribución binomial que estudiamos anteriormente es una distribución de variable aleatoria discreta. Ahora estudiaremos las distribuciones de variables aleatorias continuas. En el capítulo uno se señaló que una variable continua es aquella que puede tomar cualquier valor dentro de un intervalo especificado de valores asumidos por la variable. De manera que, entre cualquiera de los dos valores asumidos por una variable continua, existe un número infinito de valores.

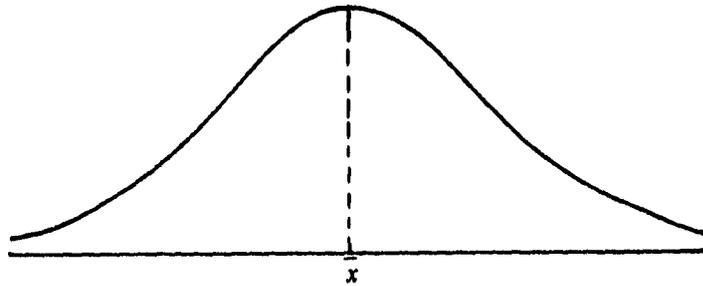
Ejemplos típicos de variables continuas son la altura y el peso de seres humanos, los errores de medición en experimentos de laboratorio, el tiempo de vida de lámparas incandescentes etc.

Para comprender mejor la naturaleza de la distribución de una variable continua, consideraremos, el histograma y el polígono de frecuencias del peso de los estudiantes de estadística del C.C.H.



Observese que el polígono de frecuencias es aproximadamente simétrico respecto a la recta paralela al eje de las ordenadas que pasa por  $\bar{X}$ .

Ahora si aumentamos progresivamente el tamaño de la muestra y el número de intervalos de clase, en el histograma disminuiría el ancho de los rectángulos, y el polígono de frecuencias se parecería cada vez más a una curva en forma de campana. A esta forma que toma el polígono de frecuencias se le conoce con el nombre de polígono de frecuencias suavizado y queda de la siguiente forma:



Si lo que acabamos de decir ocurre para cada muestra con un número suficientemente grande de datos, es de esperar que el polígono de frecuencias de la población sea muy parecido a la curva obtenida anteriormente. Esta es simétrica respecto de la recta paralela al eje de las ordenadas que pasa por el punto de mayor frecuencia de esta curva. El punto donde dicha recta corta el eje de las abscisas coincide con la media de la población.

Se dice que una población cuyo polígono de frecuencias se parece mucho a la figura anterior, tiene una distribución normal.

La distribución normal es la más importante y la más frecuentemente utilizada de toda la estadística. La fórmula para esta distribución fue publicada por primera vez por Abraham De Moivre (1667-1754). Muchos matemáticos figuran de manera prominente en la historia de la distribución normal, incluyendo a Carl Friedrich

Gauss (1777-1855). Con frecuencia a esta distribución se le nombra como distribución gaussiana en reconocimiento a sus contribuciones.

Matemáticamente hablando, la distribución normal o curva normal se define como:

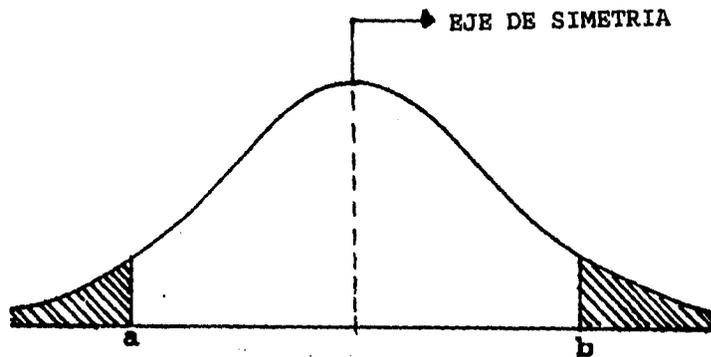
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

En donde  $\pi$  y  $e$  son, respectivamente, las constantes 3.1459 y 2.71828. Los dos parámetros de la distribución son  $\mu$  la media y  $\sigma$ , la desviación estandar, de los valores  $X$  en la población. Para nuestro estudio, nosotros no utilizaremos esta fórmula, ya que podremos encontrar en tablas los resultados que necesitaremos en la resolución de problemas.

Las siguientes propiedades son las características más importantes de la distribución normal.

1. La curva es simétrica respecto a la recta perpendicular al eje de las abscisas que corta a la curva en su punto máximo. Desde este punto la curva cae

simétricamente en las dos direcciones y se extiende sin llegar a tocar el eje de las abscisas. También se dice que la curva hacia cualquiera de los lados de la perpendicular es una imagen de espejo de la del otro lado.

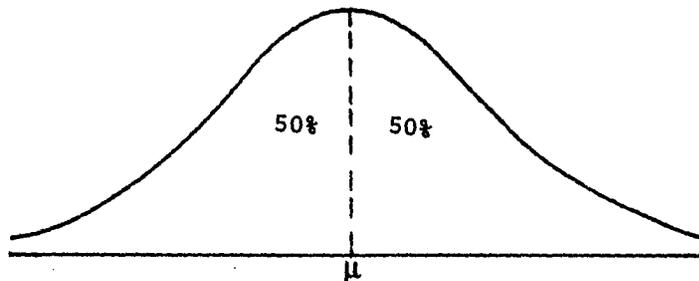


Si **a** y **b** son puntos del eje de las abscisas como se señalan en la figura, se dice que la porciones de áreas bajo la curva que quedan a la izquierda de **a** y a la derecha de **b** son la colas de campana.

2. La media, la mediana y la moda son iguales.

3. El área de la superficie limitada por la curva y el eje de las X es igual a 1. Esta característica se deduce del hecho de que la distribución normal es una distribución de probabilidad. Debido a la simetría mencionada, el 50 por ciento del área esta hacia la

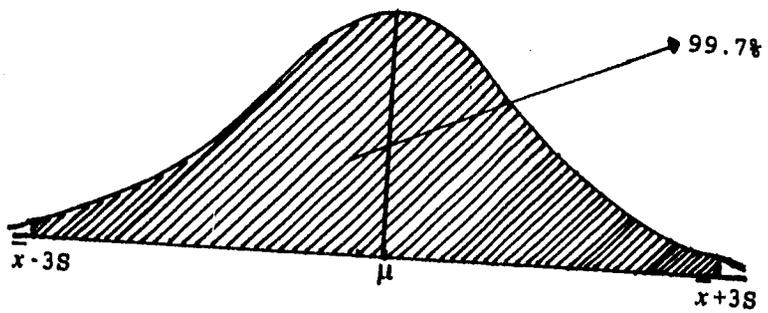
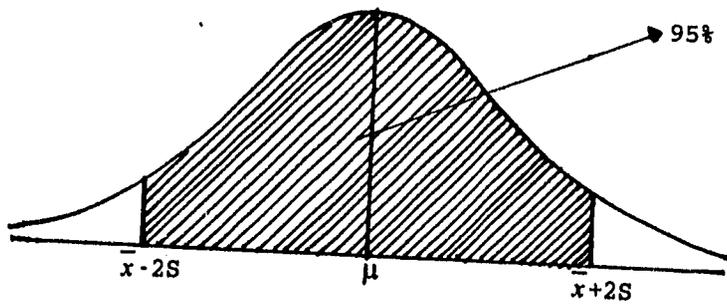
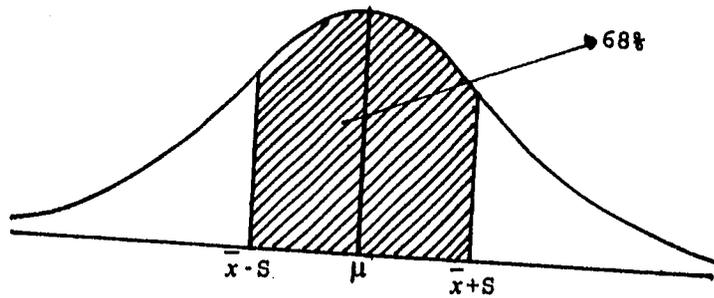
derecha de la perpendicular, trazada en la media, y el 50 por ciento restante está hacia la izquierda.



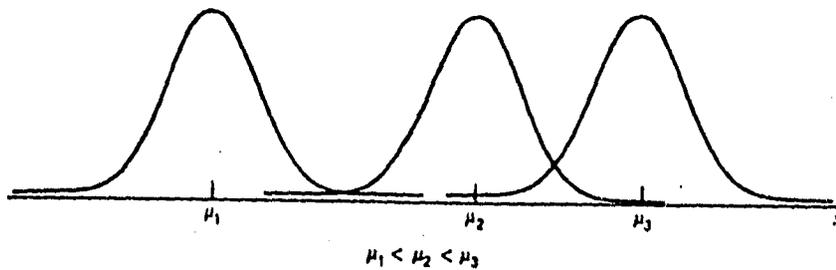
4. Si se trazan perpendiculares a una distancia de una desviación estándar a partir de la media, en ambas direcciones, el área encerrada por estas perpendiculares, el eje de las X y la curva será aproximadamente el 68% del área total.

Si se extienden estos límites laterales hasta una distancia de dos desviaciones estándar, hacia cualquier lado de la media, se encerrará aproximadamente el 95% del área.

Y extendiéndolas hasta una distancia de tres desviaciones estándar, se logrará que aproximadamente el 99.7% del área total quede encerrada. Estas áreas se ilustran en las siguientes figuras.

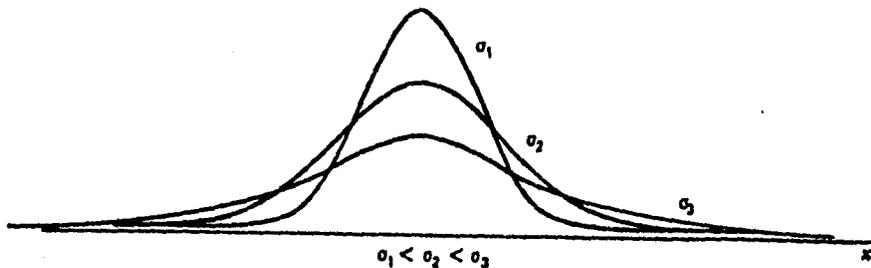


5. La distribución normal está determinada completamente por los parámetros  $\mu$  y  $\sigma$ . En otras palabras, se especifica una distribución normal distinta para cada valor distinto de  $\mu$  y  $\sigma$ . Los valores distintos de  $\mu$  trasladan la gráfica de la distribución a lo largo de los ejes de las X.



Tres distribuciones normales con diferentes medias.

Los diferentes valores de  $\sigma$  determinan el grado de aplanamiento o levantamiento de la gráfica de la distribución.



Tres distribuciones normales con diferentes desviaciones estándar.

## 8.10 LA DISTRIBUCION NORMAL UNITARIA O NORMAL ESTANDAR

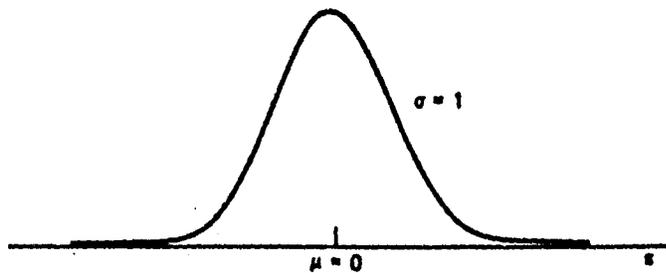
La última característica mencionada de la distribución normal implica que ésta es en realidad una familia de distribuciones en la cual un miembro se distingue de otro con base en los valores de  $\mu$  y  $\sigma$ . El miembro más importante de esta familia es la distribución normal unitaria o normal estandar, llamada así porque tiene una media igual a cero y una desviación estándar igual a uno. Esta distribución puede obtenerse a partir de la ecuación:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

haciendo que  $\mu=0$  y  $\sigma=1$ . La variable aleatoria que resulta,  $(X-\mu)/\sigma$ , se designa generalmente por la letra  $Z$ , de modo que la ecuación para la distribución normal unitaria se escribe como.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

La siguiente gráfica muestra la distribución normal unitaria.

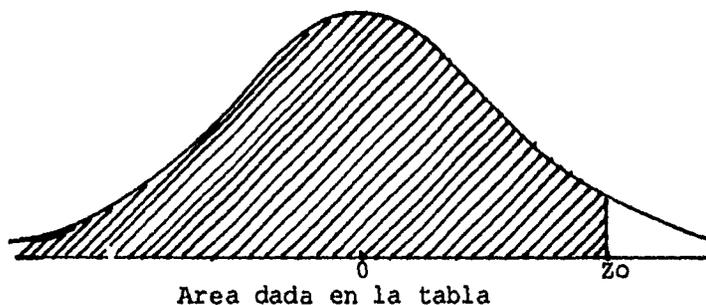


Para encontrar la probabilidad de que  $Z$  tome un valor entre dos puntos cualesquiera sobre el eje  $Z$ , por ejemplo  $Z_0$  y  $Z_1$  debe encontrarse el área limitada por las perpendiculares trazadas en estos puntos, la curva y el eje horizontal. Las áreas bajo la curva de una distribución continua se encuentran integrando la función entre dos valores de la variable. Entonces, en el caso de la normal unitaria, para encontrar el área entre  $Z_0$  y  $Z_1$  se necesita calcular la siguiente integral:

$$\int_{z_0}^{z_1} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

Afortunadamente, no tiene que llevarse a cabo esta integración, dado que se cuenta con tablas que

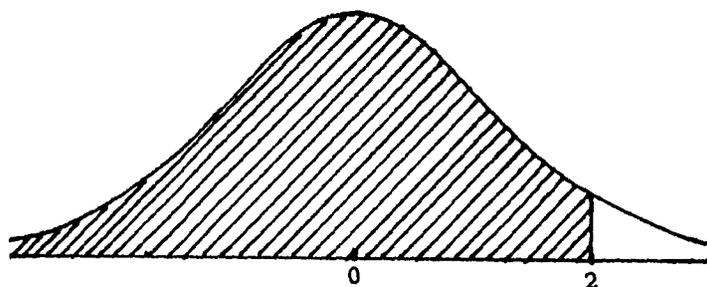
proporcionan los resultados de todas aquellas integraciones que podrían ser interesantes. La tabla que se anexa al final del capítulo es un ejemplo de estas tablas. En dichas tablas se encuentran las áreas bajo la curva entre  $-\infty$  y los valores de  $Z$  mostrados en la columna de la extrema izquierda de la tabla. El área sombreada de la siguiente figura representa el área dada en la tabla como aquellas entre  $-\infty$  y  $Z = Z_0$



#### EJEMPLOS

**EJEMPLO 1.-** Dada la distribución normal unitaria, encuentre el área bajo la curva, para arriba del eje  $Z$ , entre  $-\infty$  y  $Z = 2$ .

**Solución** Es útil trazar una gráfica de la distribución normal unitaria y sombrear el área que se desea, como se muestra.



Distribución normal unitaria que muestra el área entre  $-\infty$  y  $Z = 2$

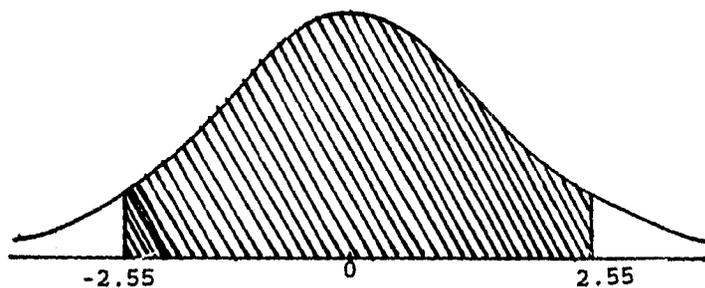
Si se localiza el  $Z = 2$  en la tabla y se lee el valor correspondiente en cuerpo de dicha tabla, se encuentra que el área deseada tiene un valor de 0.9772. Esta área puede interpretarse de varias maneras.

Como la probabilidad de que una  $Z$  seleccionada al azar de la población  $Z$  tenga un valor entre  $-\infty$  y 2.

Como la frecuencia relativa de ocurrencia (o proporción) de los valores de  $Z$  entre  $-\infty$  y 2.

O bien puede decirse que el 97.72 por ciento de las  $Z$  tiene un valor entre  $-\infty$  y 2

EJEMPLO 2. Encontrar la probabilidad de que una Z seleccionada al azar de la población de Z tenga un valor entre -2.55 y +2.55



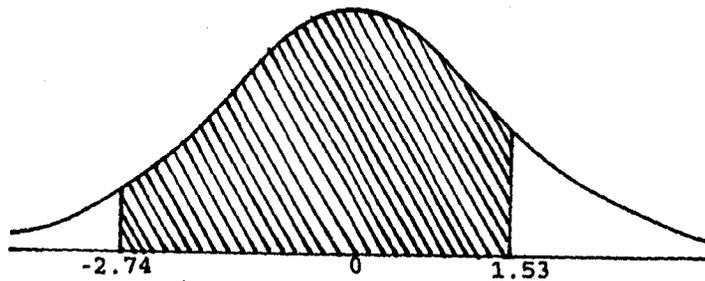
Curva normal unitaria que muestra  $P(-2.55 < Z < 2.55)$

Solución. La tabla da el área entre  $-\infty$  y 2.55, la cual se encuentra localizando el 2.5 en la columna de la extrema izquierda de la tabla y, a continuación, dirigiéndose horizontalmente hasta llegar al valor que se encuentra en la columna encabezada por 0.5. Se encuentra que esta área es de 0.9946. Si se observa la figura que se trazó, se ve que ésta es más de el área deseada. Se necesita restar de 0.9946 el área hacia la izquierda de -2.55. Utilizando la tabla se verifica que el área a la izquierda de -2.55 es 0.0054.

Por lo tanto, la probabilidad deseada es:

$$P(-2.55 < Z < 2.55) = 0.9946 - 0.0054 = 0.9892$$

EJEMPLO 3. ¿ Qué proporción de valores Z están entre -2.74 y 1.53?



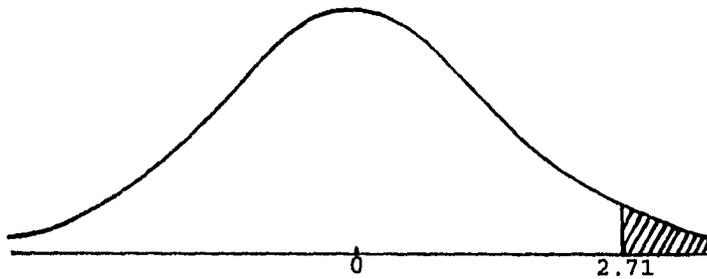
Curva normal unitaria que muestra la proporción de los valores de Z entre  $Z=-2.74$  Y  $Z=1.53$

Solución. En la tabla se encuentra que el área entre  $-\infty$  y 1.53 es 0.9370 y que el área entre  $-\infty$  y -2.74 es 0.0031.

Para obtener la probabilidad deseada, se resta 0.0031 de 0.9370. es decir:

$$P(-2.74 \leq Z \leq 1.53) = 0.9370 - 0.0031 = 0.9339$$

EJEMPLO 4. Dada la distribución normal unitaria, encontrar  $p(Z \geq 2.71)$

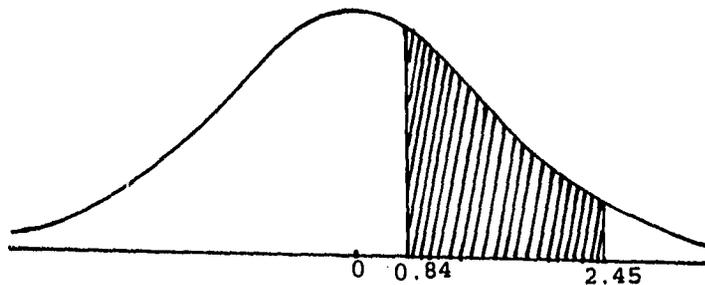


Distribución normal unitaria que muestra  $P(Z \geq 2.71)$

Solución. Se obtiene el área hacia la derecha de  $Z = 2.71$  restando el área entre  $-\infty$  y  $2.71$  de 1. Así, se tiene:

$$p(Z \geq 2.71) = 1 - P(Z \leq 2.71) = 1 - 0.9966 = 0.0034$$

EJEMPLO 5. Dada la distribución normal unitaria, encuentre  $P(0.84 \leq Z \leq 2.45)$



Curva normal unitaria que muestra  $P(0.84 \leq z \leq 2.45)$

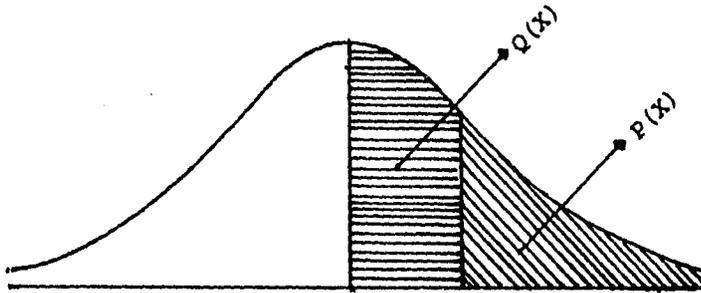
Solución. Primero se obtiene el área entre  $-\infty$  y 2.45 y, de ésta, se resta el área entre  $-\infty$  y 0.84. En otras palabras, se obtiene que:

$$P(0.84 \leq Z \leq 2.45) = P(Z \leq 2.45) - P(Z \leq 0.84)$$

$$= 0.9929 - 0.7995$$

$$= 0.1934$$

8.11 CALCULO DE AREAS DE LA DISTRIBUCION NORMAL UNITARIA USANDO LA CALCULADORA



Para encontrar áreas de la distribución normal unitaria usando la calculadora se procede a seguir los siguientes pasos.

Del ejemplo 2 Encontrar  $P(-2.55 \leq Z \leq 2.55)$

1. AC 110 FMLA  
(Recuperación de función de probabilidad normal)

P(X)	Q (X)
X ?	0

2. 2.55 EXE  
(Ingrese el valor para x)

p(x)	Q (x)
t=0.628659	

3. EXE

p(x)	Q (X)
P(X)=5.38617599	<sup>-03</sup>

4. EXE

P(X)	Q (x)
Q(X)=0.4946138	

Ahora como  $p(0 \leq Z \leq 2.55)$  es igual a  $P(-2.55 \leq z \leq 0)$   
 entonces 0.4946138 se multiplica por dos y se obtiene:  
 $P(-2.55 \leq Z \leq 2.55) = 0.9892$

Del ejemplo 3. ¿ Qué proporción de valores Z estan entre  
 -2.74 y 1.53?

$$P(-2.74 \leq Z \leq 1.53) = P(-2.74 \leq Z \leq 0) + P(0 \leq Z \leq 1.53)$$

PASOS

1. AC 110 FMLA

P(X)	Q(X)
X?	0

2. 2.74 EXE

P(X)	Q(X)
t=	0.611733

3. EXE

P(X)	Q(X)
t(x)=	3.0720166 <sup>-03</sup>

4. EXE

P(x)	Q(X)
Q(X)=	0.496927983

Por lo que  $P(-2.74 \leq Z \leq 0) = 0.4969$

Siguiendo los mismos pasos se encuentran que:

$$P(0 \leq Z \leq 1.53) = 0.43699$$

$$\text{Asi que } P(-2.74 \leq Z \leq 1.53) = 0.4969 + 0.43699 = 0.9339$$

Aplicaciones. Aunque su importancia en el campo de la  
 estadística es indiscutible, debe tenerse en cuenta que

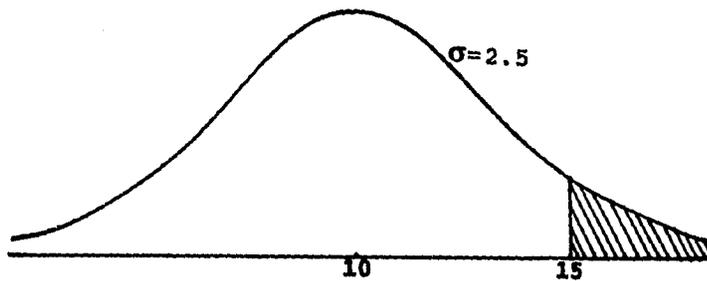
la distribución normal no es una ley a la que se apeguen todas las características medibles que se presentan en la naturaleza. Sin embargo, es cierto que muchas de esas características están distribuidas casi en forma de una distribución normal.

Veamos ahora como pueden contestarse algunas preguntas sencillas de probabilidad acerca de variables aleatorias, cuando se sabe, o se tienen motivos para suponer que, al menos, están distribuidas en forma aproximadamente normal.

#### EJEMPLOS:

**EJEMPLO 1.** Un fisioterapeuta nota que las calificaciones obtenidas en cierta prueba de destreza manual están distribuidas en forma aproximadamente normal con una media 10 y una desviación estandar de 2.5. Si un individuo seleccionado al azar realiza la prueba, ¿cuál es la probabilidad de que obtenga una calificación de 15 o más?

**Solución.** Primero se debe trazar una gráfica de la distribución y sombrear el área correspondiente a la probabilidad de interés.



Distribución normal para aproximar la distribución de las calificaciones en una prueba de destreza manual.

Si la distribución en cuestión fuera la normal unitaria con una media de 0 y una desviación estándar de 1, podría utilizarse la tabla y encontrar la probabilidad con poco esfuerzo. Afortunadamente, cualquier distribución normal puede transformarse en la normal unitaria. Lo que tiene que hacerse es transformar todos los valores de X en los correspondientes valores de Z. Esto significa que la media de X debe hacerse 0, la media de Z. En las siguientes figuras se muestran ambas distribuciones.

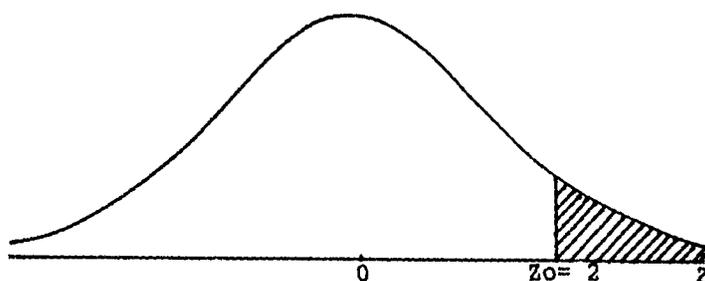
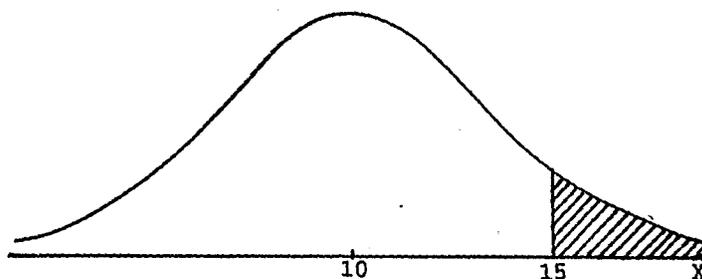
Debe determinarse qué valor de Z, por decir  $Z_0$  corresponde a un X de 15. Esto se lleva a cabo mediante la siguiente fórmula:

$$Z = \frac{X_i - \mu}{\sigma}$$

la cual transforma cualquier valor de X de cualquier distribución normal al valor correspondiente de Z en la distribución normal unitaria. Para nuestro ejemplo se tiene que:

$$Z = \frac{15-10}{2.5} = 2$$

Entonces, el valor de  $Z_0$  que se busca es 2.



Distribución normal de las calificaciones de destreza manual (X) y la distribución normal unitaria (Z).

Examinando ahora estas relaciones con más cuidado. Se observa que la distancia desde la media, 10, hasta el valor de X de interés, 15, es de  $15-10=5$ , que es una distancia de dos desviaciones estándar. Cuando se

transforman los valores X en valores de Z, la distancia del valor Z de interés a partir de su media, 0 es igual a la distancia del valor correspondiente de X a partir de su media, 10, en unidades de desviación estándar. Se vió que esta última distancia es de 2 desviaciones estándar. En la distribución Z, una desviación estándar es igual a uno y, en consecuencia, el punto sobre la escala Z localizado a una distancia de dos desviaciones estándar de 0 es  $Z = 2$  que es el resultado que se obtuvo utilizando la fórmula. Consultando la tabla, se encuentra que el área a la derecha de  $Z = 2$  es 0.0228 .

$$p(Z \geq 15) = p Z \geq \frac{15-10}{2.5} = p ( Z \geq 2 ) = 0.0228$$

Por lo que la probabilidad de que un individuo seleccionado al azar realice la prueba y obtenga una calificación de 15 o más es de 0.0228.

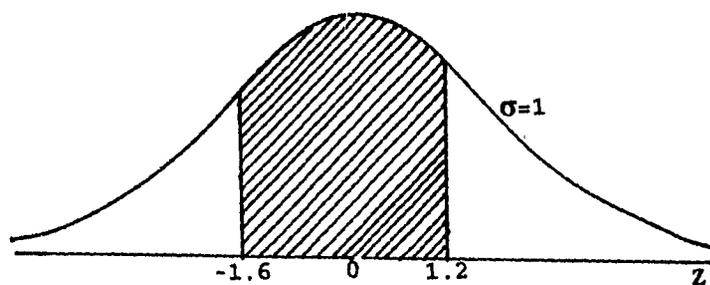
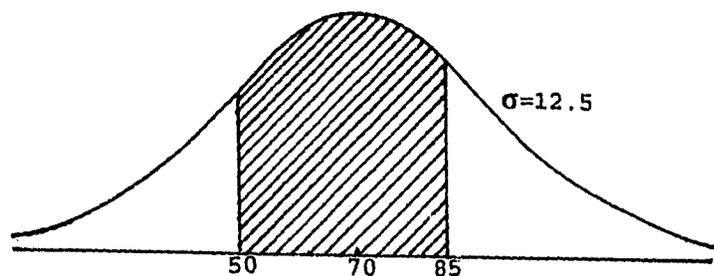
EJEMPLO 2. Supóngase que se sabe que los pesos de cierto grupo de individuos están distribuidos en forma aproximadamente normal con una media de 70 kgs. y una desviación estandar de 12.5 k. ¿Cuál es la probabilidad de que una persona seleccionada al azar de este grupo pese entre 50 y 85 Kgs.

Solución.- Las siguientes gráficas muestran la distribución de los pesos y la distribución Z a la cual se transforman los valores originales, con el fin de

determinar las probabilidades deseadas. Se encuentra que el valor Z correspondiente.

$$\text{A un } X \text{ de } 50 \text{ es. } Z = \frac{50-70}{12.5} = -1.6$$

$$\text{Y para } X=85 \text{ se tiene que } Z = \frac{85-70}{12.5} = 1.2$$



Distribución de pesos (X) y la distribución normal unitaria correspondiente (Z).

En la tabla se encuentra que el área entre  $-\infty$  y  $-1.6$  es de  $0.0548$  y que el área entre  $-\infty$  y  $1.2$  es de  $0.8849$ .

El área deseada es la diferencia entre los valores  $0.9949 - 0.0548 = 0.8301$ .

Para resumir se tiene que:

$$\begin{aligned} P(50 \leq X \leq 85) &= P\left(\frac{50-70}{12.5} \leq Z \leq \frac{85-70}{12.5}\right) = P(-1.6 \leq Z \leq 1.2) \\ &= P(-\infty \leq Z \leq 1.2) - P(-\infty \leq Z \leq -1.6) \\ &= 0.8849 - 0.0548 \\ &= 0.8301 \end{aligned}$$

#### 8.12 EJERCICIOS DE REPASO

1. Al lanzar un dado, sea  $X$  la variable aleatoria que denota al doble del número que aparezca. Hallar su distribución, su media aritmética, su histograma, su varianza y su desviación estandar.

2. Una caja contiene 8 artículos de los cuales 2 son defectuosos. Una persona selecciona 3 artículos de la caja. Hallar el número esperado de artículos defectuosos que ella saca.

3. El número de llamadas  $X$  que se reciben en un conmutador durante cualquier período de 1 minuto es una variable aleatoria y tiene la siguiente distribución de probabilidad.

$x$	0	1	2	3	4
$p(X_i)$	0.1	0.2	0.4	0.2	0.1

Hallar su media y la desviación estándar de  $X$

4. Un jugador de dominó tiene una probabilidad de ganar de  $1/4$  cuando juega. Si juega 5 veces en un torneo. Hallar la probabilidad de que gane exactamente 3 partidos.

5. Supongase que la tasa de mortalidad para cierta enfermedad es 0.10 y que la contraen 6 personas de la comunidad.

Hallar la probabilidad de que:

- a) Ninguna sobreviva
- b) Al menos tres mueran
- c) Exactamente 3 mueran
- d) Por lo menos una muera

6. El equipo de natación del C.C.H. SUR tiene  $3/5$  de ganar cuando compite. Si compite 4 veces, hallar la probabilidad de que gane:

- a) 2 competencias
- b) Una competencia por lo menos
- c) Más de la mitad de las competencias.
- d) Comprobar los resultados con la calculadora.

7. El 2% de televisores producidos por una fábrica son defectuosos. En un despacho de 3600 televisores. Hallar el número esperado de televisores defectuosos y la desviación estandar.

8. Un dado corriente se lanza 1620 veces. Hallar el número esperado de veces que sale el 6 y la desviación estandar.

9. Dada la distribución normal unitaria encuentre:

- a)  $P(Z \geq 0.55)$
- b)  $P(Z \geq -0.55)$
- c)  $P(Z \leq -2.33)$
- d)  $P(-1.96 \leq Z \leq 1.96)$
- e)  $P(-2.58 \leq Z \leq 2.58)$
- f)  $P(-1.2 \leq Z \leq 0.8)$

10. Supóngase que las edades en las que se adquiere cierta enfermedad están distribuidas en forma aproximadamente normal con una media de 11.5 años y una

desviación estándar de 3 años. Un niño acaba de contraer dicha enfermedad. ¿Cuál es la probabilidad de que el niño tenga:

- a) Entre 8.5 y 14.5 años de edad?
- b) Más de 10 años de edad?
- c) Menos de 12?

11. Para cierta porción grande de población. En un año en particular, supóngase que el número de días de incapacidad es de 5.4 con desviación estándar de 2.8 días. Hallar la probabilidad de que una persona elegida al azar tenga incapacidad de:

- a) Mayor de 6 días
- b) Entre 4 y 6 días
- c) Menor de 5 días

12. Si las capacidades de la cavidad craneana de los individuos de cierta población están distribuidos en forma aproximadamente normal con una media de 1,400 cc y una desviación estándar de 125, encuentre la probabilidad de que una persona seleccionada al azar de dicha población tenga una capacidad de cavidad craneana:

- a) Mayor de 1,450 cc
- b) Menor de 1,350 cc
- c) Entre 1,300 y 1,500 cc

13. Si las concentraciones de colesterol total para cierta población están distribuidas en forma aproximadamente normal con una media de 200 mg/100ml y una desviación estándar de 20 mg/100 ml encuentre la probabilidad de que un individuo seleccionado al azar de dicha población tenga una concentración de colesterol:

- a) Entre 180 y 200 mg/100 ml.
- b) Mayor de 225 mg/100 ml.
- c) Menor de 150 mg/100 ml.
- d) Entre 190 y 210 mg/100 ml.

RESPUESTAS A LOS EJERCICIOS DE REPASO

CAPITULO 1 EJERCICIOS 1.7 PAGINA 24

1. ESCALA NOMINAL
2. ESCALA DE RAZON
3. ESCALA ORDINAL
4. ESCALA DE NOMINAL
5. ESCALA DE RAZON
6. RAZON, INTERVALO, ORDINAL Y NOMINAL
  
7. ESCALA DE RAZON
  - a) Límite superior de la sexta clase = 899  
Límite inferior de la última clase = 400
  - b) 949.5
  - c) Límite real inferior de la última clase = 1099.5  
Límite real superior de la última clase = 1199.5
  - d) 100
  - e) 19%
  - f) 65.5%
  - g) 51.5%
  - h) 95%

CAPITULO 4 EJERCICIOS 4.10 PAGINA 105

1.  $\bar{X} = 80.6$
2.  $\bar{X} = 3.982$
3. a)  $\bar{X} = 11,875$   
b) No este promedio no es representativo debido a que un valor extremo esta muy aumentado
5.  $\bar{X} = 83.45$
6.  $\bar{X} = 81.16$

7. a)  $\bar{X} = 40.23$   
 b) Mediana = 40  
 c) Bimodal
8. a)  $\bar{X} = 114.000$   
 b) Mediana = 100,000  
 c) Bimodal  
 d) La más representativa es la mediana
9. a)  $\bar{X} = 65.1$   
 b) Mediana = 65  
 c) Moda = 65
10. a)  $\bar{X} = 18.36$   
 b) Mediana = 14  
 c) Moda = 14
11. POLLO

CAPITULO 4 EJEMPLOS 4.20 PAGINA 137

1. a)  $\bar{X} = 35.23$   
 b) Mediana = 35.63  
 c) Moda = 36.31  
 d)  $Q_1 = 26.375$   
 e)  $Q_3 = 44.15$   
 f)  $P_{10} = 17$   
 g)  $P_{90} = 52.57$
2. a)  $\bar{X} = 715$   
 b) Mediana = 708.32  
 c) Moda = 668.73  
 d)  $Q_1 = 568.46$   
 e)  $Q_3 = 860.79$   
 f)  $P_{10} = 456.02$   
 g)  $P_{90} = 974.5$
3. a)  $\bar{X} = 105.02$   
 b) Mediana = 105.07  
 c) Moda = 105.62  
 d)  $Q_1 = 97.04$   
 e)  $Q_3 = 112.75$   
 f)  $P_{10} = 88.65$   
 g)  $P_{90} = 121.16$

CAPITULO 5 EJERCICIOS 5.7 PAGINA 159

1. a)  $\bar{X} = 4$  , Mediana = 3.5 , Moda = 3

- b)  $R_2=7$   
 c)  $S^2 = 4.22$  ,  $S = 2.054$   
 d) C.V. = 51.37%
2. a)  $\bar{X} = 20.90$  , Mediana = 21 , Moda = Bimodal  
 b)  $R_2 = 12$   
 c)  $S^2 = 16.29$  ,  $S = 4.036$   
 d) C.V. = 19.31%
3. a)  $S = 3.74$  , b)  $S = 4.47$  , c)  $S = 5.13$   
 a) Tiene menor dispersión y c) Tiene mayor dispersión
- 4.-  $S = 0.47$
5. a)  $\bar{X}_2 = 78$  , Mediana = 80 , Moda = 80  
 b)  $S^2 = 113.26$   $S = 10.64$   
 c) C.V. 13.64%

CAPITULO 5 EJERCICIOS 5.16 PAGINA 180

1. a)  $\bar{X}_2 = 19$  , Mediana = 18.7 , Moda=18.42  
 b)  $S^2 = 12.46$  ,  $S = 3.53$   
 c)  $S = 0.16$  sesgada a la derecha ligeramente  
 d)  $K = 0.242$  como  $0.242 < 0.263$  la distribución es leptocúrtica.
2. a)  $S^2 = 168.036$  ,  $S = 12.628$   
 b) Sesgo = -0.085 ligeramente sesgada a la izquierda  
 c)  $K = 0.249$  como  $0.249 < 0.263$  la distribución es leptocúrtica
3. a)  $S^2 = 36,250.375$  ,  $S = 190.39$   
 b) Sesgo = 0.243 sesgada a la derecha  
 c)  $K = 0.2819$ , como  $0.281 > 0.263$  la distribución es platicúrtica.
4. a)  $S = 11.80$   
 b) Sesgo = -0.05 ligeramente sesgada al izquierda  
 c)  $K = 0.241$ , como  $0.241 < 0.263$ , la distribución es leptocúrtica
5. a)  $\bar{X} = 153.68$  ,  $S = 16.45$   
 b)  $(\bar{X}-S, \bar{X}+S) = (137.23, 170.13)$ , 14 Supermercados, 56%  
 $(\bar{X}-2S, \bar{X}+2S) = (120.78, 186.58)$ , 25 Supermercados,  
 Este intervalo cubre el 100% de los supermercados  
 por lo que no es necesario calcular  $(\bar{X}-3S, \bar{X}+3S)$   
 c) Como estos porcentajes no están próximos a lo establecidos en la regla empírica, podemos decir que los datos de la nuestra no están distribuidos de manera aproximadamente normal.

CAPITULO 6 EJERCICIOS 6.7 PAGINA 209

1. a)  $S_{xy} = 9.14$   
b)  $r = 0.35$
2. a)  $r = 0.995$   
b)  $y = 2.57 X - 262.06$
3. a)  $r = 0.25$  , la correlación es casi nula, ya que esta cerca de  $r=0$   
b)  $y = 0.014X + 6.08$
4. a)  $r = 0.87$   
b)  $y = 0.84 X + 0.31$

CAPITULO 7 EJERCICIOS 7.5 PAGINA 226

1. 4060
2. a) 45 , b) 21
3. 15890700
4. 15120
5. a) 126 , b) 28 , c) 329

CAPITULO 7 EJERCICIOS 7.11 PAGINA 248

1. a)  $5/8$  , b)  $5/8$  , c)  $1/2$
2. a)  $1/3$  , b)  $2/3$
3.  $8/216$
4.  $p(A) = 6/10$  ,  $p(B) = 3/10$  ,  $p(C) = 1/10$
5.  $p(A) = 3/4$  ,  $p(S) = 3/4$
6.  $3/15$
7.  $11/144$
8.  $4/8$

9. a)  $11/36$  , b)  $4/36$
10. 0.6
11. a)  $4/52$  , b)  $8/52$
12. a)  $1/3$  , b)  $1/3$  , c)  $1/3$
13. a)  $40/120$  , b)  $30/120$  , c)  $90/120$  , d)  $30/120$
14. a)  $1/4$  , b)  $1/4$  , c)  $2/4$
15. a)  $50/500$  , b)  $260/500$  , c)  $480/500$  , d)  $20/500$
16. a)  $7/50$  , b)  $24/50$  , c)  $26/50$
17.  $4/6$
18. a) No , b)  $8/10$
19. 0.60

CAPITULO 7 EJERCICIOS 7.16 PAGINA 315

1. a)  $p=4/220$  , b)  $56/220$  , c)  $164/220$
2.  $p=9/15$
3. a)  $p=495/2380$ , b)  $660/2380$ , c)  $120/2380$ , d)  $2375/2380$
4.  $p=15504/53130$
5.  $p=10/16$
6. a)  $3/4$  , b)  $2/4$  , c)  $7/12$
7.  $2/3$
8.  $12/30$
9. a) No , b)  $4/20$  , c)  $4/8$  , d)  $92\%$  , e)  $80\%$
10. a)  $p = \frac{6}{18} \times \frac{5}{17} \times \frac{12}{16} = \frac{360}{4896}$

$$b) p = \frac{12}{18} \times \frac{6}{17} \times \frac{11}{16} + \frac{6}{18} \times \frac{12}{17} \times \frac{5}{16} = \frac{1152}{4896}$$

$$11. p = \frac{1}{4} \times \frac{2}{3} + \frac{3}{4} \times \frac{1}{3} = \frac{5}{12}$$

$$12. p = \frac{5}{8} \times \frac{2}{8} + \frac{3}{8} \times \frac{6}{8} = \frac{28}{64}$$

13. a)  $p = 15/40$  , b)  $p = 19/40$

CAPITULO 8 EJERCICIOS 8.12 PAGINA 267

1.  $X_i$     2    4    6    8    10    12  
            $1/6$   $1/6$   $1/6$   $1/6$   $1/6$   $1/6$

$m = 7$  ,  $s^2 = 11.66$  ,  $s = 3.41$

2.  $E(x) = 3/4 = 0.75$

3.  $m = 2$  ,  $s = 1.09$

4.  $p(3) = 0.0878$

5. a)  $p(0) = 0.531$

b)  $p(3) + p(4) + p(5) + p(6) = 0.0158$

c)  $p(3) = 0.0145$

d)  $p(1) + p(2) + p(3) + p(4) + p(5) + p(6) = 0.468$

6. a)  $p(2) = 0.3456$

b)  $p(1) + p(2) + p(3) + p(4) = 0.9744$

c)  $p(3) + p(4) = 0.4752$

7.  $E(X) = 72$  ,  $s = 8.4$

8.  $m = 270$  ,  $s = 15$

9. a) 0.2912  
b) 0.2912  
c) 0.0099  
d) 0.95  
e) 0.99  
f) 0.673

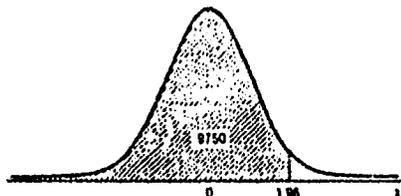
10. a) 0.6826  
b) 0.69146  
c) 0.5661

11. a) 0.41516  
b) 0.27629  
c) 0.4432

12. a) 0.34457  
b) 0.34457  
c) 0.57628

13. a) 0.3413  
b) 0.1056  
c) 0.0062  
d) 0.3829

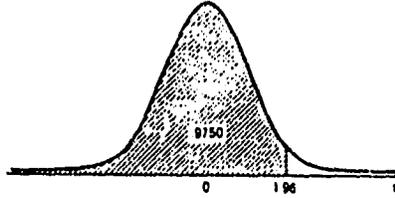
Áreas de la curva normal  $P(z \leq z_0)$  los valores en el cuerpo de la tabla son áreas entre  $-\infty$  y  $z$ .



$z$	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	0.00	$z$
-3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	-3.80
-3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	-3.70
-3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	-3.60
-3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	-3.50
-3.40	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	-3.40
-3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	-3.30
-3.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	-3.20
-3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010	-3.10
-3.00	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013	-3.00
-2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	-2.90
-2.80	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	-2.80
-2.70	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	-2.70
-2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	-2.60
-2.50	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	-2.50
-2.40	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082	-2.40
-2.30	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	-2.30
-2.20	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	-2.20
-2.10	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	-2.10
-2.00	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	-2.00
-1.90	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	-1.90
-1.80	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359	-1.80
-1.70	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446	-1.70
-1.60	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548	-1.60
-1.50	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668	-1.50
-1.40	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808	-1.40
-1.30	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968	-1.30
-1.20	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151	-1.20
-1.10	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357	-1.10
-1.00	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587	-1.00
-0.90	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841	-0.90
-0.80	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119	-0.80
-0.70	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420	-0.70
-0.60	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743	-0.60
-0.50	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085	-0.50
-0.40	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446	-0.40
-0.30	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821	-0.30
-0.20	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207	-0.20
-0.10	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602	-0.10
0.00	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000	0.00



Áreas de la curva normal  $P(z \leq z_0)$  los valores en el cuerpo de la tabla son áreas entre  $-\infty$  y  $z$ .



$z$	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	0.00	$z$
-3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	-3.80
-3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	-3.70
-3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	-3.60
-3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	-3.50
-3.40	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	-3.40
-3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	-3.30
-3.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	-3.20
-3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010	-3.10
-3.00	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013	-3.00
-2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	-2.90
-2.80	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	-2.80
-2.70	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	-2.70
-2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	-2.60
-2.50	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	-2.50
-2.40	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082	-2.40
-2.30	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	-2.30
-2.20	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	-2.20
-2.10	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	-2.10
-2.00	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	-2.00
-1.90	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	-1.90
-1.80	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359	-1.80
-1.70	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446	-1.70
-1.60	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548	-1.60
-1.50	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668	-1.50
-1.40	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808	-1.40
-1.30	.0821	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968	-1.30
-1.20	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151	-1.20
-1.10	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357	-1.10
-1.00	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587	-1.00
-0.90	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841	-0.90
-0.80	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119	-0.80
-0.70	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420	-0.70
-0.60	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743	-0.60
-0.50	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085	-0.50
-0.40	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446	-0.40
-0.30	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821	-0.30
-0.20	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207	-0.20
-0.10	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602	-0.10
0.00	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000	0.00



## BIBLIOGRAFIA BASICA

1. **PROBABILIDAD**  
Lipschutz Seymour  
Mc. Graw-Hill.  
Serie de Compendios Schaum's.
2. **ESTADISTICA**  
Lizárraga Márquez Ignacio  
Mc. Graw-Hill.  
Colección Educación Media N°8.
3. **ESTADISTICA Y METODO CIENTIFICO**  
Méndez Ramírez Ignacio  
Comunicaciones Técnicas.  
Serie Azul N°13.
4. **APUNTES DE ESTADISTICA DEL SEMINARIO SOBRE PRUEBAS DE HIPOTESIS.**  
Méndez Ramírez Ignacio  
C.C.H. Sur, octubre 1977.
5. **ESTADISTICA DESCRIPTIVA APLICADA A LAS CIENCIAS SOCIALES**  
Olguín Quiñones Fernando  
Facultad de Ciencias Politicas y Sociales, U.N.A.M.
6. **ESTADISTICA**  
Spiegel R. Murray  
Mc. Graw-Hill.  
Serie de Compendios Schaum's.
7. **¿DE CUANTAS FORMAS? COMBINATORIA**  
Vilenkin N.  
Editorial MIR.  
Mosú.
8. **BIOESTADISTICA**  
Wayne W. Daniel  
Editorial Limusa, Noriega.
9. **PROBABILIDAD Y ESTADISTICA**  
Willoughby S. Stephen  
Publicaciones Culturales.

## BIBLIOGRAFIA DE CONSULTA

1. INTRODUCCION A LA ESTADISTICA MATEMATICA  
G. Undy Yule y M. G. Kendall  
Aguilar, S.A. de Ediciones.
2. ESTADISTICA GENERAL  
Haber Audrey y Runyon P. Richard  
Addison-Wesley Iberoamericana.
3. ESTADISTICA ELEMENTAL  
Hoel G. Paul  
C.E.C.S.A.
4. PRONTUARIO DE ESTADISTICA CON EJERCICIOS  
Juárez Rodríguez Fernando  
Colegio de Ciencias y Humanidades, U.N.A.M.
5. ESTADISTICA ELEMENTAL  
Johnson Robert  
Grupo Editorial Iberoamericana.
6. MATEMATICAS E IMAGINACION  
Kasner E. y Newman J.  
C.E.C.S.A.
7. INTRODUCCION A LA ESTADISTICA  
Lincon L. Chao  
C.E.C.S.A.
8. PROBABILIDAD  
Lizárraga Márquez Ignacio  
Mc. Graw- Hill.  
Colección Educación Media Superior N°9.
9. INTRODUCCION A LA ESTADISTICA Y A LA PROBABILIDAD  
Menderhall William  
Grupo Editorial Iberoamericana.

10. INTRODUCCION A LA METODOLOGIA ESTADISTICA  
Méndez Rámirez Ignacio  
Chapingo, México.

11. ESTADISTICA  
yamane Taro  
Harla, S:A: de C:V:

12. Ex 1000/Ex 5000  
Scientific Formula 128  
Manual del propietario.