471

RELIMINAR
NSTITUTO LATINOAMERICANO DE
LANIFICACION ECONOMICA Y SOCIAL
antiago, febrero de 1966

CURSO DE ESTADISTICA BASICA PARA PROGRAMACION *

* Programa de Capacitación, Profesor Sr. Arturo Núñez del Prado B.; Ayudantes Sres. Jorge Carvajal, Sergio Chaigneau, Max Vildósola.

INDICE

I. ESTADIGRAFOS DESCRIPTIVOS

Páginas

II. NUMEROS INDICE

FINALIDAD DEL CURSO

El programa ha sido diseñado para capacitar a los participantes del Curso Básico en materias estadísticas de uso cotidiano en la planificación y la investigación económica general. Por otra parte también se ha contemplado las necesidades de instrumental estadístico que tienen otras asignaturas dentro del Programa de Capacitación.

Principalmente lo que interesa es que el planificador pueda utilizar racionalmente el aparato estadístico y pueda entenderse con estadísticos y econometristas en estudios más profundos. Sería pretencioso pensar que después de un curso de 24 horas de clase e igual número de horas de seminarios y ejercicios prácticos, el alumno haya alcanzado una cierta especialidad en este campo. No, el objetivo es proporcionarles un conjunto de conceptos que garantizarán al planificador detectar las posibilidades y limitaciones de la utilización de indicadores estadísticos, interpretando cabalmente y en su justa dimensión los estadígrafos de uso más frecuente.

Se trata de un curso intensivo, donde el trabajo principal recae en el propio alumno. Muchos conceptos y materias no serán tratados en forma teórica, si no a través de ejercicios prácticos en los seminarios.

Para facilitar el trabajo de los alumnos de manera que dispongan de un resumen de las distintas materias que se presenten a lo largo del curso, se han preparado estos apuntes. Apuntes que constituyen en buena parte resúmenes de otros textos como los Apuntes de Estadística del profesor Pedro Vuskovic, el Curso General de Estadística del profesor Enrique Cansado, Introducción a la Estadística de G.U. Yule y M.G. Kendall, y Estadística General Aplicada de F. Croxton y D. Cowden.

Estos apuntes, no son en absoluto suficientes. El alumno, para conseguir una eficiente capacitación, deberá consultar los textos citados y realizar con toda conciencia y dedicación los ejercicios y seminarios que constituyen la médula del curso.

I. ESTADIGRAFOS DESCRIPTIVOS

A. DISTRIBUCIONES DE FRECUENCIA

1. Generalidades

Un conjunto de datos o masa estadística, es susceptible de ser resumida y clasificada de acuerdo a criterios convenientes. Sea que las informaciones provengan de censos o de muestras relativamente grandes, siempre será ventajoso para el análisis, ya que difícilmente podrán obtenerse conclusiones válidas de una masa estadística no clasificada.

Los tipos de variables que preocuparán la atención del curso serán los siguientes:

a) Variables cardinales: susceptibles de medición cuantitativa, las que comprenden a:

   i) continuas: variables que pueden tomar cualquier valor dentro de un intervalo (ingresos, estaturas, distancias, etc.);

   ii) discretas: variables que sólo toman algunos valores dentro de un intervalo (número de hijos por familia, número de accidentes de tránsito por día, etc.);

b) Variables ordinales: sólo susceptibles de ordenación pero no de medición cuantitativa (grado de cultura de una persona: muy culta, regularmente culta, poco culta, inculta).

Para cada uno de estos tipos de variables, un conjunto de observaciones, puede dar origen a una distribución de frecuencias. Debe entenderse ésta, como un cuadro o tabla resumen de los datos originales.

En el caso de variables continuas será necesario fijar intervalos de frecuencia para llegar a un efectivo resumen de la información original. El punto medio de cada intervalo se denominará marca de clase y constituirá el valor representativo de cada intervalo. El número de observaciones que correspondan a cada intervalo se denominarán frecuencias absolutas.

/Una tabla

Una tabla de distribución de frecuencia para variable continua y sus símbolos correspondientes, se representa de la siguiente forma:

Ingresos de Profesionales			Número de Profesionales
Intervalos		Marcas de clase	Frecuencias Absolutas
Y'_{i-1}	Y'_i	Y_i	n_i
Y'_o	Y'_1	Y_1	n_1
Y'_1	Y'_2	Y_2	n_2
Y'_2	Y'_3	Y_3	n_3
.	.	.	.
.	.	.	.
Y'_{m-1}	Y'_m	Y_m	n_4

donde: $Y_i = \dfrac{Y'_{i-1} + Y'_i}{2}$ marca de clase

$n = \sum_{i=1}^{m} n_i$: número de observaciones

$c_i = Y'_i - Y'_{i-1}$: amplitud del intervalo

Estas tablas pueden ser de amplitud constante o de amplitud variable, dependiendo de los valores que tome c_i.

Cuando se trata de variable discreta o discontinua, la tabla de distribución de frecuencias es de la forma siguiente:

Y_i	n_i
Y_1	n_1
Y_2	n_2
Y_3	n_3
.	.
.	.
Y_m	n_m

/Cabe destacar

Cabe destacar que cuando se tiene un número grande de valores distintos de la variable, para abreviar el trabajo, con cierta arbitrariedad y con alguna pérdida de precisión, puede tratarse como una variable continua, formando intervalos de clase.

Por último, en el caso de variables no mensurables, la tabla en cuestión, tendrá una forma como la siguiente:

Variables	Frecuencia
Característica A	n_A
Característica B	n_B
.	.
.	.
Característica Z	n_Z

El estudiante advertirá que las tablas de distribución de frecuencias, facilitan enormemente el análisis. Es muy ventajoso disponer de informaciones clasificadas en intervalos o en valores específicos de la variable, ya que de esta manera es posible obtener conclusiones gruesas acerca de la variable que se investiga.

Respecto de las frecuencias, es posible y generalmente útil, presentarlas en términos relativos, calculando la proporción que del total de observaciones, corresponde a cada intervalo o marca de clase. Se denominan frecuencias relativas, y se simbolizarán por h_i:

$$h_i = \frac{n_i}{n}$$

Tanto las frecuencias absolutas como las relativas son susceptibles de acumulación respecto de los intervalos o marcas de clase. Las frecuencias absolutas acumuladas se simbolizarán por N_i y se definen:

$$N_i = \sum_{i=1}^{i} n_i$$

/Las frecuencias

Las frecuencias relativas acumuladas se simbolizarán por H_i y se definen:

$$H_i = \sum_{i=1}^{i} h_i$$

En general este tipo de frecuencias se acumulan en sentido creciente de la variable y una frecuencia acumulada N_i indica el número de casos o observaciones en que la variable toma valores a lo sumo iguales a Y_i en el caso de variable discreta y a Y'_i en el caso de variable continua. Sin embargo, para ciertos análisis, también es necesario acumular en sentido inverso; de ahí que se hable de frecuencias acumuladas hacia arriba o hacia abajo.

2. Representación gráfica

En general, la representación gráfica de una tabla de distribución de frecuencias, permite visualizar con mayor claridad algunas características de la masa de datos que se investiga. Resulta bastante más fácil transmitir conclusiones a personas no habituadas a la interpretación de distribuciones de frecuencia, cuando se utilizan gráficos estadísticos.

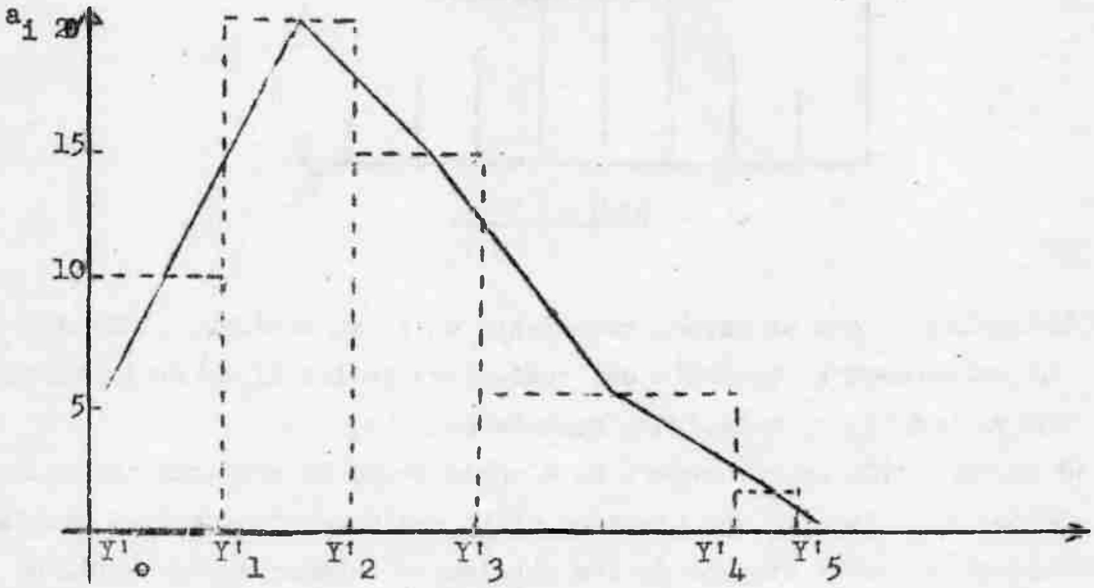
a) Representación gráfica de variable continua. Utilizando un par de ejes coordenados, en el eje de las abscisas se representará la variable que se estudia, en tanto que en el eje de las ordenadas, se representarán las frecuencias correspondientes. Recuérdese que en este tipo de variables, la frecuencia corresponde a un intervalo y por este hecho se representa mediante una superficie. Por medio de un ejemplo se aclararán estas ideas: Admítase que se tiene la siguiente tabla correspondiente a las edades de los participantes del curso básico:

Edades		Alumnos	Amplitud de Intervalo
Y'_{i-1}	Y'_i	n_i	C_i
18	22	10	4
22	26	20	4
26	30	16	4
30	38	12	8
38	40	1	2

/En vista

En vista de que la amplitud más frecuente es 4, puede elegirse ésta como amplitud unitaria; así el cuarto intervalo tendrá dos veces la amplitud unitaria elegida y el quinto intervalo tendrá la mitad de dicha amplitud. La representación gráfica se hará de la siguiente manera:

Gráfico N° 1



En el gráfico, para calcular la altura de cada rectángulo, se planteó la relación de superficie siguiente:

$$\text{superficie} = \text{base x altura}$$

$$n_i = c'_i \cdot a_i$$

donde c'_i es la amplitud unitaria elegida con objeto de diseñar un gráfico adecuado. Perfectamente pudo haberse trabajado con las amplitudes originales, pero habría sido algo más laborioso.

Este tipo de gráficos recibe el nombre de histogramas y la línea quebrada que une los puntos medios de los lados superiores de los rectángulos recibe el nombre de polígono de frecuencias.

b) Representación gráfica de variable discreta. En este caso la frecuencia correspondiente a cada valor de la variable estará representada por una barra vertical.

/Gráfico N° 2

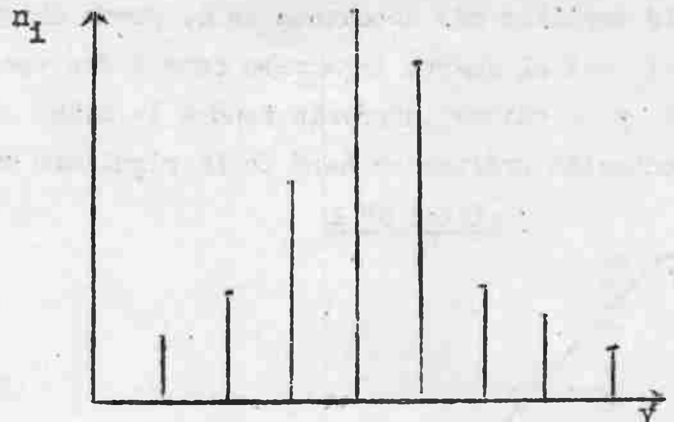


Gráfico Nº 2

Naturalmente que se pueden construir, en forma similar, gráficos que relacionen la variable con cualquiera de los tipos de frecuencias que se han visto, relativas, acumuladas, etc.

A continuación se presentará un ejemplo donde se seguirán todos los pasos para llegar a una tabla de distribución de frecuencias completa.

Supóngase que se dispone de las siguientes informaciones acerca de los sueldos de los obreros de una fábrica (en dólares por mes):

68	48	53	73	100	80	40	55	65	95	85	35	110	120	60
90	70	40	80	100	70	50	55	70	65	45	80	60	90	50
55	60	30	110	110	90	70	60	45	65	80	85	90	68	72
50	40	45	90	105	108	35	45	50	70	82	84	66	38	48

Una de las primeras decisiones es determinar el número de intervalos que tendrá la tabla. Para ello es necesario considerar el objetivo que se persigue con el estudio de la variable, qué tipo de diferenciaciones o agrupamientos interesaría conocer. Por otra parte, es necesario considerar el recorrido de la variable, es decir, el menor y mayor valor entre los datos que se analizan. Por último, el número de observaciones, de manera que los diferentes intervalos tengan frecuencias en alguna medida significativas. Cuando existan valores escasos muy alejados de lo que podría llamarse una concentración central, puede optarse por dejar los intervalos extremos abiertos. Supóngase que tomando en cuenta las consideraciones

/anteriores se

anteriores se decide clasificar los datos originales en 9 intervalos de amplitud constante. Dado que la diferencia entre los valores extremos (30 y 120) es de 90, la amplitud de los intervalos será igual a 10.

Y'_{i-1}	Y'_i	Observaciones	n_i	h_i	N_i	H_i	N_i*	H_i*	Y_i
30,0	40	~~////~~ //	7	7/60	7	7/60	60	60/60	35
40,1	50	~~////~~ ~~////~~	10	10/60	17	17/60	53	53/60	45
50,1	60	~~////~~ ///	8	8/60	25	25/60	43	43/60	55
60,1	70	~~////~~ ~~////~~ /	11	11/60	36	36/60	35	35/60	65
70,1	80	~~////~~ /	6	6/60	42	42/60	24	24/60	75
80,1	90	~~////~~ ////	9	9/60	51	51/60	18	18/60	85
90,1	100	///	3	3/60	54	54/60	9	9/60	95
100,1	110	~~////~~	5	5/60	59	59/60	6	6/60	105
110,1	120	/	1	1/60	60	60/60	1	1/60	115
			60	1					

(*) acumuladas "hacia arriba"

Para evitar situaciones ambiguas, cuando se presentan observaciones con valores de la variable que corresponden a los límites de los intervalos, puede seguirse el criterio planteado en el ejemplo de agregar un decimal a la columna de límites inferiores, aunque ésto sólo tenga utilidad para fines de clasificación de las observaciones, ya que en los cálculos que posteriormente se tratará, tales decimales son despreciados.

Con lo visto hasta el momento, es posible realizar los primeros análisis de un conjunto dado de datos. Tanto la representación gráfica como la tabulación de las distintas clases de frecuencias ayudan a ensayar los primeros juicios. Es necesario insistir sobre la necesidad de tomar en cuenta, en todas las decisiones respecto de la tabla de frecuencias, la naturaleza del fenómeno que se investiga; sobre todo en lo que se refiere al número de intervalos y a sus amplitudes: constantes o variables. Naturalmente que, una vez clasificados los datos originales, será preciso realizar análisis con mayor profundidad, utilizando instrumentos estadísticos que se detallarán en las próximas páginas.

/B. ESTADIGRAFOS DE

B. ESTADIGRAFOS DE TENDENCIA CENTRAL

Una vez que se ha conseguido la clasificación de los datos originales donde se destacan sus características más esenciales, será preciso calcular un conjunto de indicadores que caractericen en forma un tanto más precisa, la distribución que se está estudiando. Interesa en primer término disponer de estadígrafos que representen valores centrales en torno de los cuales se agrupan las observaciones. En general se los designa como promedios, y son de extraordinaria utilidad tanto en el análisis de una distribución, como en la comparación entre distribuciones.

1. Media aritmética. Es sin duda, el estadígrafo más utilizado, sobre todo en la cuantificación de variables económicas. Se simbolizará por $\bar{Y}$ ó $M[Y_i]$, y se definirá como:

$$\bar{Y} = M[Y_i] = \frac{\sum_{i=1}^{m} Y_i n_i}{n}$$

Puede observarse que a cada valor de la variable o marca de clase, se le dá una importancia o peso equivalente a la frecuencia absoluta correspondiente. Esta fórmula de cálculo es para datos agrupados en forma de una distribución de frecuencias. Cuando se desea calcular una media aritmética de datos no agrupados, todas las frecuencias absoluta serán iguales a la unidad, se simbolizará por $\bar{X}$ ó $M[X_i]$ y se definirá como:

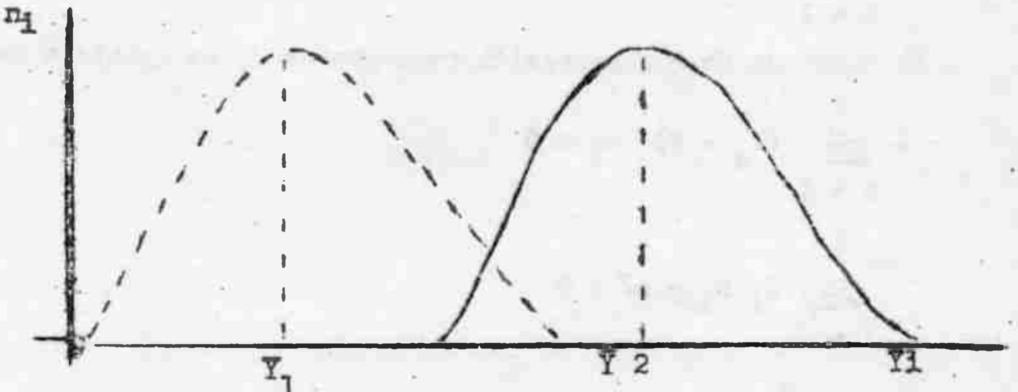
$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

En general siempre es conveniente la agrupación de frecuencias en intervalos, cuando el número de observaciones es relativamente grande. Es evidente que al resumir un conjunto de datos en un número dado de intervalos, se pierde precisión; esta pérdida estará relacionada con la amplitud del intervalo, mientras mayor sea ésta, menos preciso el cálculo. Por ello, en general, para un mismo conjunto de datos, la media aritmética obtenida de los datos originales, que es un calculo exacto, diferirá de la obtenida de una tabla de distribución de frecuencias. La razón estriba en el supuesto de uniformidad de la distribución de frecuencias dentro de cada intervalo, supuesto que generalmente no no se cumple. En todo caso esa pérdida de precisión está más que compensada por las ventajas que significa tener una tabla de frecuencias. Por lo demás en ciencias sociales, la precisión necesaria

/autoriza, dentro

autoriza, dentro de ciertos límites, a desentenderse de una rigurosidad extrema.

En el gráfico que a continuación se presenta, se tienen dos distribuciones de frecuencias (supuesto un gran número de intervalos pequeños) muy similares, y sin embargo con medias aritméticas muy distintas.



La media aritmética como estadígrafo de tendencia central, indica la posición de la distribución. Cabe advertir sobre este estadígrafo su fuerte sensibilidad en cuanto a valores extremos de la variable. Un valor muy alejado de los valores centrales, aunque poco representativo por ser único, puede hacer variar fuertemente el promedio. Por ello cuando se está utilizando este indicador en un análisis, vale la pena percatarse de la representatividad de los valores extremos, y la influencia que éstos tienen en el resultado. Muchas veces se concluye que es preferible estratificar previamente los datos originales en dos o tres categorías, realizando cálculos de medias aritméticas en forma separada para cada grupo.

a) Propiedades. Se presentarán las propiedades más importantes de la media aritmética

i) Primera propiedad: La suma de las desviaciones ponderadas de los valores de la variable respecto de la media aritmética es cero:

$$\sum_{i=1}^{m} (Y_i - \bar{Y})\, n_i = 0$$

$$\sum_{i=1}^{m} Y_i\, n_i - n\bar{Y} = 0$$

$$n\bar{Y} - n\bar{Y} = 0$$

/ii) Segunda Propiedad:

ii) Segunda propiedad: La suma de los cuadrados de las desviaciones ponderadas de los valores de la variable es un mínimo, cuando se toman respecto de la media aritmética. Se iniciará la demostración tomando desviaciones respecto a un valor cualquiera P, para luego concluir que P forzosamente tendrá que ser la media aritmética.

$$\sum_{i=1}^{m} (Y_i - P)^2 n_i \qquad \text{es mínimo}$$

La derivada de esa expresión respecto de P, se iguala a cero.

$$-2 \sum_{i=1}^{m} (Y_i - P) \; n_i = 0$$

$$\sum_{i=1}^{m} Y_i \, n_i - nP = 0$$

$$P = \frac{\sum_{i=1}^{m} Y_i \, n_i}{n}$$

Existe seguridad que igualando la primera derivada a cero se obtiene un mínimo, por que se llega a un valor concreto. Para que fuera un máximo P tendría que ser infinito.

iii) Tercera propiedad. La media aritmética de una variable más (menos) una constante es igual a la media de la variable más (menos) la constante.

$$M [Y_i \pm K] = M [Y_i] \pm K$$

$$\sum_{i=1}^{m} \frac{(Y_i \pm K) n_i}{n} = M [Y_i] \pm K$$

$$\sum_{i=1}^{m} \frac{Y_i \, n_i}{n} \pm \frac{n \, K}{n} = M [Y_i] \pm K$$

$$M [Y_i] \pm K = M [Y_i] \pm K$$

iv) Cuarta propiedad. La media aritmética de una variable multiplicada (dividida) por una constante, es igual a la constante que multiplica (divide) a la media de la variable.

$$M\left[Y_i K\right] = K\,M\left[Y_i\right]$$

$$\sum_{i=1}^{m} \frac{Y_i K n_i}{n} = K\,M\left[Y_i\right]$$

$$K \sum_{i=1}^{m} \frac{Y_i n_i}{n} = K\,M\left[Y_i\right]$$

$$K\,M\left[Y_i\right] = K\,M\left[Y_i\right]$$

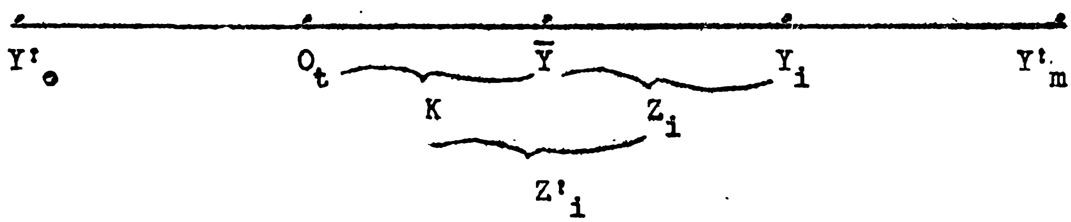
b) Métodos abreviados de cálculo. Se presentarán estos métodos más que por el ahorro de tiempo que puede significar su aplicación, porque recalcan algunos detalles sobre la media aritmética.

i) Primer método abreviado. Se trata de reducir la magnitud de la variable, en términos de desviaciones respecto de un origen de trabajo O_t elegido arbitrariamente. En cuanto a la elección de O_t vale la pena , para que el método sea realmente abreviado, tomar como origen de trabajo, un valor o marca de clase central de la distribución de frecuencias.

Se definirá esta variable reducida en la forma siguiente:

$$Z'_i = Y_i - O_t$$

Gráficamente se fijan los siguientes puntos dentro del recorrido de una variable:



Además, a las desviaciones respecto de la media aritmética, se simbolizará por Z_i, es decir:

$$Z_i = Y_i - \overline{Y}$$

La diferencia entre la media aritmética y el origen de trabajo arbitrariamente elegido se designará por K, es decir:

$$K = \overline{Y} - O_t$$

Por este hecho, observando el gráfico, puede concluirse que:

$$\overline{Y} = O_t + K$$

Se trata de encontrar una expresión para K, en función de desviaciones Z'_i, para disponer de una fórmula de cálculo. En efecto:

/Z'i =

$Z'i = Zi + K$ (multiplicando por ni)

$Z'i\ ni = Zini + K\ ni$ (aplicando sumatoria)

$$\sum_{i=1}^{m} Z'i\ ni = \sum_{i=1}^{m} Zi\ ni + nK$$

Recuérdese que en la primera propiedad de la media aritmética se demostró que:

$$\sum Zini = \sum (Yi - \bar{Y})\, ni = 0$$

Luego:

$$K = \frac{\sum Z'i\, ni}{n}$$

La fórmula de cálculo abreviado será

$$\bar{Y} = O_t + \frac{\sum Z'i\ ni}{n}$$

En la siguiente tabla de distribución de frecuencias se aplica este método:

	Ahorros			Familias	
Y'i-1	Y'i	Yi	ni	Z'i	Z'i ni
0	8	4	8	- 19	- 152
8	20	14	10	- 9	- 90
20	26	23 1/	30	0	0
26	30	28	9	5	45
30	40	35	3	12	36
			60		- 161

1/ Se elige $O_t = 23$

$$\bar{Y} = 23 - \frac{161}{60} = 20{,}32$$

ii) Segundo método abreviado. Este método en general sólo se aplica con ventaja, en el caso en que la amplitud de los intervalos sea constante. Al igual que en el anterior se trata de trabajar en términos de desviaciones, pero además en este método dichas desviaciones se expresan en unidades de intervalo (dividiéndolas por C)

Esta nueva variable es en consecuencia:

$$Z''_i = \frac{Yi - O_t}{C} = \frac{Z'i}{C}$$

de donde: $Z'i = CZ''i$

En la fórmula del primer método se reemplaza Z'i y se tiene :

$$\bar{Y} = O_t + C\ \frac{\sum Z''i\ ni}{n}$$

/Cabe destacar

Cabe destacar que este método permite obtener Z"i en forma totalmente mecánica; basta fijar el origen de trabajo para completar todos los valores de Z"i.

En el siguiente ejemplo podrá apreciarse las ventajas de esta forma de cálculo.

Rentabilidad del Capital (por cientos)				Nº de Empresas		
Y'i-1		Y'i	Yi	ni	Z"i	Z"i ni
2	-	6	4	20	-3	-60
6	-	10	8	40	-2	-80
10	-	14	12	50	-1	-50
14	-	18	16 1/	90	0	0
18	-	22	20	60	1	60
22	-	26	24	40	2	80
				300		- 50

1/ Eligiendo O_t = 16

Aplicando la fórmula del segundo método abreviado se tiene:

$$\bar{Y} = Ot + C \sum_{i=1}^{m} \frac{Z''i\ ni}{n}$$

$$\bar{Y} = 16 - 4 \frac{50}{300} = 15{,}33$$

Obsérvese que una vez fijado el origen de trabajo las Z"i se colocan en sucesión decreciente para las marcas de clase menores que O_t y en sucesión creciente para las marcas de clase mayores.

2.- Mediana(Me) Se trata de otro estadígrafo de tendencia central de aplicación muy frecuente. Se define como aquel valor de la variable que supera a no más de la mitad de las observaciones y es superado por no más de la mitad de dichas observaciones. Es un estadígrafo menos sensible que la media aritmética ante valores extremos de la variable; pudiendo calcular este estadígrafo aún en variables de tipo ordinal, como se verá en un ejemplo. Cuando las observaciones no están agrupadas en forma de una tabla de distribución de frecuencias, su cómputo es en extremo sencillo. Basta disponer los valores en orden creciente y ubicar el valor central. Por ejemplo, supóngase que se tienen ordenados los siguientes valores de gastos en consumo de 7

/ familias (en

(en dólares por mes)

40 - 47 - 60 - 70 - 78 - 80 - 90

La mediana será 70 dólares ya que este valor supera a 3 observaciones (40, 47 y 60) que no son más que la mitad (la mitad es 3,5) y a su vez es superada por 3 observaciones (78, 80 y 90) que tampoco son más de la mitad.
Cuando el número de observaciones es par, existen dos valores centrales que satisfacen la definición de mediana. Si en el ejemplo anterior se agrega una familia adicional se tiene:

40 - 47 - 60 - 70 - 78 - 80 - 90 - 180

En este caso tanto 70 como 78 son valores medianos. La mitad de las observaciones es 4 y 70 supera a 3 que no son más de la mitad y es superado por 4 que tampoco es más de la mitad, es exactamente la mitad. Igual cosa ocurre con 78; para evitar ambiguedades se toma como mediana en estos casos el punto medio entre los dos valores medianos. En el ejemplo, la mediana definitiva sería 74 dólares. Obsérvese la poca sensibilidad de este estadígrafo a los valores extremos. Al agregar la 8a observación con un valor bastante más alto que el resto, la mediana ha experimentado apenas un ligero crecimiento, es más, aunque en vez de 180 aquel valor hubiese sido de 5.000, la mediana siempre habría sido 74. En cambio no ocurre lo mismo para la media aritmética que es sumamente sensible a ese tipo de valores extremos; intente el lector su cálculo en uno y otro ejemplo, y constatará una fuerte variabilidad.
Si los datos se agrupan en una tabla de frecuencias, el cálculo de la mediana implica computar previamente las frecuencias acumuladas.

a) Variable discreta. En este caso bastará con identificar la frecuencia acumulada que es inmediatamente mayor a la mitad de las observaciones. La mediana será aquel valor de la variable que corresponda a dicha frecuencia acumulada. Ejemplo:

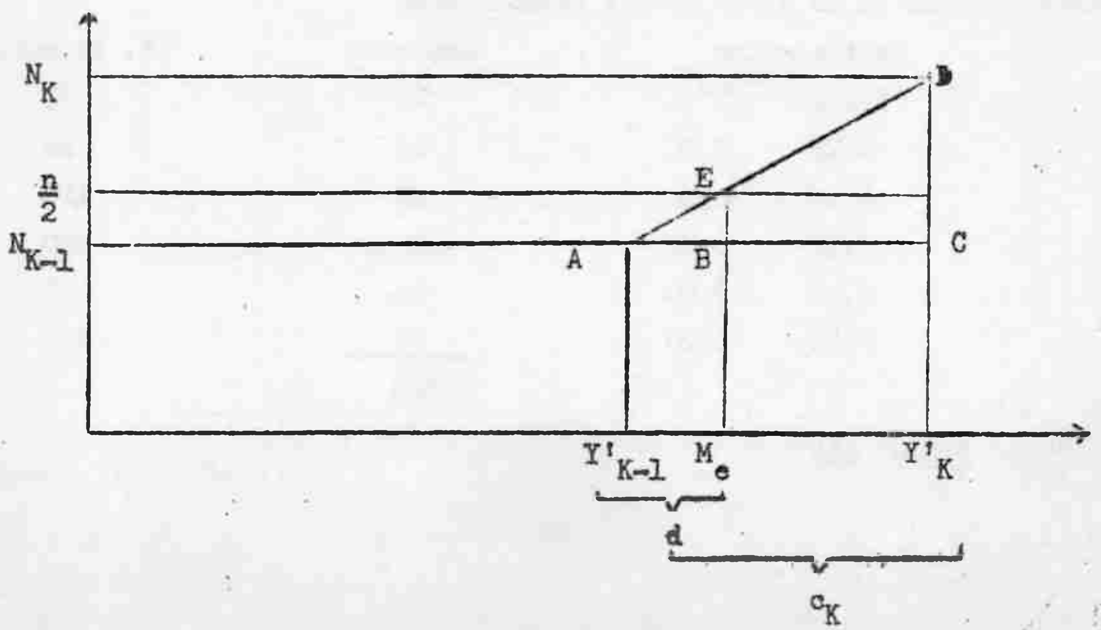
/Número de

Número de predios por persona Y_i	Número de propietarios n_i	Frecuencia acumulada N_i
1	200	200
2	160	360
3	150	510
4	100	610
5	80	690
6	40	730
7 y más	20	750
	750	

Siendo $\frac{n}{2} = 375$, la menor frecuencia acumulada que supera este valor es 510, que corresponde al valor 3 de la variable, siendo éste el valor mediano. Dicho valor supera a 360 observaciones que no es más de la mitad y es superado por 260 que tampoco es más de la mitad, satisfaciendo la definición de mediana.

b) Variable continua: En este caso el problema consiste en determinar un punto dentro del intervalo en que se halla comprendida la mediana. La identificación del intervalo en el cual se halla la mediana, es exactamente igual al caso de variable discreta: el intervalo será aquél que corresponda a la frecuencia acumulada inmediatamente superior a la mitad de las observaciones. Como se dijo, el asunto que preocupa es fijar un punto dentro de ese intervalo, que corresponde a la mediana. Para ello se adoptará el supuesto de que las observaciones se distribuyen linealmente dentro del mencionado intervalo.

Gráficamente se tiene lo siguiente:



Sea $Y'_{K-1} - Y'_K$ el intervalo donde se halla la mediana. Luego:

$$Me = Y'_{K-1} + d.$$

Será necesario encontrar una expresión para d en función de frecuencias que son los valores conocidos de que se dispone y por los cuales está determinado este estadígrafo.

Por semejanza de triángulos se tiene que:

$$\frac{AB}{BE} = \frac{AC}{CD}$$

Pero:

$$AB = d$$

$$BE = \frac{n}{2} - N_{K-1}.$$

$$AC = C_K$$

$$CD = N_K - N_{K-1} = n_K$$

Reemplazando se tiene:

$$\frac{d}{\frac{n}{2} - N_{K-1}} = \frac{C_K}{n_K}$$

de donde:

$$d = C_K \frac{\frac{n}{2} - N_{K-1}}{n_K}$$

Luego:

$$Me = Y'_{K-1} + C_K \frac{\left(\frac{n}{2} - N_{K-1}\right)}{n_K}$$

Ejemplo: La siguiente tabla muestra la distribución de los coeficientes producto-capital de 310 empresas industriales

Coeficientes		Empresas	F. Acumuladas
Y'i-1	Y'i	ni	Ni
0,15	0,20	40	40
0,20	0,30	80	120
0,30	0,42	100	220
0,42	0,50	60	280
0,50	0,70	30	310
		310	

$$\frac{n}{2} = \frac{310}{2} = 155$$

/ La menor

La menor frecuencia acumulada que supera a 155 es $N_K = 220$. Luego: $n_K = 100$, $Y'_{K-1} = 0,30$, $C_K = 0,12$ y $N_{K-1} = 120$

Reemplazando éstos valores en la fórmula:

$$Me = 0,30 + 0,12 \; \frac{155 - 120}{100} = 0,30 + 0,036 = 0,336$$

Como una extensión de este estadígrafo, es fácil ampliar el concepto a otros indicadores que dividen la masa de informaciones en otras proporciones y no sólo en mitades como lo hace la mediana.

Se tiene el caso de los cuartiles que dividen las observaciones en cuartas partes. Así, el primer cuartil Q1 es un valor de la variable que supera a no más de un cuarto de las observaciones, y es superado por no más de tres cuartos de ellas. Para identificar el intervalo en el que se halla Q_1 , habrá que determinar la frecuencia acumulada inmediatamente superior a $\frac{n}{4}$. El cálculo es similar al de la mediana:

$$Q_1 = Y'_{K-1} + C_K \frac{\left(\frac{n}{4} - N_{K-1}\right)}{n_K}$$

Para el tercer cuartil, sucede igual cosa

$$Q3 = Y'_{K-1} + C_K \frac{\left(3\frac{n}{4} - N_{K-1}\right)}{n_K}$$

Para identificar el intervalo que comprende a Q_3, habrá que constatar cuál es la frecuencia acumulada N_K, inmediatamente superior a $3\frac{n}{4}$. No es necesario detallar el segundo cuartil, porque corresponde con la mediana. En forma similar se pueden encontrar estadígrafos que dividan al total de observaciones en décimas partes (deciles) en centésimas partes (percentiles) etc. Las fórmulas correspondientes pueden deducirse por analogía con las de los cuartiles.

Así, el 7° Decil estará dado por:

$$D_7 = Y'_{K-1} + C_K \frac{\frac{7n}{10} - N_{K-1}}{n_K}$$

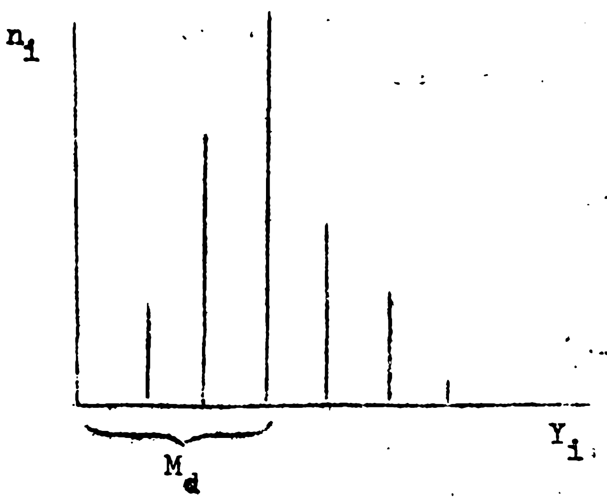
El 35 Percentil estará dado por

$$P35 = Y'_{K-1} + C_K \frac{\frac{35n}{100} - N_{K-1}}{n_K}$$

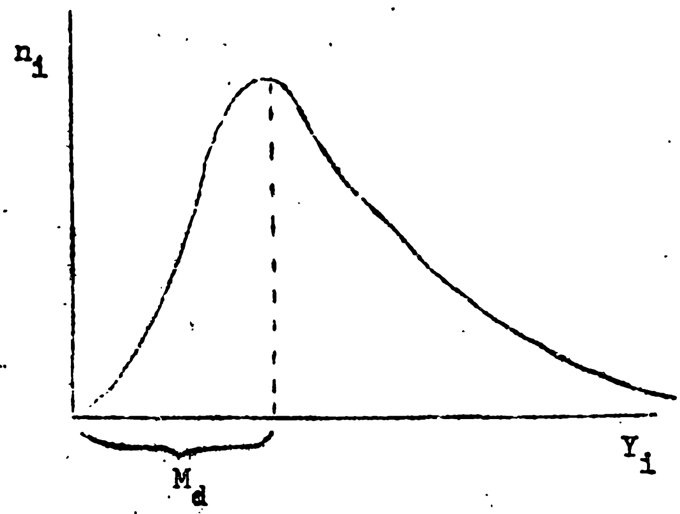
/ 3.- Valor Modal

3. Valor Modal.

Se trata de otro estadígrafo de tendencia central. Tiene un significado bastante preciso y es de extraordinaria utilidad, aunque inexplicablemente poco utilizado en estudios socioeconómicos. Se simbolizara por Md y se definirá como: aquel valor de la variable al que corresponde la máxima frecuencia. Es muy frecuente que se confunda el valor modal con la frecuencia máxima; recuérdese que es un valor de la variable y por lo mismo se le representa en el eje de las abscisas. Está dado por la frecuencia máxima, pero no se trata de una frecuencia.



Variable Discreta



Variable Continua

/a) Variable discreta

a) Variable discreta. Una vez que los datos están agrupados, es posible determinar inmediatamente el valor modal; bastará con fijar el valor de la variable que más se repite.

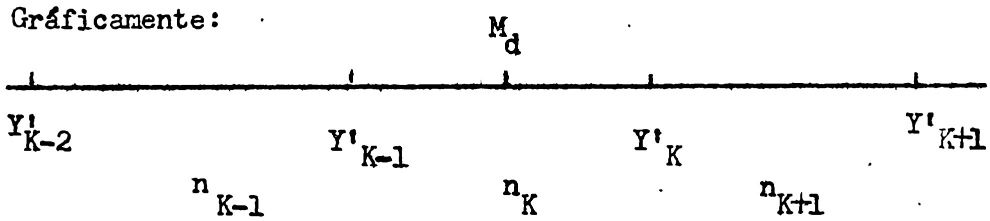
Ejemplo:

Número de cargas familiares	Número de familias
Y_i	n_i
0	80
1	120
2	210
3	380
4	180
5	60
6 o más	40
	1 070

La frecuencia máxima es 380 que corresponde al cuarto valor de la variable. El valor modal en consecuencia, es 3. Este valor modal será tanto más representativo mientras mayor es la frecuencia máxima. Se presentarán algunos casos donde el valor modal pierda significación; es el caso donde hay varios valores de las variables que tienen frecuencias similares. Es así como se califica a las distribuciones de unimodales, bimodales, multimodales, etc.

b) Variable continua. En la misma forma que cuando se calculaba la mediana, primero es necesario determinar el intervalo en el que se halla comprendido el valor modal. En este caso bastará ver cúal es el intervalo que tiene la frecuencia máxima. El siguiente paso es determinar un punto dentro de ese intervalo. Existen algunos criterios, un tanto arbitrarios, para deducir fórmulas del valor modal. Uno de tales criterios toma en cuenta la magnitud de las frecuencias de los intervalos contiguos (mayor y menor) al que comprende el valor modal. En otras palabras, dividirá al intervalo en partes inversamente proporcionales a las frecuencias de los intervalos contiguos.

Gráficamente:



$$\frac{Md - Y'_{K-1}}{Y'_K - Md} = \frac{n_{K+1}}{n_{K-1}}$$

La moda estará más cerca del intervalo contiguo que tenga mayor frecuencia.

Despejando Md de la relación anterior, se tiene:

$$Md\, n_{K-1} - Y'_{K-1}\, n_{K-1} = Y'_K\, n_{K+1} - Md\, n_{K+1}$$

pero,

$$Y'_K = Y'_{K-1} + c_K$$

$$Md \left[n_{K-1} + n_{K+1} \right] = Y'_{K-1} \left[n_{K+1} + n_{K-1} \right] + c_K\, n_{K+1}$$

$$Md = Y'_{K-1} + \frac{c_K\, n_{K+1}}{n_{K-1} + n_{K+1}}$$

Es necesario destacar que la deducción anterior, no toma en cuenta la amplitud de los intervalos contiguos. En caso de amplitudes muy diferentes, puede distorsionarse el valor de este estadígrafo.

Ejemplo:

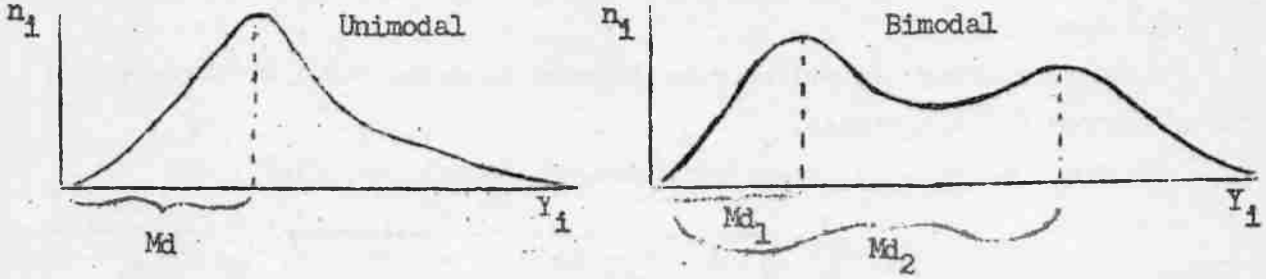
Ingresos de profesionales		Profesionales
Y'_{i-1}	Y'_i	n_i
0	20	25
20	40	45
40	60	80
60	80	60
80	100	40
100	120	15
120 y más		5

Inmediatamente se puede adelantar que la moda se encontrará en el tercer intervalo: 40 - 60. Aplicando la fórmula:

$$Md = 40 + \frac{20\ (60)}{45 + 60} = 40 + \frac{1\,200}{105} = 51{,}43$$

/Si se

Si se supone una gran cantidad de intervalos pequeños para una cierta distribución, gráficamente el valor modal estaría así representado:



Este estadígrafo al igual que la mediana, es susceptible de determinarse para variables cualitativas, ya que basta con constatar la frecuencia máxima. El siguiente ejemplo ilustra esta posibilidad:

Color de automóviles preferido por los clientes	Clientes encuestados
Blanco	18
Azul	22
Verde	40
Amarillo	25
Rojo	75

El "valor" modal en este caso es el rojo, ya que teniendo la máxima frecuencia, es el preferido por los clientes.

Al iniciar el estudio de este estadígrafo, se decía que inexplicablemente era poco utilizado en los análisis. Evidentemente que requiere, para su cómputo, más información que la media aritmética; ello podría explicar parcialmente su poco uso, pero aún disponiendo de información muchas veces se cree suficiente calcular por ejemplo un ingreso per capita y quedarse con un análisis parcial. Sin duda, para caracterizar adecuadamente una distribución, se requiere una serie de estadígrafos cuyas indicaciones se complementan. Por ejemplo, saber que dos países tienen ingresos por habitante de 150 y 200 dólares al año, puede permitir cierto tipo de conclusiones, pero saber además que los valores modales de los ingresos anuales por habitante, son de 140 y 130 dólares respectivamente, permite obtener conclusiones bastante más objetivas. Naturalmente que son necesarios muchos otros antecedentes e indicadores que se presentarán a continuación, para realizar análisis completos. Interesaba destacar la

/necesidad de

necesidad de buscar un conjunto de indicadores que permitieran analizar las distintas facetas de un fenómeno sometido a estudio.

4. Media Geométrica.

Este estadígrafo se define como la raiz de orden n del producto de los n valores de la variable.

Cuando los datos no están agrupados, su fórmula de cálculo es:

$$Mg = \sqrt[n]{x_1 \ x_2 \ x_3 \ldots\ldots\ldots\ldots\ldots x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Para fines prácticos es preferible calcular el logaritmo de la media geométrica y luego el antilogaritmo de ésta.

$$\log \ Mg = \frac{1}{n} \sum_{i=1}^{n} \log x_i$$

Si los datos aparecen agrupados, es decir, si las marcas de clase tienen frecuencias superiores a la unidad, se tendrá la siguiente fórmula.

$$Mg = \sqrt{Y_1 \ n_1 \ Y_2 \ n_2 \ Y_3 \ldots\ldots Y_m} = \sqrt{Y_1^{\,n_1} \ Y_2^{\,n_2} \ldots\ldots Y_m^{\,n_m}}$$

$$Mg = \sqrt[n]{\prod_{i=1}^{m} Y_i^{\ n_i}}$$

$$\log Mg = \frac{1}{n} \sum_{i=1}^{m} (\log Y_i) \ n_i$$

El estadígrafo que se estudia, aparte del inconveniente que representa la laboriosidad de su cálculo, tiene además la limitación de que los valores de la variable deben ser positivos para que sea susceptible de interpretación. Si algún valor de la variable es cero, la media geométrica será cero. Igualmente si existe algún valor negativo el estadígrafo toma un valor imaginario. Pese a estos inconvenientes, para cierto tipo de variables, en especial cronológicas, que sigan una tendencia exponencial, se hace indispensable su uso si se desea calcular valores intermedios, es decir, si se desea interpolar no linealmente. Por ejemplo, si cierta población en 1940 era de 2 millones y medio y en 1960 alcanza a los 4 millones, para calcular la población en 1955, sería indispensable el uso de la media geométrica, si se admite el crecimiento exponencial de la población a una tasa constante.

/ P_{1950}

$$P_{1950} = \sqrt{(2,5)\quad(4,0)} \doteq 3,15 \text{ millones}$$

$$P_{1955} = \sqrt{(3,15)\quad(4,0)} \doteq 3,55 \text{ millones}$$

5. **Media Armónica.** El último de los estadígrafos de tendencia central que se estudiará en el curso, se define como el recíproco de la media aritmética de los valores recíprocos de la variable.

Para datos no agrupados:

$$M_h = \frac{1}{\sum \frac{1}{x_i}/n} = \frac{n}{\sum \frac{1}{x_i}}$$

Para datos agrupados

$$M_h = \frac{n}{\sum \frac{n_i}{Y_i}}$$

Ejemplo: Un grupo de trabajadores construyen los primeros 120 metros de una avenida con una productividad de 12 metros diarios, en cambio los siguientes 120 metros lo hacen a razón de 18 metros por día. Se pide determinar la productividad diaria en todo el trabajo.

Si se decidiera calcular la media aritmética se tendría:

$$\bar{Y} = \frac{12 + 18}{2} = 15 \text{ metros diarios}$$

Por otra parte en los primeros 120 metros se demoran 10 días y los siguientes 120 metros lo hacen en 6,67 días; es decir, todo el trabajo lo harían en 16,67 días. Si la productividad diaria es de 15 metros, en los 16,67 días construirían un total de 250,05 metros, lo que es inconsistente ya que el trabajo total es solamente de 240 metros. Si se utiliza la media armónica en cambio:

$$M_h = \frac{2}{\frac{1}{12} + \frac{1}{18}} = \frac{72}{3 + 2} = \frac{72}{5} = 14,4 \text{ metros}$$

Trabajando con una productividad media de 14,4 metros por día, en los 16,67 días, se construirán 240 metros.

Como pudo advertirse, la media armónica se aplica en el caso en que se presenta una relación inversa entre las variables implícitas; en el caso del ejemplo se presenta una relación inversa entre la productividad y el tiempo.

$$e = p \times t$$

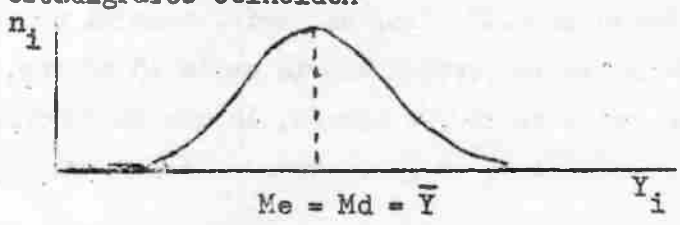
e: espacio p: productividad t: tiempo

$$p = e \cdot \frac{1}{t}$$

C - EVALUACION DE LOS ESTADIGRAFOS DE TENDENCIA CENTRAL

En más de una oportunidad se insistió sobre lo indispensable de contar con un conjunto de indicadores sobre la variable que se está estudiando. Los indicadores presentados, denominados en general como de tendencia central, tienen definiciones precisas; por ello muestran aspectos particulares del fenómeno que se estudia. Se trata de un conjunto de estadígrafos complementarios; las conclusiones a que en último término den lugar, deberán ser producto de la consideración simultánea de los valores que alcanzan dichos indicadores.

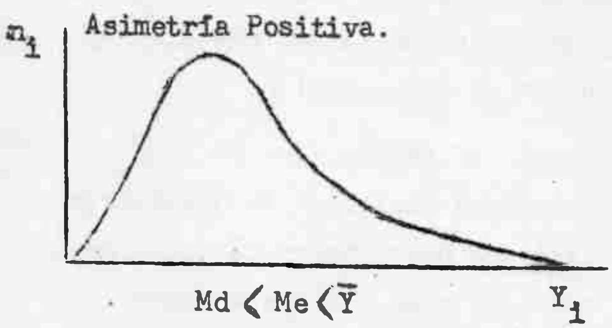
Al analizar la bondad de cada uno de estos indicadores, es preciso tener presente el volumen de observaciones tomadas en cuenta para su cálculo y las limitaciones que tiene cada uno de ellos. Siempre es conveniente complementar el análisis de las cifras, con una representación gráfica de la distribución de frecuencias de la variable. Interesa destacar la posición relativa de la media aritmética, la mediana y el valor modal. La posición relativa de estos estadígrafos, depende de la forma de la distribución. Así si la distribución es simétrica, es decir, si se observa perfecta simetría respecto de un eje central, los tres estadígrafos coinciden

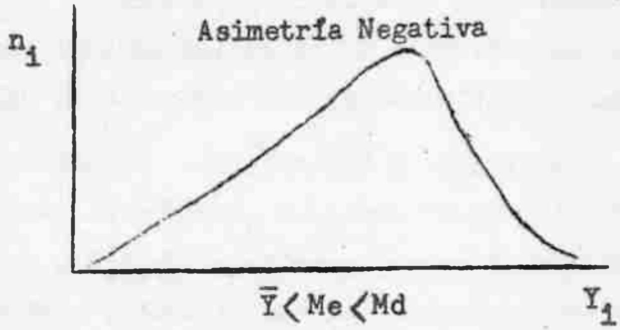


En el caso de distribuciones no simétricas; la posición relativa de los estadígrafos, depende del tipo de asimetría. Así, si la asimetría es positiva, es decir, si la distribución tiene su rama o manto más extendido hacia valores positivos de la variable, la moda será menor que la media aritmética. La mediana por el hecho de dividir la masa de observaciones en dos partes, quedará comprendida entre ambas. Si la asimetría es negativa, es decir cuando la distribución se extienda suavemente hacia valores negativos de la variable, la

/moda superará

moda superará a la media aritmética, permaneciendo la mediana por la misma razón dada en el otro caso, comprendida entre ambos indicadores. Gráficamente

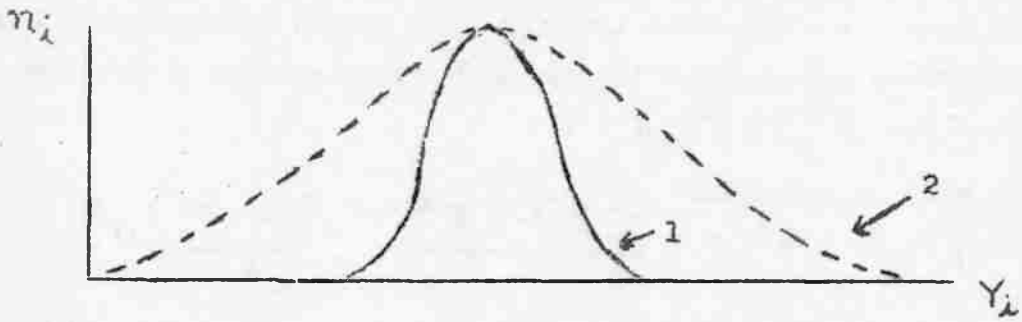




Recuérdese que la media aritmética es un estadígrafo muy sensible a valores extremos de la variable, de ahí que en un caso sea el mayor de los tres estadígrafos y en otro caso el menor de ellos. La moda, como el valor de la variable que más se repite, tiene en general una clara ubicación. Habría que agregar que su valor depende sobremanera de la amplitud de intervalo elegida y su representatividad sólo se garantiza cuando existe una clara concentración de frecuencias en un intervalo dado.

D. ESTADIGRAFOS DE DISPERSION

Una vez caracterizada la distribución a través de indicadores de tendencia central y en conocimiento del tipo de asimetría, interesa tener indicaciones acerca del grado de heterogeneidad con que la variable se distribuye en un conjunto de observaciones. Dos distribuciones pueden tener iguales estadígrafos de tendencia central, sin embargo pueden mostrar grados de dispersión diferentes, como puede observarse en el gráfico que a continuación se muestra.



Evidentemente en la primera distribución (linea continua) los valores aparecen más concentrados en torno al eje central, en tanto que en la otra aparecen mucho más dispersos. Si ambas distribuciones representaran ingresos de dos poblaciones, se concluiría que en la primera distribución los ingresos son más homogeneos, mientras que en la segunda se observaría gran disparidad

/entre ingresos

entre ingresos altos, medios y bajos.

Demás está destacar la importancia de contar con indicadores que pudieran mostrar este tipo de características en una distribución. Sobre todo en lo que se refiere a distribución de ingresos, de tanta actualidad hoy día, es indispensable contar con indicaciones adecuadas en este sentido.

1. Recorrido de la variable. Cuando se piensa en dispersión, lo primero que viene a la mente es el campo de recorrido de la variable: la diferencia entre el mayor y menor valor de ella. Si bien da una primera idea acerca de la heterogeneidad, tiene el inconveniente que sólo toma los dos valores extremos, sin considerar el conjunto de valores intermedios. Puede suceder que uno de los valores extremos, esté accidentalmente desplazado y no constituya por tanto un valór representativo; en este caso el recorrido sería exagerado y la dispersión aparecería distorsionada. Para iniciar el análisis es conveniente contemplar el recorrido, pero en ningún caso es suficiente.

2. Recorrido intercuartílico. Como una manera de obviar el inconveniente de los valores extremos que presentaba el estadígrafo anterior, se define un nuevo indicador, que toma en cuenta el recorrido entre el 1er. y 3er cuartil.

$$D_q = Q_3 - Q_1$$

Si bien es cierto que este indicador representa un adelanto respecto del anterior, no lo es menos que siempre toma dos valores de la variable, dejando de lado el resto, y en consecuencia la influencia de valores extremos puede, aunque en menor medida, originar algún tipo de deformación en cuanto al grado de dispersión.

3. Varianza. Se define este estadígrafo en virtud de la propiedad de la media aritmética que minimiza la suma de las desviaciones al cuadrado. Se simbolizará por σ^2 ó $V[Y_i]$

a) Para datos no agrupados

$$V[X_i] = \sigma^2 = \frac{\sum (X_i - \bar{X})^2}{n} = \frac{\sum X_i^2}{n} - \bar{X}^2$$

b) Para datos agrupados.

$$V[Y_i] = \sigma^2 = \frac{\sum (Y_i - \bar{Y})^2 n_i}{n} = \frac{\sum Y_i^2 n_i}{n} - \bar{Y}^2$$

/Si bien

Si bien la varianza no tiene un fin "per se" si no que se utiliza en materias que se presentarán posteriormente, da origen a un estadígrafo que sí tiene utilidad e interpretación práctica. Se trata de la desviación típica o estándar que se define como la raiz cuadrada positiva de la varianza.

$$\sigma = +\sqrt{\sigma^2}$$

Mientras más dipersa la variable, mayor será la magnitud de la desviación típica ya que mayores serán los desvíos respecto de la media aritmética, no habiendo posibilidad de compensación de desvíos por tratarse de suma de cuadrados. Este estadígrafo se expresa en las mismas unidades de la variable que se estudia, en tanto que la varianza se expresa en el cuadrado de la unidad de medida.

Como puede observarse en la fórmula, este indicador de dispersión toma en cuenta todos los valores de la variable con sus correspondientes frecuencias o pesos relativos, sin embargo, siempre es sensible a valores extremos. Por ello es conveniente, antes de calcular los estadígrafos, hacer un análisis previo de la tabl ade distribución de frecuencias, para visualizar la representatividad de valores extremos y sus posibles efectos en los valores de los estadígrafos.

c) Propiedades de la Varianza

i - Primera propiedad: La varianza de una variable a la cual se le ha sumado (restado) una constante, es igual a la varianza de la variable original.

$V[Y_i \pm K] = V[Y_i]$ Aplicando la definición de varianza a la variable $Y_i + K$, se tiene:

$$\sum_{i=i}^{m} \frac{(Y_i \pm K - M[Y_i \pm K])n_i^2}{n} = ./.$$

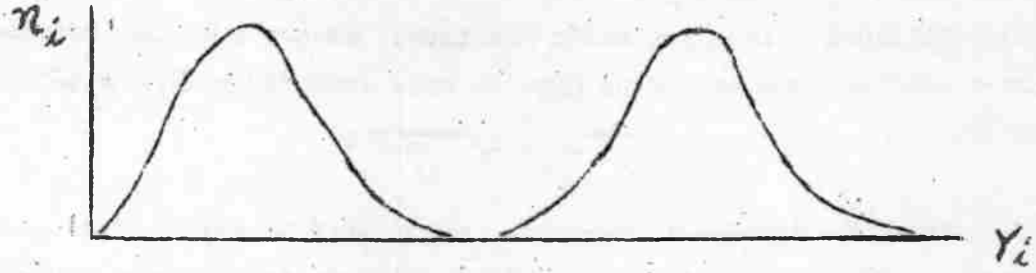
pero se vió que: $M[Y_i \pm K] = K \pm M[Y_i]$

$$\sum_{i=1}^{m} \frac{(Y_i \pm K - M[Y_i] \mp K)^2 n_i}{n} = V[Y_i]$$

$$\sum_{i=1}^{m} \frac{(Y_i - \bar{Y})^2 n_i}{n} = V[Y_i]$$

/ Gráficamente, las

Gráficsmente, las dos distribuciones tienen la misma varianza, pese a estar desplazadas en el eje de las abscisas.



ii) Segunda propiedad: La varianza del producto de una constante por una variable, es igual al cuadrado de la constante por la varianza de la variable.

$$V[K\ Y_i] = K^2\ V[Y_i]$$

$$\frac{\sum_{i=1}^{m} (K\ Y_i - K\overline{Y})^2 n_i}{} = ./.$$

$$\frac{\sum_{i=1}^{m} [K^2(Y_i - \overline{Y})]^2 n_i}{} = ./.$$

$$\frac{K^2 \sum_{i=1}^{m} (Y_i - \overline{Y})^2 n_i}{n} = ./.$$

$$K^2 V[Y_i] \qquad = K^2\ V[Y_i]$$

d) Componentes de la Varianza. En el caso en que un conjunto de datos haya sido dividido previamente en grandes categorías o estratos, es posible desglosar la varianza en dos componentes muy útiles para el análisis. Admítase que una masa de datos ha sido dividida en L estratos, cada estrato tendrá una media aritmética, una varianza y un número de observaciones que representa la importancia de cada uno de estos estratos. En este caso la variabilidad total puede deberse tanto a variabilidad dentro de cada estrato y a variabilidad entre los diferentes estratos.

i) Intervarianza: Estadígrafo que representa la variabilidad entre los estratos, se define como la varianza entre las medias de los estratos.

$$\sigma_b^2 = V[\overline{Y}_h] = \frac{\sum_{h=1}^{L} (\overline{Y}_h - \overline{Y})^2 nh}{n}$$

/donde: $\overline{Y}h$

donde: $\bar{Y}h$ es la media aritmética del estrato h

$\bar{Y}$ es la media aritmética general

nh es el número de observaciones o tamaño de cada estrato

ii) Intravarianza. Estadígrafo que representa la variabilidad dentro de los estratos. Se define como el promedio de las varianzas de los estratos.

$$\sigma_w^2 = M[\sigma h^2] = \frac{\sum_{h=1}^{L} \sigma h^2 \; nh}{n}$$

donde σh^2 es la varianza del estrato h, y n_h y n obedecen a las mismas definiciones del caso anterior.

Dado que los dos estadígrafos estudiados son partes componentes de la varianza, a continuación se presenta la correspondiente de mostración.

$$\sigma^2 = \sigma_b^2 + \sigma_w^2$$

$$\frac{\sum_{h=1}^{L} \sum_{t=1}^{nh} (Yh_i - \bar{Y})^2}{n} = \frac{\sum_{h=1}^{L} (\bar{Y}h - \bar{Y})^2 nh}{n} + \frac{\sum_{h=1}^{L} \sigma h^2 \; nh}{n}$$

pero $\sigma h^2 = \frac{\sum (Yhi - \bar{Y}h)^2}{nh}$; reemplazando.

$$\sum_{h=1}^{L} \sum_{i=1}^{nh} (Yhi - \bar{Y})^2 = \sum_{h=1}^{L} (\bar{Y}h - \bar{Y})^2 nh + \sum_{h=1}^{L} \sum_{t=1}^{nh} \frac{(Yhi - \bar{Y}h)^2 nh}{nh}$$

Elevando al cuadrado los correspondientes binomios

$$\sum_{h=1}^{L} \sum_{i=1}^{n_h} (Yhi^2 - 2\bar{Y}\, Yhi + \bar{Y})^2 = \sum_{h=1}^{L} (\bar{Y}h^2 - 2\bar{Y}\,\bar{Y}h + \bar{Y}^2)\, nh +$$

$$+ \sum_{h=1}^{L} \sum_{i-1}^{nh} (Yhi^2 - 2\,\bar{Y}h \;\; Yhi + \bar{Y}h^2)$$

Aplicando las propiedades de la sumatoria

$$\sum_{h=1}^{L} \left[\sum_{i=1}^{nh} (Yhi^2) - 2\bar{Y}\, n_h \bar{Y}h + nh\bar{Y}^2 \right] = \sum_{h=1}^{L} \bar{Y}h^2 nh - n\bar{Y}^2 +$$

$$+ \sum_{h=1}^{L} \left[\sum_{i=1}^{nh} (Yhi^2) - nh\bar{Y}h^2 \right]$$

$$/ \sum_{h=1}^{L} \sum_{i-1}^{nh}$$

$$\sum_{h=1}^{L} \sum_{i=1}^{nh} Yhi^2 - n\bar{Y}^2 = \sum_{h=1}^{L} \bar{Y}h^2 nh - n\bar{Y}^2 + \sum_{h=1}^{L} \sum_{i=1}^{nh} Yhi^2 - \sum_{h=1}^{L} nh\bar{Y}n^2$$

Simplificando términos queda:

$$\sum_{h=1}^{L} \sum_{i=1}^{nh} Yhi^2 = \sum_{h=1}^{L} \sum_{t=1}^{nh} Yhi^2$$

Luego $\sigma^2 = \sigma_b^2 + \sigma_w^2$.

Ejemplo: Piénsese en los sueldos y salarios pagados por una fábrica, que tienen una varianza de 137.440. Si las observaciones se clasifican en los estratos: Obreros, empleados administrativos, y técnicos, es posible analizar más a fondo la distribución de ingresos.

Las informaciones de que se dispone serían:

Estratos (h)	Tamaño Estrato nh	Media por Estrato $\bar{Y}h$	Varianza por Estrato σh^2
Obreros	300	400	160.000
Empleados Adm.	100	400	160.000
Técnicos	100	500	40.000

La media aritmética general será:

$$\bar{Y} = \frac{\sum_{h=1}^{L} \bar{Y}n \quad nh}{n} = \frac{2100}{5} = 420$$

$$\sigma_w^2 = \frac{\sum_{h=1}^{L} n^2 \, nh}{n} = \frac{680.000}{5} = 136.000$$

$$\sigma b^2 = \frac{\sum_{h=1}^{L} (\bar{Y}n - \bar{Y})^2 nh}{n} = \frac{7200}{5} = 1440$$

El cálculo de los anteriores estadígrafos permite concluir que la variabilidad se debe principalmente a heterogeneidad en las remuneraciones dentro de los estratos y no así a diferencias entre estratos; en otros términos las remuneraciones promedio de cada estrato, son bastante homogeneas ya que la intervarianza es pequeña, mientras que las remuneraciones dentro de cada estrato son muy heterogeneas ya que la intravarianza es bastante grande.

/e) Métodos abreviados

(Ver anexo 1)

e) Métodos abreviados de cálculo.

1) Primer método abreviado. Se trata de encontrar una fórmula que reduzca el volumen de operaciones; igual que en el caso de la media aritmética, la variable se expresará en términos de desvíos respecto de un origen de trabajo. Recuérdese que:

$$\sigma^2 = \frac{\sum_{i=1}^{m} (Yi - \bar{Y})^2 ni}{n} = \frac{\sum_{i=1}^{m} Z_i^2 \, ni}{n}$$

ya que $\qquad Zi = Yi - \bar{Y}$

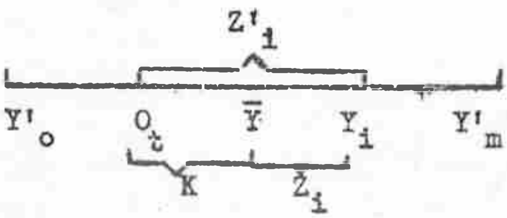
Si se desarrolla el cuadrado del binomio dentro de la sumatoria, se tiene:

$$\sigma^2 = \frac{\sum_{i=1}^{m} Yi^2 ni - 2\bar{Y} \sum_{i=1}^{m} Yi n_i + n \bar{Y}^2}{n}$$

pero $\sum_{i=1}^{m} Y_i \, n_i = n\bar{Y}$, luego:

$$\sigma^2 = \frac{\sum_{i=1}^{m} Yi^2 \, n_i}{n} - \bar{Y}^2$$

Observando el siguiente gráfico, resulta inmediata la deducción de la fórmula abreviada:



$Z'_i = Z_i + K$ Elevando al cuadrado

$Z'^2_i = Z_i^2 + 2K Z_i + K^2$ Ponderando por n_i

$Z'^2_i n_i = Z_i^2 \, n_i + 2K \, Z_i \, n_i + K^2 n_i$ Aplicando sumatoria

$$\sum_{i=1}^{m} Z'^2_i \, n_i = \sum_{i=1}^{m} Z_i^2 n_i + 2K \sum_{i=1}^{m} Z_i n_i + nK^2$$

Pero la primera propiedad de la media aritmética decia:

$$\sum_{i=1}^{m} Z_i \, n_i = 0 \quad , \quad \text{luego}$$

$$/ \sum_{i=1}^{m}$$

$$\sum_{i=1}^{m} Z'^{2}_{i} n_i = \sum Z_i^2 n_i + nK^2$$

pero $\sum_{i=1}^{m} Z_i^2 n_i = n\sigma^2$ y

$K = \sum_{i=1}^{m} Z'_i n_i$ (en el cálculo abreviado de la media aritmética)

Luego:

$$\sigma^2 = \frac{\sum_{i=1}^{m} Z'^{2}_{i} n_i}{n} - \left(\frac{\sum_{i=1}^{m} Z'_i n_i}{n}\right)^2$$

ii) Segundo mêtodo abreviado. Consiste, como se recordará, en expresar las desviaciones, en términos de unidades de intervalo (dividiendo por la amplitud del intervalo).

$$\frac{Y_i - O_t}{c} = \frac{Z'_i}{c} = Z''_i$$

es decir que $Z'_i = c\, Z''_i$ y reemplazando en la fórmula anterior se obtiene:

$$\sigma^2 = c^2 \left\{ \frac{\sum_{i=1}^{m} Z''^{2}_{i} n_i}{n} - \left(\frac{Z''_i n_i}{n}\right)^2 \right\}$$

En el siguiente ajemplo, se podrá comprobar el ahorro de tiempo que significa la aplicación de estas fórmulas. En el caso del primer método abreviado.

Impuestos por contribuyente			Número de contribuyentes			
Y'_i	Y'_i	Y_i	n_i	Z'_i	$Z'_i n_i$	$Z'^{2}_{i} n_i$
0	20	10	20	-40	-800	32.000
20	40	30	15	-20	-300	6.000
40	60	50 1/	10	0	0	0
60	80	70	8	20	160	3.200
80	100	90	5	40	200	8.000
			58		-740	49.200

1/ Ot = 50

/Reemplazando estos

Reemplazando estos valores en la fórmula, se obtiene:

$$\sigma^2 = \frac{49.200}{58} - \left(\frac{-740}{58}\right)^2 = 848,3 - 162,8 = 685,5$$

En el caso del segundo método abreviado se tiene:

Y'_{i-1}	Y'_i	Y_i	n_i	Z''_i	$Z''_i n_i$	$Z''^2_i n_i$
0	20	10	20	- 2	- 40	80
20	40	30	15	- 1	- 15	15
40	60	50 1/	10	0	0	0
60	80	70	8	1	8	8
80	100	90	5	2	10	20
			58		- 37	123

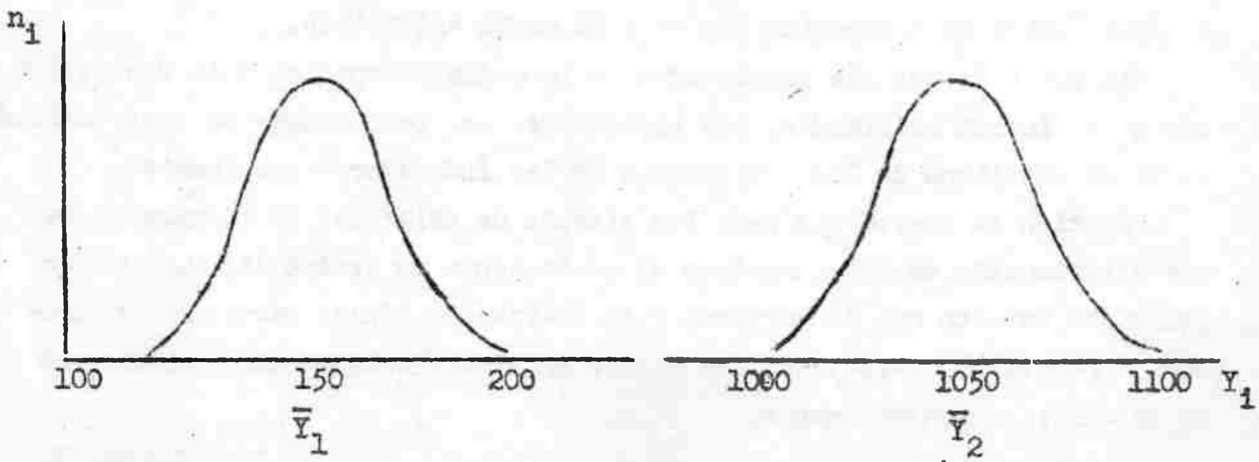
1/ Ot = 50

Aplicando la fórmula respectiva:

$$\sigma^2 = 20^2 \left\{ \frac{123}{58} - \left(-\frac{37}{58}\right)^2 \right\} = 400 \left\{ 2,12 - 0,407 \right\} = 685,5$$

4. Coeficiente de Variabilidad: Tanto la varianza como la desviación típica tienen el inconveniente de los estadígrafos absolutos, que en el caso de indicaciones sobre dispersión tiene mucha importancia, no tomar en cuenta la posición de la distribución. Estos estadígrafos, sobre todo en la comparación de distribuciones, pueden deformar las conclusiones.

Obsérvese las dos distribuciones que se tienen a continuación.



/Ambas distribuciones

Ambas distribuciones muestran la misma dispersión en torno a la media, es decir, tienen igual varianza y desviación típica; sin embargo en términos relativos, una distribución de ingresos donde el menor ingreso es 1.000 y el mayor es 1.100, es mucho más homogénea que otra distribución donde el menor ingreso es 100 y el mayor 200. En un caso la diferencia es 10 %, mientras que en el otro es de 100 %.

Surge la necesidad de disponer de un estadígrafo que tome en cuenta la posición de la distribución. Se define así el coeficiente de variabilidad, como la razón entre la desviación típica y la media aritmética.

$$C.V. = \frac{\sigma}{\bar{Y}}$$

En el ejemplo anterior, si ambas distribuciones tuvieran una desviación típica de 60 por ejemplo, los coeficientes de variabilidad serían:

$$CV_1 = \frac{\sigma_1}{\bar{Y}_1} = \frac{60}{150} = 0,4 = 40\ \%$$

$$CV_2 = \frac{\sigma_2}{\bar{Y}_2} = \frac{60}{1050} = 0,057 = 5,7\ \%$$

Estos estadígrafos permiten llegar a conclusiones más realistas y ciertas.

El coeficiente de variabilidad o desviación típica relativa como también se le llama, puede tomar valores tan grandes como se quiera, ya que no hay una relación de dependencia entre σ y $\bar{Y}$. Por otra parte en el caso de una distribución donde la media aritmetica fuera negativa no tiene sentido, para fines de calificar la dispersión, considerar el signo. Por ello este estadígrafo podría definirse como el valor absoluto del cociente entre la desviación típica y la media aritmetica.

En vista de que las propiedades de la media aritmetica y la desviación típica ya fueron analizadas, las propiedades del coeficiente de variabilidad serán el resultado de las propiedades de los indicadores componentes.

Si bien es cierto que para los efectos de calificar la dispersión de una distribución es más apropiado el coeficiente de variabilidad, no debe deducirse de esto que la varianza y la desviación típica carecen de utilidad. Por el contrario, son muy útiles en el tratamiento de materias que se estudiarán posteriormente.

/E. UTILIZACION DE

E. UTILIZACION DE INDICADORES DE LA PROGRAMACION

Es muy frecuente escuchar quejas respecto de la escasez de informaciones estadísticas básicas para los países en vías de desarrollo. Admitido este punto, es conveniente también reconocer que no siempre se hace un uso óptimo de tan escasas informaciones. Reconociendo como etapa primaria en un proceso de planificación la realización de diagnósticos, es necesario destacar la enorme utilidad de la estadística descriptiva en lo que se refiere a caracterización de fenómenos en un instante de tiempo. Por una parte, el diagnóstico implica disponer de una perspectiva historica referida a las variables estratégicas; por otra esa perspectiva historica debe complementarse con análisis cuantitativos en profundidad en períodos que presenten cambios de orientación y/o ritmo en las tendencias observadas, aparte de una cuantificación detallada en el momento "cero" de un plan. Es en esos puntos donde el instrumental de estadística descriptiva debe ser puesto a disposición del analista. Se ha insistido en la urgencia de contar con un juego de indicadores para las principales variables, cada indicador muestra una faceta de la variable estudiada, un conjunto de ellos permitirá una adecuada e integral calificación de las variables que interesan en el diagnóstico.

En general un conjunto de indicadores, permite realizar análisis de consistencia, llegando de ese modo a una primera evaluación acerca de la calidad y fidelidad de la información que se pretende utilizar.

II. NUMEROS INDICE

A. El problema general

Al definir un número índice como un indicador de la tendencia central de un conjunto de elementos que se expresa generalmente como porcentaje, saltan a la mente las limitaciones de todo estadígrafo.

El cotidiano uso que se hace de este instrumento obliga a plantear previamente algunas de sus limitaciones. Es muy útil no olvidar que se trata de un indicador que pretende reflejar el comportamiento de ciertas variables en forma aproximada, en consecuencia no se trata de una medición exacta. Por otra parte es necesario establecer que un número índice plantea una comparación ya sea en el tiempo o en el espacio, respecto de un punto de referencia denominado base del índice.

A medida que se vayan introduciendo los distintos conceptos que se refieren al conjunto de los números índice, se profundizarán estos planteamientos primarios, ya que la experiencia aconseja que el estudiante tenga ciertas reservas a medida que avanza en este campo, para evitar posteriormente una utilización indiscriminada y sin las aludidas reservas.

B. Clases de números índice

Fundamentalmente, dentro de la Estadística Económica, interesa disponer de indicadores sobre precios, cantidades y valor.

Un índice de precios será un indicador que refleje la variación de los precios de un conjunto de artículos entre dos puntos en el tiempo o en el espacio. Es el caso de un índice de costo de vida.

Un índice de cantidades, será un indicador que refleje la variación en las cantidades de un conjunto de productos entre dos puntos en el tiempo o en el espcaio. Por ejemplo un índice de producción industrial.

Por último un índice de valor indica la variación en el valor total de un conjunto de productos entre dos puntos del tiempo o el espacio. Ejemplo, índice de ventas totales comerciales.

C. Fórmulas de cálculo

Ocurre que cuando se trata de analizar la variación, por ejemplo, en el precio de un sólo artículo, no es necesario un indicador especial, basta con expresar la variación en términos porcentuales. Ejemplo:

/Período

Período	Precio del bien	Relación porcentual
1960	20	100
1961	25	125
1962	26	130
1963	30	150

Tomando como punto de referencia el precio del año 1960, y asignándole el valor 100, se calculan por una regla de tres simple, los índices correspondientes a los otros períodos. Así puede decirse que el precio del bien entre 1960 y 1965, ha experimentado un alza de 50 %. El cálculo de un índice, cualquiera que sea, referido a un sólo bien, no necesita pues de un estadígrado especial.

El problema de los números índice nace cuando se desea averiguar las variaciones de precios o cantidades de un conjunto de artículos. Considérese el siguiente ejemplo:

Bienes	Precios 1960	Precios 1964
A (metro)	10	15
B (quintal)	100	140
C (litro)	50	60
D (tonelada)	8	6
	168	221

Para resumir el incremento en los precios de este conjunto de artículos, una primera solución consistiría en sumar los precios en ambos períodos y establecer la variación porcentual entre ambos agregados. Se llegaría de esa manera a calcular un índice agregativo simple. Si se considera el año 1962 como base, se tiene:

1960 $\qquad$ 1964

100 $\qquad$ $131,5 = \frac{221}{168} \cdot 100$

Podría concluirse que el conjunto de estos precios ha variado en 31,5 % en el período. Sin embargo el método tiene dos serias limitaciones. Por una parte estará afectado por las unidades a que estén referidos los precios y por otra considera a los cuatro bienes, igualmente importantes, en circunstancias que el bien B puede ser trigo y el bien D comino; no se

/discrimina en

discrimina en este método en cuanto a la importancia relativa de cada artículo. En cuanto a las unidades de medida, considérese el ejemplo anterior, y supóngase que el precio del bien B se refiere al quintal de trigo; si se tuviera el precio del kilo el resultado del índice sería distinto:

Bienes	Precios 1960	Precios 1964
A (metro)	10	15
B (kilo)	1	1,4
C (litro)	50	60
D (tonelada)	8	6
	69	82,4

Asignándole siempre a 1960 el valor 100 como base del índice, se tiene que para 1964 el índice agregativo simple es ahora de 119,42 (100 . 82,4/69). Se llega a un resultado distinto sin que haya habido variación de precios alguna respecto del caso anterior, excepto que en ambos períodos se tomó un precio referido ahora a una unidad distinta.

En cuanto a este problema, es posible obviarlo calculando precios relativos. Se asigna a los precios de cada uno de los artículos en el período base, el valor 100 y por regla de tres simple, se calculan los correspondientes al período para el cual interesa conocer el índice. Si se observan los ejemplos anteriores, se tendrá:

Bienes	Precios 1960	Precios 1964
A	100	150
B	100	140
C	100	120
D	100	75

Obsérvese que en el bien B, sea cual fuere la unidad de medida, el incremento de su precio es de 40 %. Pero aunque en este método denominado de cifras relativas se obvia el problema de las unidades, todavía persisten problemas que exigen solución: en primer lugar respecto a la elección de un indicador de tendencia central. Se tienen los precios relativos para 1964, pero es necesario resumirlos por medio de un estadígrafo de posición: media aritmética, mediana etc. Todos ellos conducirán en

general a

general a resultados distintos. ¿Cuál de ellos tomar? Si bien en cada caso particular habrá un estadígrafo adecuado que satisfaga las necesidades de representabilidad que se tenga, en general se utiliza la media aritmética, principalmente, por la facilidad que implica su manejo algebraico. Es necesario cuidar de que no hayan valores extremos que distorcionen el estadígrafo. En el último ejemplo, si se toma la media aritmética, el índice para 1960 será de 100 y el de 1964 de 121,25, es decir, el conjunto de artículos habrá experimentado un alza de 21,25 % en el período. Obsérvese que la conclusión está referida al conjunto, ya que se trata de un promedio. Cada bien en particular acusa variaciones dispares en sus precios, incluso el bien D muestra decrecimiento.

El tomar cifras relativas en consecuencia, salva el inconveniente de las unidades de medida, pero aun subsiste el problema de la ponderación, ya que a cada artículo debe asignársele la importancia debida.

Tomando como punto de partida los precios relativos se ensayarán algunos criterios de ponderación, para obtener los índices más usuales en la investigación económica.

Si los precios relativos:

$$\frac{p_n}{p_o}$$

donde p_n es el precio en el período dado y p_o el precio en el período base, se ponderan por los valores del año base: p_oq_o, se obtiene la conocida fórmula de Laspeyres para precios (IPL), es decir:

$$IPL = \frac{\sum \frac{p_n}{p_o} \cdot p_oq_o}{\sum p_oq_o} = \frac{\sum p_nq_o}{\sum p_oq_o}$$

La sumatoria se extiende a todos los artículos considerados en el índice. Como en todo promedio aritmético, se divide por la suma de las ponderaciones.

Un índice de precios de Laspeyres debe interpretarse como el nivel que alcanzan los precios en un año dado, respecto de un año base al que se asigna el valor 100, considerando las cantidades del año base en ambos períodos. En otras palabras se trata de detectar la variación en los precios de una canasta de productos elegidos en el año base y que permanece inalterada en los períodos sucesivos.

/Este índice

Este índice, por lo tanto, tiene un significado bien concreto. El supuesto de que la canasta de productos permanezca en la realidad sin variaciones significativas, es otro problema. El analista tendrá que determinar si el supuesto se cumple o no y por consiguiente si es o no conveniente utilizar un índice de Laspeyres.

Por otra parte, si los precios relativos $\frac{p_n}{p_o}$, se ponderan por valores híbridos: $p_o q_n$, se tiene el índice de precios de Paasche (IPP), que también es de frecuente utilización:

$$IPP = \frac{\sum \frac{p_n}{p_o} \cdot p_o q_n}{\sum p_o q_n} = \frac{\sum p_n q_n}{\sum p_o q_n}$$

Obsérvese que ahora los precios están multiplicados por las cantidades del año que se calcula (q_n). Por este hecho un índice de precios de Paasche debe interpretarse como la variación de los precios de un conjunto de productos, suponiendo constantes las cantidades del año dado. En otros términos, la canasta de productos que se considera, es la del período que se calcula y se toma esta misma canasta para el año base.

Respecto de los índices de valor, por el significado simple que tienen, no requieren de deducciones especiales, ya que sencillamente es el cuociente entre los valores del año que se calcula y del año base.

$$IV = \frac{\sum p_n q_n}{\sum p_o q_o}$$

A continuación se considerará un ejemplo en que se calcularán los índices presentados:

	Año 0		Año 1		Año 2	
Artículos	p	q	p	q	p	q
A	10	4	12	5	20	3
B	4	3	4	3	5	3
C	8	10	8	12	7	15
D	20	2	30	2	40	3

p: precio q: cantidad

No hace falta el cálculo en el año 0, base del índice, ya que coinciden precios y cantidades: $p_n = p_o$ y $q_n = q_o$

/Para los

Para los índices de Laspeyres se tiene:

$$IPL\ (\text{año}1) = \frac{\sum p_1\ q_o}{\sum p_o\ q_o} = \frac{48 + 12 + 80 + 60}{40 + 12 + 80 + 40} = \frac{200}{172} = 116,3$$

$$IPL\ (\text{año } 2) = \frac{\sum p_2\ q_o}{\sum p_o\ q_o} = \frac{80 + 15 + 70 + 80}{40 + 12 + 80 + 40} = \frac{245}{172} = 142,4$$

Para los índices de Paasche, suponiendo siempre el año 0 como base del índice:

$$IPP\ (\text{Año } 1) = \frac{\sum p_1 q_1}{\sum p_o q_1} = \frac{60 + 12 + 96 + 60}{50 + 12 + 96 + 40} = \frac{228}{198} = 115,2$$

$$IPP\ (\text{Año } 2) = \frac{\sum p_2 q_2}{\sum p_o q_2} = \frac{60 + 15 + 105 + 120}{30 + 12 + 120 + 60} = \frac{300}{222} = 135,1$$

El índice de valor:

$$IV\ (\text{año}1) = \frac{\sum p_1 q_1}{\sum p_o q_o} = \frac{60 + 12 + 96 + 60}{40 + 12 + 80 + 40} = \frac{228}{172} = 115,1$$

$$IV\ (\text{año } 2) = \frac{\sum p_2 q_2}{\sum p_o q_o} = \frac{60 + 15 + 105 + 120}{40 + 12 + 80 + 40} = \frac{300}{172} = 174,4$$

En el siguiente cuadro puede apreciarse la diferencia en las indicaciones de uno y otro índice

	Indices		
Años	IPL	IPP	IV
0	100,0	100,0	100,0
1	116,3	115,2	115,1
2	142,4	135,1	174,4

Puede apreciarse que el IPL, muestra mayor crecimiento que el IPP. El primero se considera constante la canasta de productos del año base, en cambio en el segundo, se considera constante la canasta de productos del año que se calcula. Por lo tanto ambos índices indican la variación promedio de los precios bajo supuestos diferentes. Es muy frecuente confundir el significado de estos indicadores, porque no se tienen en cuenta los supuestos de uno y otro.

Con referencia a los índices de cantidad, es necesario hacer el mismo tipo de consideraciones ya que se presentan problemas similares: unidades de medida, ponderaciones, etc.

/Siguiendo el

Siguiendo el mismo criterio de los índices de precios, puede obtenerse el índice de cantidades de Laspeyres (IQL)

$$IQL = \frac{\sum \frac{q_n}{q_o} \cdot p_o q_o}{\sum q_o p_o} = \frac{\sum q_n p_o}{\sum q_o p_o}$$

El índice de cantidades de Paasche será (IQP)

$$IQP = \frac{\sum \frac{q_n}{q_o} q_o p_n}{\sum q_o p_n} = \frac{\sum q_n p_n}{\sum q_o p_n}$$

Mientras el IQL representa la variación en las cantidades suponiendo constantes los precios del año base, el IQP representa la variación de las cantidades suponiendo constantes los precios del año que se calcula. Nuevamente se insiste en la necesidad de no perder de vista estos supuestos, al interpretar un índice.

Las fórmulas presentadas son, como se dijo, las de uso más frecuente. Existen una cantidad extraordinaria de fórmulas de indices que se diferencian unas de otras según los factores de ponderación que se utilicen. Por razones de tiempo sólo se mostrarán las más conocidas.

Marshall - Edgeworth para precios

$$IPM = \frac{\sum p_n (q_o + q_n)}{\sum p_o (q_o + q_n)}$$

Indice de precios de Keyúes

$$IPK = \frac{\sum p_n (q_o \wedge q_n)}{\sum p_o (q_o \wedge q_n)}$$

donde el signo $\wedge$ es ínfimo y quiere decir que se tome la menor de las cantidades que están a sus costados.

La llamada fórmula "ideal" de Fisher que es la media geométrica de los índices de Laspeyres y de Paasche.

$$JPF = \sqrt{IPL \cdot IPP} = \sqrt{\frac{\sum p_n q_o}{\sum p_o q_o} \cdot \frac{\sum p_n q_n}{\sum p_o q_n}}$$

En estas últimas fórmulas, bastará reemplazar p por q, para obtener las fórmulas correspondientes a índices de cantidad.

/D. Pruebas sobre

D. <u>Pruebas sobre los números índice</u>

Irving Fisher, quien plantea la fórmula por él llamada ideal, propuso pruebas para calificar a los números índice.

1. Prueba de reversión de factores

La prueba se basa en un criterio de analogía: lo que es cierto para un producto, deberá ser cierto para un conjunto de ellos. Así para cualquier artículo.

precio x cantidad = valor

Esta prueba, como puede verse a continuación, no cumplen las fórmulas de Laspeyres y Paasche

IPL x IQL $\neq$ IV , en efecto

$$\frac{\sum p_n q_0}{\sum p_0 q_0} \cdot \frac{\sum q_n p_0}{\sum q_0 p_0} \neq \frac{\sum p_n q_n}{\sum p_0 q_0}$$

IPP x IQP $\neq$ IV

$$\frac{\sum p_n q_n}{\sum p_0 q_n} \times \frac{\sum q_n p_n}{\sum q_0 p_n} \neq \frac{\sum p_n q_n}{\sum p_0 q_0}$$

La fórmula de Fisher si cumple este test.

IPF x IQF = IV

$$\left(\frac{\sum \cancel{p_n q_0}}{\sum p_0 q_0} \cdot \frac{\sum p_n q_n}{\sum \cancel{p_0 q_n}}\right)^{\frac{1}{2}} \left(\frac{\sum \cancel{q_n p_0}}{\sum q_0 p_0} \cdot \frac{\sum q_n p_n}{\sum \cancel{q_0 p_n}}\right)^{1/2} =$$

$$= \frac{\sum p_n q_n}{\sum p_0 q_0}$$

Simplificando términos semejantes se tiene:

$$\frac{\sum p_n q_n}{\sum p_0 q_0} = \frac{\sum p_n q_n}{\sum p_0 q_0}$$

Sin embargo, esta prueba, también se satisface con una combinación de índices de precios de Laspeyres y cantidades de Paasche o viceversa, como se comprueba a continuación:

IPL x IQP = IV

$$\frac{\sum \cancel{p_n q_0}}{\sum p_0 q_0} \cdot \frac{\sum q_n p_n}{\sum \cancel{q_0 p_n}} = \frac{\sum p_n q_n}{\sum p_0 q_0}$$

IPP x IQL = IV

$$\left(\frac{\sum p_n q_n}{} \right.$$

$$\frac{\sum p_n q_n}{\sum p_o q_n} \cdot \frac{\sum q_n p_o}{\sum q_o p_o} = \frac{\sum p_n q_n}{\sum p_o q_o}$$

Estas dos últimas relaciones vale la pena tenerlas presente, por que se utilizan frecuentemente.

2. Prueba de reversión temporal

Nuevamente el criterio de analogía, si el precio de un producto es en el período a de 40 y en el período b de 50, en el primer período se constata que el precio es el 80 % del que se da en el período b, y en éste el 125 % del precio del período a. Lógicamente el producto de estos porcentajes debe dar 1, es decir:

$$\frac{p_n}{p_o} \times \frac{p_o}{p_n} = 1$$

Esta prueba, no la cumplen las fórmulas de Laspeyres y Paasche. En efecto:

$$IPL_{b,a} \times IPL_{a,b} \neq 1$$

donde el primer subíndice indica el período que se calcula y el segundo indica el período base.

$$\frac{\sum p_b q_a}{\sum p_a q_a} \cdot \frac{\sum p_a q_b}{\sum p_b q_b} \neq 1$$

$$IPP_{b,a} \times IPP_{a,b} \neq 1$$

$$\frac{\sum p_b q_b}{\sum p_a q_b} \times \frac{\sum p_a q_a}{\sum p_b q_a} \neq 1$$

La fórmula de Fisher, si cumple la prueba

$$IPF_{b,a} \times IPF_{a,b} = 1$$

$$\left(\frac{\cancel{\sum p_b q_a}}{\cancel{\sum p_a q_a}} \cdot \frac{\cancel{\sum p_b q_b}}{\cancel{\sum p_a q_b}}\right)^{\frac{1}{2}} \left(\frac{\cancel{\sum p_a q_b}}{\cancel{\sum p_b q_b}} \times \frac{\cancel{\sum p_a q_a}}{\cancel{\sum p_b q_a}}\right)^{\frac{1}{2}} = 1$$

3. Prueba circular

Si el precio de un producto es de 10 en el primer período de 12 en el segundo y de 18 en el tercero, se constata que en el segundo

/período el

período el precio es 120 % del precio en el primero y en el tercer período 150 % del precio que se da en el segundo. En consecuencia el precio del tercer período es 180 % del que se da en el primero:
(120 %) (150 %) = 180 %

La prueba circular no la cumple ninguna de las fórmulas analizadas. Suponiendo tres períodos para IPL se tiene:

$$IPL_{3,2}\ IPL_{2,1} \neq IPL_{3,1}$$

donde el primer subíndice indica el período que se calcula y el segundo el período base.

$$\frac{\sum p_3q_2}{\sum p_2q_2} \cdot \frac{\sum p_2q_1}{\sum p_1q_1} \neq \frac{\sum p_3q_1}{\sum p_1q_1}$$

Esta relación sólo se cumpliría en el caso en que $q_1 = q_2 = q_3$. Sin embargo, la relación planteada se cumple aproximadamente, cuando no existen diferencias significativas en las cantidades en los distintos períodos.

El estudiante podrá comprobar que la prueba circular no es cumplida, por la fórmula de Paasche ni por la de Fisher, siguiendo el mismo proceso de la comprobación para la fórmula de Laspeyres.

Respecto de estas pruebas, es conveniente aclarar que el hecho de que hayan índices que no las cumplan no es razón para dejarlos de lado. Interesa más el significado concreto del índice, teniendo sus alcances y limitaciones. Es así como la fórmula ideal de Fisher, cumpliendo con las pruebas por él planteadas, no es susceptible de clara interpretación, ya que se trata de la combinación de dos índices que por separado adquieren cabal significado pero al combinarlos presentan dificultades en su calificación.

/E. Base de

E. Base de un número índice

Al definir un número índice se ha recalcado que se trata de una comparación de dos puntos en el tiempo o en el espacio. El punto respecto del cual se hace la comparación, recibe el nombre de base de un índice y se le asigna el valor 100, para analizar las variaciones porcentuales. Respecto de la elección del período base hay que tener siempre presente el objetivo que se persigue con el índice. En general se dice que el período base debe ser un período normal. Cabe preguntarse qué se entiende por normalidad en estos casos, cuando en los países en desarrollo los cambios se están sucediendo con mucha frecuencia y la anormalidad es un denominador común. Tal vez sería más sensato al definir el período base pensar en un período en el que no existan accidentes o cambios violentos. Será necesario cambiar la base del índice cuando los supuestos planteados pierdan validez a medida que pasa el tiempo. Es el caso de los índices de costo de vida, donde es necesario cambiar la base toda vez que la estructura de consumo presente cambios significativos respecto de la que se daba en el período base.

Sobre este mismo asunto, es necesario distinguir dos tipos de base: base fija y base variable. Los índices de base fija son aquellos que mantienen como base un período fijo de referencia, en tanto que los índices de base variable son aquellos que tienen como base el período inmediatamente anterior. Teniendo un índice de base fija es posible calcular el correspondiente de base variable y viceversa; los resultados en general diferirán de los que se obtendrían a partir de los datos originales ya que las fórmulas usuales no cumplen con la prueba circular.

Ejemplo: Supóngase que el índice de Laspeyres para los precios de los materiales de construcción sea el siguiente:

	Indice base 1960 = 100
1960	100
1961	110
1962	120
1963	144
1964	180
1965	190

/El correspondiente

El correspondiente índice de base variable sería:

	Indice de Base variable
1960	—
1961	110,0
1962	109,1
1963	120,0
1964	125,0
1965	105,5

Otra operación que es muy usual respecto de los índices de base fija es la del empalme. Se trata, como dice su nombre, de empalmar índices con base distinta. Obsérvese el siguiente ejemplo, en que se tiene un índice para el período 1956-1959 con base en 1956 y otro índice de 1959 a 1962 con base en 1959 y se pretende tener una serie para todo el período 1956-1962

Años	Indice base 1956 = 100	Indice base 1959 = 100
1956	100	
1957	120	
1958	150	
1959	180	100
1960		110
1961		132
1962		150

Mediante una sencilla regla de tres, puede completarse cualquiera de las dos series, para tener el movimiento del índice en todo el período.

Años	Indice base 1956 = 100	Indice base 1959 = 100
1956	100,0	55,6
1957	120,0	66,7
1958	150,0	83,3
1959	180,0	100,0
1960	198,0	110,0
1961	237,6	132,0
1962	270,0	150,0

/Debe advertirse

Debe advertirse que este tipo de empalmes, significa tan sólo una aproximación que puede ser muy defectuosa dependiendo de la similitud de las bases y de sus supuestos. En todo caso siempre es posible recurrir a estos empalmes en tanto se tenga conciencia de sus limitaciones.

F. Utilización de los números índice

Un número índice indica la evolución de precios, cantidades y valores, para un conjunto de productos. Prestan en consecuencia la utilidad inmediata de reflejar la tendencia de los cambios y ritmos de los conceptos señalados. Ese sólo hecho ya justifica su cómputo y su periódica utilización en la investigación socio económica. Sin embargo prestan además otros servicios sobre los que es conveniente hacer algunos comentarios.

1. La deflactación: Las alteraciones en los sistemas y niveles de precios que se presentan dentro de la actividad económica, originan dificultades en la comparación de valores monetarios que corresponden a períodos de tiempo distanciados. No es mucho lo que se puede deducir de la comparación de valores nominales, es decir, de valores expresados en unidades monetarias de distinto poder adquisitivo.

   Para poder llegar a conclusiones válidas acerca del comportamiento de una variable que represente "valor", será necesario expresar los montos monetarios nominales, en unidades homogéneas. La transformación aludida recibe el nombre de deflactación, y con esta operación se pretende eliminar, exclusivamente, el efecto de alteraciones en los precios.

   El proceso de la deflactación exige disponer de un índice deflactor, es decir, de un indicador que proporcione una pauta de las alteraciones en los precios que dicen relación con la variable que se pretende deflactar. Es de fundamental importancia recordar que no existe un índice deflactor único; cada variable, estrictamente, debería tener un deflactor adecuado. La disponibilidad de sólo un reducido número de índices de precios, determina que éstos sean utilizados indiscriminadamente para una serie de propósitos. Aunque este hecho puede adolecer de errores conceptuales, muchas veces se justifica el procedimiento, porque interesa conocer un orden de magnitud antes que un valor exacto, siempre que se tenga conciencia de las limitaciones del método. En todo caso, aun dentro de las escasas disponibilidades de índices de

/precios, es

precios, es posible llegar a soluciones aceptables ya sea eligiendo racionalmente un índice de precios como deflactor o combinando y ponderando en forma adecuada dos o más indices. Este último proceso, si bien puede no conducir a soluciones ideales, al menos puede representar una disminución de las posibles distorsiones.

Antes de tener idea de este asunto de la deflactación, cuando se desea transformar unidades monetarias heterogéneas (unidades de cada período) en unidades monetarias homogéneas (unidades del período base) y permitir de este modo la comparación en el tiempo, lo primero que se piensa, es expresar los montos monetarios nominales en unidades de moneda extranjera de valor <u>más o menos</u> estable, dólares, libras, etc. Sobre este procedimiento caben algunas objeciones. Los Gobiernos tienen herramientas que les permiten fijar los tipos de cambio con las monedas extranjeras en forma arbitraria, arbitraria para los fines que aquí se comenta, y no representan generalmente las alteraciones en los niveles de precios.

En Chile hay una experiencia reciente; en el transcurso de menos de dos años, la unidad monetaria chilena ha llegado a sufrir una devaluación enorme, (de E°1,053 en noviembre de 1962 a E°3,400 en abril de 1963, por dólar norteamericano tipo de cambio corredor). Si un valor dado en escudos se expresa en dólares en ambas fechas, se concluiría que entre los dos períodos mencionados dicho valor se reduciría a la tercera parte. Si bien esto puede ser cierto para una persona que va a gastar sus ingresos en EE.UU., no lo es para quien efectúa sus desembolsos en Chile, porque en ese período, ningún índice de precios propiamente tal ha sufrido un aumento de 200 por ciento.

Por otra parte, una segunda objeción a este procedimiento, es el hecho que aun las economías más estables pueden sufrir cierto grado de inflación.

Los dos argumentos presentados descalifican, en general, la conversión a unidades monetarias extranjeras como una alternativa de la deflactación.

La mecánica de la deflactación implica decidir los montos monetarios nominales, por el índice de precios elegido como deflactor adecuado.

La razón por la que debe dividirse, estriba en la siguiente regla de tres. Si en el año n existen un valor nominal VN_n y un índice de precios IP_n, cuál sería este valor expresado en unidades monetarias

/de igual

de igual poder adquisitivo que las del año base? En otros términos, cuál sería este valor si el índice de precios no hubiera variado? El planteamiento queda así reducido:

$$\begin{array}{ll} IP_n & VN_n \\ 100 & x \end{array}$$

$$x = \frac{VN_n}{IP_n} \cdot 100 = \text{Valor real}$$

Desde otro punto de vista, se justifica la deflactación pensando en los componentes de un valor: precio por cantidad.

$$\text{Indice de precios x cantidad} = \text{Valor (100)}$$

$$\text{Cantidad} = \frac{\text{Valor (100)}}{\text{Indice de precios}} = \text{Valor real}$$

Ahora bien, la evolución de las cantidades en el tiempo esta libre, por así decirlo, de influencias monetarias directas, mostrando la evolución física, real, de una serie, que es justamente lo que se pretende con la deflactación.

Los valores reales así obtenidos, vienen expreados en unidades monetarias que tienen un poder adquisitivo del año que es la base del índice deflactor.

Con referencia a este último planteamiento, es necesario distinguir dos tipos de base. Se llamará base propiamente tal al período que corresponde al diseño del índice, donde se diseña la muestra, se establecen las ponderaciones, etc. Por otra parte, se utilizará la expresión "base aritmética" para referirse a cualquier período, que por transformación lineal, se le haya asignado el valor 100.

Los cambios aritméticos de base implican hasta cierto punto una arbitrariedad que puede originar algún tipo de perturbaciones en las comparaciones. Sus efectos serán tanto más fuertes, cuanto más distante sea la base propiamente tal del índice deflactor. Cuando exista conciencia que los supuestos de la instrucción y diseño del índice siguen siendo válidos, los cambios aritméticos de base no introducirán deformaciones significativas. Por ello, antes de deflactar será necesario decidir de qué año serán las unidades

/monetarias en

monetarias en que interesa expresar los valores reales. Es recomendable, si se cumple la condición de constancia de los supuestos de la construcción del índice, expresar los valores nominales en unidades monetarias del año más reciente, lo que se consigue mediante un índice deflactor que tenga por base el último año en el sentido cronológico. La utilidad de la sugerencia anotada radica en el hecho de que el investigador tiene una visión reciente y tal gez objetiva del sistema de precios imperante, por lo que es probable que se facilite la obtención de conclusiones.
A continuación, y a vía de ejemplo, se presentan los resultados de un proceso de deflactación:

Años	Sueldo de un empleado (U,M. de c/año) A	Indice de precios al consumidor (base 1963 = 100) B	Sueldo Real (U.M. de 1963) $\frac{A}{B} \cdot 100$
1958	400	36,5	1.096
1959	480	50,6	946
1960	600	56,5	1.062
1961	680	60,9	1.117
1962	720	69,3	1.039
1963	900	100,0	900

La deflactación presentada implica haber elegido como deflactor, el indice de precios al consumidor. Tal vez, de los valores reales, no pueda deducirse categóricamente que el empleado haya sufrido un detrimento de esa magnitud en su poder de compra. Esto hubiera ocurrido si dicho empleado gastara todo su ingreso en la forma que lo hace el empleado típico elegido como padrón en el índice de precios al consumidor Si el empleado en cuestión sólo gasta en manuntención el 60 por ciento de sus ingresos y el 40 por ciento restante lo dedica, por ejemplo, a la construcción de una vivienda, la deflactación se realizaría en dos partes: la primera parte (60 %) se deflactaría por el índice de precios al consumidor y el saldo debería deflactarse por un índice de precios de insumos de la construcción; de esta manera se obtendría la expresión real de su poder de compra, en términos del uso que dará a sus ingresos. En un principio se planteaba la necesidad de deflactar cada "variable"

/por un

por un índice de precios adecuado, en aras de simplificar la exposición. Es necesario refinar un tanto la expresión y concluir que es el uso que se dará a un monto monetario, el determinante en la elección de un deflactor óptimo. Si bien esto queda claro cuando se piensa en gastos, no lo es tanto cuando se tiene otro tipo de montos monetarios. Piénsese por ejemplo en el Producto Bruto del Sector Agrícola; para deflactarlo es necesario precisar qué se pretende. Si se pretende obtener un indicador de la evolución física del producto sectorial, tendrá que deflactarse por un índice que represente las variaciones en los precios de los componentes de dicho producto. Si se deflactara por un índice de precios al consumidor se obtendría un poder de compra real, en términos de la canasta de productos que contempla el índice deflactor.
Cuando se ha realizado una deflactación, es necesario tener presente que los valores reales así obtenidos, son simplemente aproximaciones, que serán tanto mejores, cuanto más representativo sea el índice de los precios que se cancelarán con el monto monetario. Esto no quiere decir que sea indispensable "construir" índices si no se dispone de los más adecuados, lo que se pretende es recalcar la necesidad de selección de los índices disponibles, y en todo caso, el tener presente las limitaciones que acuse una deflactación obligada por un índice que no sea el mejor. Es responsabilidad de los organismos autorizados y de los principales usuarios, la elaboración de índices de precios de la actividad económica.
La deflactación, mecánicamente es un asunto trivial, pero la elección de deflactores adecuados requiere mucha atención. Véase por ejemplo lo que sucede con los índices de producción y ventas industriales en Chile 1/. Observando las cifras siguientes debe admitirse que no existe la concordancia que debiera existir, no encontrándose razones que expliquen la diferencia. Es muy probable que el desajuste radique en la calidad del deflactor utilizado, antes que en las variaciones de stocks.

1/ "Los índices de producción industrial manufacturera y ventas reales industriales. Un comentario acerca de su comportamiento". Cesar Molestina, Mayo de 1963.

/Período

Período	Indice de producción industrial Base 1959 = 100	Indice de ventas industriales reales Base 1959 = 100
1959	100,0	100,0
1960	103,5	98,3
1961	106,1	110,6
1962	113,8	127,7
1963		
Enero	108,4	120,9
Abril	122,2	129,8
Agosto	112,8	132,1

De la simple observación de estas series surgen las contradicciones anotadas. Si se admite que el índice de producción industrial es un indicador confiable, el índice de ventas sería el que adolece de defectos. Estos defectos pueden deberse a dos causas no excluyentes: que el índice de ventas nominales no sea lo suficientemente acertado, por errores inherentes o ajenos al muestreo o que el deflactor utilizado no tiene la relación apropiada con los precios de los bienes industriales, o una combinación de ambas causas.

2. El deflactor implícito del Producto Bruto. No es necesario citar la imperiosa necesidad de disponer de un sistema de contabilidad social expresada en valores constantes o reales. Respecto de esto se tratará principalmente la deflactación del Producto Bruto. Si se piensa en los valores agregados por las diferentes ramas de actividad económica, se admitirá que el valor agregado de cada rama está relacionado con índices de precios particulares. Si se desea llegar a expresiones "físicas", en cada rama de actividad debería deflactarse por un índice de precios relacionado directamente con dichas ramas de actividad. Así el valor agregado por el sector industrial, debería deflactarse por un índice de precios de bienes industriales, el valor agregado por el sector agropecuario, se deflactaría por un índice de precios de bienes agropecuarios. De esta manera se obtendrían valores reales en cada rama que representaría una aproximación a la evolución física de los producido por cada sector. Sumando los valores reales de todas las ramas de actividad, se tendría el Producto Bruto en términos reales. Comparando los Producto Brutos

/en valores

en valores nominales y reales se obtiene el llamado deflactor implícito del Producto, que no es otra cosa que un índice general promedio de los precios que rigen en la actividad económica. Como se demostrará luego, se trata de un promedio armónico de los deflactores sectoriales.

Sean: PN_{ij} el producto nominal del sector "i" en el año "j".

IP_{ij} el índice de precios del sector "i" en el año "j"

PR_{ij} el producto real del sector "i" en el año "j"

De la deflactación resulta:

$$PR_{ij} = \frac{PN_{ij}}{IP_{ij}} \cdot 100$$

Si se suman todos los productos reales por sector en un año cualquiera "j", se tiene el Producto Bruto real en el año "j".

$$PR_j = \sum_{i=1}^{L} PR_{ij} = \sum_{i=1}^{L} \frac{PN_{ij}}{IP_{ij}} \cdot 100$$

donde i = 1, 2 .... L representa el número de sectores considerados.

Por otra parte, se tiene que el Producto Bruto real PR_j, se obtiene deflactando el Producto Bruto Nominal PN_j; por el deflactor implícito DI_j, todos los conceptos referidos a un período j, es decir:

$$PR_j = \frac{PN_j}{DI_j} \cdot 100$$

Reemplazando en la igualdad anterior, se tiene:

$$\frac{PN_j}{DI_j} 100 = \sum_{i=1}^{L} \frac{PN_{ij}}{IP_{ij}} \cdot 100$$

Despejando DI_j,

$$DI_j = \frac{PN_j}{\sum_{i=1}^{L} \frac{PN_{ij}}{IP_{ij}}}$$

que justamente corresponde con la definición de media armónica de los deflactores sectoriales considerando como ponderaciones los productos nominales de cada sector ya que:

$/PN_j =$

$$PN_j = \sum_{t=1}^{L} PN_{ij}$$

A continuación se presentará un ejemplo para ilustrar el proceso de deflactación de los productos sectoriales.

Supóngase que la actividad económica ha sido dividida en cuatro sectores para los que se dispone de las siguientes informaciones:

Producto Nominal
(unidades monetarias corrientes)

Sectores	1960	1961	1962
Minería	200	300	400
Agricultura	300	350	400
Industria	200	250	300
Servicios	500	600	700
PRODUCTO BRUTO NOMINAL	1.200	1.500	1.800

Indices de precios (base 1960 = 100)

Productos mineros	100	130	150
Productos agrícolas	100	110	120
Productos industriales	100	120	130
Servicios	100	110	120

Deflactando el producto de cada sector, por el índice de precios correspondiente, se tendrá los productos reales de cada sector:

Producto Real
(unidades monetarias constantes de 1960)

Sectores	1960	1961	1962
Agricultura	200	230,8	266,7
Minería	300	318,2	333,3
Industria	200	208,3	230,8
Servicios	500	545,5	583,3
PRODUCTO BRUTO REAL	1.200	1.302,8	1.414,1

Para calcular el deflactor implícito, recuérdese que:

$$DI_j = \frac{PN_j}{PR_j} \cdot 100$$

/Deflactor implícito

	1960	1961	1962
Deflactor implícito	100	115,1	127,3

Como se demostró, estos valores equivalen a los promedios armónicos de los índices deflactores.

Sobre la deflactación del Producto Bruto es conveniente advertir que puede prestarse a manejos caprichosos, según sea la base de los índices deflactores. En otras palabras, el Producto Bruto Real, mostrará distintas tasas de crecimiento según la base de los deflactores.

Aparentemente esto no tendría por qué ocurrir, ya que un cambio aritmético de la base del índice deflactor no afectaría las variaciones reales en el Producto Bruto. Mediante un ejemplo que pretende exagerar la situación antes que representar un hecho real, se ilustra este planteamiento.

Para abreviar, supóngase que la actividad económica ha sido dividida en dos sectores: primario y secundario. Los datos son los siguientes:

Producto Bruto Nominal (u.m. corrientes)

	1956	1962
Sector primario	1.000	1.600
Sector secundario	1.000	1.800

Indices deflactores (base 1956 = 100)

Sector primario	100	110
sector secundario	100	300

Si se calcula el Producto Bruto real, en precios constantes de 1956, se tiene:

Producto Bruto Real (u.m. constantes)

	1956	1962
Sector primario	1.000	1.450
Sector secundario	1.000	600
	2.000	2.050

Entre ambos años, el crecimiento es 2,5 por ciento, si se valora el Producto en precios de 1956. Si aritméticamente se cambia la base de los deflactores y en consecuencia el Producto se valora a precios de 1962, se llega a una variación porcentual diametralmente opuesta.

/Los nuevos

Los nuevos índices deflactores (con base en 1962) serán los siguientes:

Indices deflactores (base 1962 = 100)	1956	1962
Sector primario	91,0	100
Sector secundario	33,3	100

Deflactando los Productos nominales, por los índices anteriores se tiene:

Producto Bruto Real (u.m. constantes)		
Sector primario	1.100	1.600
Sector secundario	3.000	1.800
	4.100	3.400

Puede observarse un decrecimiento de 17 por ciento en el mismo período. Nótese que para los sectores, considerados en forma aislada, no ocurre esta incompatibilidad, ya que sólo se presenta al comparar sumas de sectores que tienen crecimientos nominales distintos y cuyos respectivos deflactores también muestran variaciones desiguales. Los resultados serán tanto más contradictorios, en uno y otro caso, cuanto más distintos sean los índices deflactores y mayores sus variaciones. Lo que ocurre es que se está sumando valores que no son rigurosamente homogéneos, dados los cambios aritméticos de la base de los índices. Los resultados a que se ha llegado no son sorprendentes si se piensa en la limitación de los cambios de base aludidos. Por otra parte la estructura del Producto Bruto nominal es distinta en ambos períodos. En el primer caso los dos índices deflactores son iguales a 100 en 1956 en circunstancias que el Sector Primario aporta con el 50 por ciento al Producto Bruto y en el 50 por ciento el Sector Secundario. En el segundo caso los dos índices deflactores son iguales a 100 en 1962 y el Sector Primario aporta con un 47 por ciento al Producto y con 53 por ciento el sector secundario. Esta es la razón de que para los sectores considerados en forma aislada, no se presenten las inconsistencias anotadas, y sí se presenten al comparar sumas deflactadas por índices distintos. La inconsistencia será mayor mientras más distintas sean las variaciones de los deflactores.

3. La proyección sobre la base de índices de cantidad. Un procedimiento corrientemente utilizado para proyectar el Producto Bruto real por sectores, es sobre la base de índices de cantidad. Consiste en hacer variar el producto de un período dado en conformidad al índice de producción correspondiente. Así, si se sabe que para 1962, el producto generado por la Industria fué de 5.000 unidades monetarias de ese año y se dispone del índice de producción industrial para los años 1962 a 1965, se puede suponer que habrá correspondencia entre las variaciones de dicho Producto sectorial y del índice de cantidad.

Años	Producto Industrial (precios de 1962)	Indice de Produc. Ind. base 1962 = 100	Estimación del Producto Ind. (precios de 1962)
1962	5.000	100	5.000
1963	—	108	5.400
1964	—	120	6.000
1965	—	135	6.750

Este tipo de estimaciones, para períodos cortos de tiempo parece ser justificada. Lo que no hay que perder de vista es que el índice de producción muestra más bien las variaciones de la Producción Global antes que del Producto (V. Agregado); la metodología de estimación presentada tiene como supuesto la constancia en los coeficientes de insumo producto, constancia que puede no ser real en una época de tanto cambio tecnológico.

G. Indices de Comercio Exterior

En lo referente al intercambio de bienes y servicios con el resto del mundo, es indispensable cuantificar indicadores acerca de precios, cantidades, valores etc, relacionados con exportaciones e importaciones. El tipo de problemas que debe enfrentarse en estos casos, son los inherentes a los números índices, más algunas precauciones que es necesario tomar por circunstancias que atañen al comercio exterior.

1. Indices de precios. En general es necesario homogenizar los datos básicos, ya que no siempre están sujetos a criterios uniformes de valuación: precios CIF, FOB, precios expresados en monedas diferentes, valores nominales de retorno u otro tipo de valorización,

/en qué

en qué momento se considera que la importación o exportación está consumada: en tránsito, en aduana etc.

Para el cálculo de índices de precios, puede utilizarse fórmulas de Laspeyres y Paasche. Sin embargo, al utilizar la fórmula de Laspeyres si algún producto deja de transarse, implica suponer que su precio bajó a cero, lo que constituye un evidente falseamiento. Habrá que cuidar que sólo se tomen en cuenta en el cálculo de los productos que se transan tanto en el período base como en el período dado. Por otra parte, no es necesario tomar esta precaución, si se aplica la fórmula de Paasche, ya que ésta se elimina automáticamente; el producto que deja de transarse y no quedará comprendido en ninguno de los factores $p_n q_n$ ni $p_n q_o$. Por lo anterior, para índices de precios se acostumbra utilizar fórmulas de Paasche. En estadísticas de comercio exterior, se acostumbra designar a estos índices como índices de valor unitario, por el tipo de informaciones que se tienen. No es el precio y la cantidad de un artículo, si no el precio promedio o valor unitario que correspondería a un artículo proveniente de una partida de artículos no totalmente homogéneos. Así si la importación de 100 automóviles de distintas marcas representa un valor CIF de 400.000 unidades monetarias, el precio promedio por automovil es de 4.000 u.m. Es en este sentido que se prefiere designar a estos índices como índices de valor unitario en vez de precios, aunque metodológicamente no hay diferencias.

2. Indices de cantidad. Nuevamente aquí hay que hacer referencia a la denominación especial de índices de quantum que se les da en comercio exterior por razones similares a las anotadas en el punto anterior.

En este caso, si un producto deja de importarse o exportarse, quiere decir que la cantidad transada bajó a cero, lo que queda ahora bien reflejado en la fórmula de Laspeyres. Por esa razón se acostumbra utilizarla en el cálculo de índices de quantum de comercio exterior. Además, el utilizar ffomulas de Laspeyres o Paasche para quantum y valor unitario respectivamente garantiza el cumplimiento de la prueba de reversión de factores planteada por Fisher. En virtud de ello, basta con conocer el valor total de exportaciones o importaciones y uno cualquiera de los dos índices mencionados, para deducir inmediatamente el otro. Recuérdese que:

/IPP x

IPP x IQL = IV

La laboriosidad que representa el cálculo de estos índices, induce a la utilización de muestras, en vez de indagaciones totales.

En el caso de índices de precios se supone que los precios de los artículos incluídos en la muestra, representan los precios de los no incluídos. Naturalmente que será necesario calcular los errores de muestreo correspondientes para hacer un racional uso del indicador.

Por lo que toca a los índices de quantum, el hecho de que en todos los períodos se disponga del valor total de las importaciones, permite cuantificar la representatividad que tenga la muestra utilizada en cada período, mediante el cuociente.

$$\frac{\sum p_n q_n \text{ (para la muestra)}}{\sum p_n q_n \text{ (para el total)}} = \text{Representatividad}$$

Posteriormente si se desea calcular un índice de quantum de Laspeyres, se cuantifica $\sum q_n p_o$ en la muestra, y para proyectar al total, se divide por la representatividad (implica la aplicación de una regla de tres simple). De esa manera se tiene la estimación del numerador de la fórmula para el total poblacional. En cuanto al denominador no hay problema alguno, ya que en cada período se dispone del total de importaciones y exportaciones. Si se desea calcular un índice de quantum de Paasche, será necesario ajustar al 100 % la expresión del denominador de la fórmula, calculada con datos muestrales.

4. Algunos indicadores económicos

Se presentarán algunos indicadores de muy frecuente uso en la literatura económica. Si bien es cierto que la mayoría de ellos serán analizados con mucha más profundidad en el curso de Contabilidad Social, es conveniente examinarlos desde el puntos de vista estadístico.

1. Indice de la relación de términos del intercambio. Se define como el cociente entre el índice de precios de exportaciones y el índice de precios de importaciones, ambos referidos a la misma base. En consecuencia este estadígrafo indica la evolución en el tiempo de la relación de precios que se dió en el período base; en otros

/términos representa

términos representa variaciones en la capacidad de compra de un volumen de exportaciones. Da la respuesta a la siguiente interrogante en cuánto ha aumentado o disminuido el número de unidades que es necesario exportar, para financiar la importación del mismo volumen de productos que se importaba en el año base. Se trata de un concepto estrictamente relativo. No se puede concluir que la relación de intercambio sea buena o mala, sino mejor o peor que la que existía en el año base. El índice de la relación de precios de intercambio queda entonces definido como:

$$IRI = \frac{\text{Indice de precios de exportaciones}}{\text{Indice de precios de importaciones}} = \frac{IPX}{IPM}$$

2. Producto e Ingreso bruto. Cuando estos conceptos se consideran en valores constantes, ambas magnitudes no coinciden. El Producto es una medida del valor de los bienes y servicios debidos al esfuerzo productivo interno. El Ingreso, tiene que ver con el intercambio con el exterior, desde el momento que parte de la producción de un país se exporta y parte de los insumos y bienes de consumo deben adquirirse en el exterior. Por consiguiente estas transacciones con el exterior afectan, por la relación de precios de intercambio, la disponibilidad de bienes y servicios que satisfacen la demanda de la población. A medida que se deteriora la relación de intercambio, hay una transferencia al exterior de parte del esfuerzo productivo interno. En valores constantes, la diferencia entre ambos conceptos, recibe el nombre de efecto de la relación de precios del intercambio, luego:

Ingreso Bruto = Producto Bruto + Efecto de la relación de precios de intercambio.

En símbolos:

$$IB = PB + ERI$$

donde:

Efecto de la relación de precios del intercambio = poder de compra de exportaciones - quantum de exportaciones.

Es decir:

$$EFI = PEX - QX$$

Además:

/Poder de

Poder de compra de exportaciones = Quantum de exportaciones x x índice de la relación de intercambio

$$PEX = QX\ (IRI)$$

Dado que el producto del Quantum de exportaciones y el índice de precios de exportaciones es equivalente al valor corriente de las exportaciones, el Poder de compra también puede definirse como:

$$PEX = \frac{\text{Valor corriente de las exportaciones}}{\text{Indice de precios de importaciones}}$$

en otros términos, el valor nominal o corriente de las exportaciones queda deflactado por el índice de precios de las importaciones. Por otra parte el Quantum de las exportaciones es el valor de las exportaciones a precios constantes.

$$QX = \sum q_n p_o$$

donde la sumatoria se extiende a todos los rubros de exportación, es decir:

$$QX = \frac{\text{Valor corriente de exportaciones}}{\text{Indice de precios de exportaciones}}$$

Reemplazando las expresiones correspondientes en la fómula de EFI se tiene:

$$EFI = QX\ (IRI) - QX = QX\ [IRI - 100]$$

A continuación se cuantificarán estos conceptos mediante un ejemplo, para hacer resaltar su importancia. Supóngase que se tienen las siguientes informaciones:

	1960	1962	1964
a) Producto Bruto (u.m. corrientes)	1.000	1.200	1.500
b) Deflactor público (base 1960 = 100)	100	115	140
c) Indice de precios de exportaciones	100	110	110
d) Indice de precios de importaciones	100	120	130
e) Valor corriente de las exportaciones	200	240	280

Para calcular el EFI, se tiene

		1960	1962	1964
f) PB real (u.m. de 1960) (a/b)		1.000	1.043	1.071
g) IRI (base 1960 = 100) (c/d)		100	92	85
h) QX (u.m. de 1960) (e/c)		200	218	255
i) QX IRI = PEX	(h.g)	200	201	217
j) EFI	(i - h)	-	- 17	- 38
k) IB real	(f - j)	100	1.026	1.033

/3. Capacidad

3. Capacidad para importar. Cuando se piensa en t´rminos de valores corrientes, la capacidad para importar no es otra cosa que el valor total de las exportaciones, más el ingreso neto de capitales extranjeros.

La capacidad para importar a precios constantes estará dada por:

	Quantum de exportaciones	QX
+	Radicación de capitales extranjeros	RC
+	Efecto de la relación de precios del intercambio	EFI
	Capacidad total de pagos sobre el exterior	
-	Remesas de utilidades e intereses	
-	Salida de capitales extranjeros	
	Capacidad para importar	CPI

La capacidad para importar a precios constantes, también puede determinarse deflactando la capacidad para importar a precios corrientes por el índice de precios de importaciones. La capacidad para importar a precios constantes es:

$$CPI = QX + QX\,[IRI - 100] + RC\ (\text{neta, a precios corrientes})$$

$$CPI = QX\,[IRI] + \frac{RC\ (\text{neta a precios corrientes})}{IPM}$$

$$CPI = QX \cdot \frac{IPX}{IPM} + \frac{RC}{IPM}$$

$$CPI = \frac{\text{Valor cte. de las export.} + RC\ (\text{neta a precios ctes})}{\text{Indice de precios de importaciones}}$$

$$CPI = \frac{\text{Capacidad para importar a precios corrientes}}{\text{Indice de precios de importaciones}}$$

4. Tipo de cambio de paridad. Como una aplicación de números es conveniente tratar la forma de convertir monedas de distintos paises a unidades homogéneas con propósitos de comparación. Utilizar los tipos de cambio oficiales para ello, puede originar distorsiones serias, en la medida que exista más de un tipo de cambio (discriminación de áreas, cambiarias), o que el tipo de cambio unico esté sobre o subvaluado.

Una alternativa teórica consistiría en la determinación de una canasta de productos, resultando el tipo de cambio entre dos monedas, como la relación de valores que sería necesario gastar para

/adquirir dicha

adquirir dicha canasta. Este método tiene inconvenientes de tipo práctico; la determinación de los artículos que conformarían la canasta significa un serio problema, sobre todo si se piensa en la enorme variación de las preferencias de los consumidores en los distintos países.

Un método que puede ser utilizado con alguna ventaja, es la proyección de un tipo de cambio base, en un período donde no se haya advertido sobre o subvaluaciones monetarias, sin cambios múltiples y en general sin accidentes que descalifiquen a ese período base.

La aludida proyección se efectúa sobre la base de las modificaciones en la relación de precios de los dos países, cuyo tipo de cambio se pretende determinar.

Sean los países A y B, y se desea estimar el número de unidades monetarias de A, por unidad monetaria de B. El tipo de cambio de paridad en un año cualquiera n, estará dado por

$$\text{Tipo de cambio de paridad año n} = \text{Tipo de cambio año base} \cdot \frac{\text{Deflactor implícito de A}}{\text{Deflactor implícito de B}}$$

Ejemplo: Supóngase que el tipo de cambio base en el año 0, fué de 5 u.m. de A por 1 u.m. de B, siendo los deflactores implícitos los siguientes:

Año	D.I. país A	D.I. país B
0	100	100
1	120	110
2	150	120
3	200	140
4	300	150

El tipo de cambio de paridad para los años siguientes será:

Año	Relación de deflactores DI_A / DI_B	Tipo de cambio de paridad
0	100,0	5,00
1	109,1	5,45
2	125,0	6,25
3	142,9	7,14
4	200,0	10,00

/Evidentemente,

Evidentemente, el método da estimaciones, que serán tanto más confiables en la medida que el tipo de cambio base haya sido elegido adecuadamente, y los deflactores implícitos sean representativos de las variaciones de precios en ambos países.

5. Transferencias implícitas. El hecho que en la actividad económica se presenten transacciones intersectoriales en que cada sector tiene precios distintos, origina cierto tipo de transferencias de producto de un sector a otro, implícitas en las transacciones que efectúan cuanto se contabilizan a precios constantes.

Aquellos sectores cuyos precios crecen menos que el promedio general de precios representado por el deflactor implícito, están transfiriendo parte de su producto hacia aquellos sectores que tienen precios que crecen más que el promedio de precios.

Evidentemente que la suma algebraica de las transferencias será nula, por cuento la ganancia de unos sectores, tiene como contrapartida la pérdida de otros.

Se definen las transferencias implícitas para cada sector, como la diferencia entre el producto valorizado a precios del sector y ese mismo producto valorizado a precios promedios representado, como se dijo, por el deflactor implícito:

Transferencias implícitas = Producto a precios corrientes - Producto a precios constantes (Deflactor implícito)

en símbolos:

$$T\ Imp = p_n q_n - p_o q_n\ (DI)$$

donde n y o son el período que se calcula y el período base respectivamente.

$$\sum_{h=1}^{L} T\ Imp\ (h) = \sum p_n q_n - \sum p_o q_n\ (DI) = 0$$

Donde L es el número de sectores considerados, h = 1, 2, 3, ... L

$$\sum_{h=1}^{L} T\ Imp\ (h) = \text{Producto nominal} - \text{Producto real} \cdot \frac{\text{Producto nominal}}{\text{Producto real}}$$

$$= 0$$

/Ejemplo:

Ejemplo:

Sectores	Producto nominal año 0	Producto nominal año 1	Indice de cantidad base 0 = 100	Producto real	Indice de precios base 0 = 100
1	500	1.080	120	600	180
2	600	800	110	660	121
3	300	350	100	300	117
4	400	420	90	360	117
		2.650		1.920	

El deflactor implícito para el año 1 será:

$$DI = \frac{\text{Producto nominal año 1}}{\text{Producto real año 1}} = \frac{2.650}{1.920} \doteq 138$$

Las transferencias implícitas por sector:

Sectores	Producto nominal $p_1 q_1$	Producto real (DI) $p_0 q_1$ (DI)	Transferencias implícitas
1	1.080	828	+ 252
2	800	911	- 111
3	350	414	- 64
4	420	497	- 77
	2.650	2.650	-

I. Etapas de la construcción de números índices

Es interesante e ilustrativo seguir cada uno de los pasos para la confección de indicadores acerca de precios o cantidades. Los pasos que se detallarán estarán referidos al diseño de un índice de costo de vida por constituir un caso bastante general y complejo.

1. Objetivo del indice: Es fundamental calificar con toda precisión el objetivo del indicador. En el caso de un índice de costo de vida, es necesario determinar a quiénes estará referido ese índice: si a la población en su conjunto, a una ciudad, a profesionales, a obreros, a campesinos, etc. De esto dependerá qué tipo de artículos conformarán la muestra de productos componentes del índice. Es imprescindible no perder de vista el objetivo básico: los usos principales que tendrá el indicador.

2. Determinación de la estructura de consumo. Es conveniente, al calcular un índice de costo de vida, clasificar los gastos: alimentación, vestuario, habitación, varios, etc. De esta manera es posible calcular índices separadamente para cada componente. Este desglose, aparte de que permite realizar análisis de las variaciones de precios en forma más detallada es muy útil en la selección de deflactores adecuados. Estos índices desglosados son susceptibles de combinación, para la obtención del índice general. Las ponderaciones de cada componente se establecen mediante una muestra. Se selecciona una muestra, de unidades familiares en el caso del costo de vida, provenientes de la población para la cual se confecciona el índice por ejemplo, la población de obreros y empleados de una ciudad. Esta muestra, en lo posible debe ser aleatoria y de un tamaño que represente fidelidad y alta confianza. En cada una de las unidades elegidas en la muestra, se lleva un registro de los gastos en bienes y servicios por tipo de bien o servicio, donde se recopilen los precios pagados y las cantidades consumidas. Vale la pena llevar este registro durante un tiempo largo, por lo general de un año, de manera de captar las variaciones en los consumos en las diferentes estaciones. De esta manera es posible llegar a obtener ponderaciones por componentes y por tipos de bienes y servicios.

/Selección

3. Selección de artículos. En las anotaciones que se hacen en los registros o "Libretas de consumo" acerca de los distintos bienes que se consumen, en general aparecerá una gran cantidad de productos, muchos de ellos obedecerán a consumos esporádicos o accidentales. Se hace necesario seleccionar los productos más importantes, de consumo habitual y representativos de las preferencias de los consumidores. Para seleccionar estos productos y servicios, no es conveniente un método aleatorio, es preferible decidir qué productos serán considerados en el índice, tomando en cuenta el volumen del gasto en cada bien o servicio. En esa forma se llegará a una lista de 100 a 300 productos para los cuales se tiene tabulada su importancia relativa o ponderación.

4. Formas de valuación. Otra etapa importante en la confección de índices de precios, es la decisión acerca de los tipos de precios que se considerarán. Sabido es que los precios varían enormemente según la fuente donde se adquieren los productos; por otra parte es corriente que para artículos considerados de primera necesidad, hayan precios oficiales y precios reales muy distantes entre sí. En el caso de índices de costo de vida los precios debieran tomarse en las fuentes donde el consumidor adquiere los productos: almacenes, ferias, etc. Sobre el particular es necesario tomar decisiones previas en forma categórica.

5. Variaciones de calidad de los bienes y servicios. Otro aspecto fundamental es la especificación precisa, hasta donde sea posible, acerca de la calidad de los productos, de manera que se pueda controlar a través del tiempo la invariabilidad de sus características. Es muy frecuente, sobre todo en los artículos controlados, que se "incrementen" los precios reales, permaneciendo fijos los precios nominales, vía una disminución de la calidad de los productos.

6. Base del índice. En general cuando se planteó el tema se advirtió sobre la necesidad de elegir como base un período donde estuvieran ausentes las modificaciones y circunstancias violentas. Ahora hay que agregar que en el caso de índices de costo de vida, el período base debe ser modificado, toda vez que la estructura de

/consumo haya

consumo haya cambiado significativamente. En los tiempos actuales, donde las innovaciones tecnológicas cambian con extraordinaria rapidez, determinando cambios en las preferencias de los consumidores la base de un índice de costo de vida no debiera tener una antiguedad superior a los 6 a 8 años.

Sin embargo, en las fórmulas de cálculo, para evitar cambios de base muy seguidos que significan altos costos, se contempla la posibilidad de introducir nuevos artículos, toda vez que se detecten cambios en las preferencias de los consumidores.

7. Elección de los métodos de cálculo. En general es necesario tomar decisiones sobre la fórmula de cálculo. Corrientemente, cuando se trata de índices de costo de vida se acotumbra utilizar fórmulas, de Laspeyres porque implica considerar constantes las ponderaciones del año base. La utilización de una fórmula de Paasche significaría un esfuerzo muy grande, ya que en cada período (generalmente cada mes) habría que reponderar "hacia atrás", aparte de que sería necesario calcular estas ponderaciones a medida que pasa el tiempo.

La decisión acerca de qué fórmula se utilizará estará condicionada en primer lugar al objetivo que se persigue con el índice y a la factibilidad de su uso desde el punto de vista del costo y la oportunidad con que se entreguen los resultados.

En Chile, la Dirección de Estadística y Censos calcula el índice de costo de vida sobre la base de una modificación a la fórmula de Laspeyres.

$$I_i = I_{i-1}\left(\frac{\sum q_{\bullet}p_{i-1}\left(\frac{p_i}{p_{i-1}}\right)}{\sum q_{\bullet}p_{i-1}}\right)$$

Esta fórmula implica el cálculo del índice en un período i, sobre la base del índice en el período anterior, afectándolo por la variación que acusen los precios. La ffomula tiene la ventaja de que se pueden introducir nuevos artículos. en términos de sus especificaciones, o cambiar la fuente de información.

III ANALISIS DE REGRESION

A. METODO DE LOS MINIMOS CUADRADOS.

Probablemente uno de los temas estadísticos de mayor utilización en la planificación, es el referente al análisis de regresión y correlación.

Es de extraordinaria utilidad conocer la forma en que están relacionadas las variables que son objeto de análisis, es decir, la función matemática capaz de representar tal relación.

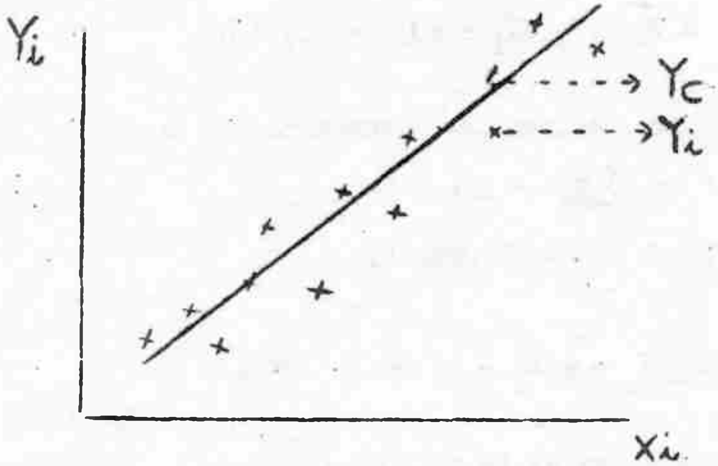
Conociendo tal función, es posible estimar el comportamiento de la variable objeto de estudio denominada variable dependiente o predictando, en términos de las variaciones de otra variable denominada independiente o predictor. De lo anterior se deduce que la regresión debe aplicarse a variables que tengan una relación lógica, es decir, que exista razonablemente dependencia entre las variables. Teóricamente, a cualquier par de variables puede encontrárseles una función matemática o ecuación de regresión que las relacione , pero sólo será de utilidad en el caso que haya una relación de causalidad entre dichas variables.

1.- Regresión simple: Se denomina de esta manera, a la metodología de obtención de ecuaciones, donde sólo intervienen dos variables: una dependiente o predictando y otra independiente o predictor. Una vez que por medio del análisis lógico se ha comprobado la existencia de una relación de causalidad entre las variables es necesario determinar cuál es la función matemática que representa adecuadamente la relación. Para ello es indispensable disponer de informaciones acerca de los valores que ha alcanzado cada una de las variables en distintos períodos, si se trata de un análisis histórico cronológico o en distintos lugares si se trata de un corte transversal en el tiempo. Con las informaciones obtenidas, que deben ser en un número suficiente para garantizar un buen ajuste, se construirá un gráfico y se podrá decidir si la función adecuada es una recta, una hipérbola, una parábola, etc.

/Una vez

Una vez que se ha decidido cuál es la función adecuada para el ajuste de regresión, es posible determinar los parámetors de la función elejida.

a) Línea recta: Si al representar los puntos en un gráfico, éstos representan satisfactoriamente una recta como en el cuadro siguiente:



es necesario calcular los parámetros o coeficientes de regresión de dicha recta.

$$Y_C = a\,x + b$$

para poder determinar los valores de a y b, se recurre al método de los mínimos cuadrados, que cumple la condición de minimizar la siguiente expresión:

$$\sum_{i=1}^{n} \left(Y_i - Y_C\right)^2$$

donde Y_i : es un valor observado

Y_C : es un valor calculado por la ecuación de regresión

n : es el número de observaciones

Si se reemplaza Y_C por $ax_i + b$ dentro de la sumatoria, es posible, derivando, encontrar los valores de los coeficientes de regresión a y b que satisfacen la condición. En efecto, llamemos Z a la

/expresión: Z=

expresión:

$$Z = \sum (Y_i - a\,X_i - b)^2$$

Se trata de derivar parcialmente respecto de cada uno de los parámetros

$$\frac{\partial Z}{\partial b} = 2 \sum (Y_i - aX_i - b)\,(-1) = 0$$

Aplicando las propiedades de la sumatoria se tiene:

$$\sum Y_i = a \sum X_i + n\,b$$

que es la primera ecuación normal.

$$\frac{\partial Z}{\partial a} = 2 \sum (Y_i - a\,X_i - b)\,(-X_i) = 0$$

Aplicando propiedades de la sumatoria

$$\sum Y_i X_i = a \sum X_i^2 + b \sum X_i$$

que es la segunda ecuación normal.

Obsérvese que se tienen dos ecuaciones normales y dos incógnitas. Se trata de un sistema de ecuaciones que permiten calcular los parámetros o coeficientes de regresión.

$$\left.\begin{array}{l} 1^{\underline{a}} \text{ Ecuación Normal: } \sum Y_i = a \sum X_i + nb \\ 2^{\underline{a}} \text{ Ecuación Normal: } \sum Y_i X_i = a \sum X_i^2 + b \sum X_i \end{array}\right\} \text{ Sistema}$$

Donde $\sum Y_i$, es la suma de los valores observados de la variable dependiente; $\sum X_i$, es la suma de los valores observados de la variable independiente y n es el número de observaciones. En este caso el sistema está formado por dos ecuaciones, porque sólo hay dos parámetros por determinar. El signo del coeficiente de regresión que corresponde con la pendiente de la recta (a), determina si la regresión es directa

/o inversa

o inversa. Si "a" es positivo, quiere decir que ante incrementos de la variable predictor, corresponde incrementos de la variable predictando. Si el signo de "a" es negativo, ante incrementos de la variable predictor habrá decrementos de la variable predictando y se dice que la regresión es inversa.

Hasta el momento se ha estado planteando una regresión de " Y en X ", es decir, considerando a Y como variable dependiente y a X como variable independiente. Donde se trataba de minimizar:

$$\sum_{i=1}^{n} \left(Y_i - Y_C\right)^2$$

Puede perfectamente plantearse una regresión de " X en Y ", donde lo que interesa minimizar es:

$$\sum_{i=1}^{n} \left(X_i - X_C\right)^2$$

siendo $X_C = a\,Y_i + b$

Las ecuaciones normales en este caso, por analogía serán:

$$\sum X_i = a \sum Y_i + nb$$

$$\sum X_i Y_i = a \sum Y_i^2 + b \sum Y_i$$

Téngase presente que los parámetros de la regresión de " Y en X ", serán distintos a los parámetros de la regresión de " X en Y ". Por ello suele distinguirse a estos parámetros de la siguiente manera:

a_{YX} : coeficiente de regresión de Y en X

a_{XY} : coeficiente de regresión de X en Y

En general, al analizar la relación de las variables cuya regresión se pretende determinar, se puede especificar cuál es la variable dependiente y cuál es la variable independiente. Una vez tomada la decisión, se denominará con Y_i a la variable dependiente o predictando y con X_i a la

/variable independiente

variable independiente O o predictor, para evitar confusiones.
A continuación se plantea un ejemplo que permitirá aclarar algunos aspectos que son difíciles de expresar literalmente. El lenguaje de los símbolos es claro y no permite malos entendidos ni interpretaciones equivocadas.

Ejemplo: En los últimos años las ventas de una empresa han crecido por razones de una intensa campaña de promoción de ventas. Las variables en cuestión han tenido el siguiente comportamiento en el tiempo.

	Ventas	Gasto en Propaganda
Años	Y_i	X_i
1958	100	10
1959	150	14
1960	200	21
1961	210	22
1962	300	28
1963	500	45
1964	600	55

Interesa determinar la función matemática o ecuación de regresión que relaciona a estas variables. Representando estos valores en un gráfico, se concluirá que la recta representa adecuadamente la relación de las variables. Para determinar los parámetros de la recta, se plantean las ecuaciones normales:

$$\sum Y_i = a\sum X_i + nb$$

$$\sum Y_i X_i = a\sum X_i^2 + b\sum X_i$$

Luego es necesario tabular los valores que interesan para reemplazar en estas ecuaciones normales. A continuación se procede con este punto.

$/Y_i \quad X_i$

Y_i	X_i	Y_iX_i	X_i^2
100	10	1000	100
150	14	2100	196
200	21	4200	441
210	22	4620	484
300	28	8400	784
500	45	22500	2025
600	55	33000	3025
2060	195	75820	7055

Las ecuaciones normales en valores serán:

$$2060 = 195\ a + 7\ b$$

$$75820 = 7055\ a + 195\ b$$

Resolviendo el sistema

$$a \doteq 11,4$$

$$b \doteq -26,1$$

La ecuación de ajuste queda en consecuencia así expresada:

$$Y_C = 11,4\ X_i - 26,1$$

Por medio de esta ecuación se puede determinar valores calculados de la variable dependiente ante cualquier valor de la variable independiente. Naturalmente que al realizar estimaciones, por ejemplo para calcular el probable volumen de ventas ante un desembolso en propaganda de 100 ($X_t = 100$) , hay que tener en cuenta el campo de validez de la regresión. No escapará a la atención del lector el hecho de que aumentos sucesivos de propaganda no siempre acarrearán mayores volúmenes de venta, porque puede llegar un momento de saturación del mercado u otra objeción por el estilo. Es necesario en consecuencia al realizar estimaciones, la verificación del cumplimiento de los supuestos implícitos en los datos disponibles. Por ello al realizar una estimación, es indispensable advertir que sólo tendrá validez, si es que se sigue manteniendo la tendencia de los puntos observados en el período histórico.

/b) Potencial

b) Potencial. Una función muy utilizada en proyecciones, por su flexibilidad, es la denominada función potencial o de elasticidad. Su expresión matemática es la siguiente:

$$Y = b\,x^{a}$$

Para determinar las ecuaciones normales se procede en forma similar al caso de la recta, linealizando previamente mediante la aplicación de logaritmos:

$$\log Y_C = \log b + a \log X_i$$

$$\log Y_C = b' + a \log X_i \qquad \text{donde } b' = \log b$$

En este caso se trata de minimizar la expresión:

$$Z = \sum_{i=1}^{n} (\log Y_i - \log Y_C)^2$$

es decir:

$$Z = \sum(\log Y_i - a \log X_i - b')^2$$

Derivando respecto de cada uno de los parámetros e igualando los resultados a cero, se obtendrán las dos ecuaciones normales.

$$\frac{\partial Z}{\partial b'} = 2 \sum(\log Y_i - a \log X_i - b')\,(-1) = 0$$

$$\frac{\partial Z}{\partial a} = 2 \sum(\log Y_i - a \log X_i - b')\,(-\log X_i) = 0$$

Aplicando las propiedades de la sumatoria a ambas derivadas, se tiene:

$$\sum \log Y_i = a \sum \log X_i - nb'$$

$$\sum \log Y_i \log X_i = a \sum(\log X_i)^2 - b' \sum \log X_i$$

que forman el sistema de dos ecuaciones normales que permitirán el cálculo de los dos parámetros. Evidentemente el método es un tanto

/laborioso cuando

laborioso cuando se tienen nuchas observaciones ya que es necesario trabajar en los logaritmos con al menos 5 decimales para evitar aproximaciones que pueden implicar serios desajustes.

c) Exponencial. Principalmente, cuando se desea calcular tasas de crecimiento tomando en cuenta todos los puntos observados en el período histórico se recurre a la función:

$$Y = a\, b^{t} \text{ donde } b = 1 + i$$

t: tiempo en períodos

Aplicando logaritmos a la anterior expresión:

$$\log Y_C = \log a + t_i \log b$$

Como en los casos anteriores interesa minimizar la expresión

$$Z = \sum_{i=1}^{n} (\log Y_i - \log Y_C)^2$$

$$Z = \sum (\log Y_i - \log a - t_i \log b)^2$$

$$\frac{\partial Z}{\partial \log a} = 2 \sum (\log Y_i - \log a - t_i \log b)(-1) = 0$$

$$\frac{\partial Z}{\partial \log b} = 2 \sum (\log Y_i - \log a - t_i \log b)(-t_i) = 0$$

Aplicando las propiedades de la sumatoria, se obtienen las dos ecuaciones normales.

$$\sum \log Y_i = n \log a + \log b \sum t_i$$

$$\sum t_i \log Y_i = \log a \sum t_i + \log b \sum t_i^2$$

El caso general de la función exponencial, es el cálculo de tasas de crecimiento cuando se considera el tiempo como variable independiente. Sin embargo, puede considerarse cualquier otra variable y ajustar la función sin hacer referencia a tasas de crecimiento.

d) Parábola: Esta conocida función , se ajusta en forma similar a los casos anteriores.

$$/\ Y = a\, x^2$$

$$Y = a\,x^2 + b\,x + C$$

Dado que la forma general contiene tres parámetros, será necesario determinar tres ecuaciones normales para determinar los valores de a, b, c. Estas tres ecuaciones normales provienen de la derivación parcial respecto de cada uno de dichos parámetros. Interesa minimizar la expresión:

$$Z = \sum_{i=1}^{n} (Y_i - Y_C)^2$$

$$Z = \sum (Y_i - a\,x_i^2 - b\,x_i - c)^2$$

Derivando respecto de a, b, y c, se tiene:

$$\frac{\partial Z}{\partial C} = 2 \sum (Y_i - a\,x_i^2 - b\,x_i - C)\,(-1) = 0$$

$$\frac{\partial Z}{\partial b} = 2 \sum (Y_i - a\,x_i^2 - b\,x_i - c)\,(- X_i) = 0$$

$$\frac{\partial Z}{\partial a} = 2 \sum (Y_i - a\,x_i^2 - b\,x_i - C)\,(- X_i^2) = 0$$

Aplicando las propiedades de la sumatoria,se tienen las siguientes ecuaciones normales:

1ª Ecuación normal: $\sum Y_i = a \sum X_i^2 + b \sum X_i + n\,C$

2ª Ecuación normal: $\sum Y_i X_i = a \sum X_i^3 + b \sum X_i^2 + C \sum X_i$

3ª Ecuación normal: $\sum Y_i X_i^2 = a \sum X_i^4 + b \sum X_i^3 + C \sum X_i^2$

En vista que en el período histórico, se tienen los valores de Y_i y X_i, es necesario tabular todas las sumatorias que aparecen en las ecuaciones normales. Resolviendo el sistema, se tiene determinado el valor de cada uno de los tres parámetros.

/e) Hipérbola

c) Hipérbola equilátera: Para el ajuste de algunas funciones de demanda, y por la propiedad que tiene de que cualquier punto de la función subtiende superficies iguales con los ejes de coordenadas, su aplicación es bastante frecuente. Su expresión matemática es:

$$Y_C = \frac{a}{x}$$

En vista que sólo tiene un parámetro, será necesario calcular una ecuación normal, minimizando la expresión:

$$Z = \sum (Y_i - Y_C)^2$$

$$Z = \sum (Y_i - \frac{a}{X_i})^2$$

Derivando respecto de a

$$\frac{\partial Z}{\partial a} = 2 \sum (Y_i - \frac{a}{X_i}) (- \frac{1}{X_i}) = 0$$

Aplicando las propiedades de la sumatoria se tiene:

Ecuación Normal $\sum Y_i/X_i = a \sum 1/X_i^2$

f) Otras funciones. Dentro de la investigación económica, a veces es preciso ajustar particulares funciones. La metodología de la obtención de ecuaciones normales es similar a los casos vistos. Por ejemplo la función:

$$\log Y_C = a\, x + b$$

Siempre se tratará de minimizar la expresión

$$Z = \sum_{i=1}^{n} (\log Y_i - \log Y_C)^2$$

$$Z = \sum (\log Y_i - a\, x_i - b)^2$$

/ donde

Donde:

$$\frac{\partial Z}{\partial b} = 2 \sum (\log Y_i - a\, x_i - b)\,(-1) = 0$$

$$\frac{\partial Z}{\partial a} = 2 \sum (\log Y_i - a\, x_i - b)\,(-X_i) = 0$$

Las ecuaciones normales serán:

$$\sum \log Y_i = a \sum X_i + n\, b$$

$$\sum X_i \log Y_i = a \sum X_i^2 + b \sum X_i$$

Resolviendo el sistema, es posible determinar el valor de los parámetros.

Siempre es conveniente seguir esta metodología, para funciones en que sus derivadas no compliquen demasiado las expresiones que aparecen en las ecuaciones normales.

2.- Regresión múltiple. Ocurre que a veces es necesario encontrar funciones en que se relacionen una variable dependiente y dos o más variables independientes, de ahí el calificativo de multiple. En este caso se adoptará una simbología especial, para designar cada una de las variables y parámetros:

X_1 : variable dependiente

X_2 , X_3 ... X_p : variables independientes

Así, si se trata de un caso de correlación múltiple donde se consideren dos variables independientes, la función se expresará de la siguiente manera:

$$X_{1.23} = a_{1.23} + b_{12.3}\, X_2 + b_{13.2}\, X_3$$

donde:

/$X_{1.23}$

$X_{1.23}$: indica la variable dependiente que se relaciona con las variables X_2 y X_3 , esa es la razón de los subíndices.

$a_{1.23}$: coeficiente de posición (término libre) del plano de regresión donde se consideran la variable dependiente X_1 y las variables independientes X_2 y X_3

$b_{12.3}$: coeficiente de regresión que multiplica a la variable X_2 , cuando además se considera la variable X_3.

$b_{13.2}$: coeficiente de regresión que multiplica a la variable X_3 , cuando además se considera la variable X_2

Es sencillo extender esta notación al caso en que se consideren 3 o más variables e independientes. En el caso de tres variables independientes (X_2, X_3, X_4) la función quedará así simbolizada:

$$X_{1.234} = a_{1.234} + b_{12.34} X_2 + b_{13.24} X_3 + b_{14.23} X_4$$

El estudiante, por analogía con el caso anterior puede interpretar cada uno de estos símbolos.

Cuando se desea ajustar una función de este tipo a una serie de datos, el método de los mínimos cuadrados implica hacer mínima la expresión.

$$\sum_{i=1}^{n} (X_{1\,i} - X_{1.23})^2$$

donde X_{1i} son los valores observados y $X_{1.23}$ son los valores calculados de la variable dependiente. Las ecuaciones normales en el caso de dos variables independientes, se obtiene minimizando la siguiente expresión:

$$Z = \sum (X_{1i} - a_{1.23} - b_{12.3} X_2 - b_{13.2} X_3)^2$$

Para ello, se deriva parcialmente respecto de cada uno de los pa' parámetros, igualando los resultados a cero.

$$\frac{\partial Z}{\partial a_{1.23}} =$$

$$\frac{\partial z}{\partial a_{1.23}} = 2\sum(X_1 - a_{1.23} - b_{12.3}X_2 - b_{13.2}X_3)^2(-1) = 0$$

$$\frac{\partial z}{\partial b_{12.3}} = 2\sum(X_1 - a_{1.23}X_2 - b_{13.2}X_3)^2(-X_2) = 0$$

$$\frac{\partial z}{\partial b_{13.2}} = 2\quad(X_1 - a_{1.23} - b_{12.3}X_2 - b_{13.2}X_3)^2(-X_3) = 0$$

Aplicando las propiedades de la sumatoria se tienen las siguientes tres ecuaciones normales que formarán el sistema para calcular el valor de cada uno de los tres parámetros.

$$\sum X_1 = b_{12.3}\sum X_2 + b_{13.2}\sum X_3 + n\, a_{1.23}$$

$$\sum X_1X_2 = b_{12.3}\sum X_2^2 + b_{13.2}\sum X_2X_3 + a_{1.23}\sum X_2$$

$$\sum X_1X_3 = b_{12.3}\sum X_2X_3 + b_{13.2}\sum X_3^2 + a_{1.23}\sum X_3$$

Tabulando los valores de las sumatorias que aparecen en el sistema, se podrá resolver para cada parámetro.

B. CONSIDERACIONES PRACTICAS

En páginas anteriores se ha detallado la metodología de obtención de ecuaciones normales por el método de los mínimos cuadrados, para el tipo de funciones de más frecuente uso en la planificación. Ahora se pretende establecer algunas consideraciones que es necesario tomar, desde el punto de vista práctico, al realizar los mencionados ajustes.

1. Respecto del tipo de función. Si se piensa que una de las principales aplicaciones de la regresión, es la proyección en el tiempo o en el espacio donde no se tengan valores de la variable en estudio, y donde no queda otra alternativa que conformarse con estimaciones provenientes de extrapolación de funciones ajustadas por regresión, deberá admitirse

/la necesidad

la necesidad de disponer de funciones sencillas que contengan un reducido número de variables y parámetros. Recuérdese que una función complicada, de muchas variables y parámetros, se parecerá más bien a una interpolación, a una función que se aproximará al mayor número de puntos observados. Para determinar tendencia no tiene sentido la interpolación. Recuérdese que para proyectar una variable dependiente, es necesario disponer de estimaciones para todas las variables independientes. Disponer de estimaciones para muchas variables independientes suele ser en extremo difícil y en todo caso existe alta probabilidad de cometer errores. En cambio una función sencilla, como las analizadas en páginas anteriores, es susceptible de representar cabalmente una tendencia de la relación de la variable dependiente con la o las variables independientes.

2. Respecto del número de observaciones. Un buen ajuste implica disponer de una cantidad significativa de puntos observados. El conjunto de puntos observados representa una muestra de la relación de las variables en el tiempo o el espacio. Mientras más grande esta muestra, es decir mientras mayor número de puntos se tenga, tendrá más representatividad y menor será la probabilidad de cometer errores. Cuando se está analizando una ecuación de regresión, una de las primeras cuestiones a aclarar debe ser el número de observaciones, para que con este antecedente se califique en parte, lo significativo de la regresión.

3. Respecto de la laboriosidad de cálculo. El estudiante comprobará a través de la realización de seminarios, la laboriosidad que representan los cálculos de regresión. En la práctica, cuando ya se tienen aclarada la parte conceptual, donde constituye una fuerte ayuda la realización de ejercicios en forma manual, es útil recurrir a los computadores electrónicos, ya que una vez entregadas las informaciones originales, en brevísimo tiempo es posible disponer de cálculos exactos, ya que los programas de regresión están previamente diseñados. Por otra parte respecto de la deducción de las ecuaciones normales, puede significar cierta demora su obtención sobre la base de las derivadas parciales. Existe una regla nemotécnica para hallar ecuaciones normales en funciones lineales

/respecto de

respecto de los parámetros. La regla es la siguiente:
Para la primera ecuación normal, multiplique la función a ajustar por el coeficiente del primer parámetro, y luego aplíquese el operador sumatoria a la función. Para la segunda ecuación, multiplíquese toda la función por el coeficiente del segundo parámetro y luego aplíquese el operador sumatoria. Así sucesivamente para todas las ecuaciones normales que se deba obtener.

Ejemplos: Se obtendrán las ecuaciones normales de la función:

$$\log Y = a \log X + \log b$$

Multiplicando ambos miembros de la ecuación por el coeficiente de log b que es uno y aplicando sumatoria, se tiene la primera ecuación normal.

$$\sum \log Y_i = a \sum \log X_i + n \log b$$

Multiplicando ambos miembros de la ecuación por log X que es el coeficiente del otro parámetro y aplicando sumatoria, se tiene:

$$\sum \log Y_i \log X_i = a \sum (\log X_i)^2 + \log b \sum \log X_i$$

que es la segunda ecuación normal. Comparando estas dos ecuaciones normales con las obtenidas por derivación parcial en la parte 1,b se concluye que son idénticas.

Si se quisiera obtener la ecuación normal de una recta que pasa por el origen, se tiene:

$Y = a\,x$ (recta que pasa por el origen ya que no tiene término libre; coeficiente de posición 0.)

Multiplicando por X_i que es el coeficiente del único parámetro y aplicando sumatoria, se tiene:

$$\sum Y_i X_i = a \sum X_i^2$$

Por el proceso de derivación, se llega al mismo resultado. En efecto:

$$Z = \sum_{i=1}^{n} (Y_i - Y_C)^2 = \sum (Y_i - aX_i)^2$$

$$/ \frac{\partial Z}{\partial a} = 2$$

$$\frac{\partial Z}{\partial a} = 2 \sum (Y_i - a X_i)(-X_i) = 0$$

$$\sum Y_i X_i = a \; X_i^2$$

En las páginas siguientes, se tratarán conceptos referentes al análisis de correlación, conceptos que permitirán cuantificar el grado de asociación entre las variables que se estudian y la validez de las proyecciones a través de las ecuaciones de regresión.