SERIE PUBLICACIONES MISCELANEAS No. 147

MANUAL DE ESTADISTICA DESCRIPTIVA

Víctor Quiroga Mag.Sc.

IICA
INSTITUTO INTERAMERICANO DE CIENCIAS AGRICOLAS – OEA
DIVISION DE PROCESAMIENTO DE DATOS

Programa de Información Agropecuaria del Istmo Centroamericano
PIADIC

San José, Costa Rica
Abril 1977

SERIE PUBLICACIONES MISCELANEAS No. 147

MANUAL DE ESTADISTICA DESCRIPTIVA

Víctor Quiroga Mag.Sc.

IICA

DIVISION DE PROCESAMIENTO DE DATOS

Programa de Información Agropecuaria del Istmo Centroamericano
PIADIC

San José, Costa Rica
Abril 1977

CONTENIDO DE FIGURAS

Figura No.		Página

CONTENIDO DE CUADROS

CONTENIDO

Página

Página

Página

Página

Página

Página

INTRODUCCION

1. SISTEMA.-

Referido a sus componentes elementales, es una colección de celdas intercomunicadas por canales que controlan el flujo de la información. Es la dramatización de la interacción de variables insumo y variables producto. Es un arreglo de las partes, que trabajando juntos, realizan exitosamente un conjunto de operaciones para la consecución de los propósitos del todo.

En los diferentes sistemas, se utilizan representaciones gráficas, extraídas de la teoría general de los gráficos, de ella destacamos:

1.1 DIAGRAMA COMPLETO. Es un conjunto de nudos comunicados entre sí; n nudos, generan un gráfico con $n(n - 1)/2$ canales de comunicación. Así para $n = 5$; el número de canales es 10 i:e:$n(n-1)/2 = 5(4)/2 = 10$.

1.2 DIAGRAMA DE ARBOL. La condición para recibir este nombre es que debe tener exactamente n nudos y $n - 1$ canales de comunicación.

1.3 DIAGRAMA DE CADENA. Es fundamentalmente un diagrama de árbol, pero sin ramificaciones.

1.4 DIAGRAMA DE CICLO. Es la representación gráfica con n nudos y $n - 1$ canales de comunicación.

1.5 DIAGRAMA DE RUTAS. Cuando los canales de comunicación, indican la dirección y el sentido del flujo de la información.

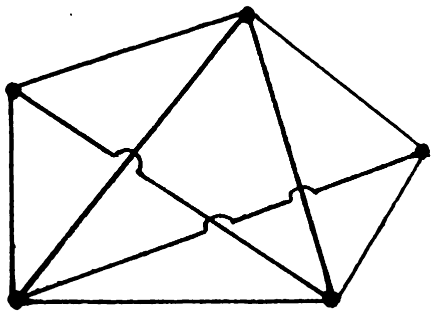


Figura 1. Diagrama Completo

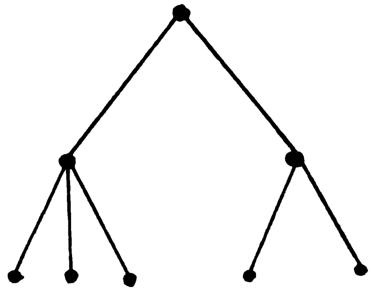


Figura 2. Diagrama de Arbol

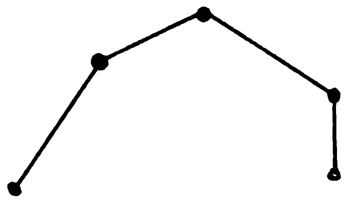


Figura 3. Diagrama de Cadena

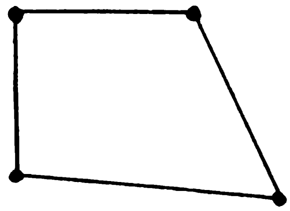


Figura 4. Diagrama de Ciclo

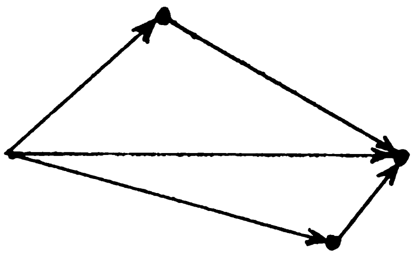


Figura 5. Diagrama de Rutas

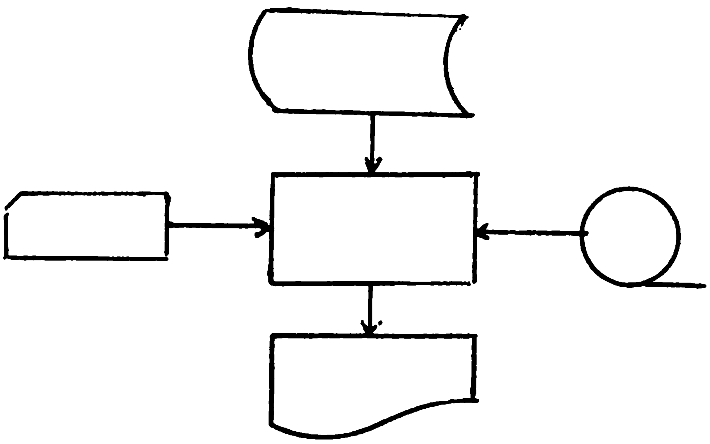


Figura 6. Sistema de Procesamiento de datos

Con los elementos definidos anteriormente, se esquematiza un Sistema de Procesamiento de Datos en la Figura 6.

1.6 ILUSTRACION DE SISTEMA. El Método Científico tal como lo define O. Kempthorn, se puede expresar simbolicamente a través del concepto de Sistema que a continuación se esquematiza.

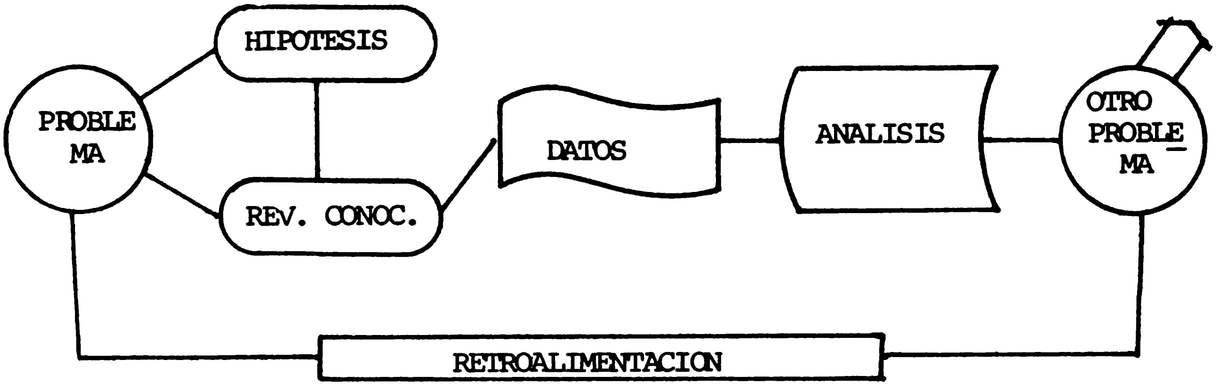


Figura 7. Método Científico aplicado a la secuencia lógica en la solución de situaciones problemáticas.

1.7 METODO CIENTIFICO. Es la aplicación de la lógica y la objetividad al estudio de los fenómenos naturales. La suposición básica, reside en la credibilidad de los fenómenos; es decir, la relación causa-efecto. La Figura 7 esquematiza el Método Científico considerándolo como un sistema, aunque este esquema no es idéntico al fenómeno natural, es suficientemente válido por gozar de la característica de la repetibilidad a través del proceso de retroalimentación y de su eficacia.

Siguiendo en forma lógica el flujo de información de la anterior figura, se distinguen los siguientes componentes del Método Científico.

1.71 SITUACION PROBLEMATICA. Es la especificación o tipificación del problema en estudio.

Ejemplo, los títulos de tesis ó los trabajos de investigación.

1.72 FORMULACION DE OBJETIVOS. Generalmente se plantean en forma de hipótesis nulas, con el expreso fin de rechazar tal suposición. Corresponde a los fines, metas y objetivos de una investigación científica.

1.73 REVISION DE CONOCIMIENTOS. Es la revisión de literatura; de naturaleza formal cuando se efectúa con base en documentación bibliográfica y de naturaleza informal si se utiliza la comunicación personal, epistolar ó experiencia propia.

1.74 RECOLECCION DE DATOS. Se diseña la forma apropiada de conseguir información, mediante: diseño de experimentos, cuando el experimento es controlable; diseño de muestreo, cuando el error no es controlable. Recopilación indirecta, utilizando censos, mapas cartográficos, estadísticas históricas, etc., y también por simulación.

1.75 ANALISIS DE LA INFORMACION. Fundamentalmente se estima, los parámetros del fenómeno en estudio para diferentes niveles de precisión.

1.76 INFERENCIA. Corresponde a la discusión y conclusiones de la investigación teniendo en cuenta el carácter predictivo de los resultados experimentales. Aquí se podría originar otra nueva situación proble

mática.

1.8 VARIABLE. Es el nombre simbólico que se asigna a las características sujetas a medición o recuento en un experimento particular como color, sexo, estatura, peso, distancia, volumen, cantidad, etc.

1.81 VARIABLES DISCRETAS. Cuando toma un número finito de valores en escala contable; así, una familia puede tener 0, 1, 2, 3, etc., niños, donde cada uno representa una unidad discreta cardinal.

1.82 VARIABLES CONTINUAS. Cuando toma infinito número de valores intermedios dentro de la amplitud de 2 valores extremos; no es posible expresar el valor exacto de una medición, independientemente de la precisión del instrumento de medida. v: gr: el peso de cierto estudiante podría ser 70 kilos ó 70,5, ó 70,542 kilos. Los valores numéricos de este tipo de variables son siempre aproximados.

1.83 VARIABLES CUALITATIVAS. Toman valores no enumerables como el color, sexo, días de la semana, etc; así, en el estudio del color de las flores de frijol en cruzamientos mendelianos; el color, es la variable en estudio.

1.84 VARIABLES CUANTITATIVAS. Cuando toma valores de la escala ordinal como el peso, estatura, cantidad, velocidad, etc., por su naturaleza se subdividen en variables cuantitativas contínuas y cuantitativas discretas.

1.9 POBLACION. Es la totalidad de los valores que toma una variable cualquiera; en un bosque de pinos sp., el diámetro de cada uno de

los árboles constituye el elemento componente de la población; en términos de variable, se habla de la población de diámetros (X_i), usualmente de tamaño N muy grande, que generalmente tiende a infinito, se esquematiza así:

$$X_i=\{X_1, X_2, X_3, \ldots\ldots, X_N\}$$

Se representa simbolicamente mediante una curva simétrica; Figura 8, donde μ y σ^2 son ciertos parámetros del fenómeno estudiado.

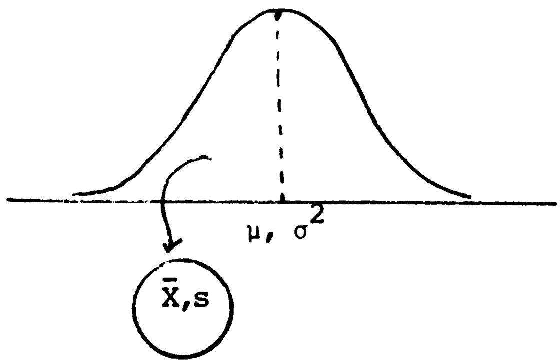


Figura 8. Representación esquemática de población y muestra.

1.10 MUESTRA. Es una parte de la población, de tamaño finito n y usualmente pequeño. En el ejemplo anterior Figura 8, la muestra podría estar formada de n = 25 observaciones (diámetro) caracterizada por su media y desviación estandar. Simbolicamente, se esquematiza así:

$$X_i=\{X_1, X_2, X_3, \ldots\ldots, X_n\}$$

1.11 PARAMETRO. Son medidas típicas, características o representativas de las poblaciones, generalmente de valores desconocidos pero es-

timables con base en la información de una muestra v:gr: media de la población, varianza de la población, etc., se representa por letras griegas (μ, σ^2) .

1.12 ESTIMADORES. Medidas típicas de una muestra extraída de una población, v:gr: el promedio aritmético de la muestra, varianza de la muestra, error estandar de la muestra, etc., se representan por letras latinas ($\bar{X}$, s)

Mediante la inferencia estadística,podemos afirmar con cierto grado de precisión que el promedio aritmético es la imagen del parámetro ($X \rightarrow \mu$).

Para que los estimadores sean fieles reflejos de los valores paramétricos, es condición que la muestra sea representativa de la población, esto se consigue eligiendo los componentes de la muestra en forma aleatoria.

1.121 EJEMPLO 1. Un experimento con 6 tratamientos se aplicará a 6 macetas (A, B, C, D, E, F) que contienen plántulas de maíz; Figura 9. ¿Cómo se asignarían los tratamientos a estos recipientes tal que la distribución sea aleatoria?

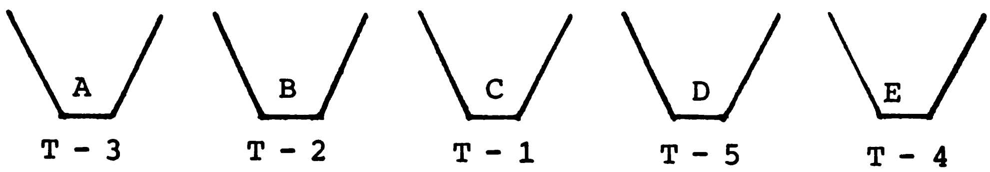


Figura 9. Asignación aleatoria de los tratamientos a 6 macetas.

1.13 SORTEO MEDIANTE TARJETAS. Se enumeran 6 tarjetas consecutivamente de 1 hasta 6. Se mezclan cuidadosamente y se extraen una a una anotando los números presentes en las tarjetas; el valor de la primera tarjeta (3) es el tratamiento que se aplicará a la maceta A (tratamiento 3), en forma similar se asignan los otros tratamientos. El propósito de esta aleatorización es eliminar en lo posible la subjetividad del investigador.

1.14 TABLA DE NUMEROS ALEATORIOS. Fundamentalmente, es una tabla que contiene aproximadamente, la misma proporción de los 10 dígitos 1, 2,....,9, 0;distribuídos mediante procesos mecánicos o electrónicos,como en el caso de tablas generadas por computador, algunas tablas son de 5000, 10000 ó mayor cantidad de dígitos y se presentan arregladas por hileras y columnas así:

Tabla 1. Tabla de 300 números al azar, obtenido con ayuda de un computador.

COLUMNAS

HILERAS		00000 01234	00000 56789	11111 01234	11111 56789
	00	43055	23063	06491	02206
	01	66358	16302	20856	48345
	02	72416	42123	42388	14107
	03	49531	08430	16156	72239
	04	27545	28847	00128	96631
	05	39385	00541	85602	90277
	06	21712	51422	73636	74897
	07	12065	30810	05546	01975
	08	26807	32064	37553	75246
	09	13145	28462	72306	64051
	10	00354	06944	64920	93641
	11	26842	46301	12836	48058
	12	92286	31056	96053	35658
	13	01959	56901	06428	60896
	14	94553	21770	54773	72208

1.15 USO DE LA TABLA. Paso No. 1 Determinar el número de dígitos aplicable al tamaño del experimento. En el ejemplo anterior es uno.

Paso No. 2 Definir el punto de partida en la tabla aleatoria. Sea el No. 3 de la hilera 00 y columna 09.

Paso No. 3 Anotar siguiendo la dirección de columna la cifra inicial y siguientes a esta como sigue 3 -2 -3 -0 -7 -1 -2 -0 -4 -2 -4 etc.

Paso No. 4 Asignar estos números a cada uno de los tratamientos A -B -C -D -E -F; eliminando las repetidas y las irrelevantes (7 -8 -9 -0); es decir:

3 - 2 - ~~3~~ - ~~0~~ - ~~7~~ - 1 - ~~7~~ - 5 - 4 - 6 - ~~4~~

A B C D E F

Usualmente los experimentos van en bloques, por lo que la asignación se hace en forma similar en cada bloque. Si una columna es insuficiente se toma las siguientes hasta completar la extracción.

1.152 EJEMPLO 2. Un directorio telefónico tiene N = 3540 abonados. Se desea extraer de esta población una muestra de tamaño n = 18.

a) Número de dígitos aplicables a este experimento: 4.

b) Punto de partida en hilera 06 y columnas 01, 02, 03, 04 es decir: 1712.

c) Lista de cifras: 1712 - 2065 - 6807 - 3145 - 0354 - etc.

d) La muestra representativa se presenta en la tabla 2.

Tabla 2. Muestra completamente al azar de tamaño n = 18 con la ayuda de la tabla de números aleatorios.

MUESTRA	ABONADO	MUESTRA	ABONADO
1	1712	13	3206
2	2065	14	2846
3	3145	15	694
4	354	16	3105
5	2286	17	2177
6	1959	18	3064
7	2306	19	2208
8	1630	20	3423
9	843	21	161
10	2884	22	1856
11	54	23	2736
12	3081	24	55

CAPITULO 2

2. DISTRIBUCION DE FRECUENCIAS.-

2.1 DATOS NO CLASIFICADOS. El Cuadro 1. presenta una colección de datos sin organización alguna; es una lista, un conjunto, una masa de información de alturas de 40 estudiantes de San José.

Cuadro 1. Alturas de 40 estudiantes de San José (Cm).

138	125	140	144	163	173	140	142
164	149	147	168	120	142	135	155
150	157	136	156	154	147	161	156
132	146	148	138	165	135	145	145
144	158	152	175	146	153	135	128

Estos datos definitivamente no dan idea alguna acerca de la estatura de los estudiantes, por lo que reciben el nombre genérico de datos no clasificados y son de escaza utilidad.

2.2 ORDENACION. El primer paso, consiste en organizar o arreglarlos en orden ascendente; Cuadro 2, o en forma descendente; es usual, organizar la información de la más pequeña a la mayor.

Cuadro 2. Alturas ordenadas de menor a mayor de 40 estudiantes

120	135	140	144	147	152	156	164
125	135	140	145	147	153	157	165
128	136	142	145	148	154	158	168
132	138	142	146	149	155	161	173
135	138	144	146	150	156	163	175

2.3 DISTRIBUCION DE FRECUENCIAS. Con el objeto de conocer algunos valores típicos o representativos del conjunto, se organizan los datos según el Cuadro 3; es decir, se distribuye la información en categorías o clases preestablecidas.

Esta presentación de datos en clases con sus frecuencias de ocurrencia recibe el nombre de Distribución de Frecuencias.

Cuadro 3. Distribución de frecuencias de las alturas de 40 estudiantes de San José (cm).

CLASES	REGISTRO O PUNTUACION	FRECUENCIA (fi)	PUNTO MEDIO CLASE
120 - 126	//	2	123.5
127 - 133	//	2	130.5
134 - 140	//// ////	8	137.5
141 - 147	//// //// //	10	144.5
148 - 154	//// //	6	151.5
155 - 161	//// //	6	158.5
162 - 168	////	4	165.5
169 - 175	//	2	172.5

2.31 CLASES. Los 40 estudiantes fueron divididos en 8 grupos o categorías: 120 - 126, 127 - 133, ..., 169 - 175, que reciben el nombre de clases. El número total de clases se determina en función de la escala y magnitud de la información; usualmente de 8 a 12 clases, pudiendo exceder estos límites si el volumen de información es extremadamente grande o muy pequeño.

2.32 LIMITES DE CLASE. Son los 2 valores numéricos de cada clase;

el de la izquierda recibe el nombre de límite inferior de clase; el de la derecha, límite superior de clase.

Límite inferior ------- 127 - 133 ------- Límite superior

2.33 INTERVALO DE CLASE. Es la diferencia entre dos límites superiores o inferiores consecutivos, se simbolizan por la letra mayúscula I.

I = Límite inferior de clase 2 - Límite inferior de clase 1 =

$127 - 120 = 7$

I = Límite superior de clase 4 - límite superior de clase 3 =

$147 - 140 = 7$

2.34 TAMAÑO DE CLASE. Se obtiene dividiendo la amplitud o rango de la información entre el número de clases deseado, así:

amplitud :$175 - 120 = 55$

tamaño de clase: $55/8 \approx 7$

2.35 PUNTO MEDIO DE CLASE. Es el valor numérico que corresponde exactamente a la mitad del tamaño de clase. Es la semidiferencia de los límites de clase agregada al límite inferior; así, el punto medio de la primera clase es $(127 - 120)/2 + 120 = 123.5$.

2.36 FRECUENCIA. Es el número de registros o puntuaciones incluídos en la clase considerada.

2.4 REPRESENTACION GRAFICA. Se realiza mediante histogramas y polígonos de frecuencia; se usa la escala de abscisas para las clases y la escala de ordenadas para las frecuencias. Tiene la ventaja de resaltar las características típicas de los datos, facilitando la comparación visual de ellos.

2.41 HISTOGRAMA DE FRECUENCIAS. Está formado de rectángulos, Figura 10; donde la base es el tamaño de clase 120 - 126.9 y la altura es la frecuencia de clase considerada. La suma de áreas de los rectángulos, representan el 100% de las observaciones.

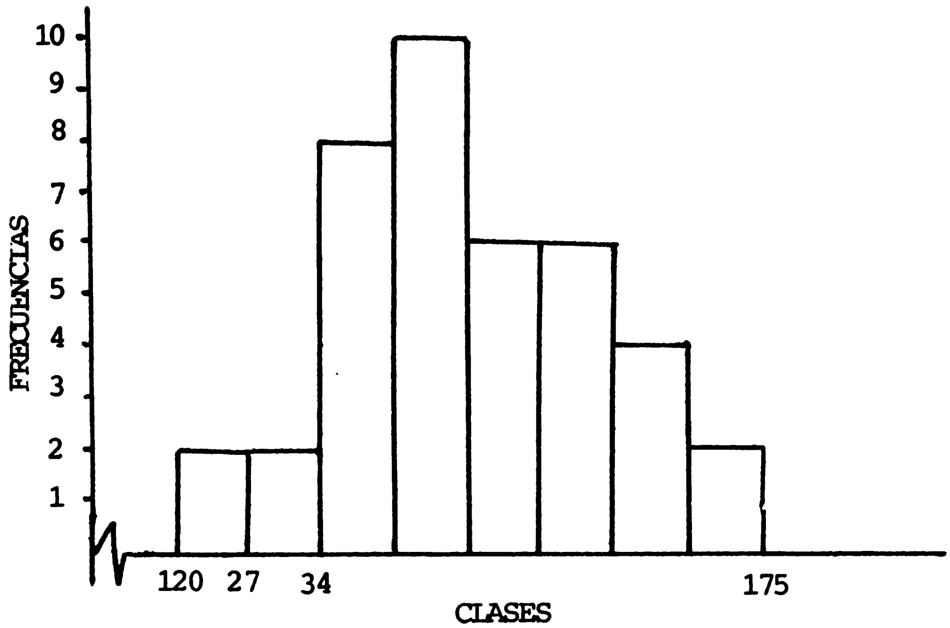


Figura 10. Histogramas de frecuencias de las alturas de 40 estudiantes (intervalo de clase igual).

Si los intervalos son desiguales, los rectángulos que representan a las clases desiguales, deberán construirse proporcionalmente a las frecuencias, tal que la suma de área sea siempre equivalente al 100% de las observaciones individuales.

Sea el Cuadro 4 con clases desigualmente espaciadas y su representación gráfica en la Figura 11. Notar que para determinar la frecuencia de la última clase, se divide la frecuencia 4 entre el número de clases.

Cuadro 4. Distribución de frecuencias desigualmente espaciadas .

CLASES	FRECUENCIAS
10 - 20	5
20 - 30	10
30 - 40	15
40 - 80	4

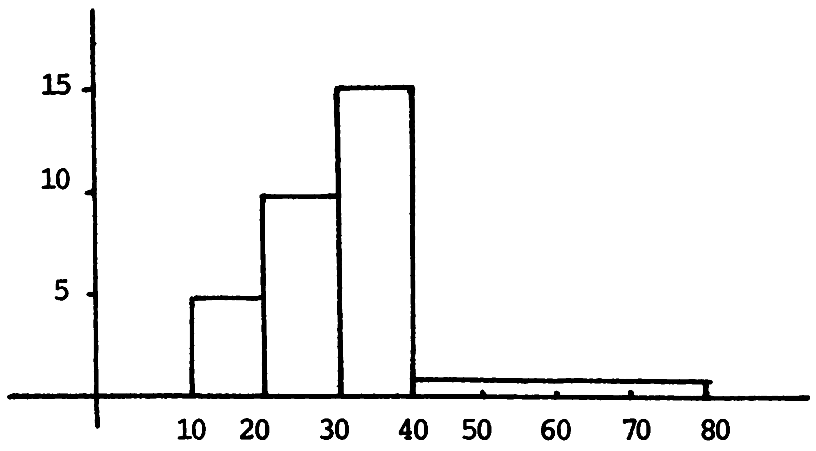


Figura 11. Histograma de frecuencia para intervalos desiguales.

2.42 POLIGONO DE FRECUENCIAS. La Figura 12, presenta un polígono de frecuencias que se grafica por intersección del punto medio de clase y su frecuencia respectiva; es una línea poligonal cerrada tal que el área inscrita equivale al 100% de las observaciones, el cierre se consigue utilizando clases extras por exceso y defecto de la amplitud de los datos y con frecuencia cero.

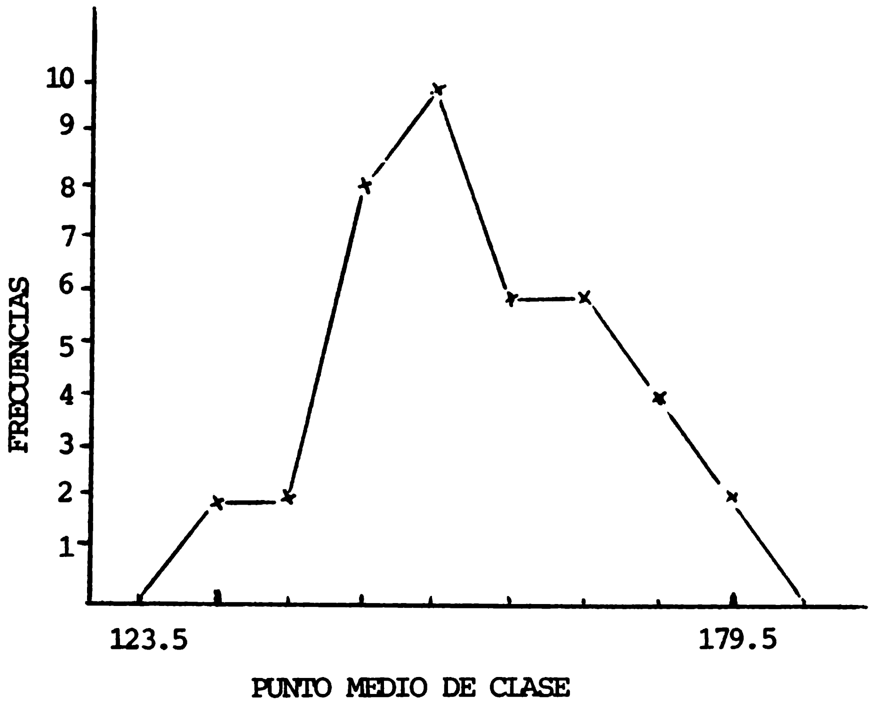


Figura 12. Polígono de frecuencias de las alturas de 40 estudiantes.

2.5 FRECUENCIAS ACUMULADAS ASCENDENTES. Es el número de observaciones con valor numérico inferior al límite superior de la clase considerada. Cuadro 5; es decir, cuántos alumnos miden menos de 127 cm? Cuántos menos de 134?, etc.

2.6 FRECUENCIAS ACUMULADAS DESCENDENTES. Es el número de observaciones con valor igual ó superior al límite inferior de la clase considerada; responde a las preguntas: cuántos alumnos miden 120 cm ó más de 120 cm?, cuántos 127 cm?, etc.

Cuadro 5. Frecuencia acumulada ascendente y descendente de las alturas de 40 estudiantes.

		FRECUENCIAS	
CLASES	ABSOLUTA	ACUMULADA ASCENDENTE	ACUMULADA DESCEND.
120 - 126	2	2	40
127 - 133	2	4	38
134 - 140	8	12	36
141 - 147	10	22	28
148 - 154	6	28	18
155 - 162	6	34	12
162 - 168	4	38	6
169 - 176	2	40	2

2.7 OJIVA ASCENDENTE Y DESCENDENTE. Es la representación gráfica de las frecuencias ascendentes y descendentes respectivamente como se muestra en la Figura 13.

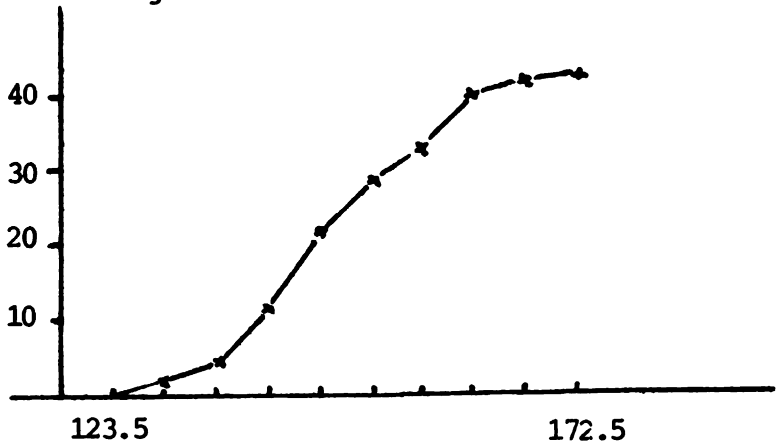


Figura 13. Ojiva ascendente de la estatura de 40 estudiantes (cm).

2.8 FRECUENCIAS RELATIVAS. Son las frecuencias absolutas transformadas proporcionalmente al 100% del total; se obtiene dividiéndose las

frecuencias de clase entre el total de las frecuencias. Cuadro 6. La frecuencia relativa de la primera clase es: (100 x 2)/40 = 5; las frecuencias relativas acumuladas ascendentes: 5 + 5 = 10; 10 + 20 = 30; 30 + 25 = 55; las frecuencias acumuladas descendentes: 100 - 5 = 95; 95 - 5 = 90, etc.

Cuadro 6. Frecuencias relativas de las alturas de 40 estudiantes

				FRECUENCIAS RELATIVAS	
CLASES	Xi	Fi	F.R.	ACUMULADA ASCEND.	ACUMULADA DESC.
120 - 126	123.5	2	5.0	5.0	100.0
127 - 133	130.5	2	5.0	10.0	95.0
134 - 140	137.5	8	20.0	30.0	90.0
141 - 147	144.5	10	25.0	55.0	70.0
148 - 154	151.5	6	15.0	70.0	45.0
155 - 161	158.5	6	15.0	85.0	30.0
162 - 168	165.5	4	10.0	95.0	15.0
169 - 175	172.5	2	5.0	100.0	5.0

2.9 OJIVAS PORCENTUALES ASCENDENTE Y DESCENDENTE. Son las frecuencias acumuladas ascendentes y descendentes, expresadas en unidades relativas. Se grafican en forma similar a las frecuencias acumuladas, cambiando unicamente la escala de Y.

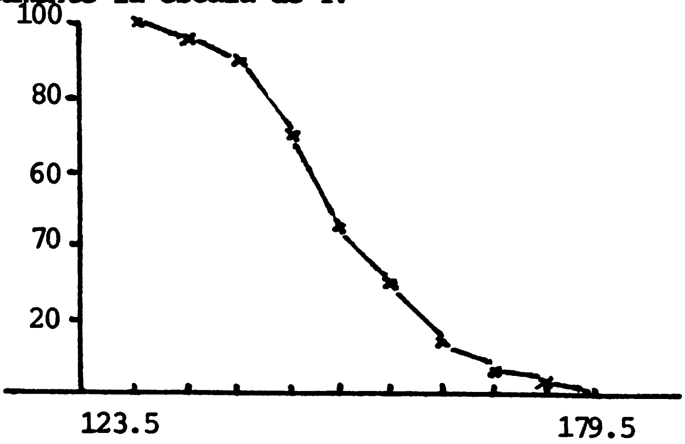


Figura 14. Ojiva porcentual y descendente de las alturas de 40 estudiantes.

CAPITULO 3

3. ESTIMADORES DE POSICION DE DATOS SIN CLASIFICAR.

3.1 MEDIA ARITMETICA. Sea la variable X_i, que toma los valores particulares $X_1, X_2, X_3, \ldots\ldots\ldots, X_n$, la media aritmetica ó simplemente promedio se define así:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots\ldots X_n}{n} = \frac{\Sigma X_i}{n}$$

Donde:

$\bar{X}$ = media aritmética

n = tamaño de muestra

$X_i = X_1, \ldots\ldots\ldots\ldots, X_n$

ΣX_i = Sumatoria sobre los valores de i (1,...., n)

3.11 EJEMPLO 1. La media aritmética de los números 3, 5, 7, 8, 15, 10 es:

$$\bar{X} = \frac{3 + 5 + 7 + 8 + 15 + 10}{6} = \frac{48}{6} = 8$$

3.2 MEDIA PONDERADA. Si los valores que toma la variable X_i no son de la misma importancia, es válido asignar 'pesos' ó 'ponderaciones' $W_1, W_2, \ldots\ldots, W_n$ como medida de esta importancia; en este sentido, se define la media ponderada, así:

$$\bar{X}_p = \frac{W_1 \cdot X_1 + W_2 \cdot X_2 + \ldots\ldots W_n \cdot X_n}{W_2 + W_2 + \ldots\ldots\ldots W_n} = \frac{\Sigma W_i X_i}{\Sigma W_i}$$

Donde:

$\bar{X}_p$ = media ponderada

W_i = pesos o ponderaciones

n = muestra

X_i = variable

3.21 EJEMPLO 2. El exámen final de estadística, se evalua asignando a la prueba final 3 veces el valor de las pruebas parciales. Las calificaciones parciales y finales de cierto alumno son: 8.6, 8.4, 8.5, 9.6 respectivamente, su promedio es:

$$\bar{X} = \frac{(1).(8.6)+(1).(8.4)+(1).(8.5)+(3).(9.6)}{1+1+1+3} = \frac{54.3}{6} = 9.05$$

3.3 PROPIEDADES DE LA MEDIA ARITMETICA. La suma de las desviaciones de la variable X_i con respecto a su media $\bar{X}$ es siempre cero.

$$\Sigma(X_i - \bar{X}) = \Sigma X_i - n\bar{x}$$

$$= \Sigma X_i - \cancel{n}\frac{\Sigma Xi}{\cancel{n}}$$

$$= 0$$

3.31 EJEMPLO 3. Las desviaciones del ejemplo 1 son:

MUESTRA Xi	DESVIACION $(Xi-\bar{X})$	D. CUADRATICA $(Xi-\bar{X})^2$
3	(3 - 8)	$(-5)^2$
5	(5 - 8)	"
8	(8 - 8)	"
7	(7 - 8)	"
15	(15 - 8)	"
10	(10 - 8)	2^2
Σ: 48	0	87
$\bar{X}$: 8	0	14.5

La suma de las desviaciones cuadráticas con respecto a $\bar{X}$, es siempre un valor mínimo, menor que la suma de los desvíos con respecto a un valor diferente de la media.

3.32 EJEMPLO 4. La suma de las desviaciones cuadráticas del ejemplo 1 con respecto a la media $\bar{X} = 8$, es un mínimo.

$$(-5)^2 + (-3)^2 + (0)^2 + (7)^2 + (2)^2 = 87 \text{ (mínimo)}$$

La suma de las desviaciones cuadráticas del mismo ejemplo, con respecto a otro valor cualquiera A = 9, no es un mínimo porque 94 es mayor que 87.

$$(3-9)^2 + (5-9)^2 + (7-9)^2 + (15-9)^2 + (10-9)^2 = 94 \text{ (no es mínimo).}$$

3.4 REPRESENTACION GRAFICA. En la Figura 15 en un eje decimal se representa la escala X_i del ejemplo 1.

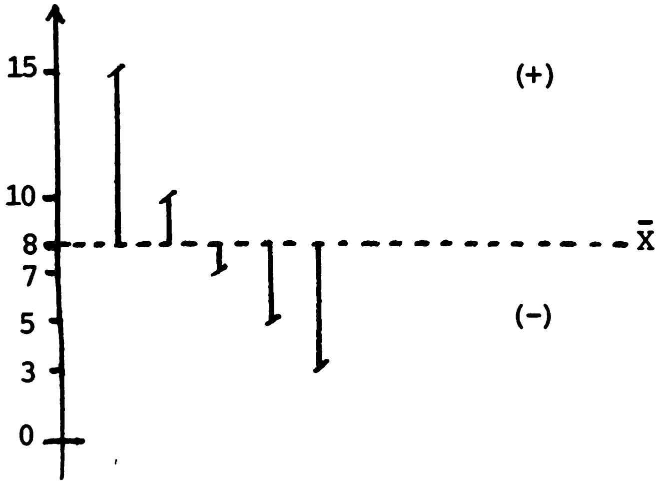


Figura 15. Propiedades de la media aritmética

3.5 MEDIANA. Es el valor numérico de la variable que divide en 2 partes una masa de información ordenada en forma ascendente. Esta defi nición es consistente para variables de tamaño n siempre impar.

Para variables cuyo tamaño es un número par, la mediana es el promedio de los elementos centrales.

3.51 EJEMPLO 5. Sean los valores de X_i: 3, 4, 7, 8, 9, 10, 13; la mediana es 8.

Sean los valores de X_i: 3, 3, 4, 4, 6, 8; la mediana es 4.

3.6 CUANTILES. Por generalización del concepto de mediana, cuantil es la partición de un arreglo en un definido número de partes. Si el a rreglo se parte en dos el cuantil recibe el nombre particular de mediana.

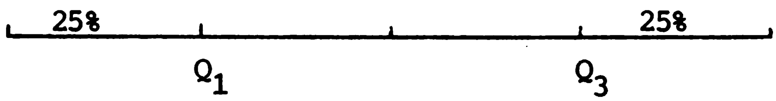
Partición es dos = mediana = Me

Si el arreglo se fracciona en 4 partes, la partición recibe el nombre de cuartil.

Partición en 4 = cuartiles = Qi; i = 1, 2, 3.

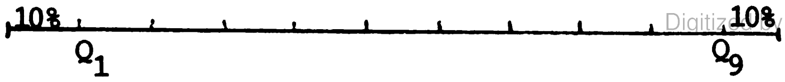
Los cuartiles definen dos colas con 25% de la información en cada una, la porción central con 50%, esquemáticamente se tiene:

Q_4 =



Si el arreglo se fracciona en 10 partes, la partición recibe el nombre de decil, siendo relevantes el 1° y el 9° decil, por dejar en las 2 colas el 20% de la información.

Q_{10} =



En forma similar un arreglo fraccionado en 100 partes recibe el nombre de percentil, son de particular interés el 1er. y 99 percentil por dejar en cada cola 1% de la información.

3.7 MODA. Es el valor numérico que se presenta más veces, el más frecuente, el más común del arreglo.

3.71 EJEMPLO 6. El conjunto {3, 4, 7, 8, 3, 5, 5, 3} tiene moda 3.

3.8 MEDIA ARITMETICA DE DATOS CLASIFICADOS SEGUN SUS FRECUENCIAS.

Cuando los datos se presentan agrupados por clase con sus frecuencias respectivas, Cuadro 7, también es posible el cálculo de la media aritmética.

Cuadro 7. Datos clasificados incluyendo frecuencias absolutas y punto medio de clase

CLASES	X_i	fi	$fi.X_i$
120 - 126	123.5	2	247
127 - 133	130.5	2	261
134 - 140	137.5	8	1100
141 - 147	144.5	10	1145
148 - 154	151.5	6	909
155 - 161	158.5	6	951
162 - 168	165.5	4	662
169 - 175	172.5	2	345
Σ:		40	5620.0

3.81 MEDIA UTILIZANDO PUNTO MEDIO DE CLASE. Es el cociente de la sumatoria del producto de las frecuencias por sus puntos medio y la sumatoria de frecuencias absolutas.

$$\bar{X} = \frac{\Sigma fiXi}{\Sigma fi} = \frac{5620.0}{40} = 140.50$$

3.9 MEDIANA DE DATOS CLASIFICADOS. Se calcula de acuerdo a la relación.

$$Me = Li + \frac{\frac{N}{2} - faa}{fa} \cdot I$$

Donde:

Li = Límite inferior de las clases que contiene la mediana

N = Suma de frecuencias

faa = Frecuencias acumuladas, hasta la anterior clase

fa = frecuencia absoluta de la que contiene la mediana

I = Intervalo de clase

3.91 EJEMPLO 7. Calcular la mediana de los datos del Cuadro 7.

$$Me = 141 + \frac{\frac{40}{2} - 12}{10} (7) = 151.6$$

3.10 CUANTILES DE DATOS CLASIFICADOS.

$$Cuantil = Li + \frac{\frac{Bi.N}{Ki} - faa}{fa} \cdot I$$

Donde:

Li = límite inferior del cuantil considerado

ki = Participación deseada, así, Ki = 2 implica la mediana, k_i = 4 implica cuartil, Ki = 10 decil, etc.

B_i = grado dentro de la partición; así si K = 4:

B_1 = Primer cuartil = 1

B_2 = Segundo cuartil = 2

B_3 = Tercer cuartil = 3

Si K = 10 :

B_1 = Primer decil = 1

B_2 = Segundo decil = 2

B_9 = Noveno decil = 9

Si K = 100:

B_1 = Primer percentil = 1

B_2 = Segundo percentil = 2

$\vdots$

B_{99}= Noventa y nueve percentil = 99

3.101 EJEMPLO 8. El primer cuartil de los datos clasificados que se presentan en el Cuadro 5, se calcula de la siguiente manera:

$$C_1 = 134 + \frac{\frac{(1).(40)}{4} - 4}{8} . (7) = 139.25$$

3.102 EJEMPLO 9. El noveno decil de los datos clasificados que se presentan en el Cuadro 5, se calcula así:

$$D_9 = 162 + \frac{\frac{9(40)}{10} - 34}{4} .(7) = 165.5$$

3.11 MODA DE DATOS CLASIFICADOS. Se define por la expresión:

$$Mo = Li + \frac{fp}{fa + fp} \cdot I$$

Donde:

fa = frecuencia de la clase anterior a la que contiene la máxima frecuencia.

fp = frecuencia de la clase posterior a la que contiene la máxima frecuencia.

3.111 EJEMPLO 10. Calcular la moda de los datos clasificados que se presentan en el Cuadro 7.

$$Mo = 141 + \frac{6}{(8 + 6)} \cdot 7 = 144.0$$

3.12 OTRAS MEDIAS

3.121 MEDIA ARMONICA. Es el recíproco de la media aritmética de los recíprocos de los términos.

$$\bar{X}_A = \frac{1}{\frac{1}{n}\left(\frac{1}{X_1} + \frac{1}{X_2} + \ldots + \frac{1}{X_n}\right)} \text{ ó } \bar{X}_A = \frac{1}{\frac{1}{n}\left(f_1 \frac{1}{X_1} + f_2 \frac{1}{X_2} + \ldots\right)}$$

3.122 MEDIA GEOMETRICA. Es la raíz enésima del producto de las observaciones:

$$\bar{X}_g = \sqrt[n]{\pi X_i} \text{ ó } \bar{X}_g = X_1^{f_1} \cdot X_2^{f_2} \ldots\ldots X_n^{f_n}$$

3.123 EJEMPLO 11. Calcular la media armónica y geométrica de los números 3, 5, 7, 8, 10.

$$\bar{X}_A = \frac{1}{\frac{1}{5} \cdot \left(\frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{8} + \frac{1}{10}\right)} = 5.55$$

$$\bar{X}_g = \sqrt[5]{(3) \cdot (5) \cdot (7) \cdot (8) \cdot (10)} = \frac{1}{5} \log 8400 = 6.09$$

CAPITULO 4

4. ESTIMADORES DE DISPERSION.-

4.1 VARIANZA. Sea la variable X_i, con valores particulares X_1, X_2, X_3, ..., X_n; la varianza de este conjunto se define así:

$$s^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots. + (X_n - \bar{X})^2}{n - 1} = \frac{\Sigma(X_i - \bar{X})^2}{n - 1}$$

Donde:

$\bar{X}$ = media aritmética

n = tamaño de muestra

s^2 = varianza

n-1 = grados de libertad

4.11 EJEMPLO 1. La varianza de los números 10, 5, 7, 8, 10; es:

$$s^2 = \frac{(10 - 8)^2 + (5 - 8)^2 + (7 - 8) + (8 - 8)^2 + (10 - 8)^2}{4} = 4.50$$

La anterior definición es de poca utilidad, porque el promedio no siempre es exacto; sin embargo, con una pequeña transformación aritmética en el numerador de la definición de varianza, se obtiene:

$$\begin{aligned}
\Sigma(X_i - \bar{X})^2 &= \Sigma(X_i^2 - 2X_i\,\bar{X} + \bar{X}^2) \\
&= \Sigma X_i^2 - 2\Sigma X_i\bar{X} + \Sigma\bar{X}^2 \\
&= \Sigma X_i^2 - 2\Sigma X_i\,\frac{\Sigma Xi}{2} + n\,\bar{X}^2 \\
&= \Sigma X_i^2 - 2\,\frac{(\Sigma Xi)^2}{n} + n\,\frac{(\Sigma Xi)^2}{n^2} \\
&= \Sigma X_i^2 - \frac{(\Sigma Xi)^2}{n}
\end{aligned}$$

Esto da origen a una nueva definición de varianza en función de los valores cuadráticos de la variable X_i.

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

4.12 EJEMPLO 2. La varianza de los números 10, 5, 7, 8, 10, es:

X_i	X_i^2
10	100
5	25
7	49
8	64
10	100
Σ: 40	338

$$s^2 = \frac{338 - (40)^2/5}{4} = 4.50$$

4.2 VARIANZA DE DATOS CLASIFICADOS. Cuando los datos se presentan agrupados en clases ó categorías con sus frecuencias respectivas la varianza se define así:

$$s^2 = \frac{\Sigma fiXi^2 - (\Sigma fiXi)^2/\Sigma fi}{\Sigma fi - 1}$$

Cuadro 8. Datos clasificados con sus frecuencias absolutas y punto medio

CLASES	Xi	fi	fiXi	fiXi^2
120 - 126	123.5	2	247	30504.5
127 - 133	130.5	2	261	34060.5
134 - 140	137.5	8	1100	151250.0
141 - 147	144.5	10	1445	208802.5
148 - 154	151.5	6	909	137713.5
155 - 161	159.5	6	951	150733.5
162 - 168	165.5	4	662	109561.0
169 - 175	172.5	2	345	59512.5
Σ:		40	5920.0	882138.0

4.21 EJEMPLO 3. Con los datos del Cuadro 7 calcular la varianza.

$$s^2 = \frac{882138 - (5920)^2/40}{39} = 153.28$$

4.3 VARIANZA DE DATOS CLASIFICADOS Y PUNTOS MEDIOS DE CLASE CODIFICADOS. Para datos clasificados con frecuencias y puntos medios de clase codificados Di según la definición Di = (xi - A)/I; donde Xi son los puntos medios de clase, A el punto medio de la clase central e I el intervalo de clase, la varianza se define así:

$$s^2 = \frac{\Sigma fiDi^2 - (\Sigma fiDi)^2/n}{\Sigma fi - 1} \cdot I^2$$

4.31 EJEMPLO 4. Calcular la varianza de los datos que se presentan en el cuadro siguiente:

Cuadro 9. Datos clasificados con sus frecuencias y punto medio de clase codificado.

CLASES	Xi	Di = Xi - 151.5/7	fi	FiDi	FiDi^2
120 - 126	123.5	-4	2	- 8	32
127 - 133	130.5	-3	2	- 6	18
134 - 141	137.5	-2	8	-16	32
141 - 147	144.5	-1	10	-10	10
148 - 154	151.5	0	6	0	0
155 - 161	159.5	1	6	6	6
162 - 168	165.5	2	4	8	16
169 - 175	172.5	3	2	6	18
Σ:		0	40	-20	132

$$s^2 = \frac{132 - (-20)^2/40}{39} \cdot 7^2 = 153.28$$

4.4 DESVIACION ESTANDAR. Es la raíz cuadrada de la varianza y se define por:

$$s = \sqrt{\frac{\Sigma (Xi - \bar{X})^2}{n - 1}}$$

4.5 COEFICIENTE DE VARIACION. Es el cociente de la desviación standar y la media, expresado en porciento.

$$C.V. = \frac{s}{\bar{X}} \times 100$$

Este estimador es independiente de la unidad de medida por lo que puede compararse la variabilidad de dos muestras, provenientes de poblàciones diferentes.

4.51 EJEMPLO 5. Calcular los coeficientes de variación de los pesos y estaturas de los estudiantes de Agronomía.

ALUMNOS	PESO	TALLA	
1	180	179	cm
2	140	168	cm
'	'	'	'
'	'	'	'
'	'	'	'
ni	175	170	cm
$\bar{X}$ =	165	172	
s =	4.5	8.4	

$$C.V. = \frac{4.5}{165} . 100 = 2.73 \quad \text{(Peso)}$$

$$C.V. \quad \frac{8.4}{172} . 100 = 4.82 \quad \text{(Estatura)}$$

Esto quiere decir que las tallas son más variables que los pesos; existe más dispersión en las tallas.

Cuando los datos se presentan en un cuadro de análisis de varianza el coeficiente de variación se calcula así:

$$C.V. = \frac{\sqrt{CME}}{\bar{X}} \cdot 100$$

Donde:

CME = Cuadrado medio del error

$\bar{X}$ = Gran media

4.6 VARIABLE TIPIFICADA NORMAL O STANDARD (Z). Para comparar muestras o elementos componentes de una misma muestra, se realiza transformación de X_i a la variable tipificada normal Z_i.

$$Z_i = \frac{X_i - \bar{X}}{s}$$

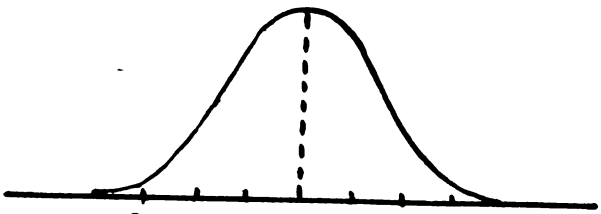
4.61 EJEMPLO 6. Se tiene 2 series de observaciones correspondientes a las notas obtenidas en las materias de matemáticas y biología. Se pregunta en qué materia se destacó AA?, para poder dar una respuesta debemos transformar a Z_i todos los valores observados.

OBSERVADOS	MATEMATICAS	BIOLOGIA
1	75.0	70.0
2	80.0	85.0
3	95.0	100.0
⋮	⋮	⋮
n	95.0	100.00
$\bar{X}$	72.0	90.0
s	8	15
AA	76	92

$$Z\ (\text{mat.}) = \frac{76 - 72}{8} = 0.50$$

$$Z\ (\text{biol}) = \frac{92 - 90}{15} = 0.13$$

En forma gráfica tenemos:



RESPUESTA.- El alumno rinde más en matemáticas

La variable normalizada Z tiene media 0 y varianza 1.

4.62 EJEMPLO 7. Una población hipotética, está definida mediante la variable Y_i que toma unicamente 6 valores:

$$Y_i = \{Y_1, Y_2, Y_3, \ldots\ldots, Y_N\} = \{4, 3, 5, 5, 4, 3\}$$

La media y varianza de esta población se calcula así:

$$\mu = \frac{1}{N} \Sigma Y_i = 4.0$$

$$\sigma^2 = \frac{1}{N} \Sigma (Y_i - \bar{Y})^2 = 0.6666$$

La variable Y_i se normaliza a Z_i y se obtiene:

$$Z_i = \{Z_1, \ldots\ldots, Z_N\} = \{0.0, -1.225, 1.225, 1.225, 0.0, -1.225\}$$

La media y varianza de esta variable normalizada es:

$$\mu = \frac{1}{N} \Sigma Z_i = 0.0$$

$$\sigma^2 = \frac{1}{N} \Sigma (Z_i - \bar{Z})^2 = 1.0$$

CAPITULO 5

5. ANALISIS COMBINATORIO.-

Es la técnica que permite determinar o calcular el número de resultados posibles de cualquier experimento, sin enumeración directa.

5.1 PRINCIPIO FUNDAMENTAL DEL RECUENTO. Si un suceso A puede presentarse de n, maneras o formas, si un segundo suceso puede presentarse de n_2 maneras y si aún es posible un tercer suceso que puede presentarse de n_3 maneras y así sucesivamente. La realización del suceso A seguido del suceso B y de C puede hacerse de $n_1 \,.\, n_2 \,.\, n_3 \,........\, n_k$ maneras, es decir:

Realización total = $n_1 \times n_2 \times n_3 \times ...... \times n_k$

5.2 FACTORIAL. El factorial de un número se denota por n! y se define como el producto de la sucesión de números de 1 hasta n.

$n! = n.(n-1).(n-2) \,.......\, 2 \,.\, 1$

$1! = 1$

$0! = 1$

Para valores grandes de n el factorial se evalúa utilizando la aproximación de Sterling.

$n! = \sqrt{2\pi n} \,.\, n^n . e^{-n}$

5.21 EJEMPLO 1. Evaluar el factorial 10 .

$$n! = \sqrt{2\pi n}\; n^{n}.e^{-n}$$

$$= \sqrt{2\pi 10}\quad 10^{10}.\ e^{-10}$$

$$= \sqrt{20\pi}\quad 10^{10}.\ e^{-10}$$

$$\log\ n = \log\ (\sqrt{20\pi}\ 10^{10}\ .\ e^{-10}$$

$$= \frac{1}{2}\log 20 + \frac{1}{2}\log \pi + 10 \log 10 - 10 \log e$$

$$= \frac{1}{2}\ (1.3010) + \frac{1}{2}\ (0.4972) + 10 - 10\ (0.4343) = 6.6061$$

$$n! \simeq 4000000.0$$

5.3 PERMUTACIONES. Es el número de maneras de seleccionar o elegir r objetos de una muestra de tamaño n, es importante tener en cuenta la composición o arreglo de la muestra, se define:

$${}_{n}P_{r} = \frac{n!}{(n - r)!}$$

$${}_{6}P_{4} = \frac{6!}{2!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = 360$$

5.4 COMBINACIONES. Son listas, arreglos o selecciones de n objetos o cosas tomados r a la vez, a diferencia de permutaciones el orden es irrelevante ya que tiene en cuenta la composición de la selección, se define así:

$$\binom{n}{r} = \frac{n!}{(n-r)!\ r!}$$

5.41 EJEMPLO 3. De cuántas maneras se puede escoger un comité compuesto de 3 alumnos, sabiendo que potencialmente son elegibles 8?

$$\binom{8}{3} = \frac{8!}{(8-3)!\ 3!} = 56$$

5.5 COEFICIENTES DEL BINOMIO. Dada la expresión $(q + p)^2$, se trata de calcular los coeficientes de cada uno de los términos de su expansión. Se puede obtener siguiendo una de las definiciones siguientes:

5.51 TRIANGULO DE PASCALL.

						1						
					1		1					
				1		2		1				
			1		3		3		1			
		1		4		6		4		1		
	1		5		10		10		5		1	etc.

5.511 EJEMPLO 4. Determinar los coeficientes en la expresión siguiente:

$$(q + p)^4 = 1q^4 + 4q^3p + 6q^2p^2 + 4qp^3 + 1p^4$$

5.52 COEFICIENTES DEL BINOMIO. Se utiliza el símbolo $\binom{n}{r}$ para evaluar los coeficientes del binomio, n es el grado del binomio y, $o < r<n$; y $\binom{n}{r} = n!/(n - r)!\,r\,!$

5.521 EJEMPLO 5. Evaluar los coeficientes del binomio $(q + p)^4$.

$$(q + p)^4 = \binom{4}{0}q^4 + \binom{4}{1} q^3p + \binom{4}{2} q^2p^2 + \binom{4}{3} qp^3 + \binom{4}{4} p^4$$

$$\binom{4}{0} = \frac{4\,!}{(4.0)\,!\;0\,!} = 1; \quad \binom{4}{1} = \frac{4\,!}{(4-1)\,!\,1\,!} = 4; \quad \binom{4}{3} = 4; \quad \binom{4}{4} = 1$$

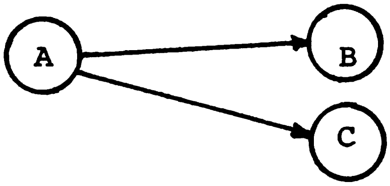
El primer coeficiente se calcula con $r = 0$, el segundo con $r = 1$ y así sucesivamente.

5.522 EJEMPLO 6. Calcular el coeficiente del octavo término del binomio $(q + p)^{10}$.

$$\binom{10}{7} = \frac{10!}{(10-7)!\,7!} = 120$$

5.6 PROBABILIDAD. Una gran parte de los fenómenos naturales son de naturaleza determinística; es decir, dada una situación inicial A necesariamente ocurrirá B.

Pero también, existen fenómenos naturales donde dada la situación inicial A, puede ocurrir B o puede ocurrir C



Este resultado deja de ser determinístico dado que la ocurrencia del suceso B está asociado a cierta probabilidad y la ocurrencia de C también está asociado a otra probabilidad.

5.7 DEFINICION CLASICA DE PROBABILIDAD. Es la relación aritmética de casos favorables y casos posibles o totales, es decir, es el cociente que en forma abreviada se expresa así:

$$p = \frac{\text{casos favorables}}{\text{casos posibles}}$$

5.71 EJEMPLO 7. al lanzar una moneda, ¿cuál es la probabilidad de obtener cara?

$$p = \frac{1}{2} = 0.5$$

5.8 PROBABILIDAD COMO FRECUENCIA. La probabilidad se define también como la frecuencia de ocurrencia del fenómeno estudiado a medida que el número de repeticiones o tamaño de muestra tiende a infinito.

5.81 EJEMPLO 8. Se lanza una moneda 80 veces y se observa la ocurrencia de 35 caras. La frecuencia de este suceso es:

$$f = \frac{35}{80} = 0.44$$

En otro experimento, se lanza una moneda 800 veces y se observa la ocurrencia de 385 caras. La frecuencia de este suceso es:

$$f = \frac{385}{800} = 0.48$$

Cuando n aproxima a infinito ($n \to \infty$), esta ocurrencia recibe el nombre de probabilidad.

$$f = p$$

5.9 ESPACIO MUESTRAL. El conjunto Ω de los resultados posibles de un experimento cualquiera recibe el nombre de espacio muestral. El suceso A es un subconjunto del espacio muestral Ω, el evento imposible es el subconjunto vacío designado por $\emptyset$.

5.91 EJEMPLO 9. Se lanzan dos dados uno de color blanco y otro azul, el espacio muestral es el conjunto universal que a continuación se presenta (36 resultados posibles)

	6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)
	5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
DADO	4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
AZUL	3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
	2	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
	1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
		1	2	3	4	5	6

DADO BLANCO

Es decir: $\Omega = \{(1,1)\ ,\ (1,2)\ ,\ \ldots\ldots\ldots,\ (6,5)\ ,\ (6,6)\}$

Si definimos por A el evento "la suma de puntos en los 2 dados es 7". Tendremos el subconjunto: $A = \{(1,6),\ (2,5),\ (3,4),\ (4,3),\ (5,2),\ (6,1)\}$ es una parte de Ω .

Si definimos por B el evento 'la suma de los dados es 15' se tiene el subconjunto:

$\emptyset = \{.\}$ imposible

5.92 EJEMPLO 10. Si se lanzan 6 monedas el conjunto universal del número de caras (espacio muestral) es:

$\Omega = \{0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6\}$

Si definimos el evento A: 'al menos 4 caras' se obtiene el subconjunto:

$A = \{4,\ 5,\ 6\}$ es una parte de Ω

Si definimos el evento B: 'el número de caras es 6' se obtiene el subconjunto:

$B = \{6\}$ es una parte de Ω

5.10 SUBCONJUNTOS Y PROBABILIDADES. Si a cada suceso (A, B, etc.) que es parte del conjunto universal Ω, se asocia un número positivo entre cero y la unidad, se define, la probabilidad del suceso A, B, etc, y se designa por P.

$0 < P < 1$ es decir $p(A) > 0$

La probabilidad del suceso seguro es 1, es decir:

$p (\Omega) = 1$

de esta forma la probabilidad de la no ocurrencia del evento A se calcula como el complemento del evento A.

p (de no A) $= 1 - p(A)$

5.11 PROBABILIDAD TOTAL. Si A y B son sucesos incompatibles, la probabilidad de ocurrencia del evento A o del evento B ($A \cup B$) es igual a la suma de sus probabilidades.

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

5.111 EJEMPLO 11. Si A es el evento: 'extraer un 8 de una baraja de 52 cartones y B es el evento 'extraer un 10 de la misma baraja'. Cuál es la probabilidad de obtener un 8 o un 10 en una sola extracción?

$$p(A \cup B) = p(A) + p(B) = \frac{4}{52} + \frac{4}{52} = 0.15385$$

Si A y B son sucesos compatibles, la probabilidad de ocurrencia del evento A o del evento B, es igual a la suma de sus probabilidades parciales menos la probabilidad de la ocurrencia simultánea de A y B ($A \cap B$).

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

5.112 EJEMPLO 12. Si A es el evento: 'extracción de un 8' de una baraja de 52 cartas y B es el evento: 'extracción de un corazón'. Cuál es la probabilidad de obtener un 8 o un corazón en una sola extracción?

$$p(A \cup B) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = 0.30769$$

5.12 PROBABILIDAD COMPUESTA. La probabilidad de ocurrencia simultánea de 2 eventos compatibles independientes A y B ($A \cap B$) es igual al producto de sus probabilidades.

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

5.121 EJEMPLO 13. Si A es el evento: '6 en el primer lanzamiento' y B es el evento: '6 en el segundo lanzamiento' de un dado. Cuál es la probabilidad de 6 en ambos lanzamientos?

$$p(A \cap B) = p(A) \ . \ p(B) = \frac{1}{6} \ . \ \frac{1}{6} = 0.02778$$

5.13 PROBABILIDAD CONDICIONAL. La probabilidad de ocurrencia simultánea de 2 eventos compatibles dependientes A y B ($A|B$), es igual al producto de la probabilidad de ocurrencia de A por la probabilidad de ocurrencia de B dado que ocurrió el evento A, (A/B).

$$\text{prob } (A \cap B) = p(A) \ . \ p(B/A)$$

5.131 EJEMPLO 14. Un frasco contiene 7 frijoles blancos y 3 frijoles negros. Si A es el evento: 'primer frijol extraído es blanco' Si B es el evento: 'segundo frijol extraído es negro' A y B son dependientes desde que la extracción es sin reemplazo.

$$\text{prob } (A \cap B) = p(A) \cdot p(B/A) = \frac{7}{10} \cdot \frac{3}{9} = 0.233$$

5.14 VARIABLE ALEATORIA. Si la variable X toma los valores X_1, X_2, X_3, ..., X_n con probabilidades de ocurrencia: p_1, p_2, p_3, ..., p_n; se dice que X es una variable aleatoria con dominio X (espacio muestral) y rango que va desde p_1 hasta p_n (P_i).

Por lo que la probabilidad de que la variable aleatoria X igual X_i será siempre P_i, es decir:

$$\text{Prob } (X = X_i) = P_i$$

Distribución de probabilidad de la variable aleatoria. La variable aleatoria X tiene distribución de probabilidades si cumple con:

i) X debe pertenecer al conjunto universal X_1, X_2, ..., X_n

ii) Las probabilidades de cada una de ellas serán: P_1, P_2, ..., P_n.

iii) La suma de la probabilidad P_i deberá ser 1.0.

5.15 ESPERANZA MATEMATICA. El promedio de la variable aleatoria X recibe el nombre de esperanza matemática y se define así:

$$E[X] = \frac{\Sigma P_i X_i}{\Sigma P_i} = \Sigma P_i X_i$$

5.151 EJEMPLO 15. Un experimento consiste en tirar una moneda equilibrada 6 veces, definir los conceptos anteriores.

VARIABLE ALEATORIA X	PROBABILIDADES $P(X = X_i)$	PRODUCTO $X.P(X = X_i)$
0	0.015625	0.00000
1	0.093750	0.09375
2	0.234375	0.46875
3	0.312500	0.93750
4	0.234375	0.93750
5	0.093750	0.46825
6	0.015625	0.09375
Σ:	1.000000	3.00000
DOMINIO DE ESPACIO MUESTRAL	RANGO DE X	

$$E\left[X \right] = \Sigma P_i X_i = 3.0$$

La esperanza matemática de una constante es otra constante.

$$E\left[c \right] = c$$

Un ejemplo típico, es la esperanza matemática de la media.

$$E\left[\bar{X} \right] = \mu$$

CAPITULO 6

6. DISTRIBUCION DISCRETA.-

6.1 DISTRIBUCION BINOMIAL. Es aplicable unicamente a experimentos repetidos, con 2 resultados posibles donde p es la probabilidad de que se presente un suceso cualquiera (probabilidad de éxito) y $q = 1 - p$ la probabilidad de que no se presente el mismo suceso en la misma prueba (probabilidad de fracaso).

La probabilidad de X_i éxitos de este suceso en n repeticiones de la misma prueba se evalua mediante la distribución binomial.

$$\text{prob } (X = X_i) = \binom{n}{X_i} p^{X_i} q^{n - X_i}$$

Donde:

$X_i = 0, 1, 2, 3, \ldots\ldots\ldots, n$

n = No. de repeticiones

p = prob (éxito)

q = prob (fracaso)

Es decir, las probabilidades de $X_i = 0, 1, 2, \ldots\ldots, n$ corresponden a los términos sucesivos del desarrollo del binomio de Newton.

$$(q + p)^n = \binom{n}{0} q^n + \binom{n}{1} q^{n-1} p + \binom{n}{2} q^{n-2} p^2 + \ldots\ldots\ldots + \binom{n}{n} p^n$$

La variable aleatoria discreta X cuya ley de probabilidad viene definida por las realizaciones $X_i = 0, 1, 2, 3, \ldots\ldots, n$ con probabilidades $p_0, p_1, p_2, p_3, \ldots\ldots, p_n$ respectivamente, donde cada $p_i = \binom{n}{X_i} p^{Xi} q^{n-Xi}$ es una variable con distribución binomial (variable binomial).

6.11 EJEMPLO 1. HALLAR la probabilidad de conseguir 5 caras ($X_i = 5$), si se tira una moneda equilibrada 6 veces.

$$\text{prob } (X = 5) = \binom{6}{5} p^5 q = \frac{6}{64} = 0.09375$$

6.2 PROPIEDADES DE LA DISTRIBUCION BINOMIAL. El promedio de la variable aleatoria binomial es:

$$E[X] = np \text{ puesto que } E[X] = \Sigma p_i X_i / \Sigma p_i = \bar{X}$$

La varianza y desviación estandar son respectivamente:

$$VAR(X) = npq$$

$$\sigma(X) = \sqrt{npq}$$

6.21 EJEMPLO 2. Una moneda supuestamente equilibrada se tira 2 veces, calcular la media y desviación estandar de la variable binomial X.

$$\bar{X} = E[X] = n \cdot p = 2.0.5 = 1.0$$

de otra forma:

$$\bar{X} = \Sigma X_i P_i = (0)(1/4) + (1)(1/2) + (2)(1/4) = 1.0$$

$$\sigma(X) = \sqrt{npq} = \sqrt{(2)(0.5)(0.5)} = 0.7071$$

6.3 REPRESENTACION GRAFICA. Cuando $p = q$ la distribución binomial es simétrica pero si $p < 0.5$ la distribución es asimétrica a la derecha y pronunciadamente asimétrica a la derecha, para p muy pequeño. De otro modo, es decir, si $p > 0.5$ la asimetría es a la izquierda

6.31 EJEMPLO 3. Graficar las distribuciones de $(q + p)^4$, para $p = 0.5$, $p = 0.20$ y $p = 0.10$.

$$(0.5 + 0.5)^4 = 0.5^4 + 4(.5)^3(.5) + 6(.5)^2(.5)^2 + 4(.5)^1(.5)^3 + (.5)^4$$

$$.0625 + .25 + .375 + .25 + .0625$$

$$(0.2 + 0.8)^4 = 0.2^4 + 4(.2)^3(.8) + 6(.2)^2(.8)^2 + 4(.2)\ (.8)^3 + (.8)$$

$$= .0016 + .0256 + .1536 + .4096 + .4096$$

$$(0.1 + 0.9)^4 = 0.1^4 + 4(.1)^3(.9) + 6(.1)^2(.9)^2 + 4(.1)\ (.9)^3 + (0.9)^4$$

$$= .0001 + .0036 + .0486 + .2916 + .6561$$

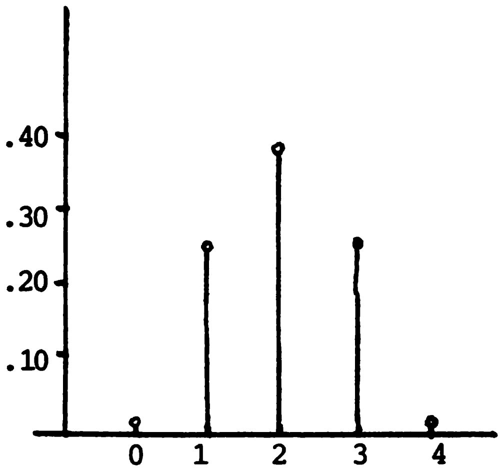


Figura 16. Binomial p = 0.5

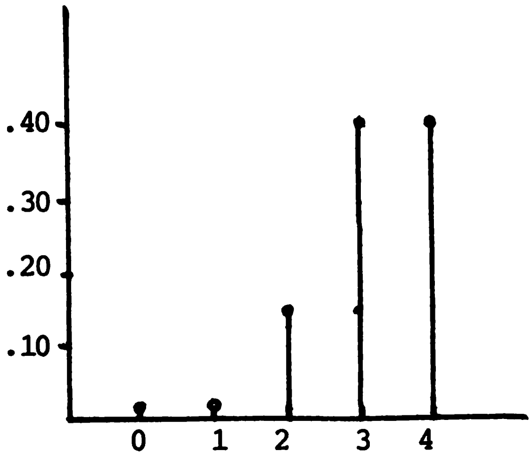


Figura 17. Binomial p = 0.2

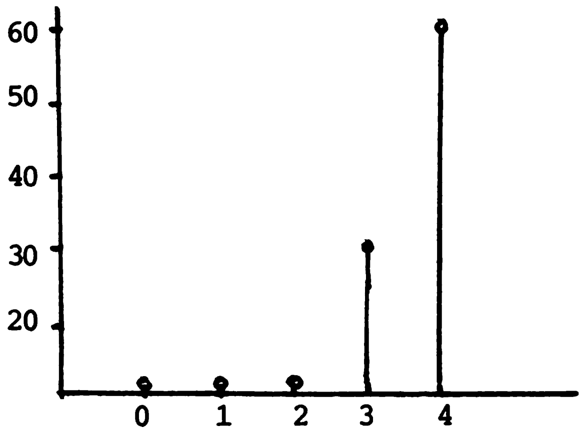


Figura 18. Binomial p = 0.1

6.4 DISTRIBUCION MULTINOMIAL. Es una generalización de la binomial. Se aplica a eventos E_1, E_2, E_3, ..., E_k que ocurren con probabilidades P_1, P_2, P_3, ..., P_k respectivamente; entonces, la probabilidad de ocurrencia de los eventos E_1, E_2, E_3, ..., E_k exactamente X_1, X_2, X_3, ..., X_k veces se evalua mediante la distribución multinomial.

$$\text{Prob}(X = X_1, X_2, X_3 \ldots, X_k) = \frac{n!}{X_1! X_2! X_3! \ldots X_k!} = p^{X_1} \cdot p^{X_2} \cdot p^{X_3} \cdot p^{X_k}$$

Donde:

$$n = X_1 + X_2 + X_3 + \ldots\ldots\ldots + X_k$$

6.41 EJEMPLO 4. Sea una ruleta formada por 13 números negros, 13 rojos y uno verde. Cuál es la probabilidad de obtener en 8 jugadas consecutivas 4 negros, 3 rojos y 1 verde? Cuál es la probabilidad de obtener exactamente 8 rojos en 8 jugadas consecutivas?

$$p \text{ (negro)} = \frac{13}{27} = .48148$$

$$p \text{ (rojos)} = \frac{13}{27} = .48148$$

$$p \text{ (verde)} = \frac{1}{27} = .037037$$

$$\text{prob } (X = 0 \text{ , } X = 8 \text{ , } X = 0) = \frac{8!}{0!\ 8!\ 0!} = \left(\frac{13}{27}\right)^0 \left(\frac{13}{27}\right)^8 \left(\frac{1}{27}\right)^0 = 0.00288$$

Propiedades de la multinomial

Promedio de la variable aleatoria

$$E\left[X_i \right] = np_i \qquad\qquad i = 1, 2, 3, \ldots\ldots\ldots$$

$$VAR\left[X_1 \right] = np_i q_i$$

6.5 DISTRIBUCIÓN DE POISSON. Se utiliza en experimentos donde la probabilidad de éxito P es pequeña, usualmente $P < 0.05$ y además n grande. Como ocurre en los fenómenos: número de accidentes, número de insectos, malezas, tráfico de insectos, emisión de partículas radioactivas, objetos defectuosos en una línea de producción, etc., en general estas reciben el nombre de fenómenos naturales raros.

La probabilidad de X_i éxitos se evalua mediante la distribución de Poisson.

$$\text{Prob } (X = X_i) = \frac{e^{-\lambda} \lambda^X}{X}$$

Donde:

$X_i = 0, 1, 2, \ldots\ldots\ldots,$ etc.

$e \;= 2.7182818$

$\lambda \;= n \,.\, p$

$X \;=$ Factorial de la variable aleatoria

La variable aleatoria X cuya ley de probabilidad viene definida por las realizaciones $X_i = 0, 1, 2, \ldots\ldots,$ etc; con probabilidades P_0, P_1, P_2, ..., etc., respectivamente, donde cada $P_i = e^{-\lambda} \lambda^{Xi} / X_i$, es una variable aleatoria con distribución de Poisson (variable Poisson).

6.51 EJEMPLO 5. La probabilidad de encontrar una mata de tomate infectada por virus X es 0.002; si se tiene un experimento con 1000 matas de tomate, cuál es la probabilidad de que ninguna esté afectada?; cuál es la probabilidad de que exactamente 3 matas estén infectadas?

$$P\ (X = 0) = \frac{e^{-\lambda}\ \lambda^0}{0} \qquad \lambda = n\ .\ p = 2.0$$

$$P\ (X = 0) = 2.718^{-2} = 0.13533$$

$$P\ (X = 3) = \frac{e^{-2}\ 2^3}{3} = 0.180447$$

6.52 EJEMPLO 5.2. Se sabe que el número promedio de vehículos que circulan por la Avenida Central por minuto entre las 11 y las 12 a.m. es de 4.8. Calcular la probabilidad de que entre las 11:00 y las 11:01 no pase ningún vehículo; pasen dos, tres, cuatro, cinco vehículos.

$$P\ (X = 0) = \frac{e^{-\lambda}\ \lambda^0}{0} = 0.00823 \qquad \lambda = 4.8$$

$$P\ (X = 3) = \frac{e^{-\lambda}\ \lambda^3}{3} = 0.151691$$

$$P\ (X = 4) = \frac{e^{-\lambda}\ \lambda^4}{4} = 0.182029$$

$$P\ (X = 5) = \frac{e^{-\lambda}\ \lambda^5}{5} = 0.349481$$

$$P\ (X = 6) = \frac{e^{-\lambda}\ \lambda^6}{6} = 0.139798$$

Propiedades de la distribución de Poisson.

El promedio de la variable aleatoria es:

$$E\ [\ X\ ] = \lambda = np$$

La varianza de la misma es:

$$V\ [\ X\ ] = \lambda$$

La desviación estandar es:

$$\sigma = \sqrt{\lambda}$$

Representación gráfica:

Para una variable Poisson con $\lambda = 1.5$, $\lambda = 2.0$ ó $\lambda = 4.8$ se tiene los casos siguientes:

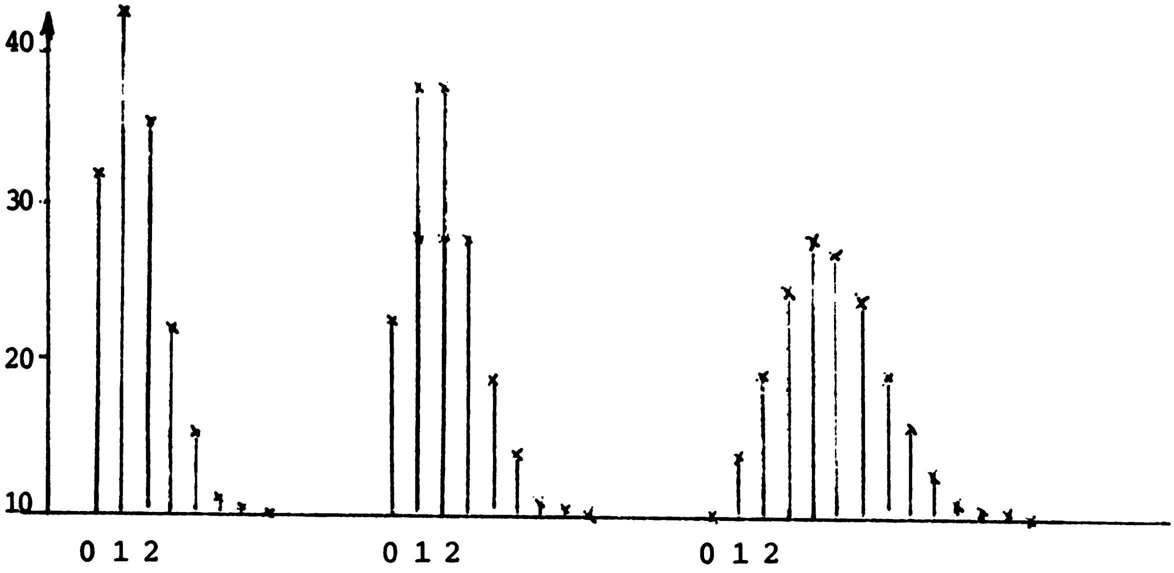


Figura 19. Distribución Poisson para diferentes valores de λ

6.6 DISTRIBUCION UNIFORME. Sea la variable aleatoria X_i que toma los valores X_1, X_2, ..., X_n con valores específicos 1, 2, 3, ..., n con probabilidades $P_1 = P_2 = P_3 = ... = P_n = \frac{1}{n}$ se define

$$p\ (X = X_i) = \frac{1}{n}$$

6.61 EJEMPLO 6. Se tira un dado, cuál es la probabilidad de obtener un uno, dos, tres, cuatro, cinco, seis?

La desviación estandar es:

$\sigma = \sqrt{\lambda}$

Representación gráfica:

Para una variable Poisson con $\lambda = 1.5$, $\lambda = 2.0$ ó $\lambda = 4.8$ se tiene los casos siguientes:

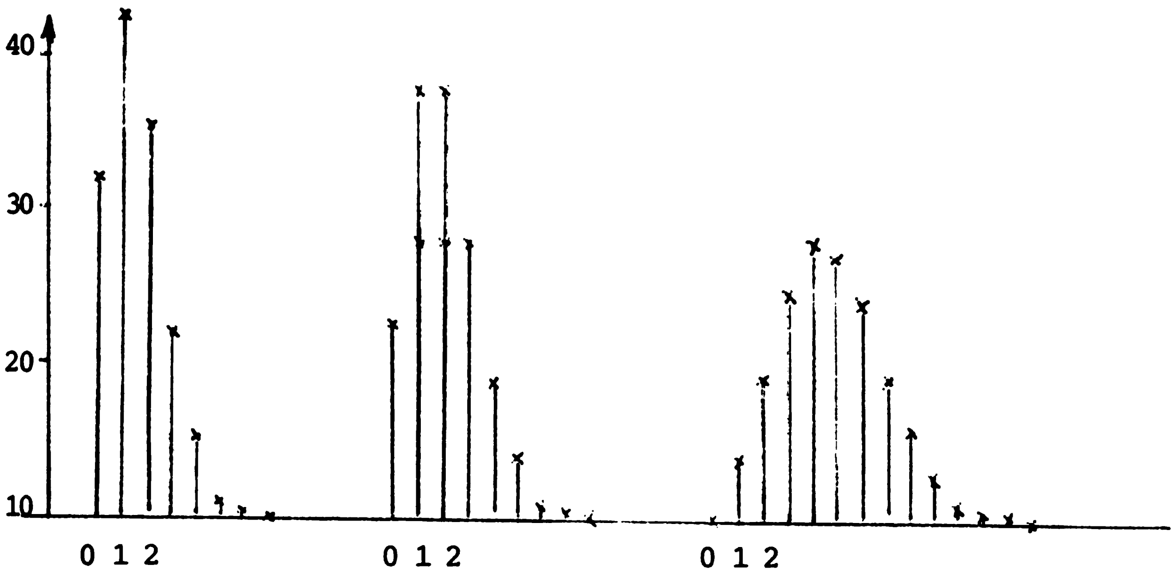


Figura 19. Distribución Poisson para diferentes valores de λ

6.6 DISTRIBUCION UNIFORME. Sea la variable aleatoria X_i que toma los valores X_1, X_2, ..., X_n con valores específicos 1, 2, 3, ..., n con probabilidades $P_1 = P_2 = P_3 = ... = P_n = \frac{1}{n}$ se define

$p\ (X = X_i) = \frac{1}{n}$

6.61 EJEMPLO 6. Se tira un dado, cuál es la probabilidad de obtener un uno, dos,tres, cuatro, cinco, seis?

$$\text{prob }(X = 1) = \frac{1}{n} = \frac{1}{6}$$

$$\text{prob }(X = 2) = \frac{1}{6}$$

$$\text{prob }(X = 6) = \frac{1}{6}$$

Representación gráfica:

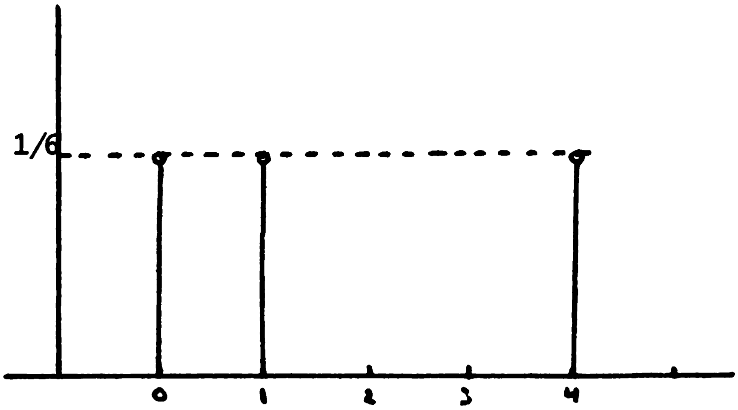


Figura 20. Distribución uniforme con $p = 1/6$

6.7 DISTRIBUCION GEOMETRICA. Sea la variable aleatoria X_i con valores X_1, X_2, X_3, ..., X_n que define el número de fracasos que proceden a la obtención del primer éxito binomial, las probabilidades individuales se evaluan por la expresión:

$$P\ (X = X_i) = q^X p \qquad X_i = 0,\ 1,\ 2,\ 3,\ \ldots\ldots$$

$$q = 1 - p$$

6.71 EJEMPLO 7. En Turrialba la probabilidad que ocurra una tormenta en cualquier día de abril (verano) es 0.30. Cuál es la probabi-

lidad que la primera tormenta ocurra el 2 de abril? Cuál es la probabilidad que la primera tormenta ocurra el 25 de abril?

$$p\ (X = 1) = (0.70)^{1}\ (0.30) = 0.210$$

$$p\ (X = 24) = (0.70)^{24}(0.30) = 0.00006$$

Representación gráfica:

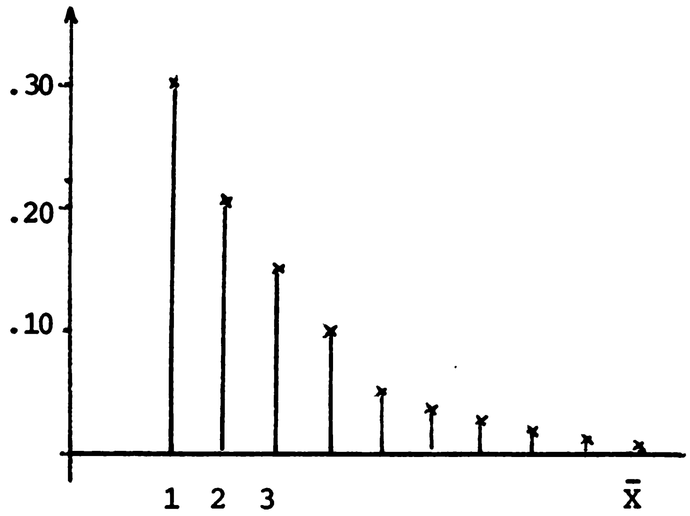


Figura 21. Distribución geométrica

6.8 DISTRIBUCION BINOMIAL NEGATIVA. Es una generalización de la geométrica, donde la variable aleatoria X_i denota el número de exactamente X + K intentos requeridos para producir exactamente K éxitos, se define mediante la ley:

$$\text{prob}\ (X = X_i) = \binom{X + K - 1}{K - 1} p^K q^X \qquad X = 0, 1, 2\ldots\ldots.. \quad K > 0$$

6.81 EJEMPLO 8. Un banco contiene 7 bolas negras y 3 bolas blancas. Cuál es la probabilidad de que se requieran 5 extracciones para

obtener 3 bolas negras? 4 extracciones 3 bolas negras?

$$\text{prob } (X = 5) = \binom{4}{2} (0.7)^3 \ (0.3)^2 = 0.185 \qquad X = 3;3 + 1;3 + 2 \text{ etc.}$$

$$\text{prob } (X = 4) = \binom{3}{2} (0.7)^3 \ (0.3)^1 = 0.309 \qquad K = 3$$

Representación Gráfica:

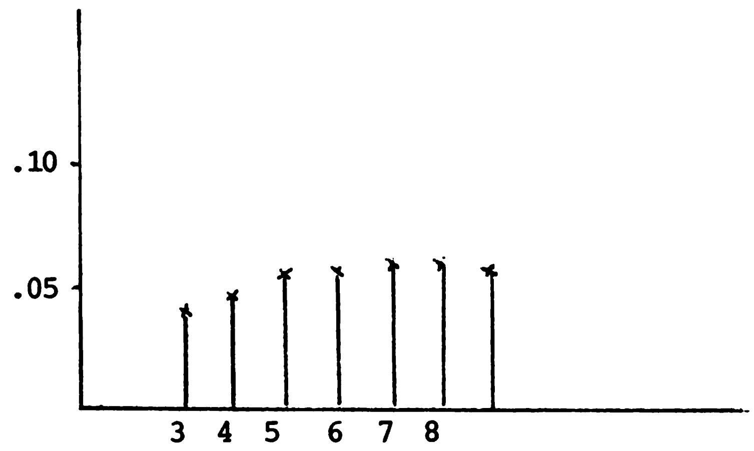


Figura 22. Distribución binomial negativa

6.9 DISTRIBUCION HIPERGEOMETRICA. Un experimento consiste en extraer n bolitas a la vez (de una sola extracción) de un frasco que contiene a bolitas negras y b bolitas blancas. La probabilidad de obtener exactamente X bolitas negras de las r extraídas es:

$$p \ (X = X_i) = \frac{\binom{a}{Xi} \cdot \binom{b}{n - Xi}}{\binom{a + b}{n}} \qquad X_i = 0, 1, 2, \ldots\ldots, n$$

$$n = \text{bolitas extraídas}$$

6.91 EJEMPLO 9. Un frasco contiene 20 frijoles de ellos 12 son rojos y 8 son blancos; si se extrae de una vez 6 frijoles, cuál es la probabilidad de que 3 sean rojos y 3 blancos? Cuál es la probabilidad de que 1 sea rojo y 5 blancos?

$$p\ (X = 3) = \frac{\binom{12}{3} \cdot \binom{8}{3}}{\binom{20}{6}} = .3178$$

$$p\ (X = 1) = 0.01733$$

Representación Gráfica:

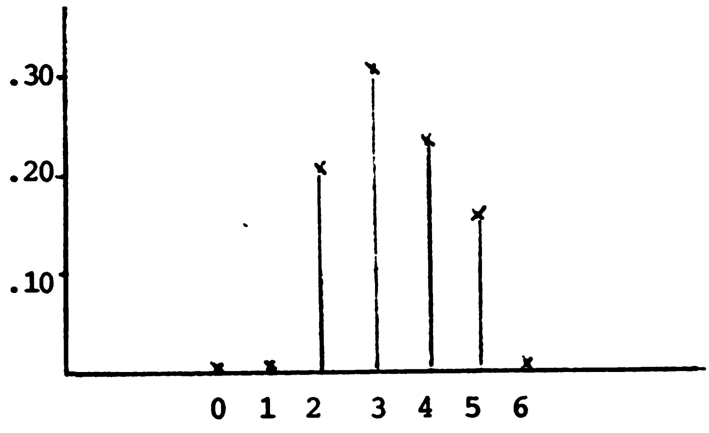


Figura 23 Distribución hipergeométrica

CAPITULO 7

7. DISTRIBUCION CONTINUA.-

Se dice que una distribución es contínua si la variable aleatoria estudiada es contínua en un rango de variación infinito; es decir, una observación puede tomar cualquier valor real positivo o negativo.

7.1 DISTRIBUCION NORMAL. Recibe también el nombre de Gauss, Laplace o normal por la forma de campana de su distribución de frecuencias. En la práctica los fenómenos naturales son más fáciles de visualizar mediante una representación simbólica (curva de campana) del espacio muestral de la variable estudiada.

Si la variable aleatoria X_i toma valores entre $-\infty < X < \infty$, se dice que tiene distribución normal si su función de densidad de probabilidad se define mediante:

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Donde:

X = Variable aleatoria

μ = Parámetro media

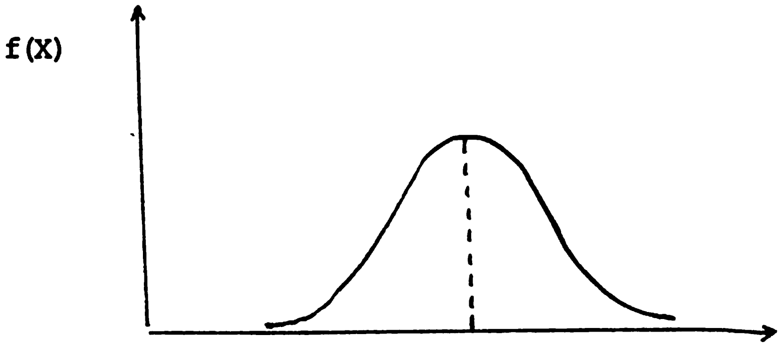
σ^2 = Parámetro variación

e = 2.71828

π = 3.1416

La mayor parte de los fenómenos naturales y por tanto un gran número de variables aleatorias tienen una distribución que se aproximan a la distribución normal.

Representación gráfica de la normal.

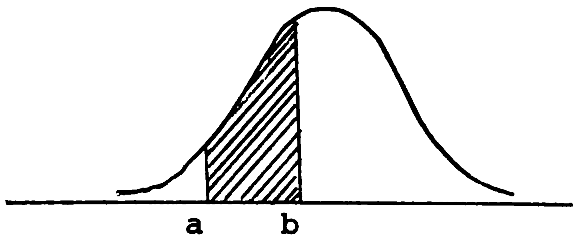


$$-\infty < X < \infty$$
$$-\infty < \mu < \infty$$
$$\sigma^2 > 0$$
$$Y = f(X)$$

Figura 24. Distribución normal de X

7.11 EJEMPLO 1. Cuál es la probabilidad de que la variable aleatoria X_i este entre a y b ?

$$\text{prob } (a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\, dx$$



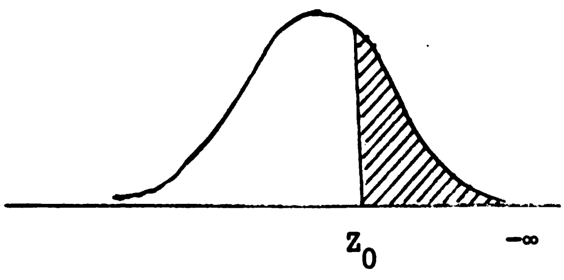
7.2 DISTRIBUCION DE Z. Aunque la forma de la curva es igual para la mayoría de los fenómenos naturales (pesos, alturas, producciones, etc.), cada uno depende de los parámetros μ y σ^2.

Para reunir bajo una curva patrón estos fenómenos y poder utilizar una sola tabla de integrales se realiza la transformación de la variable original X_i a otra Z_i según la regla:

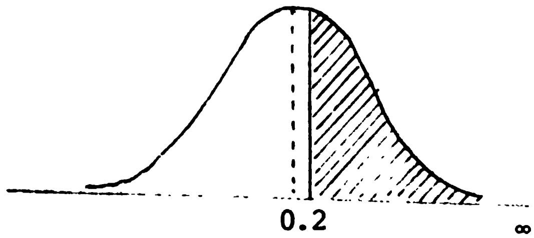
$$Z_i = \frac{X_i - \mu}{\sigma}$$

Esta variable aleatoria tiene media $\mu = 0$ y varianza = 1, como se demostró en el ejemplo 7 del capítulo 4.

7.3 TABLA DE LA INTEGRAL DE LA DISTRIBUCION DE Z. Se encuentra tabulada en gran parte de los textos de estadística, así Steel & Torrie, página 434 presenta la integral de la variable Z a la derecha de cualquier Z_0, tal que prob $(Z > Z_0) = \int_{Z_0}^{\infty} f(X)dx$ es decir; equivalente a la región sombreada.



7.31 EJEMPLO 2. i) prob $(Z > Z_0)$



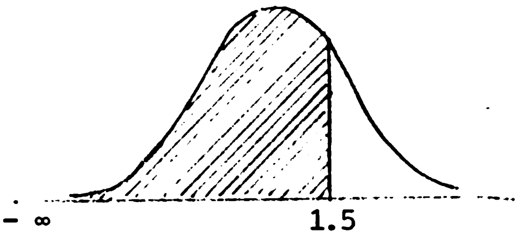
prob $(Z > 0.2) = \int_{.2}^{\infty} f(X)dx = 0.4207$ (probabilidad alta).

Cuál es la probabilidad de que Z sea mayor que $Z_0 = 2.66$?

prob $(Z > 2.66) = \int_{2.66}^{\infty} f(X)dx = 0.0039$ (probabilidad baja).

ii) prob $(Z < Z_0)$

Cuál es la probabilidad que Z sea igual o menor que $Z_0 = 1.5$?



Prob $(Z < 1.5) = \int_{-\infty}^{1.5} f(X)dx$

Las integrales a la izquierda de Z no están tabuladas, pero la curva es simétrica y el área total bajo la misma es 1.0 entonces:

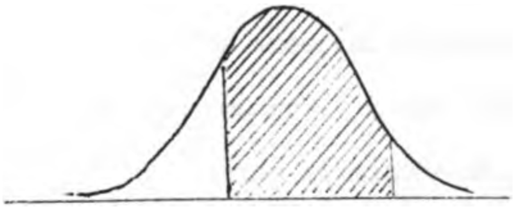
Prob $(Z < 1.5) = 1.0 - \text{prob}\ (Z > Z_0) = 1.0 - \int_{1.5}^{\infty} f(X)dx = 0.9332$

Cuál es la probabilidad que Z sea igual o menor que $Z_0 = 0.33$

Prob $(Z < 0.33) = 1.0 -$ prob $(Z > 0.33) = 1.0 - \int_{.33}^{\infty} f(X)dx = 0.6179$

iii) Prob $(Z_0 < Z < Z_1)$

Cuál es la probabilidad que Z esté en el intervalo $Z_0 = -1.5$ y $Z_1 = 2.1$?



-1.5 2.1

Prob $(-1.5 < Z < 2.1) = 1 - \int_{1.5}^{\infty} f(X)dx + \int_{2.1}^{\infty} f(X)dx$

Por simetría es mejor definir:

prob $(-1.5 < Z < 2.1) = 1 -$ (prob $(Z > 1.5)$ + prob $(Z > 2.1)$)

$= 1 - (0.0668 + 0.0179) = 0.9153$

7.32 EJEMPLO 3. Suponer que los pesos de los estudiantes de Biología se distribuyen en forma normal con media 84 y varianza 126.

El profesor de Cultura Física desea saber la probabilidad de obtener alumnos con peso igual o superior a 88. Es decir prob $(X > X_i)$ (se deberá realizar previamente la estandarización de X_i).

Se conoce:

$\mu = 84$

$\sigma^2 = 126$

$$Z_0 = \frac{X - \mu}{\sigma} = \frac{88 - 84}{11.225} = 0.36$$

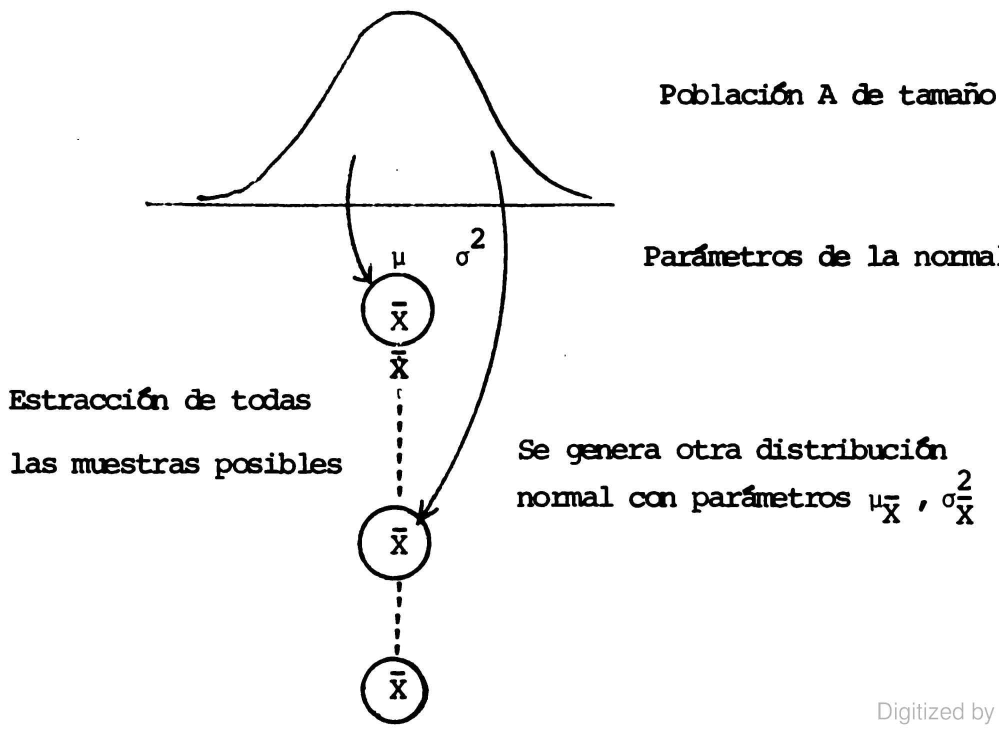
Prob $(X > X_0)$ = prob $(Z > Z_0)$ = prob $(Z > \frac{X - \mu}{\sigma}$ = prob $(Z > 0.53)$ =

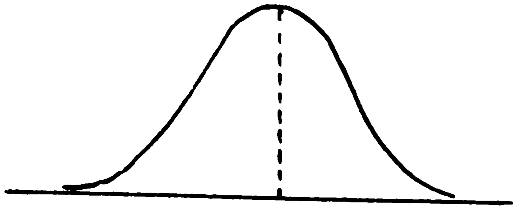
$f(X)dx$- .3594 (prob alta o medianamente alta).

7.4 DISTRIBUCION DE MEDIAS. Sea una variable aleatoria X_i proveniente de cierta población A definida por $X_i = \{X_1, X_2, X_3, \ldots\ldots\ldots\ldots, X_N\}$. Donde N tiende a infinito o bien es un número grande. Si de esta población se extraen todas las muestras posibles de tamaño n y se calcula el estimador $\bar{X}$ para cada muestra. Estos estimadores definen una nueva población de medias que recibe el nombre de Distribución de Medias, con media $\mu_{\bar{X}}$ igual a μ y varianza $\sigma^2_{\bar{X}}$ igual σ^2/n es decir:

$$\boxed{\mu_{\bar{X}} = \mu} \qquad \boxed{\sigma^2_{\bar{X}} = \sigma^2/n}$$

En forma esquemática se puede visualizar así:



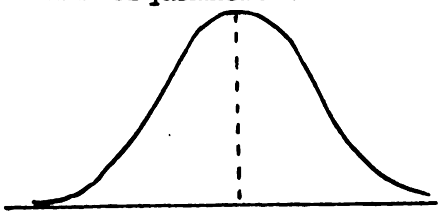


Población de medias

$\mu_{\bar{X}}$, $\sigma^2_{\bar{X}}$

7.41 EJEMPLO 4. Una población hipotética B exageradamente pequeña está formada por N = 4 individuos. Determinar el número posible de muestras que se pueden extraer de tamaño n = 2. Generar una nueva población y calcular los parámetros de estas 2 poblaciones.

En forma esquemática:



Población B con N = 4

B = { 2, 1, 2, 3 }

μ σ^2 Parámetros

$\bar{X}$

Estracción de todas las muestras n = 2 (16)

(2,2)	(1,2)	(2,2)	(3,2)
(2,1)	(1,1)	(2,1)	(3,1)
(2,3)	(1,3)	(2,3)	(3,3)
(2,2)	(1,2)	(2,2)	(3,2)

MEDIAS DE LAS MUESTRA

2.0	1.5	2.0	2.5
1.5	1.0	1.5	2.0
2.5	2.0	2.5	3.0
2.0	1.5	2.0	2.5

i) Parámetro de la media de la población B.

$$\mu = \sum_{i}^{N} X_i/N = (2 + 1 + 2 + 3)/14 = 2.0$$

ii) Parámetro de la varianza de la población B.

$$\sigma^2 = \sum_{i}^{N} (X_i - \mu)^2/N = \frac{(2 - 2)^2 + (1-2)^2 + (2-2)^2 + (3-2)^2}{4} = 0.5$$

iii) Parámetro de la media de la población de medias (la población).

$$\mu_{\bar{X}} = \frac{2.0 + 1.5 + ... + 2.5}{16} = 2.0$$

iv) Parámetro de la varianza de la población de medias.

$$\sigma^2_{\bar{X}} = \frac{(2.0 - 2.0)^2 + (1.5 - 2.0)^2 + ... + (2.5 - 2.0)^2}{16} = 0.25$$

$$= \frac{0.5}{2} = \frac{\sigma^2}{n}$$

Los anteriores resultados verifican y demuestran 2 teorías importantes de la Estadística; la esperanza matemática y el límite central.

$$\mu_{\bar{X}} = \mu = E \mid \bar{X} \mid$$

$$\sigma^2_{\bar{X}} = \sigma^2/n$$

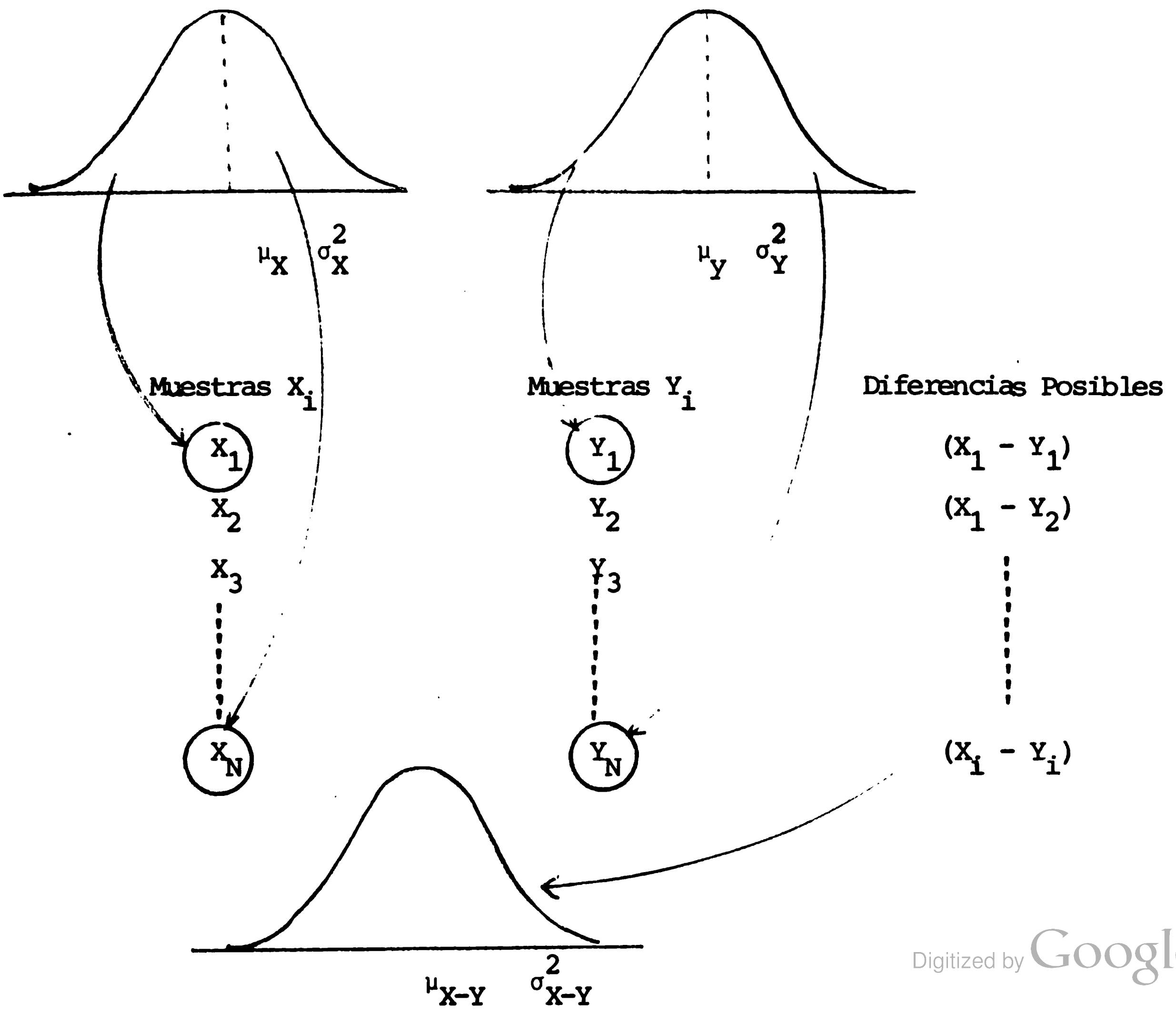
7.5 DISTRIBUCION DE DIFERENCIAS. Sea la variable X_i que defina cierta población con ámbito $X = \{X_1, X_2, X_3, ..., X_N\}$ y sea Y_i otra variable que define otra población con ámbito $Y = \{Y_1, Y_2, Y_3, ..., Y_N\}$ Para c/u de los valores de X_i se definen todas las diferencias posibles

con Y_i es decir: $X_1 - Y_1$, $X_1 - Y_2$, ... $X_1 - Y_N$, ... $X_2 - Y_1$, ..., $X_i - Y_i$. Con estas diferencias se define una tercera población de diferencias entre 2 informaciones con la media y varianza definida así:

$$\boxed{\mu_{X-Y} = \mu_X - \mu_Y}$$

$$\boxed{\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y}$$

Esquematicamente se tendría:



7.51 EJEMPLO 5. Suponer la población A compuesta por 3 fincas donde $X_i = \{2, 1, 3\}$ la población B está compuesta de 2 fincas con $Y_i = \{2, 3\}$. Todas las diferencias posibles que se pueden generar de ambas poblaciones son:

Diferencias posibles
1 - 2 = - 1
1 - 3 = - 2
2 - 2 = 0
2 - 3 = - 1
3 - 2 = 1
3 - 3 = 0

i) Parámetro de centralización.

$$\mu_X = \frac{2 + 1 + 3}{3} = 2.0$$

$$\mu_Y = \frac{2 + 3}{2} = 2.5$$

ii) Parámetro de variabilidad.

$$\sigma_X^2 = \frac{(2 - 2)^2 + (1 - 2)^2 + (3 - 2)^2}{3} = 0.6666$$

$$\sigma_Y^2 = \frac{(2 - 2.5)^2 + (3 - 2.5)^2}{2} = 0.25$$

iii) Parámetro de centralización de las diferencias.

$$\mu_{X-Y} = \frac{(-1) + (-2) + (0) + (-1) + (1) + (0)}{6} = -0.5$$

Del (i) se tiene: $\mu_X - \mu_Y = 2.0 - 2.5 = -0.5$

es decir: $\mu_{X-Y} = \mu_X - \mu_Y$

iv) Parámetro de variabilidad de las diferencias.

$$\sigma^2_{X-Y} = \frac{(-1.0-(0.5))^2 + (-2.0-(-0.5))^2 + \dots + (0.0-(-0.5))^2}{6} = 0.9166$$

7.6 DISTRIBUCION DE DIFERENCIAS DE MEDIAS. Se X_i la variable que define cierta población con ámbito $\Omega = \{X_1, X_2, X_3, \dots\dots, X_N\}$. Sea y_i otra variable que define otra población cuyo ámbito es $\Omega = \{Y_1, Y_2, \dots\dots, Y_N\}$.

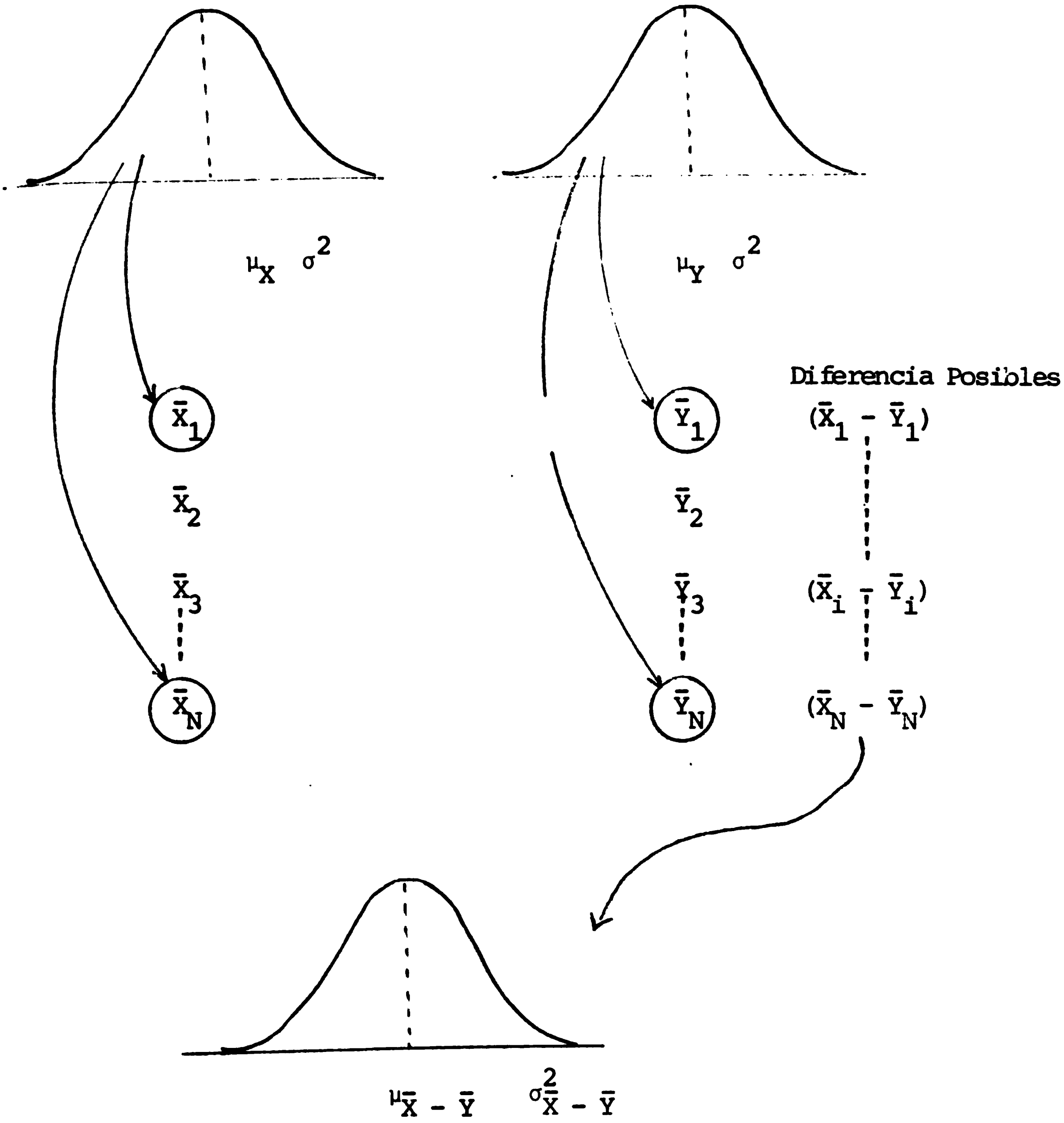
De la primera población, se extraen todas las muestras posibles de tamaño n, y para cada muestra se calcula el estimador $\bar{X}$; de la otra población: también se obtienen todas las muestras posibles de tamaño n_2, para cada muestra se calcula el estimador $\bar{Y}$. Con los estimadores se forman todas las diferencias posibles tal que se defina una tercera población de diferencias de medias con ámbito $\Omega = \{\bar{X}_1 - \bar{Y}_1 , \bar{X}_1 - \bar{Y}_2 , \bar{X}_1 - \bar{Y}_N, \dots\dots, \bar{X}_2 - \bar{Y}_1 , \dots\dots, \bar{X}_i - \bar{Y}_i , \dots\dots, \bar{X}_N - \bar{Y}_N\}$.

La media y varianza de esta población generada es:

$$\boxed{\mu_{\bar{X}-\bar{Y}} = \mu_X - \mu_Y}$$

$$\boxed{\sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}}$$

En forma gráfica:



7.61 EJEMPLO 6. Suponer la población A compuesta de 2 fincas donde $X_i = \{4, 2\}$; suponer otra población B compuesta también de 2 fincas donde $Y_i = \{0, 2\}$. Todas las diferencias posibles entre promedios de las 2 poblaciones son:

Muestras		Promedios			Diferencias posibles entre medias			
X	Y	$\bar{X}$		$\bar{Y}$	$\bar{X} - \bar{Y}$			
(4;4)	(0;0)	4		0	4	3	3	2
(4;2)	(0;2)	3		1	3	2	2	1
(2;4)	(2;0)	3		1	3	2	2	1
(2;2)	(2;2)	2	etc.	2	2	1	1	0

i) Parámetros de centralización.

$$\mu_X = \frac{4 + 2}{2} = 3.0$$

$$\mu_Y = \frac{0 + 2}{2} = 1.0$$

ii) Parámetro de variabilidad.

$$\sigma^2_X = \frac{(4-3)^2 + (2-3)^2}{2} = 1.0$$

$$\sigma^2_Y = \frac{(0-1)^2 + (2-1)^2}{2} = 1.0$$

iii) Parámetro de centralización de diferencia de medias.

$$\mu_{\bar{X} - \bar{Y}} = \frac{4 + 3 + \ldots\ldots + 1 + 0}{16} = 2.0$$

También $\mu_X - \mu_Y = 3.0 - 1.0 = 2.0$

$$\mu_{\bar{X}} - \mu_{\bar{Y}} \quad \mu_X - \mu_Y$$

iv) Parámetro de variabilidad de las diferencias entre medias.

$$\sigma^2_{\bar{X} - \bar{Y}} = \frac{(4-2)^2 + (3-2)^2 + \ldots. + (1-2)^2 + (0-2)^2}{16} = \frac{16}{16} = 1.0$$

Por otra parte: $\sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}} = \dfrac{1.0 + 1.0}{2} = 1.0$

CAPITULO 8

8. PRUEBAS DE HIPOTESIS.-

8.1 DECISIONES ESTADISTICAS. En la práctica, se tiene que tomar decisiones acerca de las poblaciones o específicamente acerca de los parámetros poblacionales, partiendo de la información de la muestra. Estas decisiones reciben el nombre de decisiones estadísticas.

8.2 HIPOTESIS ESTADISTICAS. Para tomar decisiones, previamente se realiza determinado supuesto o conjetura acerca de los parámetros que se investiga. Estos supuestos pueden ser o no ser ciertos y reciben el nombre de hipótesis estadísticas, así si se quiere tomar una decisión sobre la bondad de un dado, (si un dado está cargado o no), se formula la hipótesis de que cualquier cara es igualmente probable de obtener (P = 1/6); o si se quiere decidir sobre si un tratamiento es mejor que otro, se formula la hipótesis de la no existencia de diferencia entre tratamientos, o bien la hipótesis de la superioridad de un tratamiento con respecto a otro.

8.3 PROPOSITO DE LA EXPERIMENTACION. La experimentación se realiza con el objeto de obtener información, para obtener evidencias o pruebas de que ciertas afirmaciones (hipótesis estadísticas) son verdaderas o falsas.

Una prueba experimental arroja un resultado particular (estimador $\bar{X}$), de una familia grande de resultados posibles; es un evento particular del conjunto universal de resultados posibles.

= Experimento 1, Experimento 2, ..., Experimento_N

= $\bar{X}_1$, $\bar{X}_2$, ..., $\bar{X}_N$

8.4 ROL DE LA ESTADISTICA. La finalidad de los métodos estadísticos es condensar la información de una muestra pequeña o grande, o la información proveniente de un experimento particular a dos o más valores numéricos que reciben el nombre de estimadores (usualmente media, variancia) los que son capaces de describir las muestras y útiles para estimar los parámetros de la población estudiada.

Las distribuciones generadas de la normal (t, F, X^2, Z, etc.), provee los medios para calcular probabilidades de ocurrencia de cierto evento particular; con base en esta probabilidad, se acepta o rechaza la hipótesis en prueba. Si la probabilidad de ocurrencia de ese evento es pequeña ($P < 0.05$) se rechaza la hipótesis planteada, porque actualmente ocurrió un evento cuya probabilidad es muy baja, el resultado obtenido de esta prueba es significativo, en consecuencia, la población, representada en este caso por la muestra experimental responde a otra hipótesis (hipótesis alterna). Por otra parte, si la probabilidad del evento no es suficientemente pequeña ($P > 0.05$) se acepta la hipótesis planteada (hipótesis nula). El investigador en esta situación está en condiciones de predecir que la hipótesis es verdadera.

8.5 SUPOSICIONES DE LA EXPERIMENTACION. (MUESTRA). Se diseñan y conducen los experimentos, teniendo en cuenta que la muestra o el experimento será representativa de la población en estudio (elección aleatoria de los componentes de la muestra, según el Capítulo I).

Los datos por sí solos deberán ser suficientes como para arrojar evidencia en pro o en contra de la hipótesis que se prueba.

La respuesta resultante será una variable cuantitativa contínua, aunque existen respuestas que se describen con variables cuantitativas discretas y algunas veces variables cualitativas.

8.6 CARACTERISTICAS DE LOS DATOS DE UNA MUESTRA. A pesar de la fluctuación de las observaciones o mediciones individuales ellas tienden a distribuirse alrededor de un valor típico, central, como la media, mediana o modo. En efecto, se supone que el fenómeno en estudio tiene una distribución normal (Capítulo 3).

Es notable que, observaciones repetidas sobre el mismo material o extraídas al azar de una misma población o conseguidas en un experimento no son idénticas para sus diferentes características. Esta falta de identidad entre individuos pertenecientes a una misma población se llama variabilidad o dispersión (Capítulo 4).

8.7 VENTAJAS DE LA REPETICION O TAMAÑO DE MUESTRA. De la distribución de medias obtenidas de la normal se infiere que la repetición mejora la precisión de los estimadores y por ende la prueba estadística a medida que aumenta el tamaño de muestra; la variabilidad de las medias es menor, ya que ella es inversamente proporcional al número de repeticiones u observaciones, es decir:

$$s_{\bar{X}}^2 = \frac{s^2}{n} \quad ó \quad s_{\bar{X}} = \frac{s}{\sqrt{n}}$$

8.8 ALEATORIZACION. Ninguna muestra o tratamiento debe ser favorecido ni accidentalmente ni a propósito; con ello, se asegura la obtención de estimadores válidos e imparciales que permite utilizar la teoría de probabilidades en la prueba de hipótesis. (Las muestras deben ser representativas).

8.9 METODOLOGIA DE LA PRUEBA DE HIPOTESIS. Se aplica el método científico y debe definirse claramente:

i) Población y parámetros de la investigación
ii) Hipótesis estadísticas a probar
iii) Distribución aplicable al experimento
iv) Obtención de datos (información numérica)
v) Nivel de significación
vi) Inferencia
vii) Conclusión

8.10 ILUSTRACION DE PRUEBAS DE HIPOTESIS.

EJEMPLO 1. El peso medio de 50 estudiantes que tomaron parte en pruebas atléticas es de 78 kgs con una desviación estandar de 2.5 kg.; mientras que otros 50 estudiantes que no muestran interés en actividades atléticas tienen media y desviación estandar igual a 77.5 y 2.8 kg.

Se desea probar la hipótesis de que los estudiantes que participan en pruebas atléticas son más pesados que los otros.

METODOLOGIA:

i) Hipótesis a probar:

$$H_0 : \mu_A = \mu_\beta$$

$$H_A : \mu_A > \mu_\beta$$

La desigualdad de la hipótesis alterna conlleva el uso de la distribución teórica, considerando una cola (prueba de una cola).

ii) Criterio de Prueba. Se utiliza la distribución normal, desde que el tamaño de las 2 muestras es grande (n > 30), y por otra parte, se conocen las varianzas: $(\sigma_A^2 , \sigma_B^2)$.

$$Z_0 = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sigma_{\bar{X}_A - \bar{X}_B}}$$

iii) Aritmética. Fundamentalmente se calcula: $\bar{X}_A - \bar{X}_B$, $\sigma_{\bar{X}_A - \bar{X}_B}$, Z_0.

$$\bar{X}_A - \bar{X}_B = 78 - 77.3 = 0.70$$

$$\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\sigma_A^2}{N_1} + \frac{\sigma_B^2}{N_2}} = \sqrt{\frac{2.5^2}{50} + \frac{2.8^2}{50}} = 0.53$$

$$Z_0 = \frac{0.7}{0.53} = 1.32$$

iv) Nivel de significación. Es usual considerar el nivel $\alpha = 0.05$ la distribución de Z para $\alpha = 0.05$.

v) Decisión. Para pruebas de 1 cola y un nivel de significación $\alpha = 0.05$ se rechazaría la hipótesis nula si el Z_0 experimental estuviese a la derecha del valor tabular $Z = 1.65$, el presente resultado estadístico es 'no significativo' ($p > 0.05$); se decide aceptar la hipótesis nula (Z está a la izquierda de $Z = 1.65$).

vi) Conclusión. Las alturas promedio de los 2 grupos de estudiantes son iguales y la pequeña diferencia observada en favor del primer grupo se debe al azar.

EJEMPLO 2. CASO I. Se planteó cierto experimento en 24 parcelas para probar el efecto de la presencia o ausencia de K en el rendimiento de papa.

Cuadro 1. Producción de papa: Kg/Parcela (2 tratamientos)

n	Sin K A	Con K B	A^2	B^2
1	10.0	18.0	100.00	324.00
2	13.5	14.2	182.25	201.64
3	12.4	22.5	153.76	506.25
4	11.3	13.0	127.69	169.00
5	12.8	15.0	163.84	225.00
6	12.0	16.5	144.00	272.25
7	11.5	19.5	132.25	380.25
8	12.5	17.0	156.25	289.00
9	12.4	19.5	153.76	380.25
10	11.6	21.0	134.56	441.00
11	12.0	22.5	144.00	506.25
12	12.5	17.5	156.25	306.25
Σ:	144.5	216.2	1748.61	4001.14
$\bar{X}$:	12.04	18.01		

METODOLOGIA:

Condiciones previas:

Tamaño de muestra: $n_A = n_B$

Varianza : $\sigma_A^2 \neq \sigma_B^2$

i) Hipótesis:

$$H_0 : \mu_A = \mu_B$$

$$H_A : \mu_A \neq \mu_B$$

ii) Criterio de prueba: 't de Student', por la magnitud de n.

$$\pm t_0 = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}}$$

iii) Cálculo de estimadores: s_A^2 , s_B^2 , $\bar{X}_A$, $\bar{X}_B$, t_0 , $s_{\bar{X}_A - \bar{X}_B}$

$$s_A^2 = \frac{\Sigma A^2 - (\Sigma A)^2/n}{n - 1} = \frac{1748.61 - (144.5)^2/12}{11} = \frac{8.59}{11} = 0.78$$

$$s_B^2 = \frac{\Sigma B^2 - (\Sigma B)^2/n}{n - 1} = \frac{4001.14 - (216.2)^2/12}{11} = \frac{105.94}{11} = 9.63$$

$$\bar{X}_A = \frac{\Sigma A}{n} = \frac{144.5}{12} = 12.04$$

$$\bar{X}_B = \frac{\Sigma B}{n} = \frac{216.5}{12} = 18.01$$

$$s_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{s_A^2}{n} + \frac{s_B^2}{n}} = \sqrt{\frac{0.87}{12} + \frac{9.63}{12}} = 0.93$$

Luego:

$$t_0 = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}} = \frac{12.04 - 18.01}{0.93} = 6.42$$

iv) Nivel de significación: $\alpha = 0.05$ (es el nivel más usado)

$t_{0.05}$ con (n - 1) G.L. = 2.201

v) Decisión: Rechazamos la H_o.

vi) Conclusión. La diferencia detectada en estas dos muestras es atribuible al efecto benéfico o perjudicial del tratamiento investigado (K)

EJEMPLO 3; CASO II. Se planteó cierto experimento en 24 parcelas para probar el efecto de la presencia o ausencia de K en el rendimiento de papa.

Cuadro 2. Producción de papa: Kg/parcela (2 tratamientos)

n	Sin K A	Con K B	A^2	B^2
1	20.0	24.0	400.00	576.00
2	24.0	28.0	576.00	784.00
3	21.0	25.0	441.00	625.00
4	22.0	25.0	484.00	625.00
5	23.0	27.0	529.00	729.00
6	24.0	27.5	576.00	756.25
7	22.5	28.0	506.25	784.00
8	22.0	26.0	484.00	676.00
9	21.5	26.0	462.25	676.00
10	20.0	24.5	400.00	600.25
11	22.0	26.5	484.00	702.25
12	24.0	28.5	576.00	812.25
Σ:	266.0	316.0	5918.50	8346.00
$\bar{X}$:	22.16	26.33		

METODOLOGIA:

Condiciones previas:

Tamaño de muestra: $n_A = n_B$

Varianza : $\sigma_A^2 = \sigma_B^2$

i) Hipótesis:

$$H_o : \mu_A = \mu_B$$

$$H_A : \mu_A \neq \mu_B$$

ii) Criterio de Prueba: 't de Student'por el tamaño de n.

$$\pm t_0 = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}}$$

iii) Cálculo de estimadores: s_A^2 , s_B^2 , $\bar{X}_A$, $\bar{X}_B$, t_0 , s_C^2 , $s^2_{\bar{X}_A - \bar{X}_B}$

$$s_A^2 = \frac{\Sigma A^2 - (\Sigma A)^2/n}{n - 1} = \frac{5918.5 - (266)^2/12}{11} = 2.02$$

$$s_B^2 = \frac{\Sigma B^2 - (\Sigma B)^2/n}{n - 1} = \frac{8346 - (316)^2/12}{11} = 2.24$$

$$\bar{X}_A = \frac{\Sigma A}{n} = \frac{266.0}{12} = 22.16$$

$$\bar{X}_B = \frac{\Sigma B}{n} = \frac{316.0}{12} = 26.33$$

$$s_C^2 = \frac{s_A^2 + s_B^2}{2} = \frac{2.02 + 2.24}{2} = 2.13$$

$$s_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{2s_c^2}{n}} = \sqrt{\frac{2 \times 2{,}13}{12}} = 0.60$$

Luego:

$$t_0 = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}} = \frac{22.16 - 26.33}{0.60} = 6.95$$

iv) Nivel de significación. $\alpha = 0.05$ (es el nivel más usado)

$t_{0.05}$ con $(n_A - 1) + (n_B - 1)$ G.L. = 2.074

v) Decisión. Rechazamos la H_o.

vi) Conclusión. La diferencia entre promedios observado es atribuible al efecto de tratamiento (K), por haberse conseguido un resultado significativo.

EJEMPLO 4; CASO III. Se planteó cierto experimento en 26 parcelas, de las cuales 4 parcelas han sido consideradas perdidas; el objeto es probar el efecto de la presencia o ausencia de K en el rendimiento de papa.

Cuadro 3. Producción de papa: Kg/parcela (2 tratamientos)

n	Sin K A	Con K B	A^2	B^2
1	8.0	6.0	64.00	36.00
2	9.0	6.5	81.00	42.25
3	8.5	7.0	72.25	49.00
4	9.4	6.5	88.36	42.25
5	9.3	6.4	86.49	40.96
6	8.4	7.1	70.56	50.41
7	8.5	7.2	72.25	51.84
8	8.6	6.2	73.96	38.44
9	8.0	6.3	64.00	39.69
10	8.5		72.25	
11	9.0		81.00	
12	8.5		72.25	
13	8.4		70.56	
Σ:	112.1	59.2	968.93	390.84
$\bar{X}$:	9.3	6.57		

METODOLOGIA:

Condiciones previas:

Tamaño de muestra: $n_A \neq n_B$

Varianza : $\sigma_A^2 \quad \sigma_B^2$

i) Hipótesis:

$$H_O : \mu_A = \mu_B$$

$$H_A : \mu_A \neq \mu_B$$

ii) Criterio de Prueba: 't de Student', por el tamaño de la muestra n.

$$\pm t_0 = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}}$$

iii) Cálculo de estimadores: s_A^2 , s_B^2 , $\bar{X}_A$, $\bar{X}_B$, t_0 , s_c^2 , $s_{\bar{X}_A - \bar{X}_B}$

$$s_A^2 = \frac{\Sigma A^2 - (\Sigma A)^2/n}{n - 1} = \frac{968.93 - (112.1)^2/13}{12} = 0.19$$

$$s_B^2 = \frac{\Sigma B^2 - (\Sigma B)^2/n}{n - 1} = \frac{390.84 - (59.2)^2/9}{8} = 0.18$$

$$\bar{X}_A = \frac{\Sigma A}{n} = 9.31$$

$$\bar{X}_B = \frac{\Sigma B}{n} = 6.57$$

$$s_c^2 = \frac{(n_A - 1)\ s_A^2 + (n_B - 1)\ s_B^2}{(n_A - 1) + (n_B - 1)} = \frac{12(0.19) + 8(0.18)}{20} = 0.19$$

$$s_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{s_c^2}{n_A} + \frac{s_c^2}{n_B}} = \sqrt{\frac{0.19}{13} + \frac{0.19}{9}} = 0.18$$

Luego:

$$t_0 = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}} = \frac{9.31 - 6.57}{0.18} = 15.22$$

iv) Nivel de significación. $\alpha = 0.05$ (es el nivel más usual).

$t_{0.05}$ con $(n_A - 1) + (n_B - 1)$ G.L. = 2.086

v) Decisión. Rechazamos la H_o.

vi) Conclusión. La diferencia detectada en estas dos muestras es atribuible al efecto benéfico o perjudicial del tratamiento investigado (K).

EJEMPLO 5; CASO IV. Se planeó cierto experimento en 28 parcelas; de las cuales 4 parcelas han sido consideradas perdidas; el objeto es probar el efecto de la presencia o ausencia de K en el rendimiento de papa.

Cuadro 4. Producción de papa: Kg/parcela (2 tratamientos).

n	Sin K A	Con K B	A^2	B^2
1	3.2	4.5	10.24	20.25
2	3.5	4.2	12.25	17.64
3	3.4	4.1	12.56	16.81
4	3.6	4.6	12.96	21.16
5	3.7	4.7	13.69	22.09
6	3.4	4.2	11.56	17.64
7	3.3	4.1	10.89	16.81
8	8.5	4.5	72.25	20.25
9	3.4	4.5	11.56	20.25
10	3.4	4.4	11.56	19.36
11	3.6		12.96	
12	3.7		13.69	
13	3.2		10.24	
14	3.1		9.61	
Σ:	53.0	43.8	225.02	192.26
$\bar{X}$	3.79	4.38		

METODOLOGIA:

Condiciones previas:

Tamaño de muestra: $n_A \neq n_B$

Varianza : $\sigma_A^2 \neq \sigma_B^2$

i) Hipótesis:

$H_0 : \mu_A = \mu_B$

$H_A : \mu_A \neq \mu_B$

ii) Criterio de Prueba. 't de Student', hallada a través de las medias ponderada de los t_0 proveniente de cada muestra; ya que aún no existe una distribución exacta que aproxime a un valor t_0 calculado.

$$t_0 = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}} \quad ; \quad t' = \frac{W_A . t_A + W_B . t_B}{W_A + W_B}$$

iii) Cálculo de estimadores:

$$s_A^2 \,,\ s_B^2 \,,\ \bar{X}_A \,,\ \bar{X}_B \,,\ t_0 \,,\ t' \,,\ t_A \,,\ t_B \,,\ W_A \,,\ W_B \,,\ s_{\bar{X}_A} - \bar{X}_B$$

$$s_A^2 = \frac{\Sigma A^2 - (\Sigma A)^2/n}{n - 1} = \frac{225.02 - (53)^2/14}{13} = 1.88$$

$$s_B^2 = \frac{\Sigma B^2 - (\Sigma B)^2/n}{n - 1} = \frac{192.26 - (43.80)^2/10}{9} = 0.05$$

$$\bar{X}_A = \frac{\Sigma A}{n} = \frac{53.0}{14} = 3.79$$

$$\bar{X}_B = \frac{\Sigma B}{n} = \frac{43.8}{10} = 4.38$$

$$W_A = \frac{s_A^2}{n_A} = \frac{1.88}{14} = 0.13$$

$$W_B = \frac{s_B^2}{n_B} = \frac{0.05}{10} = 0.005$$

t_A = 2.160 (de la tabla, con (n_A - 1) G.L. y $\alpha = 0.05$

t_B = 2.262 (de la tabla, con (n_B - 1) G.L. y $\alpha = 0.05$

$$t' = \frac{W_A \cdot t_A + W_B \cdot t_B}{W_A + W_B} = \frac{0.13(2.160) + 0.005(2.262)}{0.13 + 0.005} = 2.1637$$

$$s_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} = \sqrt{\frac{1.88}{14} + \frac{0.05}{10}} = 0.367$$

Luego:

$$t_0 = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}} = \frac{3.79 - 4.38}{0.367} = 1.607$$

iv) Nivel de significación. $\alpha = 0.05$ (es el nivel más usado)

$t' = 2.1637$

$t_0 = 1.607$

v) Decisión. Aceptamos la hipótesis nula (H_o), ya que el resultado no es significativo.

vi) Conclusión. La diferencia observada entre promedios es atribuible únicamente a errores de muestreo o variabilidad natural, y no al efecto del tratamiento (K).

EJEMPLO 6; CASO V. Se planteó cierto experimento con 12 hojas de tabaco, para evaluar el efecto de 2 cepas de virus (X, Y) sobre mitades de hoja.

Cuadro 5. Magnitud del daño

n	X	Y	D = X - Y	$(D - \bar{D})^2$
1	113.5	120.5	-7.00	638.07
2	118.5	90.5	28.00	94.87
3	120.5	105.5	15.00	10.63
4	132.5	110.5	22.00	13.99
5	124.5	90.5	34.00	247.75
6	134.5	112.5	22.00	13.99
7	135.5	140.5	-5.00	541.03
8	145.5	105.5	40.00	472.63
9	160.5	130.4	30.10	140.19
10	170.5	150.5	20.00	3.03
11	146.5	135.5	11.00	52.71
12	174.5	165.5	9.00	85.75
Σ:			219.10	2314.64
$\bar{X}$			18.26	

METODOLOGIA:

$$H_O : \mu_D = 0$$

$$H_A : \mu_D \neq 0$$

ii) Criterio de Prueba: 't de Student', por la magnitud de n.

$$t_9 = \frac{\bar{D}}{s_{\bar{D}}}$$

iii) Cálculo de estimadores: $\bar{D}$, $s_{\bar{D}}$, t_0 , s_D

$$\bar{D} = \frac{\Sigma D}{n} = \frac{219.10}{12} = 18.26$$

$$s_D = \sqrt{\frac{\Sigma(D - \bar{D})^2}{n - 1}} = \sqrt{\frac{2314.64}{11}} = 14.51$$

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{14.51}{\sqrt{12}} = \frac{14.51}{3.46} = 4.19$$

Luego:

$$t_0 = \frac{\bar{D}}{s_{\bar{D}}} = \frac{18.26}{4.19} = 4.36$$

iv) Nivel de significación. $\alpha = 0.05$

$t_{0.05}$ con $(n - 1)$ G.L. = 2.201

v) Decisión. Rechazamos la H_o.

vi) Conclusión. La diferencia es atribuible al efecto del tratamiento (cepas).

Entre los objetivos del Convenio IICA/ROCAP (PIADIC), se destaca la preparación de manuales con metodologías, normas y procedimientos uniformes, compatible y comparables para el manejo de información agrícola de la Región.
Esta publicación es el resultado de la colaboración de la División de Procesamiento de Datos, el Programa de Información Agropecuaria del Istmo Centroamericano; los costos de impresión fueron pagados por PIADIC.

Se destaca el agradecimiento al personal de la D.P.D. que colaboró en la elaboración del manual.

FECHA DE DEVOLUCION

5 SEP 1988			
0 5 JUN 1991			

A
-147

Título Manual de Estadística Descrip-
tiva

Fecha Devolución	Nombre del solicitante
5 SEP 1988	yadira Gonzál
0 5 JUN 1991	M. Bla

EDITORIAL IICA