"INFERENCIA ESTADISTICA EN LA ADMINISTRACION DE PERSONAL".

SEMINARIO DE INVESTIGACION ADMINISTRATIVA
QUE PARA OBTENER EL TITULO DE
LICENCIADO EN ADMINISTRACION DE EMPRESAS
P R E S E N T A N :

FERNANDO ROSALES JIMENEZ
MARIO SERRANO CAPACETA

GUADALAJARA, JALISCO

UNIVERSIDAD AUTONOMA DE GUADALAJARA

INCORPORADA A LA

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESCUELA DE ADMINISTRACION, CONTABILIDAD

Y ECONOMIA

"INFERENCIA ESTADISTICA EN LA ADMINISTRACION DE PERSONAL".

SEMINARIO DE INVESTIGACION ADMINISTRATIVA

QUE PARA OBTENER EL TITULO DE

LICENCIADO EN ADMINISTRACION DE EMPRESAS

P R E S E N T A N :

FERNANDO ROSALES JIMENEZ

MARIO SERRANO CAPACETA

DIRECTOR DEL SEMINARIO

LIC. JUAN BAUTISTA LACAYO MARENCO

GUADALAJARA, JALISCO 1974

A mis Padres:
FERNANDO ROSALES OCAMPO
JOSEFINA JIMENEZ DE ROSALES.

A mis Hermanas.

A mis Padres:
ALEJANDRO SERRANO ZUÑIGA
MARIA DEL CARMEN CAPACETA
DE SERRANO
con cariño y gratitud.

A mis Tíos:
LUCIANO JIMENEZ CAMACHO
ERNESTINA CAPACETA
DE JIMENEZ.
con afecto y gratitud.

A mis Hermanos:
YOLANDA DEL CARMEN
ALEJANDRO.

A nuestra Universidad.

Al Lic. JUAN BAUTISTA LACAYO MARENCO:

Queremos hacer patente nuestro agradecimiento, ya que por medio de sus vastos conocimientos y atinados consejos nos guió para la realización de este estudio.

I N D I C E

INTRODUCCION

Indudablemente que la elaboración de un estudio a nivel profesional, representa una labor árdua, analítica, sistematizada, ética y de gran profundidad objetiva. Con fundamento en éstas afirmaciones, y más aún, con la emoción y deseo que nos embarqa en poder proyectar una investigación fructífera, trascendental y de actualidad, es por ello que en el presente seminario de investigación administrativa que hemos desarrollado con miras a la obtención del título profesional de Licenciado en Administración de Empresas, se exponen las técnicas y teorías estadísticas más usuales sobre la administración científica del personal.

La moderna dirección de empresas en la actualidad, ha sido testigo de un cambio radical en el modo de estudiar y resolver las cuestiones de organización del trabajo humano para fines productivos. Se debe ésto fundamentalmente a la idea, que cada vez cobra mayor número de adeptos, de que los problemas organizacionales deben abordarse con un criterio y un método científicos, en lugar de aproximarse a los mismos para improvisar soluciones que deben rectificarse una y otra vez, con grandes gastos y positivos trastornos en el desarrollo del trabajo.

Obviamente, con la mira puesta en éstos deseos, y reflexionando detenidamente sobre de qué medios podríamos valernos y auxiliarnos en la inteligencia de por obtenerse resultados lo más objetiva y cuantitativamente referente al comportamiento humano en el desempeño de sus actividades ocupacionales, lo hemos decidido mediante una evaluación concienzuda y sistematizada, valiéndonos de los instrumentos estadísticos más prácticos para la administración efectiva del personal. Estimándose que son escasísimas y contadas las empresas que en la actualidad, y particularmente el departamento de relaciones industriales hacen uso de éstos medios estadísticos.

De conformidad con lo antes mencionado, la gerencia de hoy en día depende vitalmente del uso de datos estadísticos para el control de las complejas actividades desempeñadas por su personal.

La importancia de éstos aspectos para la administración efectiva y especializada, es decisiva para la marcha organizativa, planificadora y estructural de las empresas, con miras a coadyuvar en la coordinación de sus propios elementos.

Como ya se ha señalado, las empresas mexicanas actuales no están utilizando debidamente éstas herramientas, y es por eso que se propone demostrar en el presente seminario de investigación la utilidad y conveniencia de las mismas, para acelerar el progreso tecnológico y encauzar las relaciones de la parte directriz con el personal y el adiestramiento del mismo para éstos propósitos.

Las técnicas estadísticas que aquí se abordan se muestran objetivamente, y se pueden utilizar fructíferamente, para determinar la rotación del personal, los despidos, abandonos, permisos, calificación de méritos, medición de la moral, evaluación de sueldos y salarios, en la selección, y en la capacitación, etc.

En ésta investigación, denominada "Inferencia Estadística en la Administración de Personal" se hace un estudio completo y se evalúan prácticamente las medidas de tendencia central y de dispersión, se hace un análisis de correlación y predicción.

Concluyentemente, la mira fundamental del presente estudio, ha sido la de ofrecer una perspectiva sobre la inferencia estadística en la administración de personal, en el que se combinen verdades fundamentales con realidades prácticas, tendientes al logro de beneficios empresariales en todos los órdenes.

Fernando Rosales Jiménez.

Mario Serrano Capaceta.

CAPITULO I.

MEDIDAS DE TENDENCIA CENTRAL O PROMEDIOS.

COMENTARIOS FUNDAMENTALES.—

Indudablemente que en ocasiones la estadística ha sido llamada la "ciencia de los promedios". Los promedios pueden ser útiles y proporcionar información, pero éstos también pueden ser dudosos Los promedios no deben darse a menos que sean representativos del conjunto de datos de los cuales fueron obtenidos.

Los promedios han sido definidos como una medida de tendencia central poraue ellos han sido designados para medir el valor central alrededor del cual un conjunto de datos tiende a agruparse.

El cálculo de un tipo simple de promedio frecuentemente, resulta inadecuado para describir la manera como un conjunto de datos tiende a concentrarse.

Consecuentemente varios tipos de promedio han sido desarrollados.

Los tres tipos de promedio más usados para nuestros propósitos en éste Seminario de Investiqación, son los siguientes:

a) LA MEDIA ARITMETICA.
b) LA MEDIANA.
c) EL MODO.

La media aritmética es el promedio ordinario. La mediana es simplemente el valor del término central en una serie de valores. El Modo es el valor poseído por el mayor número de observaciones en la serie.

A continuación presentamos un ejemplo sencillo de éstos promedios:

TERMINOS	SERIE (a)	SERIE (b)
1	3	3
2	5	5
3	5	6
4	5	7
5	6	7
6	6	7
7	9	8
8	11	9
9	13	11
TOTAL...........	63	63

La media aritmética de la serie (a) se calcula dividiendo el total (63) entre el número de términos (9), de lo que obtenemos un resultado 7.

La mediana, el valor del término central es 6.

El modo, el valor ue ocurre más frecuentemente, es 5.

Contrastando, la media aritmética, la mediana y el modo para la serie (b) son el mismo valor 7.

Posteriormente daremos una explicación amplia de los mismos.

Por lo tanto, antes de discutir cómo calcular la media, la mediana y el modo para series con muchos términos es necesario examinar el manejo de grandes cantidades de cifras.

DISTRIBUCION DE FRECUENCIAS.

Grandes cantidades de datos son virtualmente incomprensibles cuando se presentan sin ningún ordenamiento.

Antes de interpretarlos debe ser hecha una gráfica o determinar uno o dos números que puedan considerarse representativos del todo.

Entonces es necesaria alguna vía de clasificación o arreglo de los datos para hacer más conveniente la tarea de interpretarlos como grupo.

Una posibilidad obvia, puede ser ordenar los datos en órden de magnitud, del más bajo al más alto, en forma de serie.

LISTA PRIMITIVA	SERIE DE DATOS
21	16
22	17
17	18
20	19
21	20
27	21
16	21
24	21
22	22
23	22
19	22
22	23
25	24
21	25
18	27
ETC.	ETC.

Con tal ordenación, resulta mucho más fácil distinguir los bajos o altos valores, además es un paso necesario para el cálculo de la mediana.

Un procedimiento mejor puede ser: ordenar por magnitudes todos los diferentes valores y después indicar para cada valor el número de veces que se repite, como se muestra a continuación:

ORDENACION SIMPLE POR FRECUENCIAS

TERMINOS	FRECUENCIAS	TERMINOS	FRECUENCIAS
16	1	23	1
17	1	24	1
18	1	25	1
19	1	27	1
20	1		
21	3		
22	3		

Se hace evidente de inmediato que esta forma de ordenamiento facilita marcadamente la interpretación. Podemos condensar más aún nuestros datos por el establecimiento de clases dentro de la distribución de frecuencias.

La clase o intervalo tiene límites, superior e inferior y el valor de la frecuencia indica el número de términos contenidos en la clase o intervalo correspondiente.

INTERVALO	FRECUENCIA
16——18	3
19——21	5
22——24	5
25——27	2

16 —— 18 es la clase o intervalo
16 es el límite inferior
18 es el límite superior.

Es importante que una distribución de frecuencias sea hecha con el número de clases apropiado. Si se usan pocas clases los datos primitivos se comprimirán de tal manera que la información que aportan será escasa.

También, si son usadas muchas clases, algunas de ellas contendrán tan pocos términos que la distribución de frecuencias será irregular. Una guía para determinar el número de clases adecuada la proporciona la tabla siguiente:

NUMERO DE TERMINOS	NUMERO DE CLASES
5	2
10	4
25	6
50	8
100	10
200	12
500	15
1000	15

Como veremos más adelante, el punto medio de cada clase, es generalmente usado para representar la clase. El punto medio de una clase, lo determina el promedio del límite inferior y superior.

Por lo cual debemos seleccionar los límites de clase, de tal manera que el punto medio coincida tanto como sea posible con la concentración de términos que representa.

CLASE	PUNTO MEDIO	FRECUENCIA
96 — 100	98	3
101 — 105	103	8
106 — 110	108	11
111 — 115	113	10
116 — 120	118	3

MEDIA ARITMETICA

La media aritmética de una serie de valores o números es igual a la suma de éstos dividida entre el número de ellos. Esta es, el promedio, al que el uso popular se refiere más frecuentemente.

	SERIE X
	3
	4
	5
	6
	7
TOTAL	25

la expresión algebráica es $\overline{X} = \frac{\Sigma X}{N}$ ó media aritmética

$$(\overline{X}) = \frac{\Sigma \text{ (SUMA) de X (los térmnos 3, 4, 5, 6, 7)}}{\text{N (Número de términos en la serie)}} = \text{dividido entre}$$

ó $\frac{25}{5} = 5$

Para la distribución de frecuencia simple, calculamos la media de la manera siguiente:

VALOR	FRECUENCIA	
X	f	fX
10	2	20
11	1	11
12	4	48
13	10	130
14	13	182
15	17	255
16	8	128
17	3	51
18	1	18
19	1	19
	N = 60	ΣfX = 862

N = número de términos o frecuencias.

ΣfX = valor de la suma de los productos de las frecuencias por sus valores.

Aplicando la Fórmula

$$\overline{X} = \frac{\Sigma fX}{N} \text{ ó media } = \frac{\text{Suma de los productos de las frecuencias por los valores}}{\text{número de frecuencias}}$$

$$\frac{862}{60} = 14.37 \text{ (media aritmética)}$$

Para una distribución de frecuencias con clases, calculamos la media de la manera indicada anteriormente, si usamos el punto medio para representar los valores contenidos por la clase.

CLASE	FRECUENCIA f	PUNTO MEDIO X	fX
5-24	4	14.5	58.0
25-44	6	34.5	207.0
45-64	14	54.5	763.0
65-84	22	74.5	1,639.0
85-104	14	94.5	1,323.0
105-124	5	114.5	572.5
125-144	7	134.5	941.5
145-164	3	154.5	463.5
TOTAL	75	....	5,967.5

Usando la misma fórmula $\overline{X} = \frac{\Sigma fX}{N} = \frac{5,967.5}{75} = 79.6$

La media aritmética es: 79.6

En adición a los métodos anteriores, hay un método corto para encontrar la media. La clave para esta simplificación es la siguiente propiedad de la media: la suma (algebráica) de la desviación del valor de cada término respecto del valor de la media es siempre igual a cero. Este puede ser demostrado en la forma siguiente:

SERIE	DESVIACION DE LA MEDIA
3	— 2
4	— 1
5	0
6	+ 1
7	+ 2
TOTAL = 25	TOTAL = 0

$$\text{Media} = \overline{X} = \frac{\Sigma X}{N} = 5$$

El conocimiento de ésta característica nos permite calcular la media como sigue:

1. Estimar o tratar de adivinar la media.
2. Computar las desviaciones de los términos de la media estimada.
3. Computar la media de las desviaciones. Esta llega a ser un factor de corrección.

4. Corregir la media estimada paso (1) sumándole algebráicamente la media de las desviaciones calculadas con el paso 3.

Supongamos, como ejemplo, que estimamos la media de la serie anterior en 4. Entonces:

SERIE	DESVIACION DE LOS TERMINOS RESPECTO DE LA MEDIA ESTIMADA
3	— 1
4	0
5	+ 1
6	+ 2
7	+ 3
	TOTAL = 5 (suma algebráica)

$$\text{MEDIA DE LAS DESVIACIONES} = \frac{5}{5} = 1 \text{ (factor de corrección)}$$

MEDIA ESTIMADA + FACTOR DE CORRECCION = MEDIA
4 + 1 = 5

Ahora apliquemos esta técnica a una distribución de frecuencias, donde los puntos medios de las clases son usadas para representar los valores de las clases respectivas.

1. Seleccionar cualquier punto medio de clase como media estimada; sin embargo, los cálculos se simplifican si el punto medio escogido es de la clase con mayor número de frecuencias.
2. Computar las desviaciones de los puntos medios de clase respecto de la media estimada (nótese que estas desviaviaciones son en términos de clases).
3. Multiplicar las desviaciones por sus respectivas frecuencias en la clase.
4. Sumar algebráicamente las desviaciones ponderadas en el paso 3; dividir esta suma entre N (número de frecuencias) para obtener la media de las desviaciones ponderadas.
5. Convertir la media (paso 4) a términos de unidades multiplicando ésta por la amplitud de la clase (límite superior a límite inferior). Este resultado llega a ser el factor de corrección.

6. Corregir la media estimada (paso 1) por la adición (algebráicamente) —es decir— respetando los signos negativos o positivos) de la media de las desviaciones en términos de unidades (paso 5).

Usaremos los siguientes símbolos:

$\overline{X}d$ = media estimada.
D = desviación de los puntos medios de clase respecto de la media estimada.
i = amplitud de clase (intervalo).
La media estimada Xd para este cuadro es 74.5 el cual es el punto medio de la clase 65-84.

CLASE (int. 20)	FRECUENCIAS f	PUNTO MEDIO X	DESVIACION DE CLASE D	FRECUENCIA por DESVIACIONES fD	
5- 24	4	14.5	— 3	— 12	
25- 44	6	34.5	— 2	— 12	
45- 64	14	54.5	— 1	— 14	
					— 33
65- 84	22	74.5	0		
85-104	14	94.5	+ 1	+ 14	
105-124	5	114.5	+ 2	+ 10	
125-144	7	134.5	+ 3	+ 21	
145-164	3	154.5	+ 4	+ 12	
					— 57
TOTAL	75	....	...		+ 19

$$\text{Media} = \text{media estimada} + \left(\frac{\text{Suma del producto de las frecuencias multiplicadas por las desviaciones}}{\text{Número de frecuencias}}\right) i$$

$$\overline{X} = \overline{X}d + \left(\frac{\Sigma\ fd}{N}\right) i = 74.5 + \left(\frac{19}{75}\right)\ 20$$

$$74.5 + 5.07 = 79.6 \text{ (media aritmética)}$$

$$\overline{X}$$

RENOVACION DE LA MANO DE OBRA

Obviamente a continuación se muestra un ejemplo sencillo sobre el empleo de los promedios.

La renovación de la mano de obra es la relación de los salidos de la empresa o de sus sustituciones al número total de empleados. El método más sencillo para calcularla, exacto hasta cierto punto, consiste en dividir el número total de salidos durante el período en cuestión por el número medio de empleados en la nómina durante el mismo. Así, si en el transcurso de un año abandonan el trabajo 1,000 operarios, en un personal cuyo número medio asciende a 2,000, la renovación de la mano de obra es:

$$\text{R (renovación)} = \frac{\text{S (salidos)}}{\text{P (promedio del personal)}}$$

$$R = \frac{1{,}000}{2{,}000} \text{ o sea, } 0{,}5 \text{ o el } 50\%$$

Muchas compañías, al calcular la renovación de la mano de obra, la dividen en varios grupos, según las causas, como despidos, abandono voluntario, licencia, etc., con un análisis bastante completo de las razones para el abandono voluntario. Algunas compañías establecen una distinción entre la renovación evitable y la inevitable, poniendo en esta última clase las salidas debidas a la reducción del personal o a la muerte de los trabajadores. Análogamente, algunas compañías registran por separado la renovación de los empleados antiguos y los nuevos, de los hombres y las mujeres, de los trabajadores calificados y los inexpertos, y para otros factores de diferencia.

Si la renovación es elevada debido al gran número de empleados que abandonan voluntariamente la empresa, deben hallarse las causas y corregirlas. La poca paga, las malas condiciones de trabajo, la inseguridad en el empleo, la mala reputación de la compañía en la comunidad, la existencia de dirigentes de maneras dominantes, el tratamiento injusto, la imposibilidad de ascender, la retención sobre tareas pesadas y desagradables, las disputas sobre la paga o las ganancias, la posibilidad de encontrar mejores labores en otras empresas y la posibilidad de trabajar en un turno de día en lugar de por la noche, con solo algunas de las muchas causas de abandono voluntario del trabajo. Es conveniente celebrar una entrevista amistosa de salida con los operarios que se despiden voluntariamente. De esta manera es posible, a veces, convencer algunos empleados que sea conveniente retener para que sigan en la casa corrigiendo las causas de desagrado y haciéndoles ver que a la larga resultarán beneficiados continuando en la empresa. Deben investigarse todas aquellas causas que puedan imputarse a defectos de la compañía, a los inspectores o a otros factores controlables o introducir las medidas correctivas necesarias para eliminar las quejas. La pérdida de la moral es una de las causas principales y puede deberse a circunstancias o acontecimientos muy variados. Si no se corrigen, se irán acentuando cada vez más y a la larga producirán trastornos más graves en la empresa.

MEDIANA

En series de datos que incluyen valores muy altos o muy bajos, el valor de la media puede no ser representativo de los valores que se presentan más frecuentemente. Esto es a menudo cierto para series de ingresos y precios, donde pocos valores relativamente altos causan que la media tiendan a ser más alta que los valores comunes. En tales circunstancias, la mediana, o valor central, de la serie puede ser una cifra representativa para describir el conjunto de los datos.

La mediana puede ser simplemente definida como el valor del término central en una serie en la que todos los términos han sido ordenados de acuerdo con su magnitud. Puesto que la mediana es a menudo calculada para una distribución de frecuencias, la mejor definición en general sería: la mediana es aquel punto, bajo y sobre el cual hay un número igual de términos. La abreviatura usual para la mediana es Md.

En una serie simple:

22
24
25
29
30
32
34
48

$$\text{La Mediana es} = \frac{29 + 30}{2} = 29.5$$

Por tanto vemos que la mediana es determinada por su posición, más que por el valor de los términos en la serie. El cambio de cualquiera de las tres primeras o de las tres últimas cifras, no afecta al valor de la mediana. Eso contrasta con la media, ya que cualquier cambio en el valor de los términos afecta el valor de ésta.

En seguida damos un ejemplo de cálculo de la mediana para una distribución de frecuencias.

CLASE (intervalos 5)		FRECUENCIA f
12.5 — 17.4		9
17.5 — 22.4		14
22.5 — 27.4		16
27.5 — 32.4		13
33.5 — 37.4		8
37.5 — 42.4		6
42.5 — 47.4		4
	TOTAL	70

La fórmula usada es: $Md = Lm + \left(\frac{N/2 - F}{fm}\right) i$

Donde:

Md = mediana
Lm = límite inferior de la clase en que cae la mediana
N = número de frecuencias
F = número de frecuencias en todas las clases bajo Lm
Fm = número de frecuencias en la clase en que cae la mediana
i = intervalo (amplitud entre los límites superior e inferior)

Primero podemos determinar, por medio de una inspección en la columna de frecuencias, que el punto medio de las frecuencias N/2 ó 70/2 = 35) cae dentro de la clase con límites

22.5 y 27.4

Una vez hecho esto, podemos proceder a emplear la fórmula.

lm = 22.5
N = 70
F = 9 + 14 = 23
Fm = 16
i = 5

$$22.5 + \frac{70/2 - 23}{16} \times 5 =$$

$$22.5 + \frac{35 \quad 23}{16} \times 5 =$$

$$22.5 + \frac{12}{16} \times 5 =$$

$$22.5 + \frac{60}{16} =$$

$$22.5 + 3.80 = 26.3$$

$$Md = 26.3$$

ESTUDIOS DE ACTITUDES

Igualmente con el objeto de dar un poco más de claridad a este aspecto tan importante, se comente un estudio con relación a la mediana enfocado a las actitudes.

Puede obtenerse una medida o valuación de la moral por medio de los estudios de actitudes y que han llevado a cabo muchas compañías en su esfuerzo por averiguar lo que piensan los trabajadores y los dirigentes de la organización para lo cual trabajan. En un estudio de esta clase realizado en 30 compañías por Arthur Kolstad (Actitudes of Employes and Their Supervisors, Personnel, vol. 20) se halló que la buena moral entre los inspectores no significa por necesidad que exista también entre los trabajadores, y que no en todos los casos son más favorables las actitudes de los inspectores que la de los trabajadores. Los cuestionarios sometidos eran bastante parecidos para las 30 compañías, pues tenían cierta cantidad de detalles comunes, y aunque se emplearon modelos distintos para los inspectores y para los trabajadores, muchas de las consultas eran idénticas. Las contestaciones se escribieron sobre las hojas en reuniones de los dos grupos respectivos de cada compañía, pero sin que estuviera presente ningún ejecutivo. Las respuestas eran secretas y no iban firmadas, ni era posible identificarlas, y las depositaron los empleados en urnas especiales.

Damos a continuación algunos resultados típicos de las puntuaciones de moral en las 30 compañías, que muestran cómo los individuos calificaban a éstas:

Empleados ordinarios	Inferior 58	Superior 80	Promedio 68.6%
Personal directivo	Inferior 63	Superior 87	Promedio 75.6%

El 34 por ciento de todos los empleados ordinarios tenían puntuaciones de moral superiores al promedio de los inspectores.

El 29 por ciento de los inspectores tenían puntuaciones de moral inferiores a las del promedio de los empleados ordinarios.

Considerando organización por organización, la mayor diferencia positiva entre los dos grupos de calificaciones fué de 14 puntos (en 2 compañías, en las que las calificaciones fueron respectivamente; promedio para los inspectores 74 y empleados ordinarios 70).

La mayor diferencia negativa fué de 6 puntos (inspectores 66, empleados ordinarios 72).

La concordancia entre las puntuaciones de empleados inspectores y empleados ordinarios (coeficiente Pearson) fué +0,70 + 0,06.

Hubo una concordancia elevada en la pregunta:

En su opinión ¿existen otras compañías que traten mejor a sus empleados que ésta?

( ) Todas las demás son mejores.
( ) La mayor parte de las otras son mejores.
( ) Sobre poco más o menos lo mismo.
( ) Esta compañía es mejor que la mayoría de las otras.
( ) Esta compañía es la mejor de todas.

Las proporciones de los inspectores que llenaron este estado fueron calificadas entre 48 por ciento y 96 por ciento; las de los empleados ordinarios entre el 29 por ciento y el 94 por ciento.

En lo que respecta el punto Cuando se presentan dificultades en mi trabajo me siento en completa libertad para hacer preguntas, la proporción de las respuestas afirmativas fué para los inspectores 45-96% y para los trabajadores 53-87%.

MODO

Una tercera medida más frecuentemente usada de tendencia central es el modo, el cual es definido como el valor que en una serie se presenta con mayor frecuencia. Este valor puede ser determinado directamente en una serie de datos no agrupados (si un valor aparece más a menudo que cualquier otro), o podemos calcularlo aproximadamente de una distribución de frecuencias de los datos. En la serie de datos no agrupados siguiente: El modo es 12

5
8
10
12 12
13
14

porque éste aparece más frecuentemente que cualuier otro número.

Una aproximación de este valor puede ser obtenido de una distribución de frecuencias si conocemos la media y la mediana, con la siguiente fórmula:

Modo = 3 mediana — 2 media

o Modo = 3 veces la mediana menos 2 veces la media.

Otro método usado es la siguiente fórmula:

$$L\ mo + \frac{d\ 1}{d + d2} \times i$$

Donde:

L mo = límite inferior de la clase en que el modo cae

d 1 = diferencia entre las frecuencias de la clase del modo y la clase inferior

d 2 = diferencia entre las frecuencias de la clase del modo y la clase superior

i = intervalor (amplitud entre límites)

Supongamos que tenemos la siguiente distribución:

CLASE	FRECUENCIA
1.50 — 1.99	3
2.00 — 2.49	5
2.50 — 2.99	7
3.00 — 3.49	4
3.50 — 3.99	2

Entonces:

$$2.50 + \frac{2}{2 + 3} \times .50 =$$

$$2.50 + \frac{2}{5} \times \frac{.50}{1} =$$

$$2.50 + \frac{1.00}{5} =$$

$$2.50 + .20 = 2.70 \text{ el modo.}$$

CUANDO USAR LA MEDIA, LA MEDIANA Y EL MODO

¿Por qué es necesario tener tres medidas diferentes para indicar la tendencia central de un grupo de datos? La respuesta es que cada una se adapta mejor a ciertos usos, esto es, en algunos casos una puede ser la más representativa de un conjunto, mientras en otros casos la medida puede ser más apropiada. Ordinariamente la media es usada cuando la distribución es aproximadamente normal. En cambio si tienen preponderancia valores altos o bajos, la media puede dar una impresión incorrecta de la tendencia central de los datos. En tales circunstancias el modo o la mediana son más adecuados. Consideremos por ejemplo los siguientes ingresos anuales de cinco personas:

$ 800.00 $ 900.00 $ 850.00 $ 750.0 $ 5000.00

La media para estos cinco ingresos es:

$$\frac{\$\,800.00 + \$\,900.00 + \$\,850.00 + \$\,750.00 + \$\,5000.00}{5} =$$

$ 1,660.00

Esta cifra, aunque ha sido bien calculada, no es típica del grupo como un todo, porque ha sido *marcadamente afectada por un ingreso* de $ 5,000.00, el cual es *considerablemente mayor que los otros cuatro. La mediana de los ingresos es $ 850.00 y este valor* es más típico del grupo que el ingreso medio de $ 1.660.00. *Si se* dispusiese de un gran número de datos para el cálculo, sería interesante conocer el modo, o ingreso más común.

CAPITULO II.

MEDIDAS DE DISPERSION O DESVIACION.

ASPECTOS BASICOS.—

Desde luego, es obvio que una medida de tendencia central, por sí sola no puede describirse más de una de las características importantes de un conjunto de datos. Es esencial conocer la extensión de la variación entre cada uno de los términos y el punto de tendencia central. Al describir por ejemplo, la distribución de salarios semanales para una industria dada, no es suficiente conocer solamente el promedio de salarios. Para propósito de análisis es igualmente si no más importante, el conocer la magnitud de las diferencias individuales en salarios semanales o cuán heterogéneos son éstos. En otras palabras, nos gustaría saber si en la escala casi todos los obreros ganan el promedio, o están cerca de él, o si ésta contiene una gran proporción de altos o bajos ingresos. Esta condición en una distribución de frecuencias (conjunto de datos), se ha llamado: dispersión, amplitud, desviación y variación. Entre las varias vías para describir esta característica cuantitativamente, tenemos las percentiles y desviaciones.

PERCENTILES

Si en una distribución dada eliminamos el 25% inferior y el 25% superior, tenemos el 50% medio de la distribución al cual llamamos amplitud inter-cuartila. Esta es la parte de la distribución que generalmente es más significativa. La amplitud de la semi-cuartila o desviación cuartila (DC), la cual es la mitad de la distancia entre la cuartila superior de la inferior, es una de las medidas de variación más frecuentemente usadas, especialmente para propósitos de pruebas.

El DC es útil cuando los valores de los extremos de una distribución tienden a ser irregulares. La mediana, la cual separa el 50% bajo de los valores del 50% alto, es igual a la segunda cuartila. La primera y la tercera cuartila pueden ser calculadas usando la fórmula para la mediana de una distribución de frecuencias con pequeñas modificaciones.

El único cambio requerido es en la fracción $\frac{N}{2}$ donde N es el número total de frecuencia en la distribución. Puesto que el propósito de $\frac{N}{2}$ fué únicamente para dividir la distribución en dos grupos iguales de valores, $\frac{N}{4}$ separa el cuarto inferior de los tres cuartos superiores de valores, y $\frac{3N}{4}$ separará el cuarto superior de los tres cuartos inferiores de valores.

Así que las fórmulas para C_1 y C_3 son:

$$C_1 = L_1 + \frac{N/4 - F}{f_1} \times i$$

$$C_3 = L_3 + \frac{3N/4 - F}{f_3} \times i$$

L = límite inferior de la clase donde cae la cuartila.
f = número de frecuencias en la clase donde la cuartila cae.
N = número total de frecuencias.
F = suma de las frecuencias abajo de la clase que contiene la cuartila.

Usando la misma distribución de frecuencias que utilizamos para la mediana, vamos a hacer lo siguiente:
Primero calculamos N/4 lo cual es 70/4 = 17.5, por inspección notamos que la primera clase (12.5-17.4) tiene 9 frecuencias y la segunda 14; lo cual da un total de 23 para las dos primeras clases. Consecuentemente, el valor de C_1 (bajo el cual hay 17.4 frecuencias) debe estar en la segunda clase (17.5-22.4). Ahora podemos usar las fórmulas:

$$C_1 = L_1 + \frac{N/4 - F}{f_1} \times 1 = 17.5 + \frac{70/4 - 9}{14} \times 5$$

$$= 17.5 + \frac{17.5 - 9}{14} \times 5 = 17.5 + \frac{8.5}{14} + 5 = 20.5$$

Similarmente C_3

$$C_3 = L_3 + \frac{3N/4 - F}{f_3} \times i = 32.5 + \frac{3(70)/4 - 52}{8} \times 5$$

$$= 32.5 + \frac{52.5 = 52}{8} \times 5 = 32 + \frac{0.5}{8} \times 5 = 32.8$$

Entonces la amplitud de la inter-cuartila es 32.8 — 20.5 = 12.3 Este intervalo incluye el 50% de la distribución. Ocasionalmente puede ser deseable dividir una distribución en más de cuatro segmentos; por ejemplo una distribución puede ser dividida en decilas. La fórmula necesaria para el cálculo de la decila inferior es:

$$D_1 = L_1 + \frac{N/10 - F}{f_1} \times i$$

La amplitud de la interdecila incluye el 80% de los valores; es decir excluye el 10% más alto y el 10% más bajo de la distribución. En contraste con la amplitud de la inter-cuartila, la cual contiene sólo el 50% de los valores.

PERCENTILES

En programas de pruebas encontramos, que una puntuación no ponderada tiene poca significación a menos que ésta sea interpretada en términos de su locación en una distribución de otras puntuaciones hechas por otras personas. Si una prueba consiste en 75 cuestiones muy fáciles, una calificación de 65 podría estar cerca del extremo inferior de la distribución, por tanto debe considerarse como una calificación baja.

En el otro extremo, si una prueba consiste en 75 cuestiones muy difíciles, una caificación de 65 podría estar en el punto superior o muy cerca de la distribución y entonces debe considerarse una calificación alta. En otras palabras, una calificación no ponderada puede ser alta o baja, dependiendo de la distribución de puntuaciones, a la cual pertenezca.

Un método conveniente y ampliamente empleado para la interpretación de una calificación no ponderada, es el uso de rangos de percentiles.

Una percentila puede ser definida como el número que indica el porcentaje del grupo total igual o inferior a la puntuación en cuestión.

Por ejemplo, si en una prueba el 65% del grupo obtiene 129 o menos, la puntuación de 129 estaría en la percentila 65, o tendría el rango de percentila 65. Se hace notar que la percentila 50 es igual a la mediana. Los manuales que acompañan a las pruebas normalizadas incluyen usualmente tablas de percentiles que hacen posible la conversión de las calificaciones en rangos de percentilas.

PROMEDIO DE DESVIACION (PD)

La variación de los valores en una distribución depende llanamente de la cantidad en que cada término se desvía de la medida de tendencia central. Para describir la variabilidad de una distribución de frecuencias podemos determinar la extensión en que cada término se desvía de la media de la serie, considerando todas éstas diferencias (desviaciones) como positivas, podemos calcular la media de éstas desviaciones. La media de las desviaciones es llamada la desviación promedio. Esta medida es relativamente fácil de interpretar. Por ejemplo, al decir que la desviación media en peso para un grupo de personas es 5 kilos, queremos decir que en promedio los individuos en el grupo difieren en peso del individuo medio del grupo en 5 kilos.

El método del cálculo es simple. Consideramos una serie de cinco términos con valores de 2, 4, 6, 6 y 7. Las desviaciones de la media 5 son las siguientes:

2	3
4	1
6	1
6	1
7	2

$$8/5 = 1.6$$

La desviación promedio de la media es igual a 1.6 (PD).

La fórmula básica para el P.D. es: $\text{P.D.} = \dfrac{\Sigma\ |x|}{N}$

en la cual $|x|$ representa la desviación de cualquier término de la media ignorando los signos — es decir, + más o — menos.

Para calcular el P.D. de una distribución de frecuencia hacemos un pequeño cambio en la fórmula:

$$\text{P. D.} = \frac{\Sigma\ f\ |x|}{N}$$

Por ejemplo:

Clase	Punto Medio (X)	Frecuencia (f)	(fX)	Desviaciones de la Media $\lvert x \rvert$	$f \lvert x \rvert$
48 — 52	50	2	100	— 21.2	42.4
53 — 57	55	3	165	— 16.2	48.6
58 — 62	60	5	300	— 11.2	56.0
63 — 67	65	9	585	— 6.2	55.8
68 — 72	70	10	700	— 1.2	12.0
73 — 77	75	12	900	3.8	45.6
78 — 82	80	7	560	8.8	61.6
83 — 87	85	2	170	13.8	27.6
88 — 92	90	3	270	18.8	56.4
93 — 97	95	1	95	23.8	23.8
		N=54	3,845	....	$f \lvert x \rvert = 429.8$

$$\text{Media} = \frac{\Sigma fX}{N} = \frac{3,845}{54} = 71.2 \qquad \text{P.D.} = \frac{\Sigma f \lvert x \rvert}{N} = \frac{429.8}{54} = 7.96$$

Una característica importante de la desviación promedio es que aproximadamente el 57.5% de los términos en una distribución cae dentro de la extensión de la media ± la desviación promedio.

Esta característica es pertinente para distribuciones que tienen una mayor concentración de datos cerca del centro de su extensión; ésto no necesariamente para distribuciones que tienen una mayor concentración en los extremos.

Aunque para muchos propósitos en los negocios la desviación promedio es suficiente, muy a menudo la desviación estandar es más útil.

DESVIACION ESTANDAR (D. E.)

La desviación estandar es la más importante de las medidas de dispersión o variabilidad. Usualmente al dar a conocer el valor de la media para un conjunto se acompaña de la D.E., para indicar la variabilidad de los datos.

La desviación estandar de una distribución puede definirse como la raíz cuadrada de la media de las desviaciones cuadradas de la media de la distribución. Esta puede determinarse enconotrando la diferencia entre cada término y la media de la distribución elevando al cuadrado éstas diferencias individuales, sumando la desviaciones cuadradas y dividiendo la suma entre N, y entonces extrayendo la raíz cuadrada del resultado. La fórmula fundamental para la desviación estándar (σ) es:

$$\sigma = \sqrt{\frac{\Sigma X^2}{N}}$$

El método del cálculo consiste en los siguientes pasos fundamentales:

1. Computar las desviaciones de los términos respecto a la media.
2. Elevar al cuadrado las desviaciones encontradas en el paso 1.
3. Sumar los cuadrados de las desviaciones.
4. Calcular la media de los cuadrados de las desviaciones.
5. Calcular la raíz cuadrada de la media de los cuadrados de las desviaciones.

El siguiente problema simplificado, ilustra como la desviación estandar es calculada directamente de una serie simple de valores.

SERIE X	DESVIACION $(X - \overline{X} = x)$	x^2
2	— 6	36
10	+ 2	4
12	+ 4	16
4	— 4	16
14	+ 6	36
8	0	0
6	— 2	4
Media = 56/7 =	0	$\Sigma X^2 \cdot 112$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{56}{7} = 8$$

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}}$$

$$\sigma = \sqrt{\frac{112}{7}} = \sqrt{16} = 4$$

$\overline{X}$ = media
N = números de técnicos
X = términos en la serie
Los cuadrados y raíces cuadradas se encuentran fácilmente en las tablas de raíces cuadradas.
Además se puede calcular con regla de cálculo, máquina calculadora o por el método simple pero largo, el cual puede encontrarse en cualquier texto de álgebra o estadística.
Una fácil manera de calcular la desviación estándar (0) es:

X	X^2
2	4
10	100
12	144
4	16
14	196
8	64
6	36
56	560

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - \frac{(X)^2}{N}}$$

$$= \sqrt{\frac{560}{7} - \frac{(56)^2}{7}}$$

$$= \sqrt{80 - 64} = \sqrt{16} = 4$$

Podemos encontrar X^2 directamente en una tabla de cuadrados, sumando directamente en una calculadora y substituímos la suma en la fórmula.
Cuando disponemos de una calculadora los cuadrados pueden suministrarse directamente en la máquina el cálculo de la desviación estándar para una distribución de frecuencias, puede verse en el siguiente ejemplo:

Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6
Clase	Puntos Medios	Frecuencias	Desviación de Clase	Col. 3 X Col. 4	Col. 4 X Col. 5
	X	f	d	fd	fd^2
48 — 52	50	2	—5	—10	50
53 — 57	55	3	—4	—12	48
58 — 62	60	5	—3	—15	45
63 — 67	65	9	—2	—18	36
68 — 72	70	10	—1	—10	10
73 — 77	75	12	0	0	0
78 — 82	80	7	1	7	7
83 — 87	85	2	2	4	8
88 — 92	90	3	3	9	27
93 — 97	95	1	4	4	16
		N = 54	. . .	Σ fd = —41	$\Sigma fd^2 = 247$

i = intervalo
N = número de frecuencias
d = desviación de la media
media asumida es 75
(Nota: cualquier punto medio puede usarse).

$$\sigma = i \sqrt{\frac{\Sigma fd^2}{N} - \frac{(\Sigma fd)^2}{N}}$$

$$= 5 \sqrt{\frac{247}{54} - \frac{(41)}{54}}$$

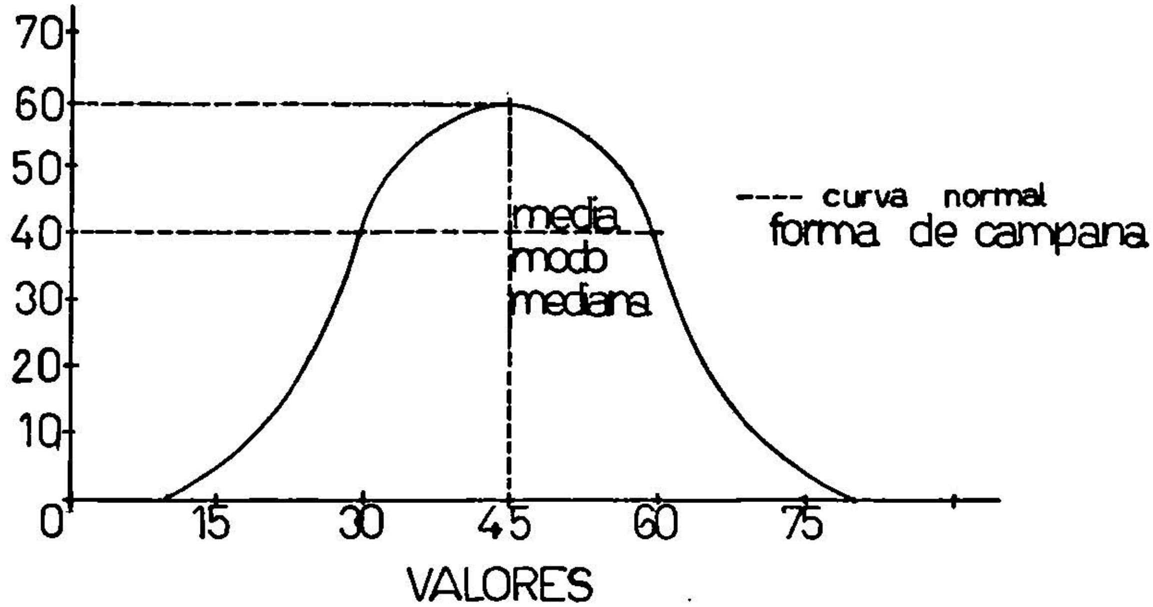
$$\sigma = 5 \sqrt{4.574 - .576} = 9.95$$

DISTRIBUCION NORMAL

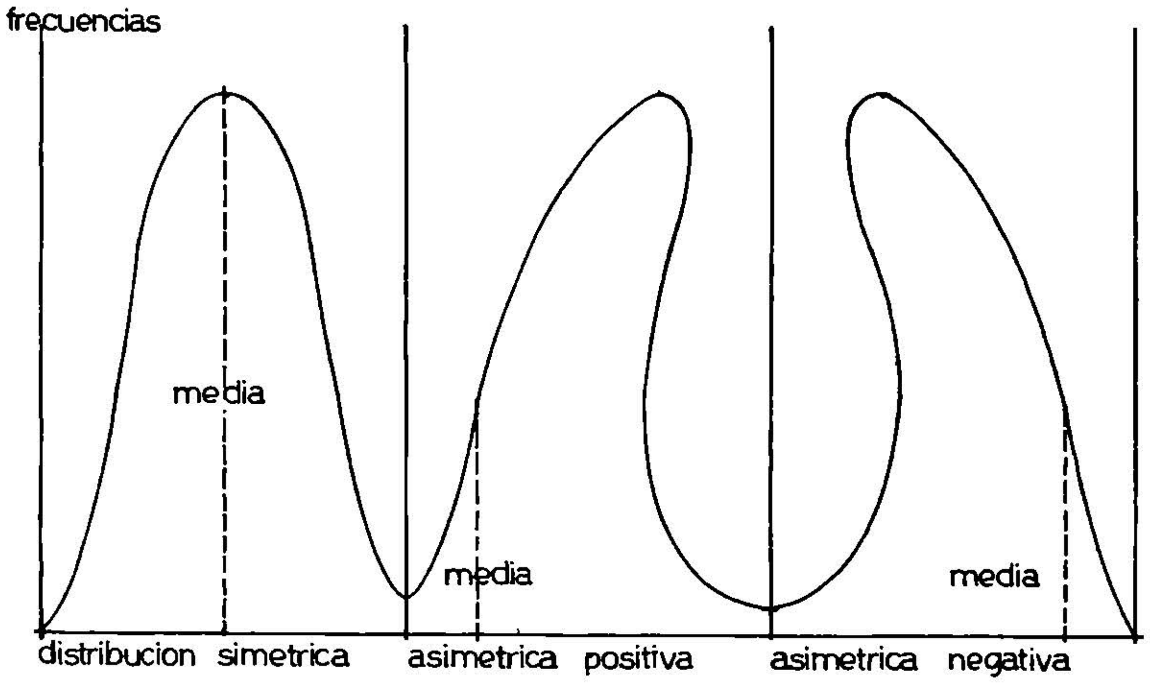
La desviación estándar es más útil y confiable cuando se calcula para términos que están normalmente distribuídos como se ilustra en las gráficas siguientes:

En tales distribuciones la amplitud de la media más o menos la desviación estándar $(M \pm 0)$ incluye el 68.27 por ciento de los términos; $(M \pm 2\ 0)$ incluye el 95.45 por ciento de los términos; $(M \pm 3\ 0)$ incluye el 99.73 por ciento de los términos en la distribución.

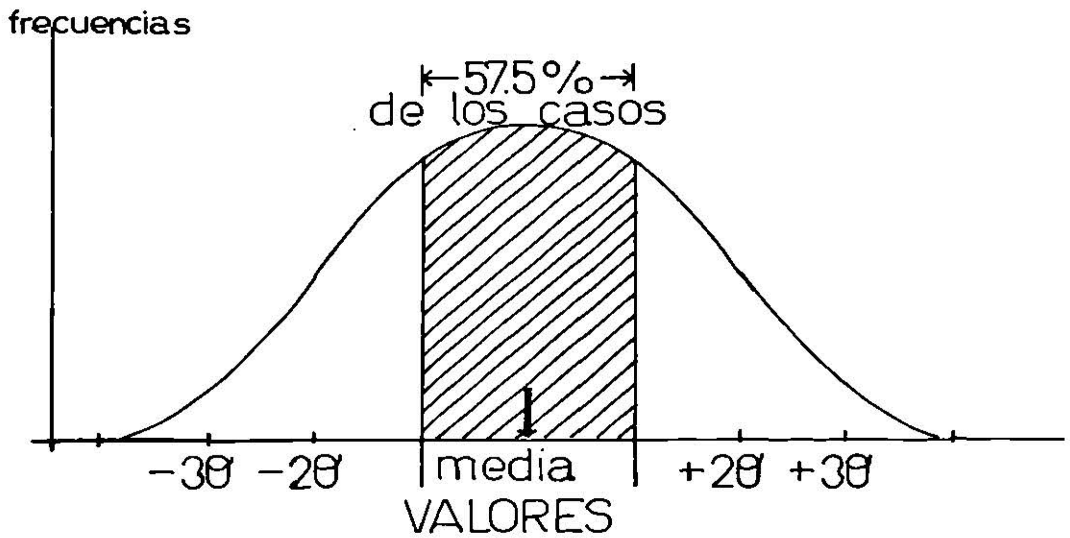
Una distribución normal de frecuencias:



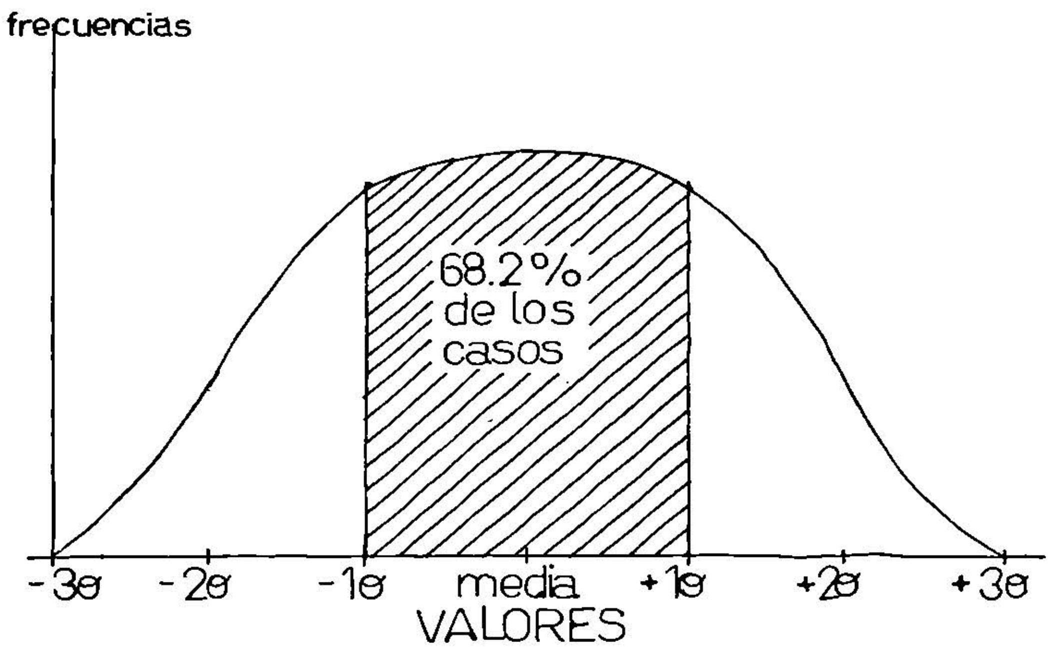
Cuando tenemos una serie de datos en la cual la mayoría de los términos se encuentran alrededor de la media, y en la que los valores extremos disminuyen gradualmente, tendemos a aproximarnos a una curva de campana. La curva indica que 60 casos tienen un valor de 45. 40 casos tienen un valor de 30; 40 casos tienen un valor de 60; 10 casos tienen un valor de 15 y 10 casos tienen un valor de 75. Cuando la distribución de los términos ni es simétrica, entonces tenemos curvas asimétricas o agusadas.



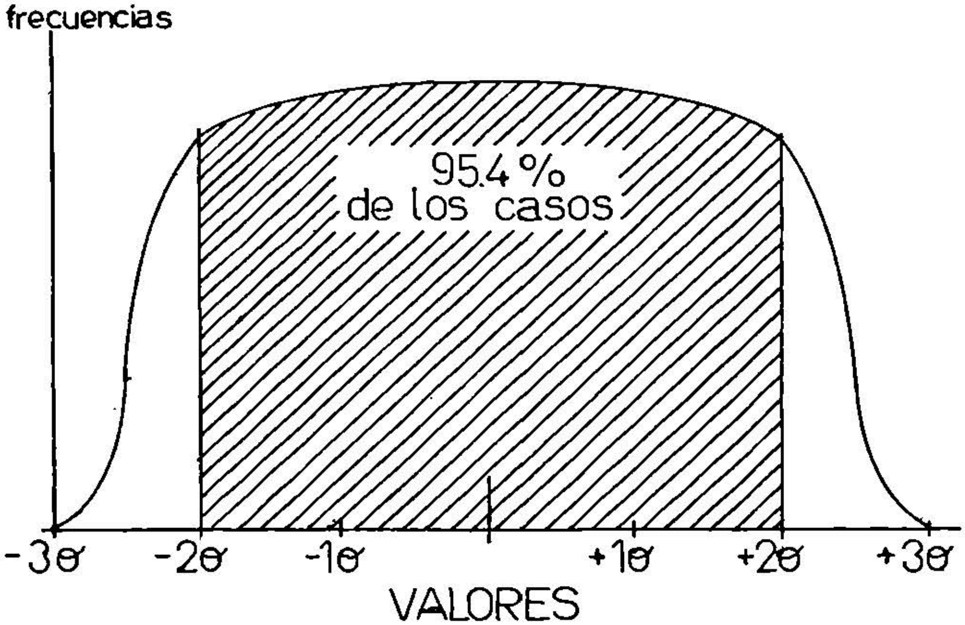
Se dice que una curva es asimétrica cuando le falta simetría, esto es, los términos tienden a concentrarse hacia uno u otro extremo de la distribución.



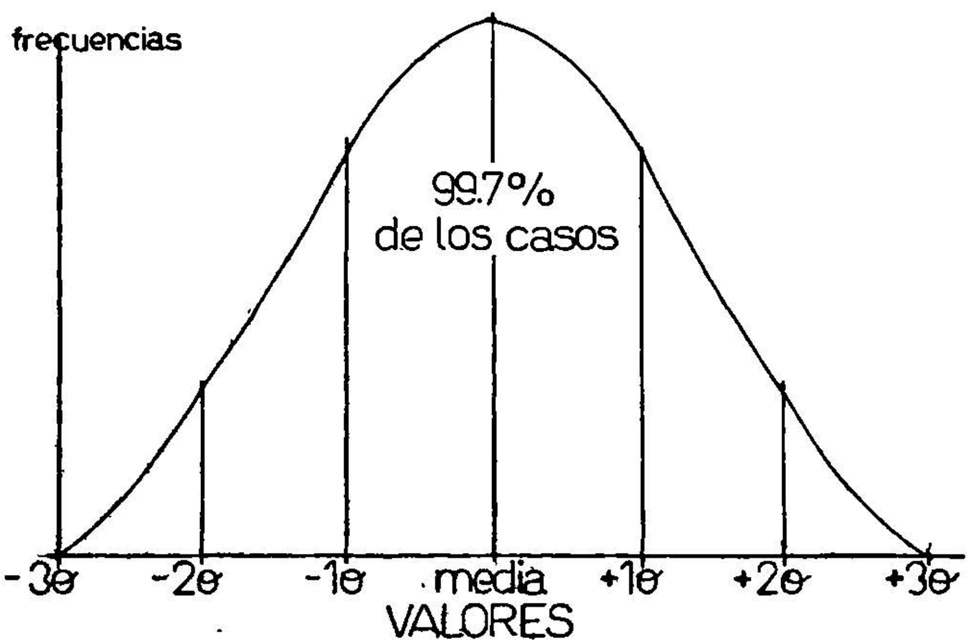
Area de la curva normal, por la magnitud de la media más o menos la desviación promedio.



La curva y amplitud de la media lo (una desviación estándar).



La curva y amplitud de la media ± 2σ.



La curva normal y amplitud de la media ± 3σ.

La desviación estándar para una serie dada tiene un valor mayor que la desviación promedio, porque se incrementa la importancia de los términos de los extremos, al cuadrarse. La desviación promedio tendrá un valor aproximado de cuatro quintos de la desviación estándar cuando la distribución es más o menos normal.
Puesto que la desviación estándar es un valor mayor que la desviación promedio, un área descrita por la media aritmética más ó menos la desviación estándar, incluirá una proporción mayor de términos que un área similar descrita por la desviación promedio. Si la distribución es normal la media aritmética más o menos la desviación estándar incluirá 68.2% de los términos. Un área basada en la desviación promedio incluirá el 57.5% de los valores.
Es ovbio que pocas distribuciones de datos de negocios llegan a adoptar la forma de una curva normal.
Sin embargo, las distribuciones frecuentemente se aproximan al patrón normal, al grado de ser utilizables medidas tales como la desviación promedio y la desviación estándar. Ha sido reconocido por los psicólogos, que todas las medidas de características y capacidades humanas se aproximan más ó menos a la forma de campana. Esto es de suma importancia para la administración de personal.

COMPARACION DE PUNTUACIONES

La desviación estándar tiene otra función de muchas utilidades; ésta puede ser usada en la comparación de puntuaciones individuales de distribución distintas. Por ejemplo, supongamos que dos inspectores de los departamentos A y B, los cuales trabajan en distintas inspecciones, encuentran respectivamente 45 y 89 defectos durante una semana de trabajo. ¿Cómo podremos comparar la eficiencia de estos dos empleados? Salta a la vista de inmediato que la comparación directa de las cifras 45 y 89 no es válida, porque los dos trabajos de inspección pueden ser muy diferentes.
Para hacer la comparación, debemos primero calcular la media de todos los defectos descubiertos por los inspectores en el departamento A y la media de los defectos descubiertos por los inspectores en el departamento B.
Supongamos que estas medias son respectivamente 38 y 95. Por lo tanto el inspector del departamento A está 45-38 = 7 piezas arriba de la media para este departamento, y el inspector del departamento B está 89-95" —6 ó 6 piezas abajo de la media de los ins-

pectores de este departamento. Entonces podemos decir que el inspector del departamento A está por encima del promedio de habilidad para el trabajo y que el inspector del departamento B, está abajo del promedio. ¿Pero cómo podremos definir su distancia relativa al promedio? Para contestar a ésta pregunta debemos calcular las D.E. de las dos distribuciones, y determinar cuántas D.E. están arriba o abajo del promedio de nuestros inspectores.

Supongamos que la D.E. de los inspectores del departamento A es de 5.5 piezas, entonces nuestro primer inspector será $\frac{45-38}{5.5}$ = 1.27 D.E. o (O) arriba de la media. Si la D.E. para los inspectores del departamento B es 9.5 el inspector que encontró 89 piezas está $\frac{89 - 95}{9.5}$ = .63 ó .63 D.E. abajo de la media.

La desviación de una puntuación respecto de la media de una distribución expresada en O o D.E., resulta una medida que puede ser comparada con otras medidas de otras distribuciones igualmente calculadas. De tal manera podemos decir que nuestro primer inspector está dos veces arriba del promedio en términos de una escala comparable en tanto que el segundo está abajo. Las puntuaciones calculadas de esta manera son conocidas como puntuaciones SIGMA. La fórmula para la puntuación O es:

puntuación O =

$$\frac{\text{puntuación no ponderada — media de puntuaciones no ponderadas.}}{\text{D.E. de puntuaciones no ponderadas.}}$$

Otra manera de usar la desviación estándar para comparar distintos grupos de datos es la siguiente:

Supongamos que deseamos comparar el salario por hora entre dos industrias:

Industria A: media = $1.50 D.E. = 0.20
Industria B: media = 1.50 D.E. = 0.15

Usando la fórmula $\frac{\text{D.E.}}{\text{M}} \times \frac{100}{1}$ tenemos:

$$\frac{0.20}{1.50} \times 100 = 13.3 \qquad \frac{0.15}{1.50} \times 100 = 10$$

La variación relativa de salarios por hora fue mayor en la Industria A.

ASIMETRIA

Ocasionalmente puede ser deseable comparar el grado de asimetría en dos o más series de datos. Una medida de asimetría relativa puede ser computada por la división del intervalo entre la media y el modo entre la desviación estándar

la fórmula usada es: Asimetría $= \frac{3\ (\text{Media - Modo})}{\text{D.E.}}$

Se dice que las distribuciones con distorsiones hacia los valores altos tienen asimetría positiva y viceversa.

CAPITULO III.

ANALISIS DE CORRELACION Y PREDICCION.

REFLEXIONES IMPORTANTES.—

Cabe decir que el análisis basado en correlación es ampliamente usado en muchos campos, incluyendo la administración de personal. Es indudable que dentro de la administración ésta tiene muchas aplicaciones tales como: pruebas de personal, control de calidad, sistemas de valuación de puestos, relaciones entre cotizaciones de las acciones, dividendos, utilidades al personal, cambios de dinero, etc. La correlación puede aplicarse en el análisis de mercado para estimar las ventas potenciales en nuevos territorios, etc.

Aunque el análisis de correlación tiene aplicaciones muy variadas el objetivo básico es siempre el mismo: predicción del valor de una variable con base en el valor de otra variable.

Cuando el análisis se limita a dos variables, el método usado es llamado análisis de correlación simple. Supongamos por ejemplo que un examen diseñado para medir la habilidad ejecutiva es administrado cada año a un grupo considerable de prospectos para ejecutivos, en una compañía grande, y que al fin de un período de cinco años, se prepara un diagrama de dispersión que muestra las relaciones entre las puntuaciones obtenidas en el examen y la historia de la actuaciún de cada hombre en la compañía. Supongamos también que ésta relación es alta y positiva. Asumiendo que los hombres estudiados representan más o menos al prospecto medio del ejecutivo, este examen podría usarse en años subsecuentes para predecir al seleccionar nuevos elementos cuáles fracasarán o lograrán éxito como ejecutivos.

También se facilita la predicción de ventas al determinar la relación entre las ventas y otros factores. Por ejemplo: si las fluctuaciones entre las ventas tienen una relación cerrada con los ingresos, las predicciones de ingresos hechas por economistas gubernamentales o privados pueden usarse como base para la predicción de ventas.

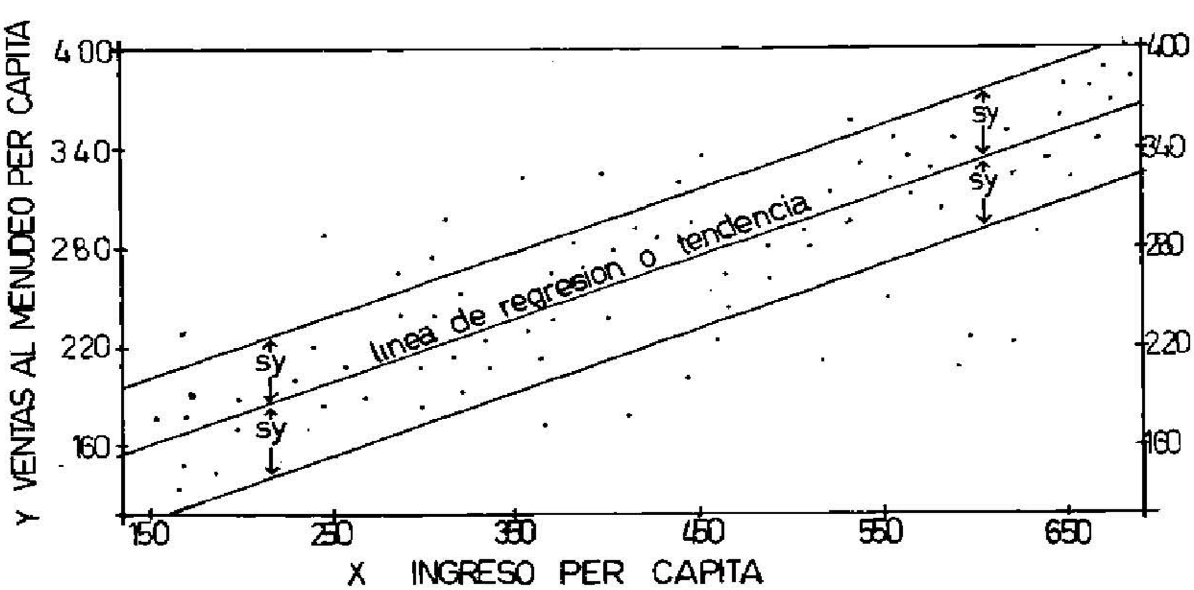
Cuando llega a hacerse necesario examinar la influencia de varias variables, en lugar de una, para comprender el movimiento de grupos de datos, usamos el método del análisis múltiple de correlación.

Una de las razones principales para efectuar análisis de correlación de datos económicos, de personal y de negocios, es la de permitir al ejecutivo estimar puntuaciones de pruebas; costos, ventas o precios, que pueden ser relacionadas fundamentalmente. Es posible eliminar parte del elemento de adivinanza en las decisiones del ejecutivo, si la relación entre la variable que él está estimando y aquella en la que basa su estimación, es cerrada y no ha tenido cambio apreciable respecto de lo que ha sido en el pasado. Tenemos entonces una técnica de estadística que puede ser muy valiosa cuando es adecuadamente aplicada.

DETERMINACION VISUAL DE CORRELACION

Diagramas de dispersión.

La naturaleza de las relaciones entre dos variables puede ser fácilmente estudiada por una persona sin conocimientos estadísticos por medio de la presentación de datos en un diagrama de dispersión. Supongamos por ejemplo que deseamos estudiar la naturaleza de las relaciones entre las ventas al menudeo por persona y el ingreso per cápita en una localidad dada. Para hacerlo, podemos preparar un diagrama que muestre las ventas en una escala y los ingresos en la otra. La horizontal o escala de las X es usada para mostrar los valores de la variable independiente, en este caso los ingresos. La vertical o escala de las Y es usada para indicar las magnitudes de la variable dependiente o ventas per cápita. La variable dependiente, depende de, o es causada por la variable independiente.



Se trazan las escalas X e Y de tal manera que puedan incluirse los valores mayores y menores de las variables que han de representarse en el diagrama. No es necesario iniciar las escalas en cero. Cada punto en el diagrama representa dos valores. Las ventas al menudeo a cada individuo y su ingreso son representadas por una simple marca; su posición, respecto de la escala horizontal representa el monto de las compras que hizo.

Dirección de correlación:

Si un incremento en una de las series es acompañado de un incremento en la otra serie, se dice que la relación es positiva o directa.

Una correlación positiva al representarse gráficamente tendrá una dirección procedente de la parte izquierda inferior que tenderá hacia la parte derecha superior, como en el ejemplo anterior. También, cuando los términos de dos variables tienden a concentrarse sobre una línea recta, decimos que hay una relación recta o lineal. La relación es lineal si la magnitud de la variación en una serie permanece en proporción constante a la magnitud de la variación en la otra serie. En ocasiones una línea curva es la expresión más apropiada de los datos. Una correlación negativa o inversa existe cuando los incrementos en una serie son acompañados por decrementos en la otra serie. Una línea representante de esta relación procedería de la parte superior izquierda de la gráfica hacia la parte inferior derecha.

En ocasiones no existe patrón para los diagramas de dispersión. Si los términos tienden a concentrarse en el centro del diagrama, y de allítienden a dispersarse en todas direcciones, decimos entonces que no existe relación entre las dos variables. Esto es, individuos altos en una serie tienden a ser ni altos ni bajos en la otra. De acuerdo con el diagrama de dispersión anterior, observamos una relación bastante estrecha entre el monto de las ventas a cada persona y el ingreso de las mismas. Esto es, entre más ganaron, más gastaron.

Con base en la relación expresada en la gráfica anterior, las ventas podrían predecirse por medio de los ingresos. Sin embargo, una línea trazada a través de los puntos en el diagrama de dispersión, facilitaría la predicción.

Para muchos fines un diagrama de dispersión y una línea trazada libremente o calculada son adecuados, para trabajos de predicción de ventas o de la actuación futura de solicitantes. No es necesario un adiestramiento avanzado en estadística y matemáticas, para la aplicación de esta técnica y la interpretación de sus resultados. Sin embargo, es posible obtener beneficios adicionales con el cálculo de medidas que describan el grado de correlación con mayor exactitud. Vamos a describir tres medidas útiles de correlación.

1) Línea de mínimos cuadrados. La de mejor ajuste a través de los puntos del diagrama de dispersión:

La ecuación básica para una recta es:

$Y = a + b X$. El símbolo Y es usado para indicar la variable por predecirse y el símbolo Y se refiere a la variable independiente.

La constante (a) representa el valor de Y cuando X es cero, y (b) representa la cantidad media de cambio en X que ocurre con cada unidad de cambio en X.

Pasos del cálculo. En la columna Y anotamos todos los valores variables: en la columna X, las variables independientes. La columna 3 contiene el producto de X Y, y la columna 4 y 5 contienen respectivamente valores en pares. Para resolver la ecuación $Y = a + b X$, usamos dos ecuaciones:

$$a = \frac{\Sigma X^2 \Sigma Y - \Sigma X \;\Sigma XY}{N \Sigma X^2 \quad (\Sigma X)^2}$$

$$b = \frac{N \Sigma XY - \Sigma X \;\Sigma Y}{N \Sigma X^2 - (\Sigma X)^2}$$

Ahora veremos cómo trabajar con estas fórmulas:

Abajo tenemos los datos de rentas e ingresos mostrados en el diagrama de dispersión anterior, con los totales (omitimos muchos de los datos, con objeto de abreviar).

Términos	1 Ventas Y	2 Ingresos X	3 XY	4 Y^2	5 X^2
1	193	315	60,795	37,249	99.225
2	373	670	249,240	138,384	448,900
3	209	387	80,883	43,681	149,769
4	271	508	137,668	73,441	258,064
etc.	etc.	etc.	etc.	etc.	etc.
67	311	597	185,667	96,721	356,409
67	17,506	29,800	8,219,479	4,858,416	14,276,826

Substituyendo tenemos:

$$a = \frac{\Sigma x^2 \Sigma Y - \Sigma x \Sigma xY}{N \Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{(14{,}276{,}826 \times 17{,}506) - 29{,}800 \times 8{,}219{,}479}{67 \times 14{,}276{,}826 - (29{,}800)^2}$$

$$= \frac{249{,}930{,}115{,}956 - 244{,}940{,}474{,}200}{956{,}547{,}342 - 888{,}040{,}000}$$

$$= \frac{4{,}989{,}641{,}756}{68{,}507{,}342}$$

$$a = 72{,}834$$

$$b = \frac{N \Sigma xY - \Sigma x \Sigma Y}{N \Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{67 \times 8{,}219{,}479 - (29{,}800)(17{,}506)}{67 \times 14{,}276{,}828 - (29{,}800)^2}$$

$$= \frac{550{,}705{,}093 - 521{,}678{,}800}{68{,}507{,}342}$$

$$= \frac{29{,}026{,}293}{68{,}507{,}342}$$

$$b = + .423696$$

Entonces $Yc = 72.83 + .42\ X$

Mientras Y representa ventas reales, Yc indica estimaciones o predicciones de ventas.

Una vez determinadas las ventas estimadas correspondientes a las cantidades de ingresos por persona, podemos trazar una línea de tendencia. Es decir, tomamos dos valores cualquiera de X y hacemos la substitución en la ecuación.

$$Yc = 72.83 + .42\ X$$

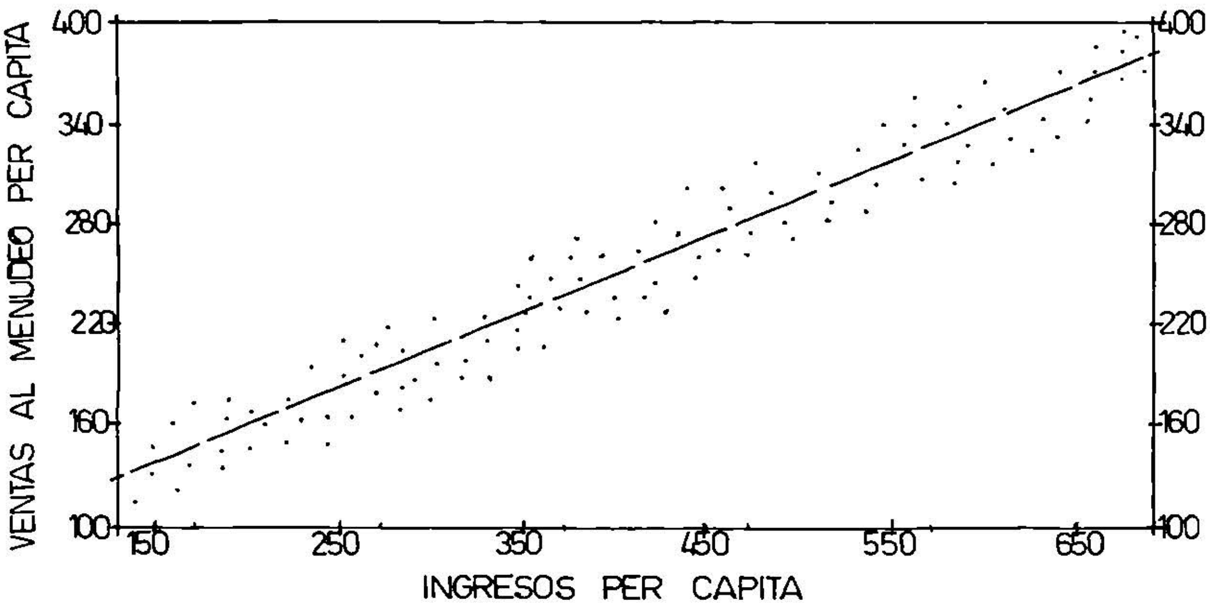
Fijamos las respuestas en el diagrama de dispersión y trazamos una línea entre los dos puntos

$$\text{Así: } Yc = 72.83 + .42\ (315)$$

y

$$Yc = 72.83 + .42\ (597)$$

Al resolver estas ecuaciones y fijar los puntos, obtenemos la siguiente línea de tendencia:



Prolongando la línea más allá de 700. podemos predecir las ventas, con la lectura correspondiente en el eje de la Y.

2 Error estándar de estimación.

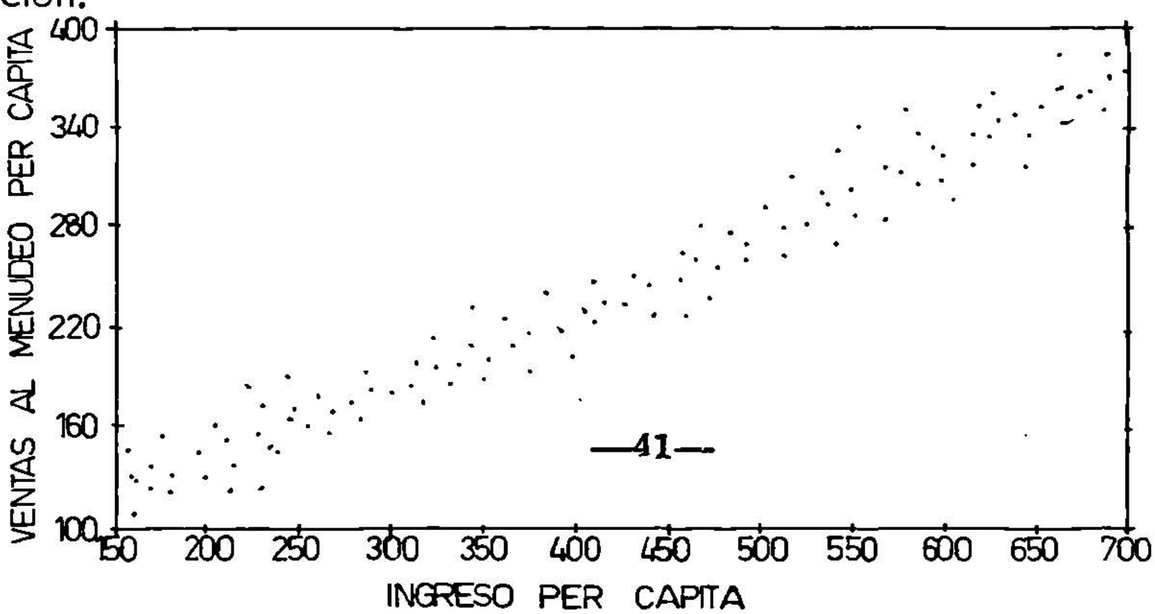
El error estándar de estimación es una medida de variación la cual guarda la misma relación con la línea de tendencia, que la desviación estándar tiene con la media aritmética. Difícilmente encontramos que los puntos del diagrama de dispersión coincidan con la línea de tendencia, porque generalmente hay más de un factor que afecta la serie bajo consideración. Debemos esperar algunos errores en todas las estimaciones hechas con las líneas de tendencia. En el caso del análisis de las ventas al menudeo en relación con los ingresos per cápita, las estimaciones en la línea de tendencia serán afectadas de error a causa de que las ventas son influenciadas por otros factores además de los ingresos per cápita. Por ejemplo, algunos miembros de la comunidad con altos ingresos, quizás hagan sus compras en una ciudad grande cercana, causando así que las ventas al menudeo de su localidad sean más bajas de lo que debían ser al juzgar por los ingresos.

Si son muy importantes los otros factores que afectan el comportamiento de Y, y no los hemos tomado en cuenta al computar la línea de tendencia, las estimaciones hechas serán pobres porque el error es grande. Por tanto es necesario medir la amplitud del error que acompaña a cada estimación.

Se recordará que aproximadamente el 68% de los términos en una serie que tiende a tener una distribución normal, caen dentro de una amplitud determinada por la media más o menos 1 desviación estándar (unidad O); y cerca de 95% dentro de 2 desviaciones estándar. El error estándar de estimación posee estas mismas características. Sin embargo la media aritmética es un valor simple, en tanto la línea de tendencia comprende una serie de valores, que forman una línea recta.

El error estándar de estimación puede ser usado para medir la amplitud de variación en cualquiera y en todos los puntos a lo largo de la línea de tendencia. En efecto, la línea de tendencia más o menos 1 error estándar de estimación representa una zona alrededor de la línea de tendencia, dentro de la cual se localiza el 68% de los términos. Usando el mismo ejemplo, encontramos que el error estándar de estimación Sy es:

Línea de tendencia más y menos un error estándar de estimación.



Cálculo del error estándar de estimación:

Un método consiste en la sustitución, en la ecuacion $Yc = a + bx$ de los valores dados de X. El resultado es por supuesto los valores calculados o estimados para Y, los cuales son unidos para formar la línea de tendencia. Las desviaciones de valores de Y respecto de los valores de Yc se computan de esta manera. Estas desviaciones son cuadradas, sumadas, divididas entre N. y al cociente se le extrae raíz cuadrada.

$$Sy = \sqrt{\frac{\Sigma (Y - Yc)^2}{N}}$$

El resultado es el error estándar de estimación.

Un método más corto de cálculo, es usando la siguiente ecuación:

$$Sy^2 = \frac{\Sigma (Y)^2 - [a \quad \Sigma (Y) + b \quad \Sigma (XY)]}{N}$$

Las constantes a y b son las constantes en la ecuación de la línea recta ajustada. N es el número de pares. Usando los mismos datos de ventas e ingresos per cápita tenemos:

$$Sy^2 = \frac{4,858,146 - (72.834)(17,506) + (.423696)\ (8,219,479)}{67}$$

$$Sy^2 = \frac{100,554}{67} = 1500.8059$$

$$Sy^2 = 38.74$$

El error estándar de estimación es \$ 38.74. Si escogemos cualquier valor Yc en la línea de tendencia, podemos decir que en 68 de cada cien casos, el valor de las compras variarán entre Yc más y menos \$ 38.74, y que 95 de cada 100 personas con un ingreso dado el valor de sus compras tendrán una amplitud de Yc más y menos \$ 77.48 ó 2 desviaciones estándar 2 (\$ 38.74).

El valor de la ecuación de tendencia, como una herramienta de predicción aumenta a medida que el valor relativo del error estándar disminuye. Cuando existe una relación perfecta entre dos series de datos, el error estándar de estimación es igual a cero.

En la misma forma que se usa la desviación estándar para medir la dispersión de datos con respecto a la media, el error estándar de estimación tiene la finalidad de describir la dispersión de los puntos con respecto a la línea de tendencia.

Coeficientes de correlación.

Hemos visto hasta ahora, la posibilidad de estimar y explicar el comportamiento de una variable económica en función de otra relacionada con ella. Estas variaciones fueron expresadas en términos de los valores originales, tales como el monto de ventas, calificaciones de pruebas o ingresos en pesos, etc.

Es deseable ahora, determinar una cifra aritmética simple, la cual indicará la extensión en que dos variables se correlacionan.

Este debe ser un número puro para que las unidades que representan los valores no la afecten. La cifra debe tener límite para poder interpretarla fácilmente. El coeficiente de correlación es un número abstracto que varía de + 1 a — 1 pasando por cero.

El signo indica si la curva de tendencia es positiva o negativa, mientras la magnitud del coeficiente indica el grado de relación. Cuando no existe ninguna relación entre las variables. r (coeficiente de correlación) es cero; cuando la relación es perfecta, este es + 1 ó — 1.

Hay muchas vías para calcular el coeficiente de correlación.

Presentaremos los dos mejores métodos en nuestra opinión.

a) Método de cuadrados mínimos

$$r = \sqrt{1 - \frac{S_y^2}{\sigma_y^2}}$$

r = coeficiente de correlación

S_y^2 = error estándar de estimación al cuadrado

σ_y^2 = desviación estándar de y al cuadrado

y = en este caso es una variable (ventas per cápita)

$\sigma_y = 65.12$

$\sigma_y^2 = 4240.5221$

$S_y = 38.74$

$S_y^2 = 1500.8059$

Usamos los mismos datos de ventas e ingresos per cápita para ilustrar el uso de estas fórmulas:

$$r = \sqrt{1 - \frac{S_y^2}{\sigma_y^2}}$$

$$r = \sqrt{1 - \frac{1500.8059}{4240.5221}}$$

$$r = \sqrt{1 - .353920}$$

$$r = \sqrt{.646080}$$

r = .8038 (coeficiente de correlación)

b) El segundo método es llamado Momento-Producto y puede ser usado directamente con los datos primitivos, antes de calcular las desviaciones estándar.

La fórmula es:

$$r = \frac{N \quad \Sigma XY - \quad \Sigma X \quad \Sigma Y}{\sqrt{[N \quad \Sigma X^2 - (\quad \Sigma X)^2]}\sqrt{[N \quad \Sigma Y^2 - (\quad \Sigma Y)^2]}}$$

$$r = \frac{(67)(8{,}219{,}479) - (29{,}800)(17{,}506)}{\sqrt{[(67)(14{,}276{,}790) - (29{,}800)^2]}\sqrt{[(67)(4{,}858{,}146) - (17{,}506)^2]}}$$

$$r = + \frac{29{,}026{,}293}{36{,}112{,}202}$$

$$r = + .8038$$

Tenemos entonces un coeficiente de correlación de .8 entre ventas e ingresos. Esto indica un alto grado de correlación. Es imposible establecer arbitrariamente, la magnitud del coeficiente de correlación que se hace necesario, para que ésta tenga significación. Mucho depende de la naturaleza de las series que son analizadas y del número de pares de valores. Un coeficiente de una magnitud dada llega a ser más significativo, a medida que aumenta el número de pares de valores del cual fue calculado.

Coeficiente de Indeterminación. Si nos interesa saber que porcentaje de casos es afectado por otros factores distintos de la variable independiente, usamos la siguiente fórmula:

$$\text{Coeficiente de indeterminación} = \frac{Sy^2}{O\,y\,a} \quad O \quad \frac{1500.8059}{4240.5221} = .3539$$

Podemos concluir entonces que otros factores además del ingreso per cápita, influyen aproximadamente en el 35% de la variación en las ventas per cápita entre los distintos ingresos.

Coeficiente de determinación. Si el coeficiente de indeterminación es .35 por ciento, parece lógico que el coeficiente de determinación sea 100 por ciento menos .35 por ciento o sea .65 por ciento. Esto es cierto y se calcula como sigue:

$$1 - \frac{Sy^{2}}{O\,y^{2}} = 1 - .3539 = .6461$$

Es obvio que se explica el 65% de la variación total en ventas per cápita, por X — diferencias en ingresos per cápita. El coeficiente de determinación es uno de los valores más útiles y significantes que encontramos en el proceso del análisis de correlación.

Simplemente, elevando al cuadrado cualquier coeficiente de correlación, obtenemos el coeficiente de determinación, el cual indica el porcentaje de casos o términos que en las dos series bajo observación se relacionan directamente.

Correlación para medir la fidelidad de una prueba. El siguiente ejemplo nos muestra cómo usar la correlación para medir la fidelidad de puntuaciones de pruebas. Naturalmente este método puede emplearse en casos similares.

X = puntuaciones de estudiantes a preguntas nones en la prueba.

Y = puntuaciones de estudiantes a preguntas pares en la prueba.

Estudiantes	Puntuaciones totales en nones X	Puntuaciones totales pares Y	X^2	Y^2	XY
A	2	3	4	9	6
B	3	2	9	4	6
C	3	2	9	4	6
D	4	5	16	25	20
E	4	3	16	9	12
F	5	7	25	49	35
G	5	6	25	36	30
H	7	8	49	64	56
I	8	9	64	81	72
J	9	5	81	25	45
Total	50	50	298	306	288

Un tercer método (no mencionado arriba) es usado para computar el coeficiente de correlación es el siguiente:

$$\Sigma X^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = 298 - 250 - 48$$

$$\Sigma Y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{N} = 306 - 250 = 56$$

$$\Sigma XY = XY - \frac{(\Sigma X)(\Sigma Y)}{N} = 288 - 250 = 38$$

$$\sigma X = \sqrt{\frac{\Sigma X^2}{N}} = \sqrt{\frac{48}{10}} = \sqrt{4.8}$$

$$\sigma Y = \sqrt{\frac{\Sigma Y^2}{N}} = \sqrt{\frac{56}{10}} = \sqrt{5.6}$$

$$r = \frac{XY}{N\sigma X\, \sigma Y} = \frac{38}{10\sqrt{4.8}\sqrt{5.6}} = \frac{38}{10\sqrt{4.8 \times 5.6}} = \frac{38}{51.84} = 0.73$$

$r = .733$

Fidelidad de la prueba entera (coeficiente de fidelidad)

$$rt = \frac{2r}{1 + r} = \frac{1.466}{1.733} = 0.855 \quad rt = \text{coeficiente de fidelidad}$$

Podemos deducir que el coeficiente + .73 es 85% confiable o fiel.

A pesar de que en nuestros ejemplos de correlación usamos datos de ingresos y ventas, se debe entender que estas técnicas se aplican a cualquiera dos series de datos que quisiésemos comparar, para determinar su relación.

El encontrar un coeficiente de correlación alto, no significa en sí mismo la evidencia de una relación entre las series. Hay muchos casos de correlación sin sentido. Por tanto es de suma importancia que la lógica de cualquier relación entre series sea determinada a base de análisis subjetivos profundos.

CONCLUSIONES

1.—Encontramos, que el creciente uso de estadísticas en la administración es parte de la tendencia hacia basar las decisiones administrativas en un fundamento tan objetivo y científico como resulte posible. Los métodos de administración modernos cada vez dependen más de datos estadísticos para obtener la información requerida de hechos acerca de sus operaciones internas y los campos más amplios de negocios y actividad económica.

2.—Un número usado para describir una serie debe ser representativo de los datos medidos por él. Para ser representativo el número debe reflejar la tendencia de las partidas individuales de la serie para concentrarse en ciertos valores centrales y ser distribuido entre ellos. Por ésta razón un número representativo se considera como una medida de la tendencia central. Más comunmente es conocido como un promedio.

3.—Los promedios basados en sus propiedades matemáticas, pueden clasificarse en dos categorías: promedios computados que comprenden la media aritmética y promedios de posición que comprenden la mediana y el modo.

4.—Las medidas de dispersión representan parámetros (o estadísticos) que pueden cuantificar la discrepancia de los valores del universo (o muestra) respecto a una medida de tendencia central. La desviación stándar es la medida de dispersión o variabilidad, comunmente usada en el análisis estadístico. Su uso es grande debido a que nos permite tomar en cuenta los elementos de variación.

5.—Se puede señalar que correlación es la medida cuantitativa del grado de asociación entre dos variables. Es decir se pretende apreciar el grado de bondad de la manera como una ecuación, describe o expresa la relación entre dos variables.

6.—La estadística es el estudio de métodos y procedimientos para coleccionar, clasificar y analizar datos y obtener inferencias científicas de tales datos. En éste caso referidos específicamente al campo de la administración científica.

7.—El Licenciado en Administración de Empresas, siendo el profesionista técnicamente más capacitado para planear, organizar, dirigir, controlar y coordinar el departamento de personal; dados sus amplios conocimientos adquiridos a través de sus estudios universitarios, deberá pugnar en forma objetiva y constructiva, por el empleo de tales técnicas, con fines productivos en cualquier tipo de organización empresarial.

B I B L I O G R A F I A

1.—ESTADISTICA DESCRIPTIVA.
Gotkin y Goldstein.
Editorial Limusa. - Willey, S.A.
México, 1972.

2.—MATEMATICAS UNIVERSITARIAS.
Brittou, Jack R. - Krieg - Rutland.
Compañía Editorial Continental, S. A.
México, 1968.

3.—INTRODUCCION A LA INFERENCIA ESTADISTICA.
Guenther William C.
McGraw - Hill Book Company.
México, 1965.

4.—METODOS ESTADISTICOS.
Ríos, Sixto.
México, 1967.

5.—ADMINISTRACION DE PERSONAL.
Pigors y Myers.
Compañía Editorial Continental, S. A.
México, 1970.

6.—ADMINISTRACION DE PERSONAL.
Ettinger, Karl E.
Herrero Hermanos Sucesores, S. A.
México, 1961.

7.—DIRECCION DE PERSONAL.
Scott, Clothier y Spriegel.
McGraw, Hill Book, Company.
México, 1964.

8.—MANEJO DE PERSONAL Y RELACIONES INDUSTRIALES.
Yoder Dale.
Compañía Editorial Continental, S. A.
México, 1967.