CONSEJO FEDERAL DE INVERSIONES

PUBLICA

Introducción A La Estadistica Económica

La presente publicación corresponde a las clases dictadas por el Prof. Enrique DIEULEFAIT en los cursos sobre Desarrollo Económico, organizados por el C. F. I.

Buenos Aires
1962

INTRODUCCION A LA ESTADISTICA ECONOMICA

La mayor parte de los problemas que afectan a la persona que debe tomar una determinación en el ámbito económico, requieren como elemento de información, a más del juicio subjetivo de valoración intangible, el apoyo de cifras numéricas correspondientes a observaciones o mediciones hechas sobre variables directamente vinculadas con los fenómenos bajo estudio.

Ha de ser el motivo central de estas líneas el presentar los primeros elementos de la metodología estadística tendientes a permitir la interpretación de una masa de información numérica y efectuar los análisis necesarios para lograr una primera interpretación del fenómeno sobre el que debe actuar.

Se verán así diversas técnicas, metodologías para el tratamiento de la información numérica cuyo valor principal consiste en proveernos de una sistemática uniforme para afrontar los problemas del cálculo numérico dentro del esquema mental o modelo matemático a que se apele para disponer de una interpretación de la evolución e implicaciones del fenómeno económico que se considere.

Los distintos Capítulos, Grandes Temas para el estudio de esta metodología son: DISTRIBUCIONES DE FRECUENCIAS - NUMEROS INDICES - SERIES CRONOLOGICAS - CORRELACION Y REGRESION, y permiten familiarizarse con las primeras herramientas necesarias para la sustitución de la aproximación subjetiva, por la valoración objetiva.

CAPITULO I

DISTRIBUCIONES DE FRECUENCIAS

Supongamos haber realizado una encuesta a los fines de caracterizar dentro de ciertas variables de interés económico a una población, interrogando a través de un procedimiento de selec - ción al azar y conforme a una metodología de selección, para asegurar la representatividad de la información, a 150 familias sobre características tales como: ocupación del jefe de la familia, extensión de la familia, edad del jefe de la familia e ingreso de la unidad familiar.

Estas cuatro características: ocupación, extensión, edad e ingreso, de las que resulta posible obtener información a través de la encuesta, son atributos de la población y, a través de los valores numéricos correspondientes a la encuesta, la caracterizan.

I - ATRIBUTOS CUALITATIVOS

Las declaraciones correspondientes a la "ocupación del Jefe de la familia", encuadradas dentro de una clasificación uniforme y conveniente a los fines de comparabilidad, pueden resumirse a través de cuadros del siguiente tipo:

OCUPACION DEL JEFE DE LA FAMILIA	FRECUENCIAS	PORCENTAJE
0. Trabajadores profesionales técnicos y similares	7	4,8
1. Trabajadores administrativos, directivos y oficinistas	24	16,1
2. Vendedores	10	6,8
3. Agricultores, pescadores, cazadores y madereros	10	6,4
4. Trabajadores en ocupaciones de minería y canteras.	8	5,1
5. Personal móvil de transportes	7	4,4
6. Artesanos, trabajadores en procesos de producción	46	30,5
7. Jornaleros	18	12,4
8. Trabajadores en servicios de protección.	1	0,7
9. Trabajadores en servicios	18	12,2
10. Ocupaciones no identificadas o no declaradas	1	0,6
TOTALES	150	100,0

Por regla general el tratamiento a que se someten las frecuencias correspondientes a los distintos atributos cualitativos se reduce al cálculo y expresión de porcentajes o frecuencias relativas, las que permiten una más fácil comparación entre poblaciones de distinta extensión al brindar una rápida información sobre las proporciones existentes en la población.

II - VARIABLES DISCONTINUAS

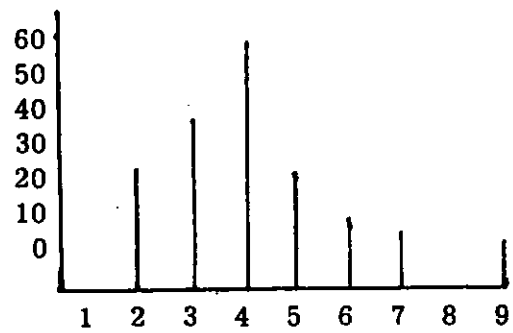
Las respuestas correspondientes al item "número de miembros de la familia" son ahora cuantitativas y se expresan según valores enteros de una variable.

Si ordenamos en un cuadro las respuestas correspondientes a las 150 familias consideradas, podríamos observar una distribución de este tipo.

Número de miembros de la familia	frecuencias	porcentajes
2	21	14%
3	36	24%
4	60	40%
5	20	13,3%
6	9	6%
7	3	2%
8	0	0%
9	1	0,6%
TOTALES	N = 150	100

GRAFICA DE BARRAS:

Denotando con x_i los valores de la variable "número de miembros de la familia", y con y_i, las correspondientes frecuencias, se tiene la gráfica de barras:



A fin de disponer de un par de medidas resumidas de nuestra distribución de frecuencias correspondientes a la variable discontínua x_i, introducimos la noción de "tendencia central" y "concentración".

Como medida de la tendencia central se emplea la media aritmética, $\bar{x}$ valor de la variable que corresponde al centro de gravedad de la distribución, y que responde a la siguiente definición:

$$\bar{x} = \frac{1}{N} \sum x_i \, y_i$$

en donde N representa la frecuencia total; x_i cada valor de la variable, e y_i las correspondientes frecuencias observadas.

Como medida de concentración se emplea la dispersión o desviación estandard, σ , medida basada en los cuadrados de los des - víos, $(x_i-\bar{x})^2$, entre los valores de la variable y la media; y que responde a la siguiente definición:

$$\sigma = \sqrt{\frac{1}{N} \sum (x_i - \bar{x})^2 \, y_i}$$

A fin de calcular numéricamente los valores de la media aritmética y desviación estándar correspondientes a nuestra distribución de frecuencias, apelamos al siguiente cuadro de cálculo:

x_i número de miembros	y_i frecuencia	$x_i . y_i$	$x_i - x$	$(x_i-\bar{x})^2$	$(x_i - \bar{x})^2 \; y_i$
2	21	42	- 1,82	3,31	69,51
3	36	108	- 0,82	0,67	24,12
4	60	240	0,18	9,03	1,80
5	20	100	1,18	1,39	27,80
6	9	54	2,18	4,75	42,75
7	3	21	3,18	10,11	30,15
8	0	0	4,18	17,47	0
9	1	9	5,18	26,83	26,83
	150	574			222,96

resultando:

$$\bar{x} = \frac{574}{150} = 3,82$$

$$\sigma = \sqrt{\frac{222,96}{150}} = \sqrt{1,48} = 1,22$$

Como resultan un tanto lento el cálculo directo de la desviación estándar a través del esquema de trabajo que provée la definición presentaremos una fórmula de cálculo que abrevia y facilita los pasos necesarios para el cálculo de la misma.

De la definición de variancia como resulta:

$$\sigma^2 = \frac{1}{N} \sum (x_i - \bar{x})^2 \; y_i$$

desarrollando el cuadrado y distribuyendo la sumatoria se tiene

$$\sigma^2 = \frac{1}{N} \sum \left[x_i^2 - 2\, x_i\, \bar{x} + \bar{x}^2 \right] y_i$$

$$= \frac{1}{N} \left[\sum x_i^2\, y_i - 2\, \bar{x} \sum x_i\, y_i + \bar{x}^2 \sum y_i \right]$$

$$= \frac{1}{N} \left[\sum x_i^2\, y_i - 2\, N\, \bar{x}^2 + N\, \bar{x}^2 \right]$$

$$= \frac{1}{N} \sum x_i^2\, y_i - \bar{x}^2$$

$$= \frac{1}{N} \sum x_i^2\, y_i - \left[\frac{1}{N} \sum x_i\, y_i \right]^2$$

resultando que puede escribirse condensadamente, simbolizando los "momentos" según m_k, según la siguiente expresión más fácil de retener en la memoria.

$$\sigma^2 = m_2 - \left[m_1 \right]^2$$

de donde resulta la fórmula de cálculo de la desviación estándar:

$$\sigma = \sqrt{m_2 - \left[m_1 \right]^2}$$

$$= \sqrt{\frac{1}{N} \sum x_i^2\, y_i - \left[\sum x_i\, y_i \right]^2}$$

Nuestro <u>cuadro para el cálculo de la media aritmética y desviación</u> estándar tendrá así, la siguiente disposición:

x_i	y_i	$x_i \quad y_i$	$x_i^2 \qquad y_i$
2	21	42	84
3	36	108	324
4	60	240	960
5	20	100	500
6	9	54	324
7	3	21	147
8	0	0	0
9	1	9	81
TOTAL	150	574	2420

resultando los valores numéricos de nuestras estadísticas:

$$\bar{x} = \frac{574}{150} = 3,82$$

$$\boxed{\bar{x} = 3,82}$$

$$\sigma = \sqrt{\frac{2420}{150} - \left[\frac{574}{150}\right]^2}$$

$$= \sqrt{16,3 - 14,59}$$

$$= \sqrt{1,54}$$

$$\boxed{\sigma = 1,23}$$

Extrañará al comparar los resultados obtenidos para la desviación estándar, el observar que los valores numéricos difieren. Esto se debe a que los redondeos efectuados afectan fuertemente al resultado calculado según la definición, por lo que, de ahora en adelante, calcularemos el valor de σ, desviación estándar, exclusivamente a través de la operatoria que provée la fórmula de cálculo.

III - VARIABLES CONTINUAS

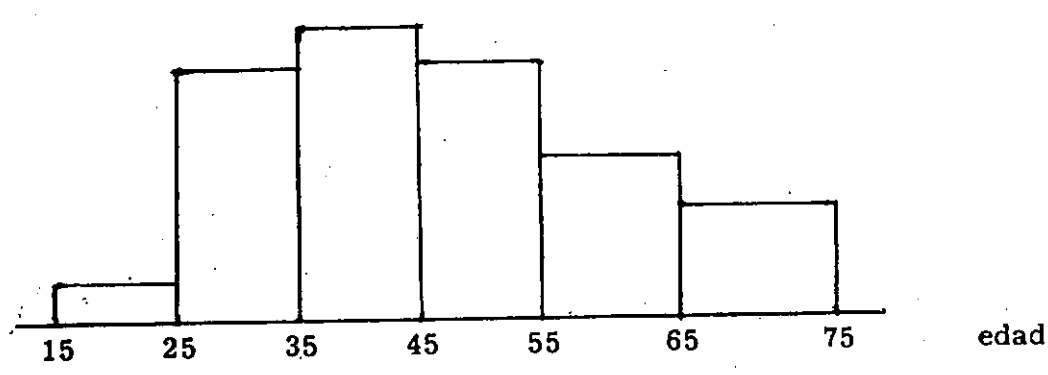
Remontándonos nuevamente a los elementos de la encuesta con que comenzáramos esta exposición toca ahora analizar los resultados correspondientes a la "edad del jefe de la familia", el que constituye un ejemplo de una distribución de frecuencias de una variable contínua.

Agrupando los valores de la variable "edad" en intervalos de edad de igual amplitud se observa la siguiente distribución:

intérvalo de edad	frecuencias
15- 25	5
25 - 35	34
35 - 45	39
45 - 55	35
55 - 65	22
65 - 75	15

HISTOGRAMA

Para efectuar la representación gráfica de esta distribución de frecuencias de una variable contínua recurrimos al HISTOGRAMA, gráfico en donde las frecuencias son representadas por las superficies de rectángulos que se levantan sobre cada intervalo de clase. En nuestro caso, dado que nuestra distribución presenta intervalos de amplitud constante -distribución monomodular- siendo las superficies de los respectivos rectángulos proporcionales a la altura se tiene la siguiente representación gráfica.



Para esta distribución de frecuencias habremos de calcular igualmente dos estadísticas o medidas resumidas de la misma; la media aritmética ($\bar{x}$), representativa de la tendencia centrál y la desviación estándar (σ) medida del grado de uniformidad entre los valores de la distribución.

Las fórmulas que se han visto para el cálculo de estas mismas estadísticas en el caso de una distribución de frecuencias de una variable discreta:

$$\bar{x} = \frac{1}{N} \sum x_i \; y_i$$

$$\sigma = \sqrt{\frac{1}{N} \sum x_i^2 \, y_i - \left[\frac{1}{N} \sum x_i \; y_i \right]^2}$$

requieren para su aplicación el efectuar el supuesto de que en cada intervalo de clase, el punto medio corresponde con el punto de densidad media. Así, de la distribución original se tiene a los fines del cálculo una distribución en donde las frecuencias y_i se suponen alojadas en los puntos medios x_i de los correspondientes intervalos.

La distribución se suele presentar en tal sentido a través del siguiente cuadro de cálculo:

intervalo de edad	punto medio x_i	frecuencias y_i	$x_i \; y_i$	$x_i^2 \; y_i$
15 - 25	20	5	100	2000
25 - 35	30	34	1020	30600
35 - 45	40	39	1560	62400
45 - 55	50	35	1750	87500
55 - 65	60	22	1320	79200
65 - 75	70	15	1050	73500
		150	6800	335200

de donde $\bar{X} = \frac{1}{N} \sum X_i \, y_i$

$$\sigma_X = \sqrt{\frac{1}{N} \sum X_i^2 \, y_i - \left[\frac{1}{N} \sum X_i \, y_i\right]^2}$$

$$= \sqrt{2234,66 - (45,33)^2}$$

$$= \sqrt{2234,66 - 2054,81}$$

$$= \sqrt{179,85}$$

$$= 13,41 \text{ años}$$

Muy frecuentemente, y a fin de abreviar el proceso de cálculo numérico se recurre al auxilio de una VARIABLE DE CALCULO. En estos casos la variable de cálculo es una variable adicional cuyos valores se ponen en correspondencia con los de la variable natural, correspondencia que se expresa a través de una fórmula de cálculo.

Sabido es que la variable de cálculo estará definida una vez determinados para la misma un origen y una unidad de medida convenientes. En tal sentido se suele adoptar como origen el punto medio del intervalo correspondiente a la mayor frecuencia, y como unidad de medida, dada la característica de distribución monomodular, a la amplitud del intervalo de clase.

Así, conservando la notación x_i , y_i y designando según X_i a los correspondientes valores de la variable de cálculo nace el siguiente cuadro para el cálculo de las medidas resumidas de nuestra distribución de frecuencias empleando para ésto una variable de cálculo.

Las fórmulas $\bar{X} = \frac{1}{N} \sum X_i \, y_i$

$$\sigma_X = \sqrt{\frac{1}{N} \sum X_i^2 \, y_i - \left[\frac{1}{N} \sum X_i \, y_i\right]^2}$$

proveerán las medidas resumidas expresadas en valores de la variable de cálculo, para lo cual será necesario apelar a las fórmulas de conversión:

$$\bar{x} = x_b + \bar{X} \quad h$$

$$\sigma_x = h \quad \sigma_X$$

en donde con x_b se designa al valor de la variable natural correspondiente al origen de la variable de cálculo y con h al ancho o amplitud del intervalo de clase.

intervalo de edad	punto medio x_i	Variable de cálculo X_i	Frecuencias y_i	$x_i \; y_i$	$X_i^2 \; y_i$
15 - 25	20	-2	5	-10	20
25 - 35	30	-1	34	-34	34
35 - 45	40	0	39	0	0
45 - 55	50	1	35	35	35
55 - 65	60	2	22	44	88
65 - 75	70	3	15	45	135
			150	124 - 44 + 80	312

de donde resulta: $\bar{X} = \frac{1}{150} \cdot 80 = 0{,}533$

$$\sigma_X = \sqrt{\frac{1}{150} \cdot 312 - \left[\frac{80}{150}\right]^2}$$

$$= \sqrt{2{,}08 - (0{,}533)^2}$$

$$= \sqrt{2{,}08 - 0{,}28} = \sqrt{1{,}80}$$

$$\sigma_X = 1{,}341$$

Valores que en términos de la variable natural resultan:

$$\bar{x} = 40 + 0,533 \times 10$$

$$= 45,3 \text{ años}$$

$$\sigma_x = 10 \times 1,341$$

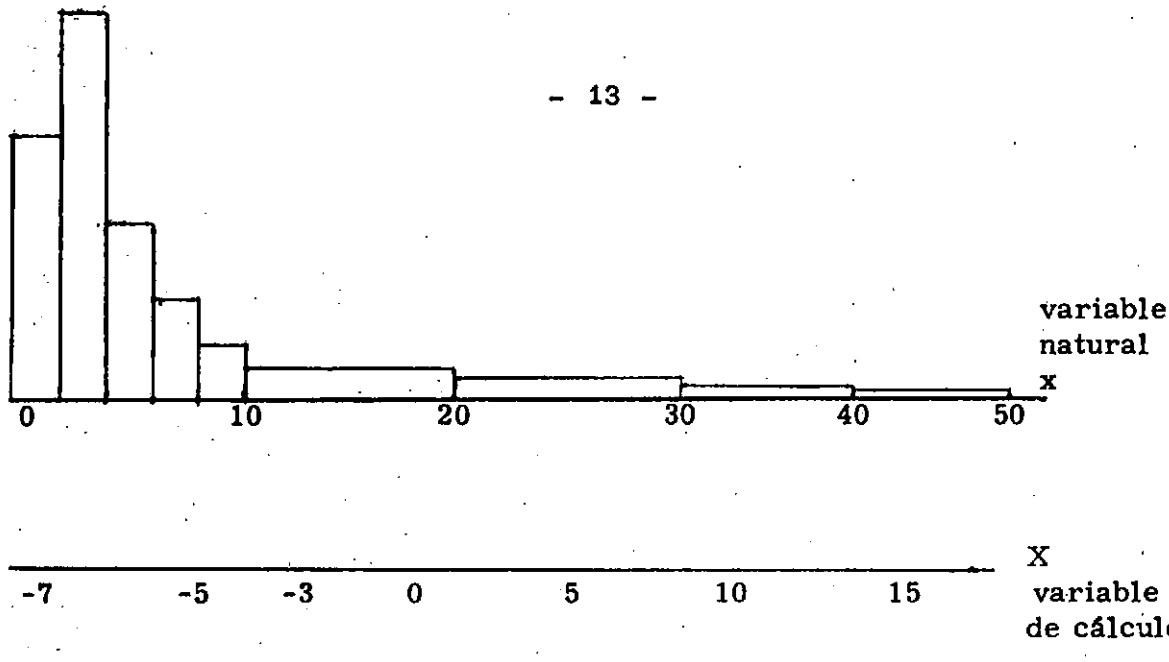
$$= 13,41$$

Se observa así que, a través del artificio de considerar una variable de cálculo resulta más sencillo el cálculo de los valores resumidos, sobre todo cuando los valores originales implican cifras de cierta magnitud.

Los valores correspondientes al item ingreso familiar de nuestra encuesta, corresponden a una distribución de frecuencias marcadamente asimétrica, para la que ha sido necesario considerar intervalos de clase de amplitud variable.

Intervalos de ingreso	Frecuencias
Menos de 2000	35
2000 - 4000	55
4000 - 6000	21
6000 - 8000	12
8000 - 10000	9
10000 - 20000	10
20000 - 30000	5
30000 - 40000	2
40000 - 50000	1
	150

Para representar gráficamente a través del correspondiente histograma esta distribución de frecuencias, dado que los intervalos de clase son de distinta amplitud, debemos retrotraernos la propiedad de los mismos, representando las correspondientes frecuencias a través de las superficies de cada uno de los rectángulos. Se obtiene así la siguiente representación gráfica:



Si se efectúa el supuesto de equirepartición dentro de cada intervalo de clase, se podrá disponer el cuadro de cálculo en forma similar a la adoptada para el caso de una variable discontínua.

ESCALA DE CALCULO

Dada la magnitud de los valores de la variable x que se observan, será conveniente introducir una variable de cálculo X, que se hace corresponder a la variable natural, centrándola frente al punto medio de un conveniente intervalo de clase y tomando como unidad de medida un módulo conveniente h. [En caso de tratarse de una distribución de frecuencias monomodular, esta cantidad h, se tomará igual al ancho del intervalo de clase.]

Así la relación entre ambas variables será:

$$x = x_o + X h$$

Correspondiendo las siguientes fórmulas de pasaje para la media aritmética y desviación estandar:

$$\bar{x} = x_o + \bar{X} h$$

$$\sigma_x = h \sigma_X$$

De esta forma, el cuadro para el cálculo de la media y desviación estándar de una distribución de frecuencias, empleando una variable de cálculo, tiene la siguiente disposición:

intervalo de ingreso	x_i punto medio	y_i frecuencias	X_i variable cálculo	$X_i\ y_i$	$X_i^2\ y_i$
menos de 2000	1000	35	-7	-245	1715
2000 a 4000	3000	55	-6	-330	1980
4000 a 6000	5000	21	-5	-105	525
6000 a 8000	7000	12	-4	- 48	192
8000 a 10000	9000	9	-3	- 27	81
10000 a 20000	15000	10	0	0	0
20000 a 30000	25000	5	5	25	125
30000 a 40000	35000	2	10	20	200
40000 a 50000	45000	1	15	15	225
TOTALES		150		-755 60 -695	5043

Resultan así los valores de media y desviación estándar en variable de cálculo:

$$\bar{X} = \frac{-695}{150} = -4,63$$

$$\sigma_X = \sqrt{\frac{5043}{150} - \left[\frac{695}{150}\right]^2}$$

$$= \sqrt{33,62 - 21,43}$$

$$= \sqrt{12,22}$$

$$= 3,49$$

y finalmente en variable natural:

$$\bar{x} = 15.000 - 4,63 \quad 2000 \quad ; \quad \sigma_x = 3,49 \quad . \quad 2000$$

$$\boxed{\bar{x} = 5.740} \qquad \boxed{\sigma_x = 7.020}$$

CAPITULO II

NUMEROS INDICES

La variable económica de utilización más general, que resuelve todos los problemas que presenta la comparación de distintos elementos, ya sea tratándose de producción, ventas, demanda, oferta, etc., y que permite ubicar en un nivel de comparabilidad a las demás variables, aparece gracias al concepto de valor, siendo unidad de medida necesaria para definirlo, el dinero.

Como esta unidad de medida del valor, que es el dinero, no tiene a través del tiempo la característica de mantenerse invariable, propiedad deseable de toda unidad de medida, resulta de primordial importancia abordar el estudio del instrumento estadístico, de los números índices, que permite conocer y cuantificar las variaciones que experimenta ésta a lo largo del proceso económico.

Todos reconocemos como de utilización corriente la estadística que se designa corrientemente bajo el nombre de índice de costo de vida, la que nos proporciona una medida de la variación del poder adquisitivo del dinero para un dado presupuesto familiar.

Esta estadística, que responde al concepto de número índice, permite llegar a una medida resumida de la evolución de un sistema de precios de un conjunto invariable de artículos, al que se designa bajo el nombre de presupuesto o mejor aún, "canasta familiar". Cuando decimos que el costo de vida presentaba en octubre de 1960 el valor 3.454,2 estamos proporcionando una medida resumida de la evolución del precio de todo un conjunto de bienes y servicios, respecto a un valor base 100 correspondiente al año 1943.

I - FORMULAS ARITMETICAS

Entra dentro de los temas de nuestro interés el observar el mecanismo que nos lleva a expresar medidas resumidas de este tipo;es decir los pormenores metodológicos de la confección de los números índices.

Si lo que se desea es disponer de una medida de la evolución a través del tiempo, del precio de un único artículo, bastará con ir considerando las razones resultantes de comparar el precio actual con el correspondiente al período base. Así, el cociente:

$$\frac{P_t}{P_o}$$

relaciona el precio de un bien en el período t, respecto al precio del mismo bien en un período que se considera como base de referencia, y al que designamos por tanto, mediante el subíndice o.

El índice:

$$I_{o,t} = \frac{P_t}{P_o}$$

correspondiente a la relación de precios de un único artículo a través del tiempo, requiere una pequeña modificación cuando se desee hacer intervenir dos o más artículos. Dentro del enfoque que brindan los cocientes o relativos de precios, al aumentar el número de items se tendrá una medida resumida de la evolución en el nivel general de precios del conjunto de items considerado, calculando la media aritmética de estos relativos de precio. - Así se tendrá:

$$I_{o,t} = \frac{1}{n} \sum \frac{p_t^i}{p_o^i}$$

en donde con p_t^i designamos el precio del bien i-ésimo en el período t, y con p_o^i el precio del bien correspondiente en el período base. Este índice,

uno de los más simples para expresar una medida resumida de la evolución en el nivel general de precios, por ende una evolución en el valor del dinero, y que se conoce bajo el nombre de promedio de relativos de precio, asigna a todos los items de nuestro conjunto igual importancia, cosa que podría muy bien no responder a la realidad. A fin de aumentar la representatividad del índice, incorporando en la estructura del mismo la magnitud relativa de la importancia de los distintos items que entran en su composición, se hace necesario recurrir al concepto de valor relacionando valores en los períodos actuales y de base. Para incorporar el concepto de valor, se hace necesario agregar al ya presente concepto de precio, una medida de cantidad para cada item, la que actuando como ponderación permite efectuar la relación de precios, ahora si respetando la importancia relativa de cada uno de los items.

Se hace necesario así, considerar como fuente de información primaria, el siguiente cuadro de precios y cantidades correspondientes a un conjunto de artículos.

	período actual		período base	
Item	precio	cantidad	precio	cantidad
1	p_t^1	q_t^1	p_o^1	q_o^1
2	p_t^2	q_t^2	p_o^2	q_o^2
⋮	⋮	⋮	⋮	⋮
n	p_t^n	q_t^n	p_o^n	q_o^n

Relacionado el valor, a los precios actuales, de las cantidades intercambiadas, respecto del valor correspondiente a los precios del período base, se tiene, según se consideren como ponderaciones, las cantidades correspondientes al período base, el índice de Laspeyres

$$I_{o,t}^{(L)} = \frac{\sum p_t^i \; q_o^i}{\sum p_o^i \; q_o^i}$$

y aquellas correspondientes al período de referencia, el índice de Passche.

$$I_{o,t}^{(P)} = \frac{\sum p_t^i \; q_t^i}{\sum p_o^i \; q_t^i}$$

en donde se han considerado como ponderaciones las cantidades intercambiadas en el período actual.

Características propias del índice de Laspeyres son las de brindar para el conjunto de items que considera, un valor resumido en la evolución del nivel general de precios, el exceso respecto al valor que idealmente correspondería si se considerase que, por imperio de las leyes económicas de demanda, las ponderaciones correspondientes a cada item deberían modificarse atento a la variación de los precios de los items individuales. Asigna este tipo de índice a aquellos items que pudieran haber presentado variaciones notables en el nivel de precios, demandas estáticas, cosa que no es de esperar en una economía sujeta a las leyes de oferta y demanda, posibilitando, al incorporar los valores correspondientes al período base, el carácter alcista del índice Laspeyres. Del mismo modo se explica la propiedad bajista del índice de Passche, el que proporciona valores resumidos inferiores a los que en justicia correspondería considerar como expresión de la variación resumida en el nivel general de precios.

A fin de obviar estos inconvenientes se recurre a índices más elaborados, que se conocen bajo el nombre de índices ideales, correspondien

do destacar entre éstos el índice de Marshall-Edgeworth:

$$I_{o,t}^{(M)} = \frac{\sum p_t^i \ (q_o^i + q_t^i)}{\sum p_o^i \ (q_o^i + q_t^i)}$$

en donde se considera como ponderaciones, al promedio aritmético de las cantidades correspondientes a los períodos base y actual y el índice de Fisher

$$I_{o,t}^{(F)} = \sqrt{I_{o,t}^{(L)} \cdot I_{o,t}^{(P)}} = \sqrt{\frac{\sum p_t^i \ q_o^i}{\sum p_o^i \ q_o^i} \quad \frac{\sum p_t^i \ q_t^i}{\sum p_o^i \ q_t^i}}$$

promedio geométrico de los índices de Laspeyres y Paasche.

II - EJEMPLOS NUMERICOS

Consideramos, a título de ejemplo, las siguientes estructuras de precios y cantidades correspondientes a los períodos anuales de 1950 y 1951 .

Años	1 9 5 0		1 9 5 1	
Item	precio	cantidad	precio	cantidad
1°	150	23	180	20
2°	210	12	210	12
3°	180	15	190	19
4°	14	27	14	31
5°	35	180	35	192
6°	70	92	75	47
7°	90	51	120	52
8°	27	83	27	90

Si aplicamos a las mismas, las expresiones vistas para nuestros distintos números índices, tendremos en cada caso:

promedio de relativos de precio:

$$I_{50,51}\% = 108$$

Índice de Laspeyres

$$I^{(L)}_{50,51}\% = 110,2$$

Índice de Paasche

$$I^{(P)}_{50,51}\% = 109,2$$

Índice de Fisher

$$I^{(F)}_{50,51}\% = 109,9$$

III - PROPIEDADES DESEABLES

No son tanto las propiedades alcista o bajista que puedan tener las fórmulas empleadas para el cálculo de los índices simples vistos, ni su ausencia en los índices compuestos, el motivo especial de preocupación al abordar el estudio de los números índices, sino las propiedades más generales de identidad, conmensurabilidad, proporcionalidad, determinación, reversibilidad y circularidad que deseablemente deben ser satisfechas por las fórmulas empleadas para proporcionar una correcta información resumida sobre las variaciones de una estructura general de precios.

Estas propiedades deseables pueden resumirse según las siguientes consideraciones:

1 - Identidad: simbolizada por la condición de que una estructura de precios referida a si misma deba dar como resultado la unidad. Así, como consecuencia de la condición de identidad se observa

$$I_{O,O} = 1$$

2 - Proporcionalidad: esta propiedad requiere que al va -

riar de un período a otro, todos los elementos de una constelación de precios en un factor de proporcionalidad k, esta variación se vea reflejada en el valor numérico índice. Así, en estas condiciones, se tendrá:

$$I_{O,1} = k$$

3 - Determinación: Según la cual el valor numérico del índice no debe hacerse infinito cuando alguno de los elementos de la estructura de precios o cantidades se anule.

4 - Conmensurabilidad: según la cual el valor numérico del índice debe ser independiente de las unidades de medida escogidas para cada item.

Estas cuatro propiedades son cumplidas por nuestros tres índices simples (salvo la de determinación, que no se ve satisfecha en el caso del promedio aritmético de relativos de precios) y naturalmente por los índices compuestos vistos, y se refieren más a los resultados a esperar de la operatoria aplicada que a la estructura propia de un mecanismo de números índices y los alcances y posibilidades de su empleo. Las dos propiedades que detallaremos a continuación tocan más al alcance de los números índices como mecanismo de proporcionar medidas resumidas para la interpretación de situaciones.

5 - Reversibilidad: por la cual el producto de los valores numéricos de los índices que resultan de intercambiar las constelaciones de precios y cantidades correspondientes al período base, por las del período en consideración y viceversa, debe dar por resultado la unidad. Esquemáticamente, la condición de reversibilidad lleva a:

$$I_{O,1} \cdot I_{1,0} = 1$$

Esta condición de reversibilidad que sólo es satisfecha por los índices compuestos de Marshall y Fischer, permite comparar la evolución de situaciones en uno u otro sentido respecto al tiempo, y la no aplicabilidad de la misma, por imperio de la expresión particular utilizado, para el cálculo del índice, limita el alcance las aplicaciones a efectuar con los mismos.

6 - Circularidad: condición que permite el encadenamiento entre índices de igual clase, aunque con distinta base. Se expresa esquemáticamente a través de la siguiente relación:

$$I_{O,1} \cdot I_{1,2} = I_{O,2}$$

no siendo satisfecha por ninguno de los índices vistos, constituye empero la propiedad más deseable de todo sistema de números índices, ya que evita el tener que recalcular series cuando es necesario proceder a considerar nuevos períodos base.

CAPITULO III

SERIES CRONOLOGICAS

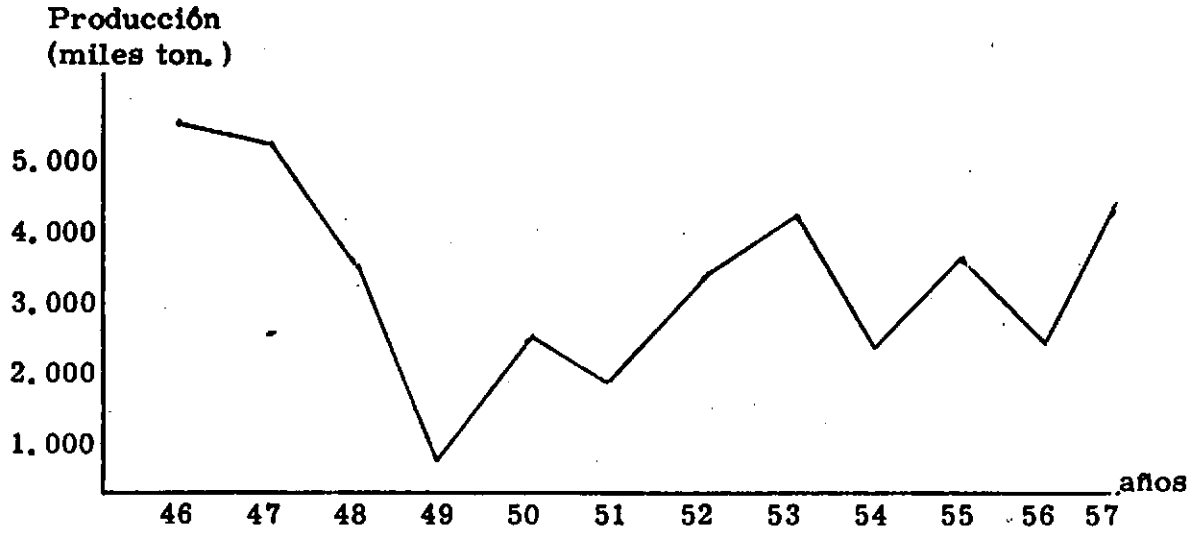
Al estudiar cuantitativamente una variable de interés económico, como ser por ejemplo estadísticas monetarias, de producción, sociológicas, financieras, o de otras actividades que hacen a la vida económica de la Nación, así también como en las razones que interpretan la pequeña economía de la empresa, nos encontramos frente a valores numéricos que responden a la evolución de un fenómeno a través del tiempo, valores de una serie cronológica o serie de tiempo que proporcionan a través de datos anuales, mensuales o correspondientes a períodos de menor duración, la información primaria con la que se podrá iniciar el estudio que lleve a la determinación del comportamiento típico e irregularidades que caracterizan a una serie de tiempo.

I - REPRESENTACIONES GRAFICAS

Uno de los primeros elementos a que se recurre para proceder al análisis de una serie cronológica es el de las representaciones gráficas, por medio del cual se llevan a trazos geométricos las cifras que, presentadas en forma de cuadro, constituyen nuestra información original. Consideremos a vía de ejemplo, la representación gráfica de los valores correspondientes a la producción de maíz en toda la república en el período 1946-1957.

Cosecha	Producción
1946	5.815
1947	5.200
1948	3.450
1949	836
1950	2.670
1951	2.040
1952	3.550
1953	4.450
1954	2.546
1955	3.870
1956	2.698
1957	4.806

A la que corresponde el siguiente gráfico:



En esta representación gráfica, denominada en escala natural o aritmética, la poligonal trazada uniendo los extremos de las barras representativas de los valores observados, permite una inmediata visualización del fenómeno en cuestión y tiene la propiedad de representar, a través de la pendiente de cada segmento, el grado de crecimiento absoluto que experimenta en cada período, nuestra serie cronológica. Así se observa que el mayor decrecimiento absoluto corresponde al período 1948/49, mientras que el mayor crecimiento absoluto se verifica en el período 1956/57.

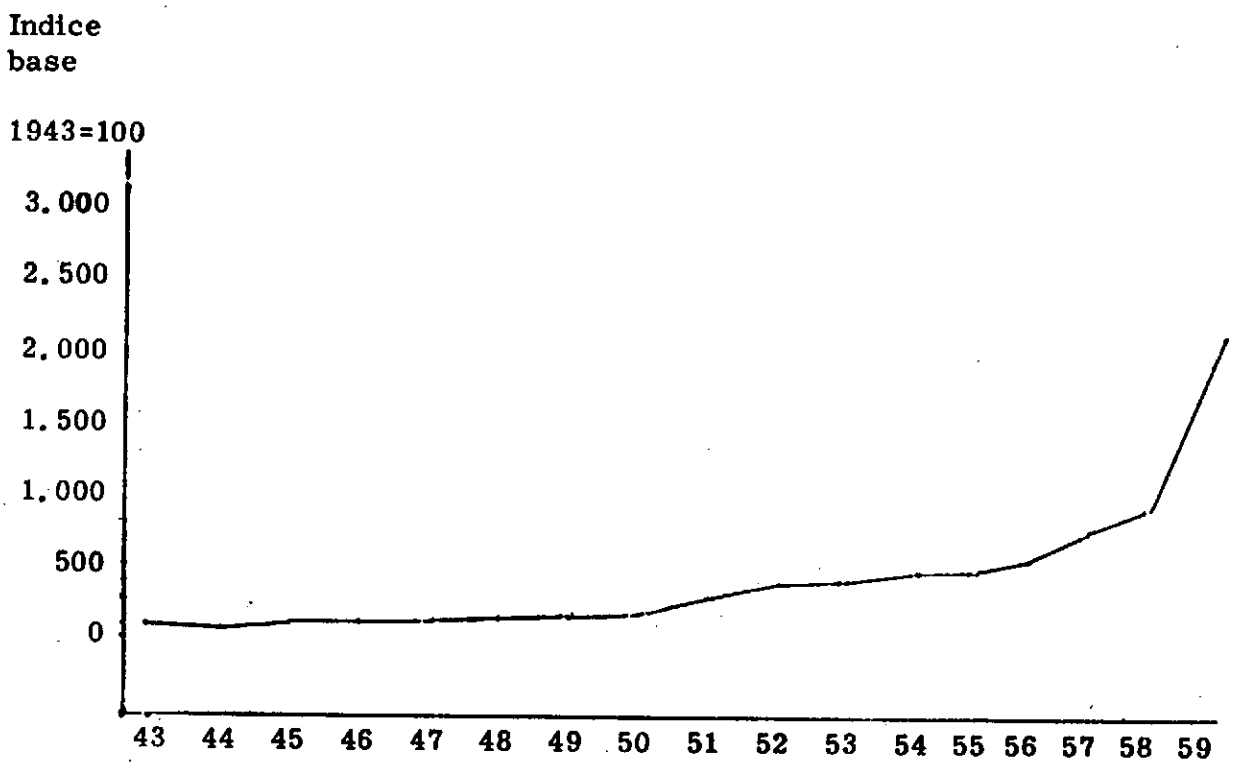
Si la evolución que experimentan los valores de la serie cronológica fuera muy pronunciada podrá convenir recurrir a la representación gráfica en escala semilogarítmica, mediante la cual se representan los valores de la variable de interés en una escala logarítmica apropiada.

Observamos para la serie Indice del costo de un nivel de vida en la Capital Federal, ambas representaciones gráficas cubriendo un período que va desde 1943, base del índice, hasta 1959.

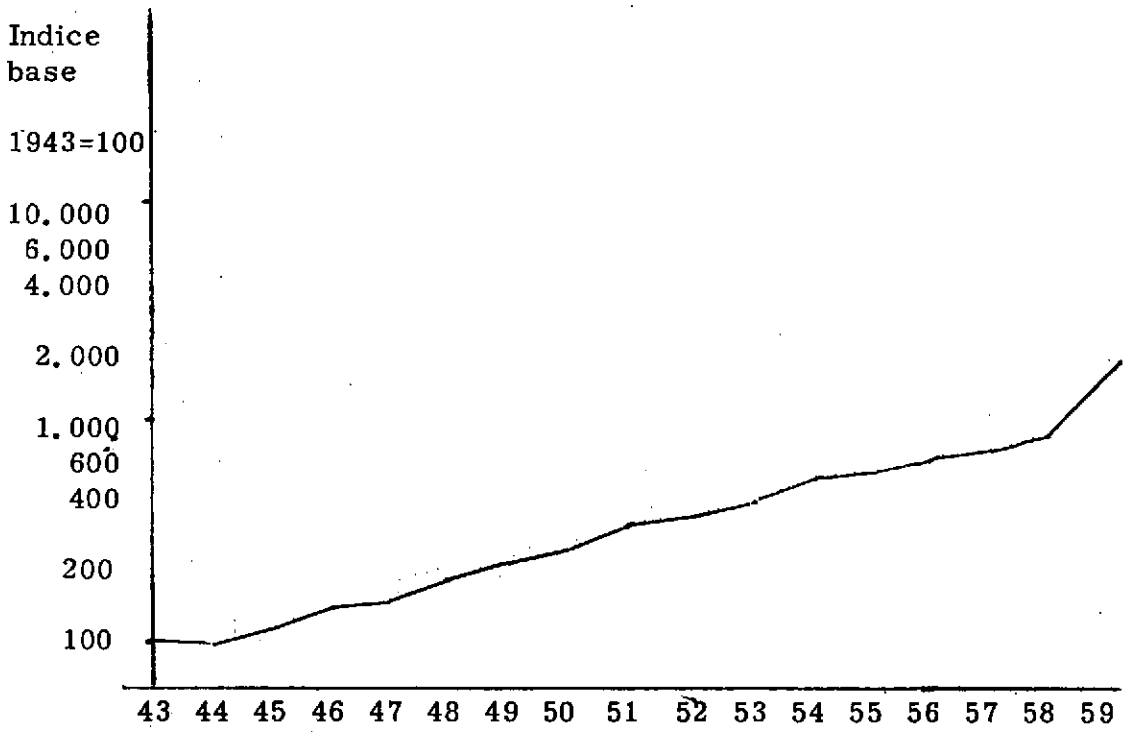
Año	Indice
1943	100
1944	99
1945	119
1946	140
1947	159
1948	180
1949	236
1950	296

Año	Indice
1951	405
1952	562
1953	585
1954	607
1955	682
1956	773
1957	965
1958	1269
1959	2713

Representación Gráfica en Escala Natural



Representación Gráfica en Escala Semilogarítmica



La pendiente de los segmentos de la poligonal sustentada por esta representación gráfica en escala semilogarítmica corresponde ahora a una medida del crecimiento relativo. Así se puede observar directamente por inspección ocular que el mayor incremento relativo corresponde al período 1958-59, en donde se observa también, como lo muestra la representación gráfica en escala aritmética, el mayor incremento absoluto.

III - TENDENCIA

La representación gráfica de los valores de una serie cronológica sugiere la existencia de una tendencia dominante en la evolución que experimentan los valores de la misma. Esta tendencia, sustentada por los valores anuales, recibe el nombre de tendencia secular y constituye un valor medio dinámico que interpreta a las variaciones experimentales del fenómeno en estudio. Cada serie de tiempo, al ser representada gráficamente en escala aritmética, sugiere la forma de la línea de tendencia que mejor la interpreta. En este sentido se observan como convenientes para ser empleados como ecuación de las líneas de tendencia, las ecuaciones lineales, polinomios de orden superior y, eventualmente, funciones exponenciales.

La elección del grado del polinomio de tendencia que mejor se adapta a la interpretación del fenómeno observado constituye en sí un problema que se resuelve fácilmente si se recurre al expediente de dibujar a mano levantada sobre la representación gráfica de los valores de la serie de tiempo, un trazo suave y contínuo que interprete la evolución del fenómeno observado. El mayor número de intersecciones que se pueda lograr superponiendo a dicha línea, una recta, determinará prácticamente el grado del polinomio de ajustamiento de tendencia que mejor interpreta el fenómeno observado.

Si simbolizamos con $P_k(t)$ al polinomio de ajustamiento de tendencia de grado k, este método empírico nos ayudará en la determinación de k.

III - AJUSTAMIENTO POR MINIMOS CUADRADOS

Resuelto el problema de la determinación del grado del polinomio de ajustamiento de tendencia, pasamos a estudiar el criterio a emplear para la determinación de sus coeficientes. Todos estos problemas de ajustamiento serán resueltos remitiéndonos al método de los mínimos cuadrados.

Consideremos, a los efectos de la determinación de la tendencia en una serie de tiempo, los valores provistos por el siguiente cuadro, correspondientes a producción en el período 1949/57 que, como integrantes de una serie más extensa, utilizáramos ya en una representación gráfica en estas líneas.

Año	Producción
1949	836
1950	2670
1951	2040
1952	3550
1953	4450
1954	2546
1955	3870
1956	2698
1957	4806

Una observación simple de la evolución correspondiente a esta serie de tiempo en el período observado, muestra que la misma puede ser adecuadamente interpretada mediante una tendencia lineal que se expresa según la siguiente ecuación:

$$P_1(t) = a + b\,t$$

de donde con P designamos los valores teóricos de tendencia que queremos determinar, siendo a y b valores convenientes de los parámetros de la ecuación de la recta.

Aceptando la representatividad de este tipo de función veremos la forma de adecuar los valores de los parámetros de la misma a nuestro conjunto de valores experimentales, siguiendo para ello el método de los mínimos cuadrados, que impone la condición de que la suma de los cuadrados de los desvíos entre los valores experimentales y los valores teóricos, provistos por la función de ajustamiento, sea un mínimo.

La expresión matemática de la condición de los mínimos cuadrados nos lleva a determinar los valores de a y b que hagan que la función

$$z = \sum \left[y_i - a - b\,t_i \right]^2$$

tenga un valor mínimo. Las condiciones necesarias para la existencia de un extremo, (mínimo en este caso) son:

$$\frac{\partial z}{\partial a} = 2 \sum_{i=1}^{n} [y_i - a - b\, t_i] \cdot [-1] = 0$$

$$\frac{\partial z}{\partial b} = 2 \sum_{i=1}^{n} [y_i - a - b\, t_i] \cdot [-t_i] = 0$$

las que nos llevan al siguiente sistema de ecuaciones normales de Gauss:

$$n\, a + b \sum t_i = \sum y_i$$

$$a \sum t_i + b \sum t_i^2 = \sum y_i\, t_i$$

cuya solución se simplifica si se acude al expediente de considerar para la expresión del tiempo una variable centrada t, para la cual

$$\sum t_i = 0$$

En términos de una variable centrada el anterior sistema se simplifica, estando los parámetros de la ecuación de la recta de tendencia dados según:

$$a = \frac{\sum y_i}{n} \quad ; \quad b = \frac{\sum y_i\, t_i}{\sum t_i^2}$$

Para atender al cálculo de estos coeficientes dispondremos los datos de nuestro cuadro original según el siguiente <u>cuadro de cálculo</u>:

Año	t_i	y_i	$y_i \quad t_i$	t_i^2
1949	-4	836	- 3344	16
1950	-3	2670	- 8010	9
1951	-2	2040	- 4080	4
1952	-1	3550	- 3550	1
1953	0	4450	0	0
1954	1	2546	2546	1
1955	2	3870	7740	4
1956	3	2698	8094	9
1957	4	4806	19224	16
		27466	18620	58

donde con y_i hemos designado nuestros valores de producción, resultando en consecuencia:

$$a = \frac{27466}{9} = 3051,77$$

$$b = \frac{18620}{58} = 321,03$$

por lo que nuestro polinomio de ajustamiento de tendencia, calculado según el método de los mínimos cuadrados, tiene por ecuación:

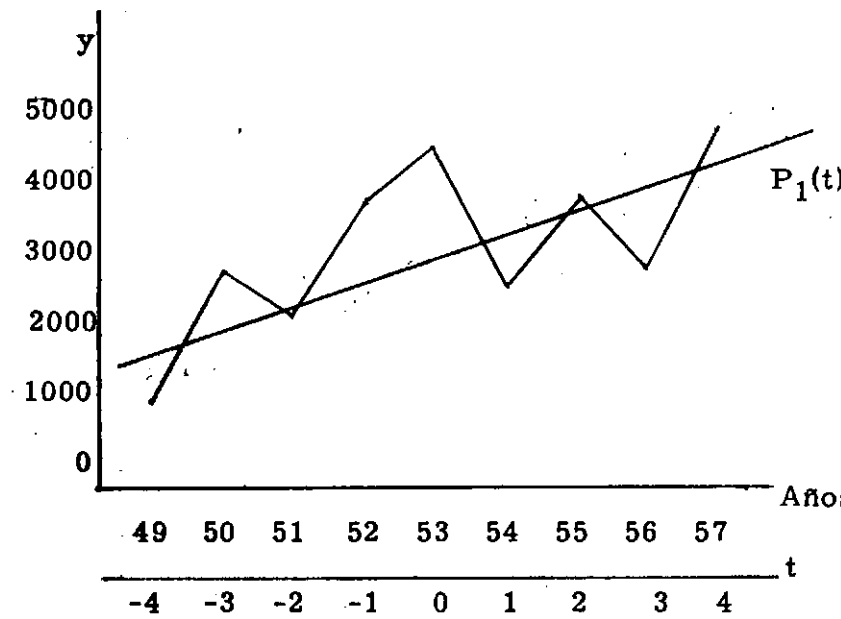
$$P_1(t) = 3051,77 + 321,03 \; t$$

La representación gráfica de la línea de tendencia así determinada, que se logra valorizando la ecuación de la misma en dos valores particulares de la variable t, y uniendo estos dos puntos mediante una línea recta, nos muestra la forma en que la misma interpreta la evolución de los valores de nuestra serie de tiempo.

III - EXTRAPOLACION

La recta de tendencia nos provee de un elemento teórico adecuado para efectuar proyecciones o pronósticos para los valores de nuestra variable y_i, mientras no se produzcan cambios fundamentales en la estructura que gobierna los valores de la misma o no actúen componentes accidentales que modifiquen circunstancialmente las condi-

ciones. Por tal causa es que se deberá trabajar con cuidado al emplear estos elementos para efectuar pronósticos, limitando los alcances de las proyecciones que se hagan mediante estos métodos.



El valor estimado por tendencia para 1958 se obtendrá valorizando el polinomio de ajustamiento para $t = 5$, valor que corresponde a 1958 en la escala de cálculo utilizada.

$$P_1(t)\ 58 = 3051,77 + 321,03 \cdot 5$$
$$= 4656,92$$

La proyección hecha de esta forma, por extrapolación de tendencia constituye una estimación puntual. Naturalmente no se esperará que el valor que se observe en 1958 coincida con el valor estimado. A fin de ampliar el alcance de estos métodos de proyección dispondremos de un mecanismo adicional que convierta nuestra proyección puntual en espacial. Así, en lugar de predecir un valor para 1958, será más razonable predecir un intervalo, dentro del cual con mayor razón se esperará se dé el valor correspondiente al año 1958.

Hemos visto que la ecuación de tendencia interpreta en media a los valores experimentales, pero no coincide exactamente con

los mismos; es decir se producen desvíos entre los valores experimentales y los valores teóricos de tendencia, desvíos que tienen las propiedades derivadas del método de los mínimos cuadrados; 1) la suma de sus cuadrados es un mínimo, 2) la suma de los mismos es igual a cero; vale decir se compensan en valor, aquellos positivos con los negativos.

Estos desvíos $\delta_i = y_i - P_k(t_i)$ caracterizan las irregularidades de nuestra serie de tiempo al apartarse de los valores de tendencia. Si calculamos el desvío estándar.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \delta_i^2}$$

y orlamos nuestra línea de tendencia con dos franjas paralelas cuyos bordes disten de ésta en 3σ, se tendrá, por las propiedades derivadas del Teorema de Tchebycheff que, con gran probabilidad [mayor que 0.89] el valor correspondiente a 1958 se encontrará dentro de los límites de esta franja. Se tiene así una estimación a través de un intervalo en donde, salvo que actúen causas ajenas a las que se han venido observando en el período considerado, habrá de producirse con una gran probabilidad el valor de 1958.

IV - COMPONENTE CICLICA

Los desvíos δ_i aludidos en la presentación de este elemento de variación que es el desvío estándar, constituyen los valores de la denominada componente cíclica de una serie cronológica y representan por construcción, las irregularidades propias de la serie cronológica respecto a sus valores explicados por tendencia. Los valores experimentales, por encima o por debajo de los valores teóricos de tenden-

cia, muestran los movimientos propios de la variable en cuestión y serán de suma utilidad cuando, a través de los métodos de correlación y regresión que estudiaremos en el capítulo IV de estas lecciones, querramos comparar la relación existente entre los movimientos de dos series de tiempo. Este estudio de comparación se realiza según el caso, a través de los valores naturales de las series en cuestión, como a través de los valores de las correspondientes componentes cíclicas, y más corrientemente a través de los valores de las correspondientes componentes cíclicas reducidas que se definen para cada serie según el cociente de dividir la componente cíclica absoluta por la desviación estándar σ que resulta de considerar los desvíos entre los valores experimentales y los teóricos calculados sobre la recta de tendencia. Esta componente cíclica reducida tiene la propiedad de hacer comparables las variaciones de cualquier serie cronológica, ya que al tomar como unidad de medida para las mismas, los correspondientes valores de la desviación estándar, lleva a una común dimensión los valores obtenidos para la componente cíclica.

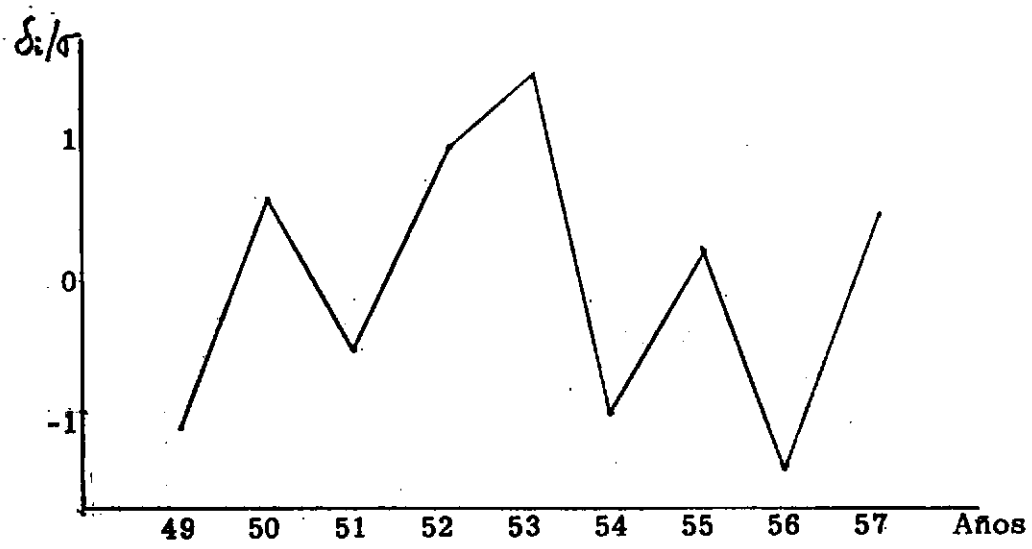
Consideremos a continuación los elementos del cuadro de cálculo necesarios para la determinación de los valores numéricos de la componente cíclica reducida, aplicados a los valores de la serie cronológica expuesta en el cuadro anterior.

Año	y	t	y t	t^2	P(t)	δi	δi^2	$\delta i/\sigma$
1949	836	-4	-3344	16	1667,66	-931,66	867.990	-1,085
1950	2670	-3	-8010	9	2088,69	581,31	337.921	0,677
1951	2040	-2	-4080	4	2409,72	-369,72	136.693	-0,430
1952	3550	-1	-3550	1	2730,75	819,25	671.171	0,954
1953	4450	0	0	10	3051,78	1398,22	1.955.019	1,628
1954	2546	1	2546	1	3372,81	-826,81	683.615	-0,963
1955	3870	2	7740	4	3693,84	176,16	31.032	0,205
1956	2698	3	8094	9	4014,87	-1316,87	1.734.166	-1,533
1957	4806	4	19224	16	4335,80	470,10	220.994	0,547
			18620	58			6.638.882	

$$\sigma = \sqrt{\frac{6.638.882}{9}} = \sqrt{737.653,5} = 858,87$$

La representación gráfica de los valores de la <u>componente cíclica absoluta</u> muestra las variaciones propias de la serie cronológica en estudio, despojadas de la componente de tendencia que acabamos de eliminar

La representación de los valores de la <u>componente cíclica</u> reducida sigue el mismo molde de variación, desprovisto empero, de unidad de medida y llevados los mismos a unidades estándar.



Son las propiedades de la componente cíclica reducida las que permitirán, a través de los mecanismos del análisis de correlación y regresión, ampliar los estudios relativos al comportamiento individual de una serie cronológica, proporcionando un instrumento más fiable en el problema de las estimaciones y pronósticos que el de simple extrapolación de tendencia.

En el estudio de las series cronológicas que, a través de estas líneas presentáramos en un aspecto muy simplificado, debe naturalmente, tenerse en cuenta las regularidades correspondientes a períodos de tiempo de menor extensión que el año, como ser meses o aún semanas, existiendo diversos métodos para la determinación de las componentes estacionales que, aunque escapan del alcance de este primer enfoque en el problema, no pueden dejar de considerarse en todo estudio en que se considere una variable suceptible de tomar valores en períodos de tiempo dentro del año calendario.

CAPITULO IV

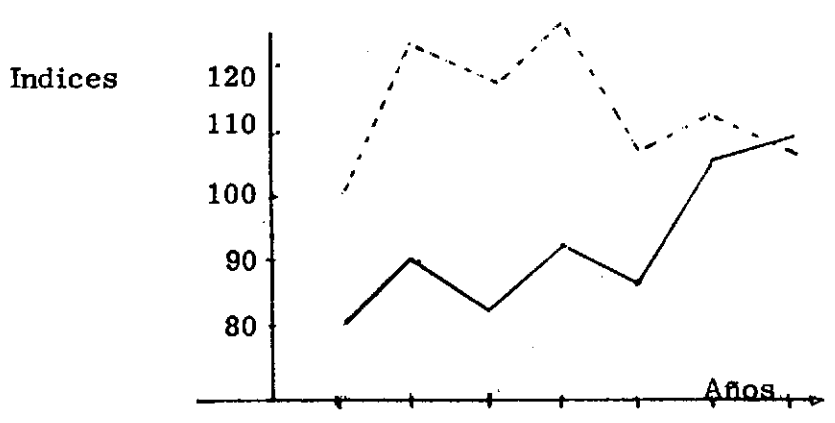
CORRELACION

En el lenguaje corriente decimos que dos variables están correlacionadas cuando las variaciones de una de ellas repercuten sobre los valores de la otra.

Consideremos, a fin de ilustrar nuestra exposición mediante un ejemplo, dos variables que presentan valores a través del tiempo;,a saber: X : salario de obreros mineros; Y precio del carbón. Valores de las mismas, expresados en números índices podrían ser los del siguiente cuadro:

AÑOS	X	Y
1951	80	100
1952	90	125
1953	81	118
1954	92	130
1955	85	105
1956	103	110
1957	108	105

Una representación gráfica de los valores de estas dos series de tiempo permitirá observar las variaciones de las mismas e iniciar, a través de ellas, una primera medida de la correlación existente entre ambas variables.



I. INDICE DE CONCORDANCIAS

Será a través del índice de concordancias que lograremos obtener una primera medida del grado de asociación existente entre ambas variables, observando las variaciones correspondientes a ambas series para cada período y determinando el número de concordancias y discordancias que se presentan en la evolución de los valores cronológicos.

Definiremos nuestro índice de concordancias según:

$$I = \frac{C - D}{C + D}$$

Como el valor numérico de este índice puede oscilar en un campo de variación comprendido entre -1 y +1, deberemos observar la significación de los valores extremos del mismo, a fin de poder asignar por comparación, una significación correspondiente a las medidas cuantitativas correspondientes a una determinada situación. El valor extremo +1, se alcanza cuando, a lo largo de todos los períodos se observan tan solo concordancias entre las variaciones correspondientes a las dos series. Haremos así corresponder a este valor extremo, la noción de muy fuerte asociación, ya que, a lo largo de todo el período de observación, los movimientos de una de las variables, acompañan en sentido a los de la otra. El valor extremo -1, que resulta cuando a través de todos los períodos se observan tan solo discordancias, indica la existencia de una marcada correlación en el sentido inverso entre ambos fenómenos. Valores intermedios de nuestro índice de concordancias corresponderán naturalmente a situaciones intermedias entre estos dos extremos, correspondiendo el valor central del índice, cero, a una situación de independencia entre las variaciones de ambos fenómenos.

En nuestro ejemplo numérico, los valores observados para el número de concordancias y discordancias son C = 5 y D = 1, siendo en consecuencia, el valor numérico de nuestro índice:

$$I = \frac{5 - 1}{5 + 1} = \frac{4}{6} = 0,66$$

valor que nos da una medida del grado de asociación directa existente entre ambas variables.

II . INDICE DE VARIACION

Si quisiéramos tener una medida más representativa del grado de asociación entre ambas variables, deberíamos incluir en nuestro índice, una medida de la intensidad de las variaciones observadas, pudiendo hacerlo a través del producto de las mismas, con el signo correspondiente, según se trate de discordancias o concordancias.

Así, una expresión más representativa del grado de asociación existente entre ambas variables se obtendría a través del índice de variación, definido según

$$I_v = \frac{\sum \Delta X_i \quad \Delta Y_i}{\sum |\Delta X_i| |\Delta Y_i|}$$

en donde con Δx_i simbolizamos la variación observada para el intervalo iésimo de variación en la variable X, y con $|\Delta x_i|$ su correspondiente valor absoluto. Necesitaremos, en consecuencia, disponer nuestros valores según el siguiente cuadro de cálculo:

Δx	Δy	$\Delta x \quad \Delta y$	$\lvert\Delta x\rvert \quad \lvert\Delta y\rvert$
10	25	250	250
-9	-7	63	63
11	12	132	132
-7	-10	70	70
15	5	75	75
2	-3	- 6	6
		584	596

por medio del cuál se calcula el siguiente valor del índice de variación

$$I = \frac{596}{584} = 0,97$$

III. <u>COEFICIENTE DE CORRELACION</u>

Una medida más representativa de la correlación existente entre las variables X e Y, puede lograrse apelando al <u>coeficiente de correlación</u> definido según

$$r = \frac{\frac{1}{n} \sum x \cdot y}{\sigma x \quad \sigma y}$$

en donde $x = X - \bar{X}$; $y = Y - \bar{Y}$, respondiendo σx así como σ_y a sus definiciones habituales.

El cuadro para el cálculo del coeficiente de correlación tendrá en consecuencia, la siguiente disposición:

	X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	x y
	80	100	- 10	- 20	100	400	200
	90	125	0	5	0	25	0
	81	118	- 9	- 2	81	4	18
	92	130	2	10	4	100	20
	85	120	- 5	0	25	0	0
	100	125	10	5	100	25	50
	102	122	12	2	144	4	24
Total	630	840			454	558	342

$\bar{X} = 90$ $\bar{Y} = 120$

resultará así, el valor numérico de nuestro coeficiente de correlación

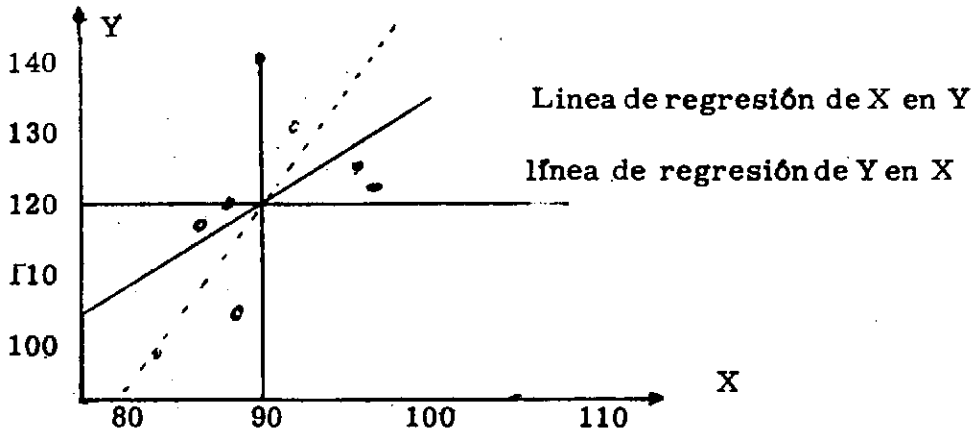
$$r = \frac{312}{\sqrt{454}\ \sqrt{558}} = 0,62$$

valor que aparte de proporcionar una información más completa del grado de asociación entre las dos variables en estudio, servirá de base para estudios posteriores.

IV . REGRESION

Hemos determinado a través del coeficiente de correlación una medida del grado de asociación entre dos variables. Enfrentaremos ahora el problema de estimar los valores de una de ellas en función de un valor dado para la otra. La medida de la correlación permite, hemos visto, determinar si ambas variables, X e Y se corresponden. Serán las ecuaciones de regresión las que permitan emplear el conocimiento de los valores de una de las variables para estimar los valores correspondientes de la otra.

Si representamos gráficamente en un diagrama de dispersión los valores correspondientes de nuestras dos variables X e Y se tendrá:



un conjunto de puntos cuyas abscisas y ordenadas corresponden a los valores de X e Y, y constituyen la llamada nebulosa de correlación. Este conjunto de puntos, que muestra la relación existente entre los valores de X y los correspondientes de Y, puede resumirse según una línea -recta- que interprete en media, las relaciones existentes. Si lo que se desea es estimar un valor de la variable Y en términos de un dado valor de la variable X, deberemos particularizar refiriéndonos a la línea de regresión de Y en X, cuya ecuación calculamos aplicando la condición de los mínimos cuadrados, siendo la misma, referida al baricentro del diagrama de dispersión

$$y = b\ x$$

en donde

$$b = \frac{\sum x\ y}{\sum x^2}$$

y referida a los valores naturales de las variables

$$\boxed{Y = Y + \frac{\sum x\ y}{\sum x^2}\ (X - \bar{X})}$$

La relación existente entre el coeficiente de correlación y el coeficiente de regresión de Y en X,

$$b = r \cdot \frac{\sigma_y}{\sigma_x}$$

permite escribir la ecuación de la línea de regresión de Y en X en términos del valor numérico obtenido para el coeficiente de correlación, resultando:

$$\boxed{Y = \bar{Y} + r \cdot \frac{\sigma_y}{\sigma_x} (X - \bar{X})}$$

fórmulas ambas que permiten estimar un valor de la variable Y en términos de un dado valor de la variable X, lo que justifica el nombre de línea de regresión de Y en X.

Como el problema inverso, estimar el valor de la variable X correspondiente a un dado valor de la variable Y, tiene igualmente sentido, deberemos ocuparnos de esta segunda línea de regresión, de X en Y, la que nos permite hacer una estimación recorriendo un camino inverso, es decir, apoyándose en un valor de la variable Y para obtener un valor estimado de la variable X. La línea de regresión que permite este segundo tipo de estimación no puede ser la misma que se obtuviera al considerar la regresión de Y en X, en la que el método de los mínimos cuadrados aplicado aseguraba un menor error en los desvíos de los valores de la variable Y. Si aplicamos nuevamente el criterio de los mínimos cuadrados a la determinación de esta segunda línea de regresión, su ecuación referida al baricentro será:

$$X = \frac{\sum x \quad y}{\sum y^2} \cdot y$$

la que puede ser expresada en términos del valor numérico del coeficiente de correlación según:

$$x = r \frac{\sigma_x}{\sigma_y} \cdot y$$

En el ejemplo que se empleara como ilustración de una medida de correlación, la ecuación de la línea de regresión de Y en X resulta:

$$y = 0,62 \cdot \frac{558}{454} \cdot x$$

$$\boxed{y = 0,76 \cdot x}$$ LINEA DE REGRESION DE Y EN X

la que se representa en el diagrama de dispersión a través de un trazo lleno.

Recíprocamente, la ecuación de la línea de regresión de X en Y será

$$x = 0,62 \cdot \frac{454}{558} \cdot y$$

$$x = 0,50 \cdot y$$

la que explicitada en términos de la variable y da por ecuación

$$\boxed{y = 2 \cdot x}$$ LINEA DE REGRESION DE X EN Y

la que se ha representado sobre el diagrama de dispersión en forma de un trazo punteado.

Veamos a través de una nueva aplicación de nuestro ejemplo numérico cómo podemos emplear este análisis de regresión para la estimación de un valor de una variable en términos de un valor conocido para otra correlacionada.

Supongamos para un determinado año conocer el valor de la variable X, salario de obreros mineros: X = 100, y querer estimar a través del mecanismo de los métodos de regresión, el valor correspondiente para la variable Y, precio del carbón, desconocido en este caso.

Expresando la ecuación de regresión de Y en X en función

de los valores naturales de las variables, se tiene:

$$Y = 120 + 0,76\ (X - 90)$$

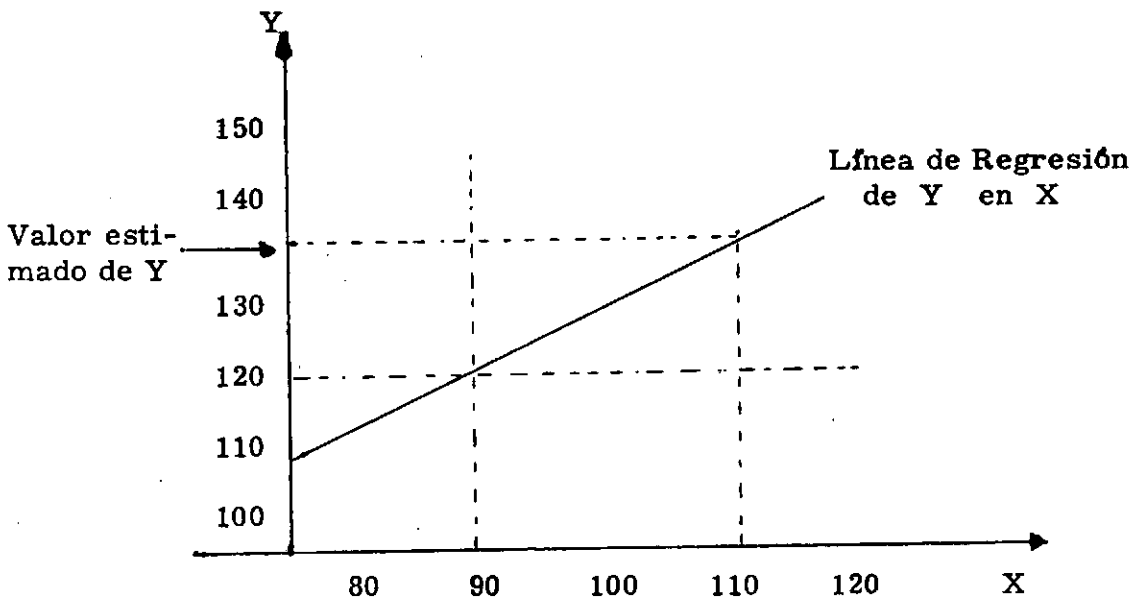
$$= 120 - 68,4 + 0,76\ X$$

$$= 51,60 + 0,76\ X$$

si en ésta se sustituye el valor X libre, por el valor conocido X = 110, se tendrá la estimación por regresión de Y dada según:

$$Y = 51,60 + 0,76\ .\ 110$$

$$= 51,60 + 83,6$$

$$Y = 135,20$$

resultado que gráficamente expresado sobre el diagrama de dispersión corresponde al siguiente esquema: