INSTITUTO LATINOAMERICANO DE PLANIFICACION ECONOMICA Y SOCIAL

Curso de Santiago
25 de julio de 1962

ANALISIS DE CORRELACION *

Pedro Vúskovič

* Reproducción de los apuntes de clases de la Escuela de Economía de la Universidad de Chile.

ANALISIS DE CORRELACION

Pedro Vuskovic

Los temas anteriores han tendido básicamente a ilustrar dos aspectos:

a) las medidas que permiten la descripción y análisis de las características de ciertas observaciones estadísticas en un momento o para un período dado. Fue lo que se designó como análisis de distribuciones de frecuencias, en el que se precisaron cuatro tipos de medidas: de tendencia central, de dispersión, de asimetría y de apuntamiento; y

b) las medidas que permiten la descripción y análisis de los cambios que experimentan esas observaciones en el transcurso del tiempo. Esto exigió en primer lugar el estudio de algunos métodos para cuantificar la magnitud de esos cambios, agrupados en general bajo el concepto de números índices; y luego propiamente la determinación estadística de los componentes de una serie cronológica; tendencia, fluctuaciones estacionales, variaciones cíclicas y accidentales.

Ha habido también necesidad de examinar algunos instrumentos imprescindibles para distintas fases del análisis, tales como los criterios y técnicas para el ajuste de curvas.

Corresponderá examinar ahora un tercer aspecto básico: el de la descripción y análisis de las relaciones existentes entre dos o más conjuntos de observaciones. Mientras el propósito de los temas anteriores ha sido la consideración de una sola variable, el análisis de correlación se ocupa del exámen conjunto de dos o más variables en forma simultánea.

Lo que en definitiva se propone la correlación es ofrecer los instrumentos estadísticos que permitan dar respuesta a dos interrogantes: si las variables que se están considerando están o no relacionadas entre sí; y si lo están, qué forma puede razonablemente aceptarse para esa relación.

Podrían también adelantarse algunos problemas prácticos en que podría ser necesario contestar a preguntas de esta índole:

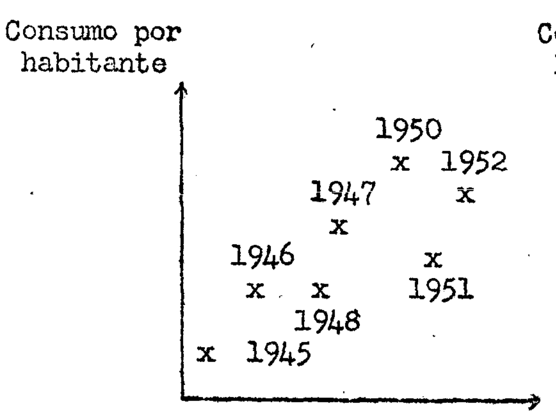
a) explicar las variaciones experimentadas por una variable dada durante un cierto período, en términos de las fluctuaciones mostradas por otras variables a las que pueda estar asociada. Por ejemplo, el consumo

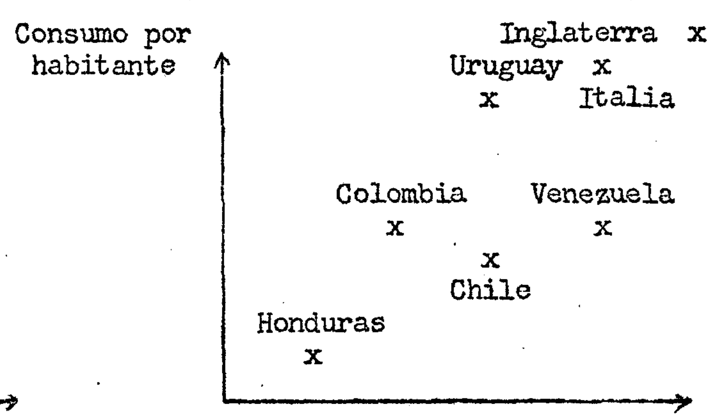
/ de un producto

de un producto cualquiera puede haberse modificado como consecuencia de un aumento en el nivel de ingreso, o de variaciones de sus precios relativos, de un incremento en la actividad de ciertas industrias, etc. El problema consistiría en tal caso en determinar estadísticamente con cuales de estas variables han estado asociados los cambios de ese consumo, o en que medida lo han estado con cada una de ellas. En otras palabras, el análisis procura poner de manifiesto cuáles son las variables determinantes experimentadas por cierta función; y,

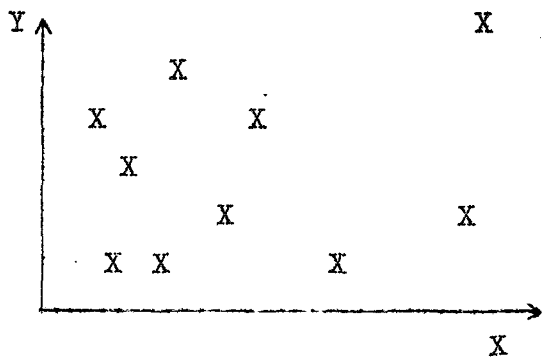
b) la finanlidad puede no ser un análisis histórico, sino obtener indicaciones de la influencia que puede tener un cambio futuro de una cierta variable sobre otras magnitudes importantes a las que está asociada. Si se admite, por ejemplo, que el ingreso por habitante crecerá durante la próxima década en una cierta proporción, sería importante poder adelantar qué modificaciones habría que esperar en la demanda de bienes y servicios de consumo (en cuánto podría aumentar la demanda de ciertos alimentos o determinadas manufacturas), en las necesidades de importación, etc.

Admítase pues que se requiere examinar el problema de calificar y cuantificar la asociación existente entre dos variables cualesquiera _X_ e _Y_. Un primer paso en el análisis podría consistir en la construcción de un diagrama de dispersión, en el que medirían los valores que simultáneamente han registrado las dos variables en cada período (o lugar): por ejemplo, el ingreso y el consumo de cierto bien en un país determinado durante varios períodos sucesivos, o el ingreso y el consumo en varios paises en un período dado:

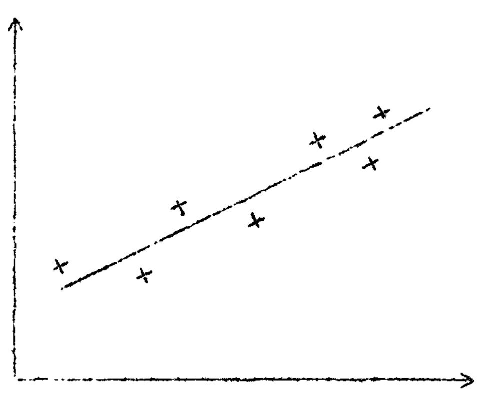


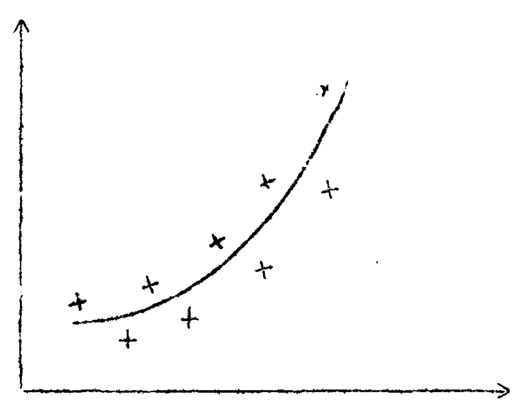


Ya esta representación podría proporcionar algunas indicaciones: si las variables fueran completamente independientes, un valor cualquiera de una de las variables tenderá a estar asociada tanto con valores pequeños como elevados de la otra variable, como si los puntos se hubieran ubicado al azar:



Si, en cambio, existe alguna relación entre las variables, los puntos tenderían a colocarse en torno a cierta línea:





Para los efectos del análisis de correlación, esta línea que ilustra la relación entre las dos variables se designa "línea de regresión", y queda definida mediante una "ecuación de regresión", la que puede determinarse mediante un ajuste siguiendo el método de los mínimos cuadrados.

La ecuación de regresión indica una forma de la relación, cuya elección está por lo demas sujeta a cierto grado de arbitrariedad. Pero no indica el grado de asociación que existe entre las dos variables que se está considerando. En la ecuación de regresión, en efecto, no hay ningún elemento que indique el grado de error que cometeríamos al tomar los valores dados por la

/ ecuación de

ecuación de regresión (los valores "calculados") en lugar de los valores efectivos. Por otra parte, una misma variable podría relacionarse con muchas otras variables; en cada caso se obtendría una ecuación de regresión determinada, pero hasta este momento no podríamos cuantificar con cual de esas variables está relacionada en forma más estrecha.

El propósito del análisis de correlación es justamente determinar el grado de asociación que existe entre determinadas variables, las que se suponen relacionadas en cierta forma. Al resultado de la medición se designa coeficiente o índice de correlación.

Cabe advertir desde ya que el coeficiente de correlación sólo mide el grado de asociación existente entre las variables consideradas, pero en ningún caso constituye una indicación de la naturaleza de la relación existente entre esas variables. En otras palabras, no prueba que exista una relación de causalidad entre ellas; es completamente neutro desde este punto de vista.

En efecto, dos o más variables pueden aparecer relacionadas por varias razones:

a) puede existir una relación de causalidad directa, de modo que las variaciones de una puedan ser consecuencia directa de las modificaciones de las otras. A la variable que está determinada por las modificaciones de las otras se designará "variable dependiente"; a las que la determinan "variables independientes";

b) la asociación puede estar motivada no por la acción directa de una variable sobre la otra, sino por el hecho de que las dos variables consideradas están influídas por una tercera, no considerada explícitamente. El análisis puede mostrar así un alto grado de asociación - y por lo tanto un elevado coeficiente de correlación - sin que exista realmente una relación de causalidad. En un período de inflación, por ejemplo, muchas series expresadas en términos monetarios muestran variaciones similares, aunque no estén relacionadas en absoluto;

c) pueden presentarse relaciones de causalidad, pero de carácter reversible, de modo que no sea una variable necesariamente "dependiente" y las otras "independientes". Tal podría ser el caso, por ejemplo, de las relaciones entre oferta, demanda y precio;

/ d) finalmente

d) finalmente, puede darse el caso de que la asociación entre dos variables sea puramente accidental. Pueden aparecer estrechamente ligadas durante cierto período - lo que daría como resultado un alto coeficiente de correlación - y luego diferir completamente en períodos siguientes, bajo condiciones similares.

Las consideraciones anteriores obligan a considerar dos fases en el análisis de correlación:

a) un exámen preliminar, basado en consideraciones lógicas, consistentes con la teoría económica o los conocimientos que se tengan sobre el fenómeno de que se trate, destinado a seleccionar variables que puedan razonablemente suponerse asociadas; y

b) el análisis propiamente estadístico, destinado a confirmar o rechazar esos supuestos, y a cuantificar el grado de asociación que exista.

La primera etapa permitiría adelantar una lista de variables que pueden considerarse como determinantes de una variable dependiente determinada, la segunda indicaría con cuáles está efectivamente asociada, y en qué magnitud lo está con cada una de ellas.

Tipos de correlación.

La correlación puede clasificarse atendiendo a dos elementos: la forma de la asociación que se admita entre las variables y el número de variables independientes que se considere.

Dos variables pueden aparecer relacionadas en muy distinta forma (lineal, exponencial, logarítmica, etc.), de modo que desde este punto de vista podrían distinguirse numerosos tipos de correlación. Se acostumbra, sin embargo, a clasificarlos en dos: correlación lineal (o mejor rectilínea) y correlación no lineal.

Por otra parte, una misma variable dependiente puede relacionarse con una sola variable independiente o simultáneamente con varias. De ahí que se distinga entre: correlación simple - una sola variable independiente - y correlación múltiple - dos o más variables independientes.

Se examinará también un tipo especial de correlación, designado correlación parcial, en el que se considera una variable dependiente y una variable independiente, pero después de eliminar estadísticamente la influencia de una o más variables independientes adicionales.

/ El concepto

El concepto de correlación. Correlación simple y lineal.

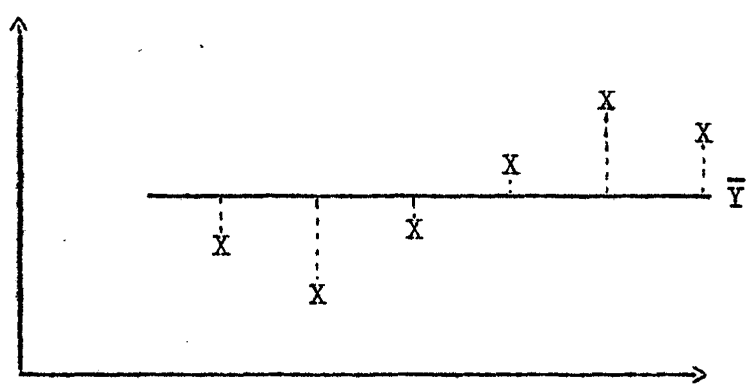
Pueden seguirse diversos caminos para ilustrar los conceptos básicos del análisis de correlación. En términos simples, tómese como punto de partida la consideración de que lo que se propone es explicar las variaciones experimentadas por una variable dependiente cualquiera. En tal caso, habrá que comenzar por precisar cómo ha variado, en cuánto ha variado; para ello, podría compararse cada uno de los valores que ha tenido con la media aritmética de los mismos. En otras palabras, se puede comenzar por cuantificar la medida en que los distintos valores han diferido de la media aritmética. Esto puede hacerse mediante el concepto ya conocido de varianza:

$$S_y^2 = \frac{\sum (Y_i - \bar{Y})^2}{n}$$

en que S_y sería la deviación típica o standard.

Para los efectos del análisis de correlación, a este concepto de varianza se designará varianza total.

Magnitudes que determinan la varianza total:



Admítase ahora que esas variaciones hayan estado determinadas o influídas en cierta medida por las variaciones experimentadas por otra variable a la que está asociada. Se podría examinar entonces cuánto de esa varianza total se puede "explicar", por la relación existente entre las dos variables. Al introducir la línea de regresión que ilustra esa relación, las diferencias entre cada valor de Y con repecto a $\bar{Y}$ pueden dividirse en dos partes:

$$Y_i - \bar{Y} = (Y_c - \bar{Y}) + (Y_i - Y_c)$$

en que Y_c son los valores calculados dados por la ecuación de regresión.

/ La primera parte

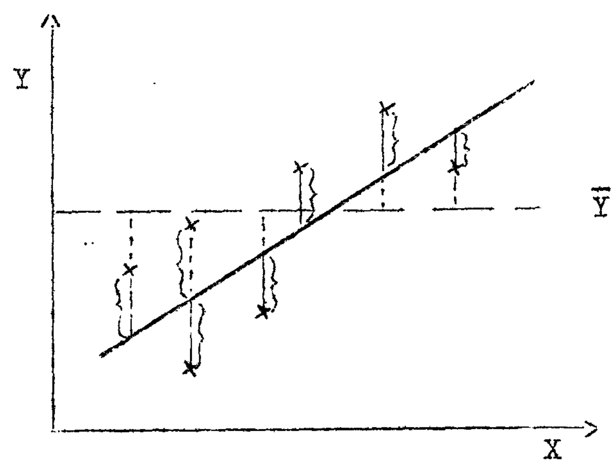
La primera parte puede considerarse como la parte "explicada" de esas diferencias por la relación entre X e Y; la otra, como una parte que esa relación no alcanza a explicar. Puede definirse así una <u>varianza explicada</u> y una <u>varianza no explicada</u>:

$$\text{varianza explicada:} \quad S_y^2 = \frac{\sum (Y_c - \bar{Y})^2}{n}$$

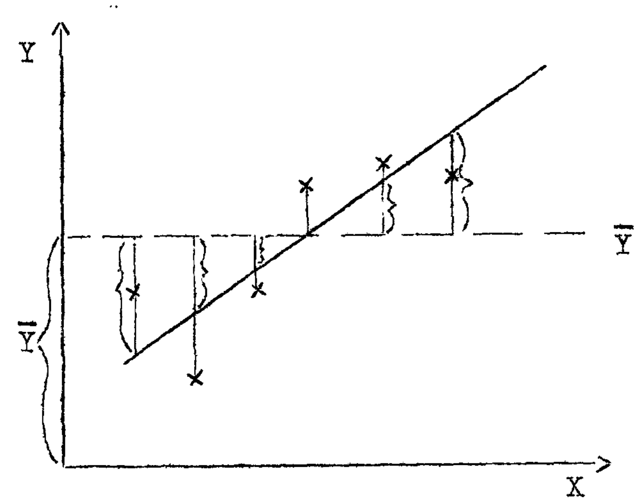
$$\text{varianza no explicada:} \quad S_{y_s}^2 = \frac{\sum (Y_i - Y_c)^2}{n}$$

He aquí las ilustraciones correspondientes:

Magnitudes que determinan la varianza no explicada:



Magnitudes que determinan la varianza explicada:



Por otra parte, la varianza total es igual a la suma de la varianza explicada y de la varianza no explicada. Es decir:

$$S_y^2 = S_{y_c}^2 + S_{y_s}^2$$

/ Esto puede

Esto puede demostrarse de la siguiente manera:

Se quiere probar que:

$$\sum (Y_i - \bar{Y})^2 = \sum (Y_c - \bar{Y})^2 + \sum (Y_i - Y_c)^2$$

Desarrollando: $\sum Y_i^2 - 2\bar{Y} \sum Y_i + n\bar{Y}^2 = \sum Y_c^2 - 2\bar{Y} \sum Y_c + n\bar{Y}^2 + \sum Y_i^2$
$- 2 \sum Y_i Y_c + Y_c^2$

o sea: $2 \sum Y_i Y_c - 2\bar{Y} \sum Y_i = 2 \sum Y_c^2 - 2\bar{Y} \sum Y_c$

si la ecuación de regresión es una recta, recuérdese que: $Y_c = aX + b$; $\sum Y_c = a\sum X + nb$

y la primera ecuación normal $\sum Y_i = a \sum X + nb$

luego: $\sum Y_c = \sum Y_i$

y sólo quedaría por probar: $2 \sum Y_i Y_c = 2 \sum Y_c^2$

de la definición de Y_c se deduce: $Y_c^2 = a^2 \sum X^2 + 2ab \sum X + nb^2$

$$\sum Y_c^2 = a^2 \sum X^2 + ab \sum X + ab \sum X + nb^2$$

$$\sum Y_c^2 = a \left[a \sum X^2 + b \sum X \right] + b \left[a \sum X + nb \right]$$

Según las ecuaciones normales: $\sum Y_c^2 = a \sum X_i Y_i + b \sum Y_i$

Por otra parte: $\sum Y_i Y_c = \sum Y_i \left[aX_i + b \right]$

$$\sum Y_i Y_c = a \sum X_i Y_i + b \sum Y_i$$

luego: $\sum Y_c^2 = \sum Y_i Y_c$

Mientras más estrecho sea el grado de asociación entre X e Y, mayor será la parte de la varianza total que queda explicada por la línea de regresión. Es precisamente en la relación que exista entre la varianza total y la varianza explicada en lo que se apoya todo el análisis de correlación.

/ Se define

Se define así un <u>coeficiente o índice de determinación</u>:

$$r^2 = \frac{s^2_{y_c}}{s^2_y}$$

y el coeficiente de correlación: $r = \pm\sqrt{\dfrac{s^2_{y_c}}{s_y^{\,2}}}$

o lo que es lo mismo: $r = \pm\sqrt{1 - \dfrac{s^2_{y_s}}{s_y^{\,2}}}$

En otras palabras, el coeficiente de determinación mide la proporción de la varianza total que queda "explicada" por la ecuación de regresión que ilustra la relación existente entre las dos variables. Otra forma de expresar lo mismo sería que indica el porcentaje de las variaciones de la variable dependiente que es atribuible a la influencia de la variable independiente.

De acuerdo con las definiciones anteriores:

$$r = \sqrt{\frac{\sum (Y_c - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}} \quad \text{o bien} \quad r = \sqrt{1 - \frac{\sum (Y_i - Y_c)^2}{\sum (Y_i - \bar{Y})^2}}$$

Si todas las observaciones quedaran ubicadas justamente en la línea de regresión, la asociación entre las dos variables sería precisa. Esto significaría una correlación perfecta. En tal caso:

$$Y_c = Y_i$$

y por lo tanto $r = \pm 1$

Se desprende pues que el valor máximo a que puede alcanzar el coeficiente de correlación es igual a la unidad.

En el sentido contrario, si no hubiera relación alguna entre las dos variables - o sea una "ausencia total de correlación" - la línea de regresión coincidiría con una paralela al eje de las X trazadas a una distancia $\bar{Y}$: en tal caso:

$$Y_c = \bar{Y} \qquad r = 0$$

/ y coincidirían

y coincidirían la varianza total y la no explicada. El valor que resultaría para el coeficiente de correlación sería cero. Se concluye así que el coeficiente de correlación puede variar desde un valor de cero, cuando hay ausencia total de correlación, hasta un valor de ± 1, cuando la correlación entre las dos variables es perfecta.

El signo del coeficiente de correlación se atribuye arbitrariamente, en forma tal que indique si las variables se modifican en la misma dirección o en sentido opuesto. Si los aumentos de la variable independiente determinan también aumentos de la variable dependiente, se le atribuye aigno positivo; en caso sontrario, signo negativo. Se habla así de correlación positiva y de correlación negativa. Entre la demanda de un bien y el ingreso, por ejemplo, cabría esperar una correlación positiva; entre la demanda y los precios relativos del mismo, correlación negativa.

Tratandose de correlación rectilínea, el signo del coeficiente de correlación coincidirá por lo tanto con el del coeficiente angular de la recta de regresión. En la correlación no lineal no siempre puede atribuirse un signo al coeficiente de correlación, ya que la variación de las dos variables puede ser en el mismo sentido durante cierto intervalo y en sentido inverso en otro intervalo, como ocurriría si se aceptara una línea de regresión parabólica. Algo similar sucede en el caso de la correlación múltiple, ya que la correlación de la variable dependiente podría ser positiva con respecto a una variable independiente y negativa con respecto a otra. Por eso, lo que en definitiva importa es el valor absoluto del coeficiente de correlación, y no su signo.

Error standard de estimación.

Como se recordará, en el análisis de distribuciones de frecuencia se definió el concepto de varianza (aquí utilizado como "varianza total") y el de desviación típica, o standard:

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

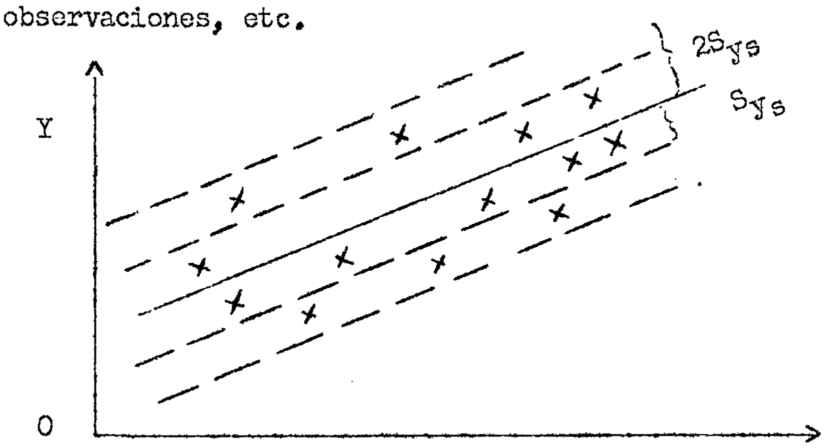
/ Esta última

Esta última constituía una indicación de la magnitud de la dispersión de los valores de la variable con respecto a la media aritmética. Se indicó también que en el caso de una distribución normal en el intervalo $\bar{X} \pm 2S$ quedaban comprendidas un 95.4% de las observaciones, en el intervalo $\underline{X} \pm 3S$ quedaban comprendidas un 99.7% de las observaciones.

Un concepto similar al de desviación standard se puede hacer extensivo al análisis de correlación, pero interesados esta vez no en la dispersión en torno a la media, sino en <u>la dispersión en torno a la línea de regresión</u>. Puesto que las desviaciones de cada valor de la variable con respecto a los valores calculados dados por la ecuación de regresión eran las magnitudes que daban origen a la varianza no explicada, puede así definirse el concepto de "error standard de estimación":

$$S_{y_s} = \sqrt{\frac{(Y_i - Y_c)^2}{n}}$$

de interpretación similar a la desviación standard, En otras palabras, cabe esperar que en los intervalos dados por la ecuación de regresión más o menos el error standard de estimación queden comprendidos alrededor de los dos tercios de las observaciones, etc.



Si se utiliza la ecuación con fines de proyección, el concepto descrito puede usarse como una indicación de probabilidad: puede así decirse que hay una probabilidad de alrededor de dos tercios de que el valor efectivo que asuma la variable en el futuro esté comprendido en el intervalo dado por la ecuación de regresión más o menos el error standard. Esto explica el nombre de "error standard de estimación" con que se le designa.

En resumen, hasta el momento la correlación estaría proporcionando los siguientes instrumentos de análisis:

a) una ecuación de regresión, la que indica la forma de relación existente entre las dos variables y cuyos parámetros permitirían estimar el valor de la variable dependiente que correspondería a un valor cualquiera de la variable independiente;

b) un coeficiente de correlación, el que indicaría el grado de asociación existente entre las dos variables. Serviría por lo tanto: i) para apreciar hasta qué punto es razonable que consideremos a una variable independiente dada como elemento de explicación o estimación de la variable dependiente; ii) para seleccionar entre varias variables independientes aquellas que mejor puedan considerarse como elementos de explicación o estimación de la variable dependiente; y

c) un error standard de estimación, el que indicaría la probable magnitud del error que se comete al aceptar los valores dados por la ecuación de regresión. Podría por lo tanto utilizarse para estimar no un valor único para la variable dependiente (dado por la ecuación de regresión) sino más bien un intervalo dentro del cual habría un grado razonable de probabilidad de que estuviera ubicado.

El cálculo práctico del coeficiente de correlación simple lineal.

Si se dispusiera de un conjunto de observaciones simultáneas para las dos variables, el cómputo del coeficiente de correlación exigirá de acuerdo con las definiciones anteriores de las siguientes etapas:

a) la determinación de la media aritmética de los valores de la variable dependiente y el cálculo de las diferencias con respecto a los valores observados;

b) la determinación de la ecuación de regresión, de los valores calculados de la variable dependiente dados por la misma y de sus diferencias con respecto a los valores observados y a su media aritmética;

c) la determinación de las sumas de los cuadros de las diferencias obtenidas, de las que podrían luego computarse el coeficiente de correlación y el error standard de estimación.

/ En otras palabras

En otras palabras, sería necesario tabular los siguientes valores:

X_i	Y_i	X_i^2	X_iY_i	Y_c	$Y_i - \bar{Y}$	$Y_c - \bar{Y}$	$Y_i - Y_c$	$(Y_i - \bar{Y})^2$	$(Y_c - \bar{Y})^2$	$(Y_i - Y_c)^2$

La tabulación de X_i, X_i^2 y X_iY_i sería necesaria para formular las ecuaciones normales que permitirían obtener la ecuación de regresión.

El procedimiento resulta así un poco laborioso, pero presenta la ventaja con respecto a métodos más abreviados de permitir el cómputo de todos los conceptos descritos (varianza total, explicada y no explicada; error standard).

Un método más abreviado puede deducirse partiendo de la definición inicial del coeficiente de determinación:

$$r^2 = \frac{\sum (Y_c - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

Desarrollando la sumatoria en numerador y denominador:

$$r^2 = \frac{\sum Y_c^2 - 2\bar{Y} \sum Y_c + n\bar{Y}^2}{\sum Y_i^2 - 2\bar{Y} \sum Y_i + n\bar{Y}^2}$$

Pero:

$$\bar{Y} = \frac{\sum Y_i}{n}$$

$$n\bar{Y} = \sum Y_i$$

$$n\bar{Y}^2 = \bar{Y} \sum Y_i$$

Por otra parte:

$$Y_c = aX + b$$

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

y siendo esta la primera ecuación normal:

$$\sum Y_c = \sum Y_i$$

/ Del mismo modo

Del mismo modo:

$$Y_c^2 = (aX + b)^2$$

$$\sum Y_c^2 = a^2 \sum X_i^2 + 2\,ab \sum X_i + nb^2$$

$$\sum Y_c^2 = a^2 \sum X_i^2 + ab \sum X_i + ab \sum X_i + nb^2$$

$$\sum Y_c^2 = a\,(a \sum X_i^2 + b \sum X_i) + b\,(a \sum X_i + nb)$$

en que los paréntesis son iguales a los segundos miembros de las dos ecuaciones normales. Por lo tanto:

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

Haciendo ahora estas sustituciones en la ecuación inicial:

$$r^2 = \frac{a \sum X_i Y_i + b \sum Y_i - \bar{Y} \sum Y_i}{\sum Y_i^2 - \bar{Y} \sum Y_i}$$

y finalmente:

$$r = \pm \sqrt{\frac{a \sum X_i Y_i + b \sum Y_i - \bar{Y} \sum Y_i}{\sum Y_i^2 - \bar{Y} \sum Y_i}}$$

Para un cómputo bastará con tabular:

X_i	Y_i	X_i^2	$X_i \cdot Y_i$	Y_i^2

donde X_i y X_i^2 serían necesarios para el cálculo de los parámetros a y b. El cómputo del coeficiente mismo sería así más sencillo, pero en cambio no se obtendrían los valores de las varianzas y del error standard.

Un ejemplo de aplicación práctica.

Como ilustración de un ejemplo práctico del cálculo y utilización del análisis de correlación simple lineal se utilizará un problema extractado del Estudio Ecoćmico de América Latina, 1949, de la CEPAL.

El problema planteado consistía en determinar si el nivel de las exportaciones en América Latina para los Estados Unidos dependía más de las variaciones en el ingreso nacional en ese país o de los precios relativos de esas exportaciones en comparación con los precios de las exportaciones que Estados Unidos efectúa a América Latina (en otras palabras, de la relación de precios del intercambio entre las dos áreas).

/ Vale la pena

Vale la pena resumir la forma en que se prepararon las series básicas:

a) Las exportaciones latinoamericanas hacia los Estados Unidos podían obtenerse de las estadísticas de importación de este país o de las estadísticas de exportación de los paises latinoamericanos. En el último caso, había necesidad de convertir las cifras en moneda nacional de cada país a dólares, para obtener los totales de la región, utilizando los tipos de cambio correspondiente. En uno y otro caso se obtenían cifras en dólares corrientes, es decir a precios de cada año; pero lo que interesaba considerar era la magnitud de las exportaciones independientemente de las fluctuaciones de precios, es decir el volúmen físico o quantum de las mismas. Para ello, hubo necesidad de deflacionar las cifras corrientes por un índice de precios, llegandose en definitiva a obtener la serie expresada en dólares a precios constantes de un año determinado.

b) Para mantener la comparabilidad, se utilizó también la serie de ingreso nacional de los Estados Unidos en dólares a precios constantes (valores corrientes deflacionados por el índice deflactor implícito del ingreso en ese país).

c) Para la relación de precios del intercambio, era necesario disponer no sólo de un índice de precios de los productos exportados por América Latina a Estados Unidos, sino también de las exportaciones de éste haciá América Latina.

Para el primero se construyó un índice de precios de los 13 principales productos exportados por América Latina a Estados Unidos, ponderados de acuerdo con el valor en dólares de sus respectivas exportaciones en el período base; para el segundo se tenía el índice de precios de exportación de productos terminados de los Estados Unidos (Statistical Abstract of the United States).

Para simplificar los cálculos, las tres series básicas se expresarán en forma de índices con base en el año 1937. Los resultados correspondientes se resumen en el Cuadro 1 (a) adjunto, en el que:

Y = índice del volumen físico de las exportaciones de A.L. hacia EE.UU.

X_1 = índice del ingreso real en los EE.UU.

X_2 = Indice de la relación de precios del intercambio entre A.L. y EE.UU.

/ La relación

La relación entre X_1 e Y se ilustra en el Gráfico 1; entre X_2 e Y en el Gráfico 2. En el primer caso, óbviamente existe una correlación positiva; en el segundo, una correlación negativa.

Las columnas correspondientes a los valores de X_1^2, X_1Y, X_2^2, X_2Y del cuadro 1 (a) se han calculado para determinar las respectivas ecuaciones de regresión:

$$Y = a_1X_1 + b_1 -$$

$$Y = a_2X_2 + b_2 -$$

Los parámetros respectivos quedan entonces determinados por:

$$\sum Y = a_1 \sum X_1 + nb_1$$

$$\sum X_1Y = a_1 \sum X_1^2 + b_1 \sum X_1$$

$$3.231.4 = 3.471.6 \quad a_1 + 29\ b_1 \qquad / \cdot 119{,}71$$

$$432.428.0 = 460.381.0 \quad a_1 + 3.471.6\ b_1$$

$$386.830.9 = 415.585.2 \quad a_1 + 3.471.6\ b_1$$

$$432.428.0 = 460.381.0 \quad a_1 + 3.471.6\ b_1$$

$$45.597.1 = 44.795.8\ a_1$$

$$a_1 = 1.018$$

$$29\ b_1 = 3.231.4 - 3.471.6\ (1.018)$$

$$29\ b_1 = -302.7$$

$$b_1 = -10.4$$

$$\underline{Y = 1.018 \quad X_1 - 10.4}$$

$$\sum Y = a_2 \sum X_2 + nb_2$$

$$\sum X_2Y = a_2 \sum X_2^2 + b_2 \sum X_2$$

$$3.231.4 = 2.808.0\ a_2 + 29\ b_2 \qquad / \cdot 96{,}828$$

$$307.089.0 = 280.882.0\ a_2 + 2.808.0\ b_2$$

/ 312.890 -

$$312.890 = 271.893\ a_2 + 2.808\ b_2$$
$$307.089 = 280.882\ a_2 + 2.808\ b_2$$

$$5.801 = -\ 8.989\ a_2.$$
$$a_2 = -\ 0.6453$$

$$29\ b_2 = 3.231.4 + 1.812.0$$
$$29\ b = 5.043.4$$
$$b_2 = 173,9$$

$$Y = -\ 0,6453\ \ X_2 + 173,9$$

Los valores calculados dados por las 2 ecuaciones de regresión se presentan en las dos últimas columnas del cuadro 1 (a). En el cuadro 1 (b) se han tabulado las diferencias correspondientes y en el cuadro 1 (c) los cuadrados de las mismas.

Esas tabulaciones permitirían ahora calcular los diferentes conceptos:

a) Ingreso y exportaciones:

Varianza total: $s_y^2 = \dfrac{\sum (Y_i - \bar{Y})^2}{n}$

$s_y^2 = \dfrac{51.842}{29} = \underline{1.787.6}$

Desviación standard: $s_y = 42.2$

Varianza explicada: $s^2_{y_c} = \dfrac{\sum (Y_c - \bar{Y})^2}{n}$

$s^2_{y_c} = \dfrac{46.520}{29} = \underline{1.604.1}$

Varianza no explicada: $s^2_{y_s} = \dfrac{\sum (Y_i - Y_c)^2}{n}$

$s^2_{y_s} = \dfrac{5.449}{29} = \underline{187.9}$

Error standard: $s_{y_s} = \underline{13.7}$

/ Coeficiente

Coeficiente de determinación:
$$r^2 = \frac{s^2_{y_c}}{s^2_y}$$

$$r^2 = \frac{1.604.1}{1.787.6} = \underline{0.897}$$

Coeficiente de correlación: $r = \underline{0.947}$

b) Relación de precios del intercambio y exportaciones:

Varianza total (la misma) $s^2_y = \underline{1.787.6}$

Varianza explicada: $s^2_{y_c} = \frac{3.803}{29} = \underline{131.1}$

Varianza no explicada: $s^2_{y_s} = \frac{48.107}{29} = \underline{1.658.9}$

Error standard de estimación: $s_{y_s} = \underline{40.7}$

Coeficiente de determinación: $r^2 = \frac{131.1}{1.787.6} = \underline{0.073}$

Coeficiente de correlación: $r = - \underline{0.27}$

De lo anterior se concluiría que la influencia que pueden tener las variaciones de la relación de precios del intercambio sobre el quantum de exportaciones en América Latina a los Estados Unidos es prácticamente insignificante. En cambio, el quantum de esas exportaciones muestra una relación muy estrecha con las variaciones del ingreso real de los Estados Unidos. En consecuencia, es este último factor el que mejor puede contribuir a explicar las modificaciones en el quantum de las exportaciones latinoamericanas, y por lo tanto el que habría que utilizar de preferencia como base para una proyección del posible volumen futuro de esas exportaciones.

Admítase a título puramente ilustrativo que se quisieran utilizar las informaciones anteriores para formular una proyección del quantum de exportaciones latinoamericanas a los Estados Unidos en 1960.

/ El primer

El primer problema que se plantearía sería estimar el nivel de ingreso de los Estados Unidos en ese año. Si arbitrariamente se admite una tendencia lineal para el mismo, podría determinarse la ecuación correspondiente mediante las siguientes ecuaciones normales:

$$\sum X_1 = a \sum X + nb$$
$$\sum X_1X = a \sum X^2 + b \sum X$$

en las que X_1 representa como antes el índice del ingreso real y X el tiempo medido en años y con origen en 1921. Efectuando los cálculos correspondientes se obtendría:

$$\sum X_1 = 3.471.6$$
$$\sum X = 406$$
$$\sum X^2 = 7,714$$
$$\sum X_1X = 56.335.8$$

y el sistema :

$$48.602.4 = 5.684\ a + 406\ b$$
$$56.335.8 = 7.714\ a + 406\ b$$

$$7.733.4 = 2.030\ a$$
$$a = 3.81$$

$$29\ b = 3.471.6 - 1.546.9 = 1.924.7$$
$$b = 66.4$$

$$X_1 = 3.81 + 66.4$$

Para 1960, $X = 39$; luego el índice del ingreso real de EE.UU. estaría dado por:

$$X_1 = (3.81) \cdot (39) + 66.4$$
$$X_1 = 215$$

Reemplazando ahora este valor en la ecuación de regresión entre quantum de exportaciones e ingreso, se obtendría:

$$Y = 1.018\ X_1 - 10.4$$
$$Y = 1.018 \cdot 215 - 10.4$$
$$Y = 218.9 - 10.4 = 208.5$$

/ Si en lugar

Si en lugar de formular esta proyección de un valor único se deseara más bien proyectar un intervalo dentro del cual existiera un grado razonable de probabilidad de que quedara comprendido el valor que se estima, podría utilizarse el error standard de estimación:

$$Y \begin{cases} & + 13.7 = \underline{222.2} \\ 208.5 & \\ & - 13.7 = \underline{194.8} \end{cases}$$

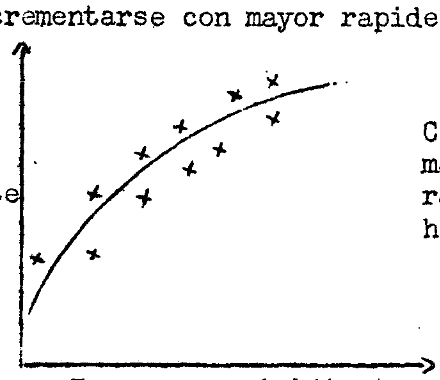
cifras que comparadas con el índice registrado en 1949 (170) indicarían un crecimiento probable hasta 1960 que fluctuaría entre 14,6 y 30,7 por ciento.

<u>Correlación no lineal.</u>

En lo anterior se ha supuesto que las variables están asociadas en la forma más simple, en la que el crecimiento unitario de la variable independiente determina siempre crecimientos de igual magnitud absoluta en la variable dependiente. En la práctica, este tipo de asociación no es siempre la más adecuada y las variables pueden aparecer relacionadas en muy diversas formas.

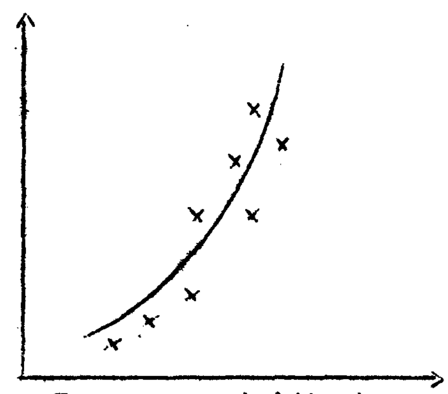
Por ejemplo, ante un aumento dado del ingreso, el consumo por habitante de productos alimenticios tiende en general a crecer con menor intensidad (y el de algunos productos alimenticios incluso a estabilizarse en determinado nivel); en cambio, el consumo de productos manufacturados o de servicios tiende a incrementarse con mayor rapidez.

Consumo de alimentos por habitante



Ingreso por habitante

Consumo de manufacturas por habitante



Ingreso por habitante

En tales casos, una hipótesis de relación lineal no sería adecuada.

/ Si se trata

Si se trata de escoger una relación no lineal, hay naturalmente múltiples posibilidades de elección. Tales posibilidades están sin embargo limitadas por la relación entre el número de observaciones disponibles, el número de parámetros de la ecuación de regresión y la significación real que tendría el coeficiente de correlación. Esto se explica por el hecho de que si sólo se dispusiera de 3 observaciones y se utilizara una ecuación de regresión de segundo grado, ésta pasaría por los tres puntos observados; en tal caso, resultaría una correlación perfecta, aún si las variables no tuvieran porque estar asociadas en absoluto. En otras palabras, el coeficiente de correlación no lineal es tanto más significativo cuanto mayor sea el número de observaciones en que se basa, y tanto menos significativo cuanto mayor el número de parámetros de la ecuación de regresión.

Como en la práctica el número de informaciones está dado o por la disponibilidad de informaciones estadísticas básicas, la significación sólo podrá incrementarse operando con curvas relativamente sencillas. La solución más común consiste en reducir las formas no lineales a una forma lineal mediante el uso de logaritmos. Se plantean así varias posibilidades para la ecuación de regresión en forma que ésta contenga sólo dos parámetros.

1) $Y = aX + b$

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

3) $\log Y = aX + b$

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

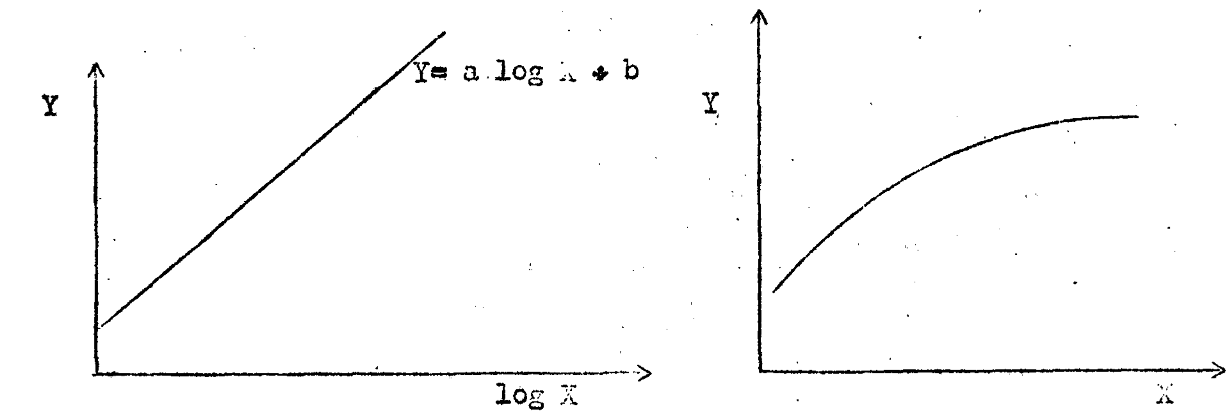
La ecuación 1) indica una relación entre variaciones absolutas de X y variaciones absolutas de Y (cada vez que X varía en una cierta cantidad, Y varía en una cantidad determinada).

La ecuación 2) indica una relación entre variaciones relativas de X y variaciones absolutas de Y (cada vez que X varía en un cierto porcentaje, Y varía en una cantidad determinada).

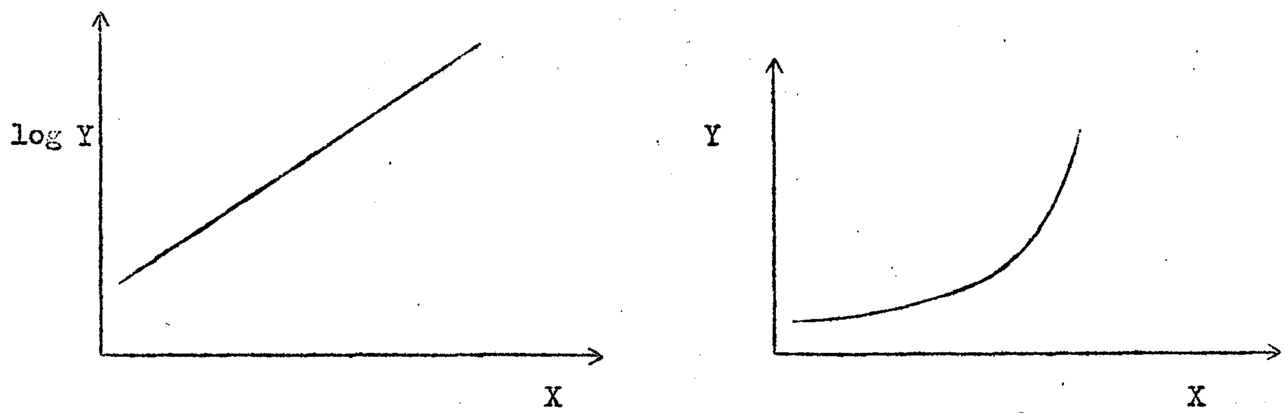
La ecuación 3) indica una relación entre variaciones absolutas de X y variaciones relativas de Y (cada vez que X varía en una cierta cantidad, Y varía en un porcentaje determinado).

La ecuación 4) indica una relación entre variaciones relativas de X y variaciones relativas de Y (cada vez que X varía en cierto porcentaje, Y varía en un porcentaje determinado).

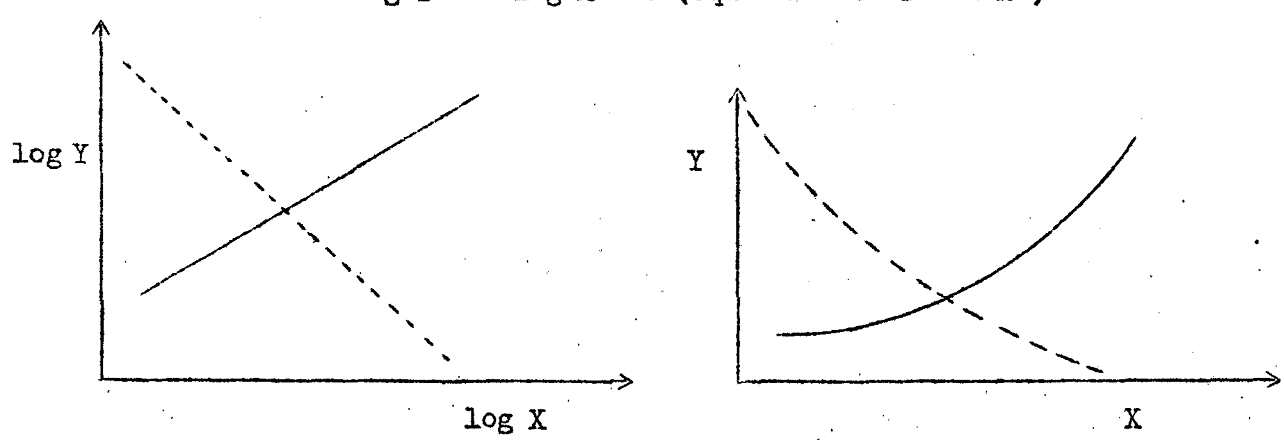
/ (Gráficos)



$\log Y = aX + b$ (equivale a: $Y = k \cdot 10^{aX}$)



$\log Y = a \log X + b$ (equivale a: $Y = k\, X^{a}$)



Si ninguna de estas expresiones concordará con la forma de la relación que muestran las dos variables, podrían utilizarse ecuaciones de regresión de segundo o tercer grado.

/ En cualquier

En cualquier caso, el coeficiente de correlación se deduce de conceptos absolutamente similares a los aplicados en la correlación lineal, es decir de la relación de varianza explicada y varianza total. Es este caso, el coeficiente de determinación:

$$\rho^2 = \frac{s^2_{y_c}}{s^2_y} \quad \text{o sea} \quad \rho^2 = \frac{\sum (Y_c - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

Para la determinación de Y_c, habría que computar las ecuaciones normales correspondientes, que serán

para la expresión 2):
$$\sum Y_i = a \sum \log X_i + nb$$
$$\sum Y_i \log X_i = a \sum (\log X_i)^2 + b \sum \log X_i$$

para la ecuación 3):
$$\sum \log Y_i = a \sum X_i + nb$$
$$\sum X_i \log Y_i = a \sum X_i^2 + b \sum X_i$$

para la ecuación 4):
$$\sum \log Y_i = a \sum \log X_i + nb$$
$$\sum \log X_i \log Y_i = a \sum (\log X_i)^2 + b \sum \log X_i$$

Del método abreviado para el cómputo del coeficiente de correlación lineal se dedujo que:
$$n\, s^2_{y_i} = \sum Y_i^2 - \bar{Y} \sum Y_i$$
y que:
$$n\, s^2_{y_c} = a \sum X_i Y_i + b \sum Y_i - \bar{Y} \sum Y_i$$

La varianza total y la varianza explicada pueden deducirse por analogía con lo anterior. Las expresiones correspondientes a cada una de las ecuaciones de regresión que se están considerando para el cómputo abreviado del coeficiente de determinación no lineal, serían las siguientes:

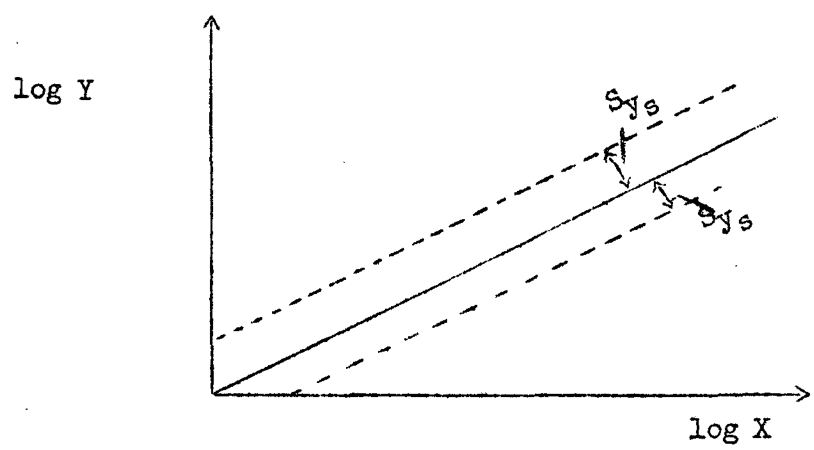
Para la ecuación 2)
$$\rho^2 = \frac{a \sum \log X_i \,.\, Y_i + b \sum Y_i - \bar{Y} \sum Y_i}{\sum Y_i^2 - \bar{Y} \sum Y_i}$$

Para la ecuación 3):
$$\rho^2 = \frac{a \sum X_i \log Y_i + b \sum \log Y_i - \overline{\log Y} \sum \log Y_i}{\sum (\log Y_i)^2 - \overline{\log Y} \sum \log Y_i}$$

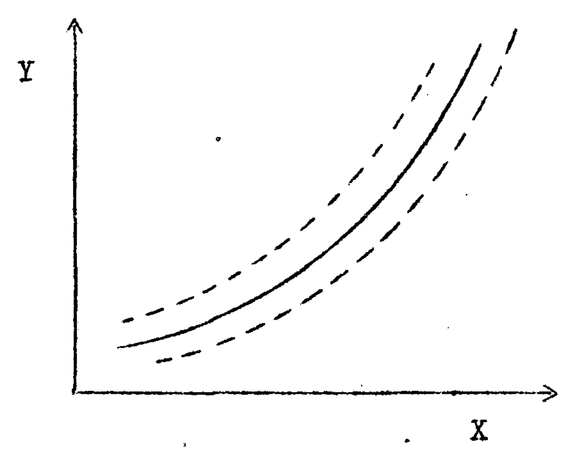
y para la ecuación 4):

$$\rho^2 = \frac{a \sum \log X_i \log Y_i + b \sum \log Y_i - \overline{\log Y} \sum \log Y_i}{\sum (\log Y_i)^2 - \overline{\log Y} \sum \log Y_i}$$

El concepto y cálculo del error standard de estimación serían también similares a lo expuesto en el caso de la correlación lineal. Es preciso examinar, sin embargo, la forma de la faja que queda definida por los intervalos $Y_i \pm S_{y_s}$. Si se utiliza una relación logarítmica (log Y = a log X + b) la faja será similar si se miden los logaritmos de las dos variables:



Pero no ocurriría igual si se representan los valores absolutos de las mismas, caso en el cual la faja se irá haciendo más amplia a medida que aumentan los valores de X, y el límite inferior estará más cerca de la línea de regresión en magnitudes absolutas:



/ En otras

En otras palabras, los valores de los límites conservarán una proporción y no una diferencia constante con respecto a los valores de Y_c. Esto puede comprobarse mediante un ejemplo numérico.

Los valores de los límites están dados por:

	antilogarit. $(\log Y_c + Z)$	y antilog $(\log Y_c - Z)$
Si $Y_c = 10$ y $Z = 0.1$	antilog $1.1 = 12.59$	antilog $0.9 = 7.94$
Si $Y_c = 100$ y $Z = 0.1$	antilog $2.1 = 125.9$	antilog $1.9 = 79.4$

Designando por S'_{y_s} el error standard en términos absolutos, se tendrá:

$Y_c - S'_{y_s}$	Y_c	$Y_c + S'_{y_{c_s}}$
7.94	10	12.59
79.4	100	125.9

Esta interpretación del error standard es perfectamente lógica, ya que es natural que el posible margen de error en términos absolutos sea tanto mayor cuanto más alto los valores de la variable. Un error standard de 10 sería grande si el valor de la variable independiente es de 50; en cambio el mismo margen de error de 10 sería insignificante para un valor de la variable dependiente igual a 1.000. Prácticamente, en este caso, se diría, por ejemplo, que hay alrededor de un 66 por ciento de probabilidad de que un valor futuro de la variable dependiente esté correspondido en el intervalo dado por Y_c más y menos un 26 por ciento de éste (en lugar de más y menos un cierto monto absoluto).

El método abreviado en una correlación de segundo grado.

Si se escoge una ecuación de segundo grado, el valor del coeficiente de determinación estará dado como en los casos anteriores por:

$$\rho^2 = \frac{\sum (Y_c - \bar{Y})^2}{\sum (Y_i - Y_c)^2}$$

El denominador será igual que antes: $\sum Y_i^2 - \bar{Y} \sum Y_i$

/ En cuanto

En cuanto al numerador: $\sum Y_c^2 - 2\bar{Y} \sum Y_i + n\,\bar{Y}^2$

o sea: $\sum Y_c^2 - \bar{Y} \sum Y_i$

Para Y_c :

$$Y_c = aX^2 + bX + c$$
$$Y_c^2 = (aX^2 + bX + c)^2$$

$$Y_c^2 = a^2X^4 + b^2X^2 + c^2 + 2\,abX^3 + 2\,acX^2 + 2\,bcX$$

y en términos de sumatoria:

$$\sum Y_c^2 = a^2 \sum X_i^4 + b^2 \sum X_i^2 + 2ab \sum X_i^3 + 2ac \sum X_i^2 + 2bc \sum X_i + nc^2$$

Descomponiendo y reordenando los términos:

$$\sum Y_c^2 = a^2 \sum X_i^4 + ab \sum X_i^3 + ac \sum X_i^2 + ab \sum X_i^3 + b^2 \sum X_i^2 + bc \sum X_i + ac \sum X_i^2 + bc \sum X_i + nc^2$$

Sacando factor común $\underline{a}$ entre los tres primeros términos, factor común $\underline{b}$ entre los tres siguientes y factor común $\underline{c}$ entre los tres últimos:

$$\sum Y_c^2 = a(a \sum X_i^4 + b \sum X_i^3 + c \sum X_i^2) + b(a \sum X_i^3 + b \sum X_i^2 + c \sum X_i) + c(a \sum X_i^2 + b \sum X_i + nc)$$

donde los tres paréntesis corresponden a los segundos miembros de las tres ecuaciones normales. Luego:

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

Resumiendo, podría anotarse entonces que:

$$\rho^2 = \frac{a \sum X_i^2 Y_i + b \sum X_i Y_i + c \sum X_i - \bar{Y} \sum Y_i}{\sum Y_i^2 - \bar{Y} \sum Y_i}$$

que sería una expresión para un cómputo abreviado del coeficiente de correlación cuando se supone una ecuación de regresión de segundo grado. Por analogía podrían deducirse expresiones similares para ecuaciones de regresión de grado más alto.

/ <u>La aplicación</u>

La aplicación práctica del concepto de elasticidad.

Si bien no relacionado directamente con el problema de correlación no lineal, es oportuno discutir aquí desde el punto de vista estadístico algunos aspectos relativos al concepto de elasticidad.

Se define un coeficiente de elasticidad como una medida de cambio proporcional experimentado por una variable Y con respecto al cambio proporcional de una variable X. Es decir:

$$\mathcal{E} = \frac{\frac{dy}{Y}}{\frac{dX}{x}} = \frac{dy}{Y} \cdot \frac{X}{dx} = \frac{dy}{dx} \cdot \frac{X}{Y} = \frac{X}{Y} \cdot Y'$$

Pero al comparar las variaciones relativas de las dos variables se están admitiendo implícitamente que éstas están asociadas de cierta manera, si bien en esa definición general no se indica explícitamente cuál es la forma de relación.

Además, se presenta aquí la necesidad de conciliar el significado matemático estricto de la definición con las necesidades de utilización práctica del concepto. Matemáticamente, el coeficiente permite determinar la variación proporcional de Y como consecuencia de cambios infinitamente pequeños de X; prácticamente, no interesan los cambios infinitesimales, sino aquellos cambios apreciables, en otras variables. En otras palabras, interesa la determinación de coeficientes de elasticidad que pueden aplicarse dentro de intervalos más o menos amplios.

Esto significa que habrá que seleccionar en la práctica formas de relación entre las dos variables que conducen a coeficientes de elasticidad constantes y no a valores del mismo que sean distintos para cada valor de X, ya que en tal caso no tendrían utilidad práctica alguna.

Si se utiliza una relación lineal: $Y = aX + b$

$Y' = a$

$$\mathcal{E} = a \cdot \frac{X}{Y} = a \cdot \frac{X}{aX + b}$$

$$\mathcal{E} = \frac{1}{1 + \frac{b}{aX}}$$

es decir que el valor de depende de los valores de X, y es por lo tanto distinto para cada valor de la variable independiente.

/ Si, en cambio

Si, en cambio, se utiliza una relación logarítmica entre las variables, se tendrá:

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

$$\frac{dy}{Y} = a \cdot \frac{dx}{X}$$

$$\frac{dy}{dx} = Y' = a \cdot \frac{Y}{X}$$

$$\mathcal{E} = a \cdot \frac{Y}{X} \cdot \frac{X}{Y}$$

$$\mathcal{E} = a$$

En este caso, se obtiene pues un coeficiente de elasticidad constante e igual al valor del parámetro a de la ecuación de regresión, que es la condición necesaria desde el punto de vista de las posibilidades de utilización práctica del concepto. Esta utilización estaría así condicionada a que pueda razonablemente admitirse una asociación de carácter logarítmico entre las dos variables; en otras palabras, sería condición necesaria que se comprobara un valor alto para el coeficiente de correlación ρ log X, log Y.

Las consideraciones anteriores exigen también una cierta forma de interpretar el valor del coeficiente de elasticidad. Frecuentemente se interpreta el coeficiente de elasticidad en la siguiente forma: si $\mathcal{E} = 1.2$, ello significa que una variación de 1 % en X determinará una variación de 1,2 % en Y; una variación de 10% en X determinará una variación de 12 % en Y; a una variación de 100% en X corresponderá una variación de 120 % en Y. Puede probarse sin embargo que en la medida en que se quiere ser consistente con la necesidad de operar con un coeficiente constante - tal interpretación no es rigurosamente exacta, especialmente para cambios grandes de X o para valores elevados de $\mathcal{E}$.

Admítase, por ejemplo, que se tenga la ecuación de regresión:

$$\log Y = 2 \log X + 0.1$$

en la que $\mathcal{E} = 2$

Tabulando para algunos valores de X:

X	log X	log Y	Y
1	0	0,1	1,26
1,1	0,04	0,18	1,514
2	0,3	0,7	5,01
3	0,477	1,054	11,33
4	0,602	1,304	20,14

/ y las

y las variaciones correspondientes:

	Aumentos porcentuales	
	X	Y
Si X varía de 1 a 1,1 :	10%	20,1%
Si X varía de 2 a 3 :	50%	121 %
Si X varía de 1 a 2 :	100%	300 %
Si X varía de 2 a 4 :	100%	300 %

De la ecuación de regresión se deducirá en cambio una interpretación estricta:

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

Si $X = X_1$: $\log Y_1 = a \log X_1 + b$

Si $X = X_2$: $\log Y_2 = a \log X_2 + b$

restando: $\log Y_2 - \log Y_1 = a(\log X_2 - \log X_1)$

de donde: $\dfrac{Y_2}{Y_1} = \left(\dfrac{X_2}{X_1}\right)^a$

y como $a = \mathcal{E}_i \quad \dfrac{Y_2}{Y_1} = \left(\dfrac{X_2}{X_1}\right)^{\mathcal{E}}$

Es decir que la variación unitaria de Y es igual a la variación unitaria de X elevada al coeficiente de elasticidad.

En el ejemplo anterior, si X varía de 1 a 4:

$$\frac{X_2}{X_1} = \frac{4}{1} = 4$$

Siendo $\mathcal{E} = 2$: $\left(\dfrac{X_2}{X_1}\right)^{\mathcal{E}} = 4^2 = 16$

luego: $\dfrac{Y_2}{Y_1} = 16$

$$Y_2 = 16 \cdot Y_1 = 16 \cdot 1,26$$

$$Y_2 = 20,16$$

/ Los métodos

<u>Los métodos de estimación de la elasticidad - ingreso de la demanda.</u>

Si bien en muchos casos puede ser necesario operar con el concepto de elasticidad-precio, en la práctica la utilización más frecuente corresponde al concepto de <u>elasticidad - ingreso de la demanda</u>. Vale la pena por lo mismo hacer aquí breve referencia a los métodos prácticos que podrían utilizarse para estimar coeficientes de esta índole.

En esencia, lo que se procura es obtener algunas indicaciones objetivas sobre la forma en que los consumidores distribuyen un aumento de su ingreso entre diversos tipos de bienes y servicios de consumo. En otras palabras, qué variación cabe esperar en el consumo de un bien determinado cuando el ingreso se modifica en cierta proporción.

En general, hay tres tipos de informaciones estadísticas básicas que pueden conducir a la estimación de coeficientes de esta índole:

a) <u>Estadísticas retrospectivas</u>. Los coeficientes de elasticidad - ingreso pueden deducirse de comparaciones históricas del nivel de ingreso y del consumo del bien de que se trate;

b) <u>Encuestas sobre ingreso y gastos familiares</u>. Para ciertos períodos, y generalmente con el propósito principal de establecer o revisar las bases de cálculos de los índices de costo de vida, se efectúan encuestas relativas a la composición de los gastos de nuestras familias. En tal caso, los coeficientes de elasticidad - ingreso de la demanda pueden estimarse comparando el gasto en un bien determinado de familias con distintos niveles de ingreso; y

c) <u>Comparaciones internacionales</u>. También los coeficientes pueden deducirse comparando para un período dado los gastos respectivos en paises con diferente nivel de ingreso.

Los tres tipos de comparaciones pueden ser útiles y no excluyentes, ya que cada uno puede aportar la consideración de determinados aspectos específicos importantes. Las series retrospectivas conducen a resultados de interpretación más completa, ya que dentro de los períodos considerados han ocurrido generalmente no sólo modificaciones del ingreso, sino también alteraciones de los precios relativos, modificaciones en la situación de oferta, etc. Desde este punto de vista, las comparaciones derivadas de encuestas efectuadas en cierto período son más homogéneas; en cambio, generalmente tales encuestas se circunscriben a ciertos sectores de la población, por ejemplo población

/ urbana

urbana, y no reflejan por lo tanto la incidencia sobre la demanda de la urbanización u otros factores, los que sí aparecen implícitamente considerados en las derivadas de estadísticas retrospectivas. Las comparaciones internacionales, finalmente, se ven a menudo influídas no sólo por el diferente nivel de ingreso, sino también por la estructura distinta de precios relativos, aparte de factores tales como hábitos o diferencias climáticas; en cambio, tienen la virtud de ofrecer una mayor perspectiva en el análisis de la demanda, mostrando por ejemplo los amplios cambios que suelen ocurrir en determinados consumos cuando un país pasa de un cierto nivel de ingreso a otro más elevado.

La ecuación de regresión y el coeficiente de elasticidad como instrumento de proyección.

Como se ha visto, la ecuación de regresión puede utilizarse directamente como instrumento para estimar valores futuros de la variable dependiente, una vez planteadas determinadas hipótesis sobre el comportamiento de la variable independiente. Cabría discutir entonces qué diferencia existiría entre utilizar la ecuación de regresión o el coeficiente de elasticidad como instrumento de proyección, por ejemplo, de la demanda de un bien en función del ingreso.

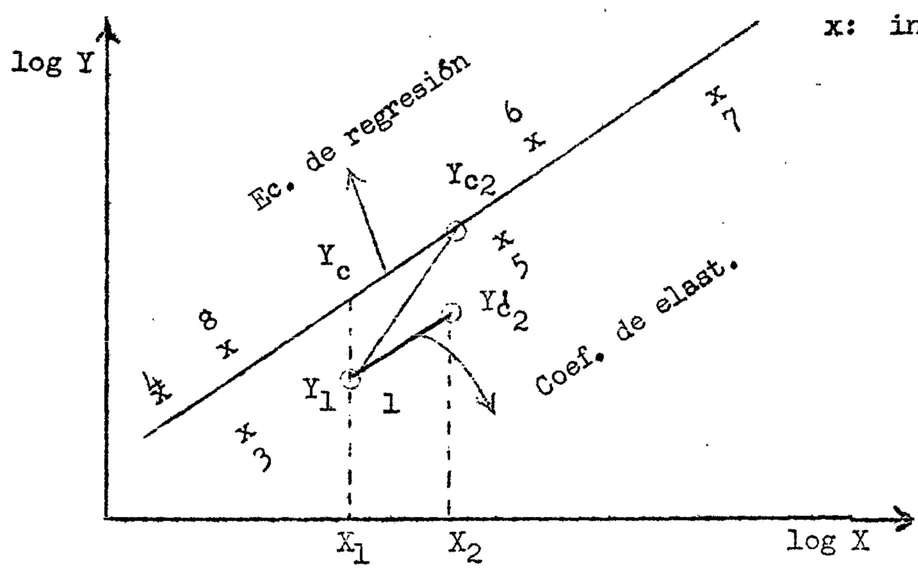
En primer término, la ecuación de regresión es de aplicabilidad mucho más general, ya que podría utilizarse cualquiera que fuese la forma de la relación que se admita entre las dos variables. Desde este punto de vista, el concepto de elasticidad no sería sino un caso particular, en el que como se ha visto se admite una relación logarítmica.

Podría suceder, sin embargo, que en determinados casos prácticos no pudiera disponerse de la ecuación de regresión correspondiente; en cambio, sí podría ser posible utilizar una estimación del coeficiente de elasticidad. Si sólo se tienen las cifras globales de consumo e ingreso para un solo período reciente, podrían por ejemplo, utilizarse coeficientes de elasticidad deducidos de las experiencias de otros países de condiciones similares.

/ Aún si se

Aún si se dispusiera de los dos instrumentos - ecuación de regresión y coeficiente de elasticidad-ingreso - y se admitiera una relación logarítmica, los resultados de las proyecciones a que conducirían uno y otro serían en la generalidad de los casos diferentes. Supóngase por ejemplo que las comparaciones correspondientes se hayan referido al consumo de determinado bien en paises con distinto nivel de ingreso (paises 1, 2, 3, ... en el gráfico siguiente, siendo el país 1 el que interesa para la proyección).

y: consumo por habitante
x: ingreso por habitante



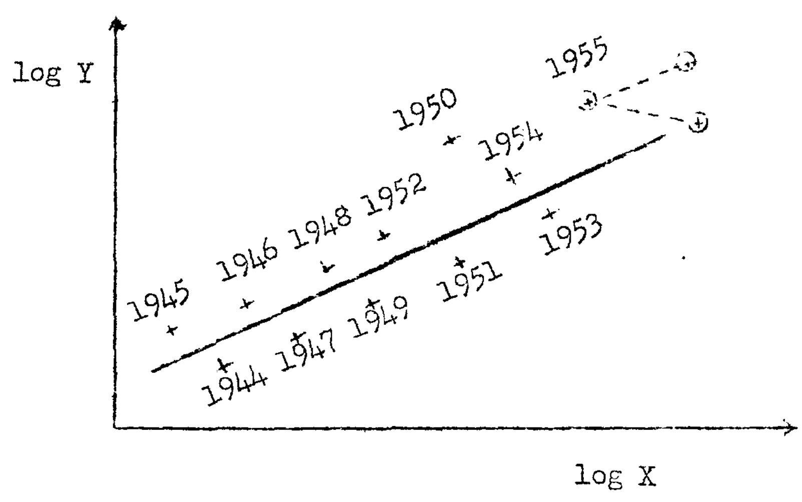
En la medida en que la relación entre las dos variables esté más alejada de la línea de regresión en el país que interesa para las proyecciones, mayor sería la diferencia a que se llegue utilizando la ecuación de regresión y el coeficiente de elasticidad como instrumento de proyección. Si, como en el gráfico anterior, la relación está en ese país por debajo de la línea de regresión, ello significaría que existe allí un consumo relativamente bajo (en comparación con el nivel de ingreso) del bien de que se trate; una estimación del consumo futuro basado en la ecuación de regresión supondría que tal situación se eliminaría, y el consumo tendería a aumentar no sólo por efecto del incremento del ingreso, sino también para superar ese retraso relativo; la utilización del coeficiente de elasticidad, en cambio, equivaldría a admitir que el consumo aumentará sólo por el efecto ingreso, pero que continuará registrándose un consumo relativamente bajo (en comparación con el nuevo nivel de ingreso).

_/ En otras palabras

En otras palabras, al aumentar el ingreso de X_1 a X_2 en la ecuación de regresión se admite un aumento del consumo de Y_1 a Y_{c_2}. En el primer caso, se supone que el nuevo nivel del consumo corresponderá exactamente al valor dado por la ecuación de regresión; en el segundo, se admite que continuará existiendo una discrepancia entre el valor teórico (dado por la ecuación de regresión y el valor efectivo proporcionalmente igual al que existiría en el período base).

Es deficil juzgar en términos generales cuál de los dos métodos podría ser más adecuado ante una situación de esta índole. Si el nivel relativamente bajo del consumo en el país considerado es atribuíble a limitaciones de la oferta, u otros factores de carácter temporal, podría ser más adecuada la proyección basada en la ecuación de regresión; si se debe, en cambio, a diferencias en hábitos de los consumidores, a factores climáticos u otros de carácter relativamente permanente, sería más adecuada la proyección basada en el coeficiente de elasticidad. Aún en el primer caso sería necesario tener en cuenta si el período a que se refiere la proyección es lo suficientemente largo como para que lleguen a eliminarse los efectos adversos de los factores temporales.

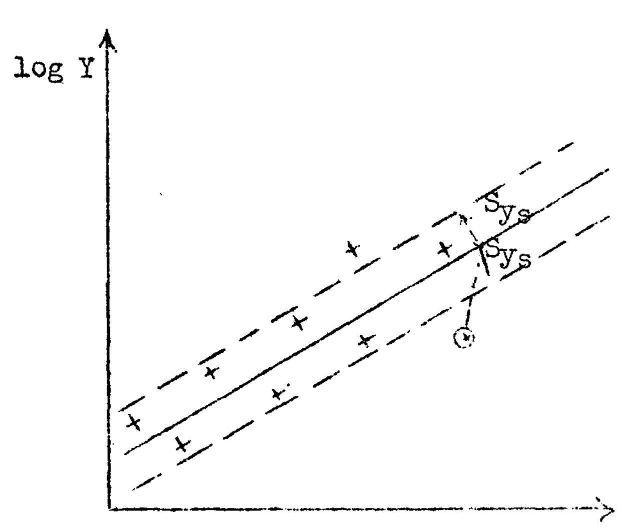
Las diferencias anotadas entre los dos métodos de proyección podrían presentarse también en el caso en que todo el análisis se hubiera basado en series cronológicas correspondientes a un mismo país, puesto que las cifras correspondientes al período que se tome como base seguramente serán diferentes de los valores teóricos dados por la ecuación de regresión.

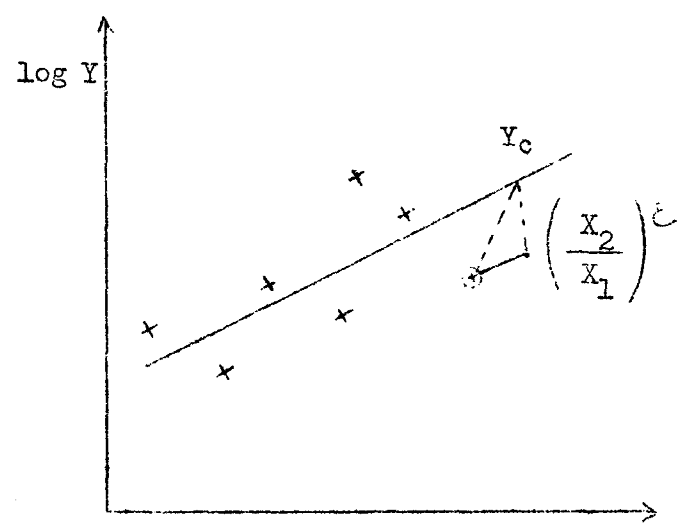


/ La interpretación

La interpretación sería sin embargo algo diferente, ya que tratándose de un mismo país, el consumo relativamente bajo (o relativamente elevado) registrado en el período base sería una mayor probabilidad atribuíble a factores de carácter temporal. En consecuencia, podría considerarse como más adecuada la proyección basada en la ecuación de regresión.

De cualquier modo, en términos generales la utilización de la ecuación de regresión y del coeficiente de elasticidad conducirá en la mayor parte de los casos a proyecciones diferentes, sin que resulte posible precisar cual de los dos tendría que considerarse más adecuada. Esto puede conducir a la proyección de un intervalo probable para la variable dependiente, basado no en la magnitud del error standard de estimación sino en la diferencia entre la proyección obtenida con la ecuación de regresión y la proyección a que conduce la aplicación del coeficiente de elasticidad.





Los gráficos anteriores ilustran esta alternativa. En el primer caso se utilizan la ecuación de regresión y el error standard de estimación para proyectar el intervalo correspondiente. Al utilizar $Y_c - S_{y_s}$ se está admitiendo no sólo que se elimina el bajo consumo relativo registrado en el período base o en el país correspondiente si se trata de una comparación internacional) sino además se estima como probable que llegue en el período cubierto por la proyección a registrarse un consumo relativamente elevado.

/ Es evidente

Es evidente que las posibilidades prácticas de que esto aconteciera son muy limitadas. En el segundo caso, se utilizan la ecuación de regresión y el coeficiente de elasticidad, y se proyecta un intervalo delimitado por estos dos valores. De este modo se estima un intervalo más amplio por debajo de la línea de regresión y ninguno por encima de ésta, lo que parecería más lógico en una situación como la supuesta en esos gráficos.

Un ejemplo de aplicación práctica del análisis de correlación no lineal y de la utilización del coeficiente de elasticidad-ingreso.

(Extractado de los estudios de la CEPAL sobre la industria del papel y celulosa en América Latina)

Naturaleza del problema: estimar la demanda futura del papel de diarios en América Latina.

Primera fase de la investigación: estimar la medida en que ese consumo está asociado con el nivel de ingresos.

Informaciones básicas: Cifras de ingreso (en dólares por habitante) y consumo de papel de diarios (en gramos por habitante) en 31 países en 1949. Las cifras correspondientes se incluyen en el cuadro 2 (a) adjunto.

Forma de la relación: se supone que las dos variables estaban asociadas en forma logarítmica, basándose en el diagrama de dispersión adjunto.

Las cuatro primeras columnas del cuadro 2 (b) contienen las tabulaciones necesarias para calcular los parámetros de la ecuación de regresión:

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

Las ecuaciones normales son:

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

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

Reemplazamos:

$$101{,}4523302 = 69{,}7938412\ a + 31\ b$$

$$236{,}3012838 = 162{,}6209746\ a + 69{,}7938412\ b$$

de donde:

$$a = 1{,}4381842$$

$$b = 0{,}0347075$$

de modo que se obtiene la ecuación de regresión:

$$\log Y = 1{,}4381842 \log X + 0{,}0347075$$

/ Las dos

Las últimas dos columnas del mismo cuadro 2 (b) tienen por objeto determinar en cada caso el valor calculado (dado por la ecuación de regresión anterior) de log Y_c.

En el cuadro 2 (c) se muestra la tabulación correspondiente a las diferencias y los cuadrados de las mismas que dan origen a la varianza total, explicada y no explicada.

Varianza total: $\sigma^2 \log Y = \frac{13,6981052}{31} = 0,442$

Varianza explicada: $\sigma^2 \log Y_c = \frac{11,3473621}{31} = 0,366$

Varianza no explicada: $\sigma^2 \log Y_s = \frac{2,3507371}{31} = 0,076$

Coeficiente de determinación : $\rho^2 \log X \log Y = \frac{0,386}{0,442} = 0,828$

Coeficiente de correlación : $\rho \log X \log Y = \sqrt{0,828} = \underline{0,91}$

Se comprueba así un coeficiente de correlación no lineal bastante elevado, si se asume una relación logarítmica entre las dos variables.

De acuerdo con las definiciones anteriores, el coeficiente de elasticidad-ingreso de la demanda de papel de diario sería:

$$\varepsilon = a$$
$$\varepsilon = 1,4381842$$

Si se quiere ahora estimar el crecimiento de esta demanda entre 1950 y 1960 como consecuencia de tasas alternativas de crecimiento del ingreso por habitante, se tendría:

Tasas anuales de crecimiento del ingreso por habitante	Monto unitario $\frac{X_2}{X_1}$	$\log\left(\frac{X_2}{X_1}\right)$	$\varepsilon \log\left(\frac{X_2}{X_1}\right) = \log\left(\frac{Y_2}{Y_1}\right)$	Monto unitario del consumo $\left(\frac{Y_2}{Y_1}\right)$	Crecimiento porcentual del consumo de papel de diarios
1 %	1,1046	0,0432050	0,0621367	1,154	15,4 %
2 %	1,2190	0,0860037	0,1236892	1,330	33,0 %
3 %	1,3439	0,1283670	0,1846154	1,530	53,0 %
4 %	1,4802	0,1703204	0,2449321	1,758	75,8 %
5 %	1,6289	0,2118944	0,3047432	2,017	101,7 %

/ Es pues

Es pues la última columna la que indica el crecimiento que cabría esperar en el consumo por habitante de papel de diarios entre 1950 y 1960 según el aumento que durante igual período experimente el ingreso por habitante. El cálculo hubo luego de efectuarse país por país, partiendo del consumo en gramos por habitante registrado en 1950, aumentándolo en el porcentaje correspondiente y multiplicándolo finalmente por la población estimada para 1960, a fin de obtener la proyección del consumo total en toneladas.

Correlación múltiple.

El análisis de correlación simple permite en último término precisar cuáles son las variables independientes que muestran una asociación más estrecha con la variable dependiente en que se está interesado; en otras palabras, qué variable independiente es la que contribuye en mayor proporción a explicar las variaciones de la variable dependiente.

Pero el hecho de que una variable independiente contribuya en menor medida que otra no significa que tenga que excluírse completamente del análisis. La explicación - o la proyección si tal es el caso - será mejor si no se descartan variables que, aunque en proporción inferior a otras, contribuyen también en cierta medida. La finalidad del análisis de correlación múltiple es examinar en forma simultánea la asociación de varias variables independientes con la variable dependiente.

En este caso, los principios son similares a los ya expuestos por la correlación simple, y el coeficiente de correlación múltiple (R) se define de nuevo como una relación entre varianza total y varianza explicada. La varianza total es siempre la misma, puesto que está definida sólo en términos de la variable dependiente, no importa cuántas y cuáles variables independientes se consideren. La varianza explicada queda determinada, como en lo anterior, por las diferencias entre los valores calculados de la variable dependiente y la media aritmética de la misma y la única diferencia está en que esta vez para la determinación de Y_c es necesario tomar en cuenta los valores de varias variables independientes. La varianza no explicada, finalmente, será también similar, puesto que no es sino un residuo de las dos anteriores.

Para mayor claridad, es conveniente introducir aquí algunos cambios en la simbología que ha venido utilizándose hasta ahora. La variable dependiente se designará por X_1 y las variables independientes que se consideran por X_2, X_3, X_4 .....

/Si se consideran

Si se consideran dos variables independientes, la ecuación de regresión aparecería de tal modo definida por:

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

Si se consideran tres variables independientes:

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

De este modo es fácil identificar un coeficiente cualquiera. Así por ejemplo, $b_{13.245}$ será el coeficiente que multiplica a X_3 en una ecuación de regresión múltiple en que además se están considerando las variables independientes X_2, X_4 y X_5.

En la ecuación de regresión podría admitirse relaciones no lineales con respecto a una o varias variables independientes. Podrían así definirse ecuaciones de la forma:

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

$$X_{1.23} = a_{1.23} + b_{12.3} X_2^2 + b_{13.2} X_3$$

Al igual que en la correlación simple, el primer paso en el análisis será la determinación de la ecuación de regresión, tabulando los valores necesarios para formar las ecuaciones normales. Si se tuvieran dos variables independientes, éstas serían:

$$\begin{aligned}
\Sigma X_1 &= n\, a_{1.23} + b_{12.3} \Sigma X_2 + b_{13.2} \Sigma X_3 \\
\Sigma X_1X_2 &= a_{1.23} \Sigma X_2 + b_{12.3} \Sigma X_2^2 + b_{13.2} \Sigma X_2X_3 \\
\Sigma X_1X_3 &= a_{1.23} \Sigma X_3 + b_{12.3} \Sigma X_2X_3 + b_{13.2} \Sigma X_3^2
\end{aligned}$$

Si se consideran tres variables independientes, las ecuaciones normales correspondientes serían:

$$\begin{aligned}
\Sigma X_1 &= n\, a_{1.234} + b_{12.34} \Sigma X_2 + b_{13.24} \Sigma X_3 + b_{14.23} \Sigma X_4 \\
\Sigma X_1X_2 &= a_{1.234} \Sigma X_2 + b_{12.34} \Sigma X_2^2 + b_{13.24} \Sigma X_2X_3 + b_{14.23} \Sigma X_2X_4 \\
\Sigma X_1X_3 &= a_{1.234} \Sigma X_3 + b_{12.34} \Sigma X_2X_3 + b_{13.24} \Sigma X_3^2 + b_{14.23} \Sigma X_3X_4 \\
\Sigma X_1X_4 &= a_{1.234} \Sigma X_4 + b_{12.34} \Sigma X_2X_4 + b_{13.24} \Sigma X_3X_4 + b_{14.23} \Sigma X_4^2
\end{aligned}$$

/ Determinados

Determinados los parámetros de la ecuación de regresión podrían determinarse los valores calculados de la variable dependiente (por ejemplo $X_{c\ 1,234}$) y luego las diferencias y cuadrados de las mismas que dan origen a las tres varianzas.

Si se consideran tres variables independientes, se definiría así el coeficiente de determinación:

$$\underset{\text{coef.de det.}}{R^2_{1.234}} = \frac{S^2 X_{c\ 1.234}}{S^2 X_1} = 1 - \frac{S^2 X_{s\ 1.234}}{S^2 X_1}$$

de donde se deduciría el coeficiente de correlación múltiple $R_{1.234}$. Necesariamente un coeficiente de correlación múltiple cuando se consideran 3 variables independientes será mayor que el que se obtiene cuando se consideran sólo 2 de ellas y éste que un coeficiente de correlación simple con respecto a sólo una de esas variables independientes.

Así por ejemplo: $R_{1.234} \geqq R_{1.23} \geqq R_{1.2}$

Esto se explica por que aún si la asociación en una nueva variable independiente fuera muy débil, ésta siempre contribuirá en alguna medida (aunque sea escasa) a explicar las variaciones de la variable independiente, y por lo tanto algo aportará a la varianza explicada.

En la medida en que se fueran agregando nuevas variables independientes, el coeficiente de correlación múltiple tenderá a acercarse a la unidad. Esto es lógico, puesto que si llegaramos a tomar en cuenta explícitamente todas las variables que influyen sobre el fenómeno en que se está interesado, la varianza explicada llegaría a ser igual a la varianza total, y se tendría un caso de correlación múltiple perfecta. Por cierto que cada vez el procedimiento de cálculo se hace más laborioso, de modo que en la práctica sólo será útil considerar un número de variables independientes suficiente como para obtener un valor adecuadamente elevado del coeficiente de correlación, y rechazar aquellas variables que incrementen ese valor en magnitud muy pequeña.

Naturalmente, no puede atribuírse signo al coeficiente de correlación múltiple, puesto que la asociación puede ser positiva respecto de algunas variables independientes y negativa respecto de otros. Este último se reflejará en el signo de los coeficientes b.

/ Correlación

Correlación parcial.

La correlación múltiple y simple se proponen en definitiva medir el grado de asociación existente entre una variable dependiente y una o varias variables independientes, conforme a un conjunto dado de informaciones estadísticas básicas. En especial en el caso de la correlación simple, cabría sin embargo mantener una reserva importante en la interpretación de los resultados: la posible incidencia de algunas variables independientes no consideradas explícitamente, y que pueden sin embargo haber influído sobre las variables dependientes e independientes. De este modo, 2 variables podrían mostrar una asociación más o menos estrecha, sólo como consecuencia de la influencia que ha ejercido una tercera variable.

Admítase por ejemplo que se desee examinar el grado de asociación entre el consumo de cierto bien y los precios relativos del mismo. Durante cierto período podría constatarse una disminución persistente de los precios relativos, acompañada de un incremento también persistente del consumo, mostrando así un alto grado de correlación; pero si durante el mismo período ha venido aumentando el nivel de ingresos, podría atribuírse a este factor - y no a los precios relativos - el incremento del consumo. O por lo menos se plantearía la duda de cuánto puede atribuírse a uno y otro de esos factores. He aquí una respuesta que no podría dar la correlación simple, la que podría indicar en los dos casos coeficientes de correlación igualmente elevados.

En otros términos, se plantea pues el problema de cómo aislar el efecto de una variable, para cuantificar sólo el efecto neto de otra. Puesto que no existen en casos como éste posibilidades de experimentación, es necesario utilizar una solución estadística, a través del análisis de correlación parcial. Lo que se propone en este caso es cuantificar el grado de asociación existente entre dos variables una vez eliminado estadísticamente el efecto de una o más variables independientes adicionales.

Concretamente, la solución consiste en medir el incremento de la varianza explicada que se logra al introducir una nueva variable independiente y compararlo con la varianza no explicada antes de introducirla. Podría así definirse un coeficiente de determinación parcial:

$$r^2_{12.3} = \frac{S^2_{X_c\,1.23} - S^2_{X_c\,1.3}}{S^2_{X_s\,1.3}}$$

de donde se obtendría el coeficiente de correlación parcial $r_{12.3}$.

/Si se

Si se consideran 3 variables independientes, habría tres coeficientes de correlación parcial:

$$r_{12.34} = \sqrt{\frac{S^2X_{c\ 1.234} - S^2X_{c\ 1.34}}{S^2X_{s\ 1.34}}}$$

y $r_{13.24}$ y $r_{14.23}$ definidos de manera similar.

De acuerdo con las definiciones anteriores, el coeficiente de correlación parcial puede confirmar o corregir los resultados de la correlación simple. Puesto que en este caso no se establece una comparación directa de los valores observados de las dos variables sino que previamente se elimina la influencia de otras variables independientes, el coeficiente de correlación parcial será necesariamente inferior al correspondiente coeficiente de correlación simple.

Así por ejemplo: $r_{12.3} \leqslant r_{1.2}$

o bien: $r_{13.24} \leqslant r_{13.2} \leqslant r_{1.3}$

Cálculos abreviados en correlación múltiple y parcial.

De acuerdo con lo visto en correlación simple para un coeficiente cualquiera 1.2 se tendría:

Varianza total: $n\,S^2_{X_1} = \sum X^2_1 - \bar{X}_1 \sum X_1$

Varianza explicada: $n\,S^2_{X_c} = a \sum X_1 + b_{1.2} \sum X_1 X_2 - \bar{X}_1 \sum X_1$

Varianza no explicada: $n\,S^2_{X_s} = \sum X^2_1 - (a \sum X_1 + b_{1.2} \sum X_1 X_2)$

Coeficiente de correlación: $r_{1.2} = \sqrt{\dfrac{S^2_{X_c}}{S^2_{X_1}}} = \sqrt{1 - \dfrac{S^2_{X_s}}{S^2_{X_1}}}$

En el caso de correlación múltiple $S^2_{X_1}$ sería la misma. Para $S^2_{X_c}$, recuerde que:

$$S^2_{X_c} = \frac{\sum (X_c - \bar{X}_1)^2}{n}$$

de donde: $n\,S^2_{X_c} = \sum X^2_c - \bar{X}_1 \sum X_1$

Si se consideran dos variables independientes:

$$X_c = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$$

$$X^2_c = a_{1.23} + b^2_{12.3} X^2_2 + b^2_{13.2} X^2_3 + 2a_{1.23} b_{12.3} X_2 + 2a_{1.23} b_{13.2} X_3 + 2b_{12.3} b_{13.2} X_2 X_3$$

Aplicando sumatoria y descomponiendo algunos términos:

$$\Sigma X^2_c = n\, a^2_{1.23} + a_{1.23} b_{12.3} \Sigma X_2 + a_{1.23} b_{13.2} \Sigma X_3 + a_{1.23} b_{12.3} \Sigma X_2 + b^2_{12.3} \Sigma X^2_2 + b_{12.3} b_{13.2} \Sigma X_2X_3 + a_{1.23} b_{13.2} \Sigma X_3 + b_{12.3} b_{13.2} \Sigma X_2X_3 + b^2_{13.2} \Sigma X^2_3$$

Sacando factor común:

$$\Sigma X^2_c = a_{1.23}(n a_{1.23} + b_{12.3} \Sigma X_2 + b_{13.2} \Sigma X_3) + b_{12.3} (a_{1.23} \Sigma X_2 + b_{12.3} \Sigma X^2_2 + b_{13.2} \Sigma X_2X_3) + b_{13.2}(a_{1.23} \Sigma X_3 + b_{12.3} \Sigma X_2X_3 + b_{13.2} \Sigma X^2_3)$$

o sea: $\Sigma X^2_c = a_{1.23} \Sigma X_1 + b_{12.3} \Sigma X_1X_2 + b_{13.2} \Sigma X_2X_3$

Por lo tanto: $n\, S^2_{X_c} = a_{1.23} \Sigma X_1 + b_{12.3} \Sigma X_1X_2 + b_{13.2} \Sigma X_1X_3 - \bar{X}_1 \Sigma X_1$

y el coeficiente de correlación : $R_{1.23} = \sqrt{\dfrac{a_{1.23} \Sigma X_1 + b_{12.3} \Sigma X_1X_2 + b_{13.2} \Sigma X_1X_3 - \bar{X}_1 \Sigma X_1}{\Sigma X^2_1 - \bar{X}_1 \Sigma X_1}}$

En forma similar para tres variables independientes se tendría:

$$R_{1.234} = \sqrt{\frac{a_{1.234} \Sigma X_1 + b_{12.34} \Sigma X_1X_2 + b_{13.24} \Sigma X_1X_3 + b_{14.23} \Sigma X_1X_4 - \bar{X}_1 \Sigma X_1}{\Sigma X^2_1 - \bar{X}_1 \Sigma X_1}}$$

Un análisis completo de un problema de correlación en que se consideran tres variables independientes conduciría así a la determinación de:

3 coeficientes de correlación simple:	$r_{1.2}$	$r_{1.3}$	$r_{1.4}$	
3 coeficientes de correlación parcial:	$r_{12.34}$	$r_{13.24}$	$r_{14.23}$	
4 coeficientes de correlación múltiple:	$R_{1.23}$	$R_{1.24}$	$R_{1.34}$	$R_{1.234}$

Para ello, habría que determinar 7 ecuaciones de regresión:

$$X_{c1.2} = a_{1.2} + b_{1.2} X_2$$
$$X_{c1.3} = a_{1.3} + b_{1.3} X_3$$
$$X_{c1.4} = a_{1.4} + b_{1.4} X_4$$
$$X_{c1.23} = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$$
$$X_{c1.24} = a_{1.24} + b_{12.4} X_2 + b_{14.2} X_4$$
$$X_{c1.34} = a_{1.34} + b_{13.4} X_3 + b_{14.3} X_4$$
$$X_{c1.234} = a_{1.234} + b_{12.34} X_2 + b_{13.24} X_3 + b_{14.23} X_4$$

/Un ejemplo

Un ejemplo de aplicación práctica del análisis de correlación múltiple y parcial.

Naturaleza del problema: determinación de los factores que influencian el consumo de acero en América Latina. Propósito y proyecciones del consumo en paises que intentaban instalar plantas siderúrgicas propias (necesidad de proyecciones adecuadas por la magnitud de las inversiones necesarias y por la gran influencia que tiene sobre los costos la escala de operación).

Observación anticipada: aumentos muy acentuados en el consumo en los paises que habían instalado plantas siderúrgicas, lo que parecía constituír en sí un factor importante (por el incentivo para creación de industrias de transformación, etc.).

Tres factores seleccionados a priori como determinantes: ingreso, nivel de inversiones en equipos y maquinarias, y actividad en la industria de la construcción. Se procuró cifras indicativas en cada una de ellas para el mayor número posible de paises latinoamericanos, para algunos de los cuales se consideró dos períodos distintos (en especial antes y después de instalar industrias siderúrgicas).

Las variables consideradas fueron así:

X_1 = consumo de acero en Kgs. de acero crudo por habitante.
X_2 = ingreso, en dólares de 1949 por habitante.
X_3 = importaciones de bienes de capital, en dólares de 1949 por habitante.
X_4 = consumo de cemento, en Kgs. por habitante.

Primer paso: tabulación de todos los datos necesarios para los cálculos siguientes: primeras 5 columnas del cuadro 3 (a) y todas las de los cuadros 3 (b), 3 (c), y 3 (d).

$$X_1 = a_{1.2} + b_{12}X_2$$
$$\sum X_1 = n\, a_{1.2} + b_{12} \sum X_2$$
$$\sum X_1X_2 = a_{1.2} \sum X_2 + b_{12} \sum X^2_2$$

$$430 = 15\, a_{12} + 2.535\, b_{12}$$
$$103.641 = 2.535\, a_{12} + 611.365\, b_{12}$$

$$a_{1.2} = 0{,}05497$$
$$b_{12} = 0{,}1693$$
$$X_1 = 0{,}05497 + 0{,}1693\, X_2$$

/Del mismo

Del mismo modo: $X_1 = 7,3135 + 1,518\ X_3$

y $X_1 = 2,679 + 0,4674\ X_4$

Para $X_1 = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$

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

$$\sum X_1X_2 = a_{1.23}\sum X_2 + b_{12.3}\sum X^2_2 + b_{13.2}\sum X_2X_3$$

$$\sum X_1X_3 = a_{1.23}\sum X_3 + b_{12.3}\sum X_2X_3 + b_{13.2}\sum X^2_3$$

$$430 = 15\ a_{1.23} + 2.535\ b_{12.3} + 211\ b_{13.2}$$

$$103.641 = 2435\ a_{1.23} + 611.365\ b_{12.3} + 51.656\ b_{13.2}$$

$$9.214 = 211\ a_{1.23} + 51.656\ b_{12.3} + 5.053\ b_{13.2}$$

$$a_{1.23} = 0,530966$$

$$b_{12.3} = 0,1110$$

$$b_{13.2} = 0,6664$$

$$X_1 = 0,530966 + 0,1110\ X_2 + 0,6664\ X_3$$

del mismo modo $X_1 = -0,5882 + 0,1134\ X_2 + 0,18148\ X_4$

y $X_1 = 3,7153 + 0,7022\ X_3 + 0,2711\ X_4$

Finalmente para:

$$X_1 = a_{1.234} + b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4$$

$$\sum X_1 = n a_{1.234} + b_{12.34}\sum X_2 + b_{13.24}\sum X_3 + b_{14.23}\sum X_4$$

$$\sum X_1X_2 = a_{1.234}\sum X_2 + b_{12.34}\sum X^2_2 + b_{13.24}\sum X_2X_3 + b_{14.23}\sum X_2X_3$$

$$\sum X_1X_3 = a_{1.234}\sum X_3 + b_{12.34}\sum X_2X_3 + b_{13.24}\sum X^2_3 + b_{14.23}\sum X_3X_4$$

$$\sum X_1X_4 = a_{1.234}\sum X_4 + b_{12.34}\sum X_2X_4 + b_{13.24}\sum X_3X_4 + b_{14.23}\sum X^2_4$$

$$430 = 15\ a_{1.234} + 2.535\ b_{12.34} + 211\ b_{13.24} + 834 b_{14.23}$$

$$103.641 = 2.535\ a_{1.234} + 611.365\ b_{12.34} + 51.656\ b_{13.24} + 197271 b_{14.23}$$

$$9.214 = 211\ a_{1.234} + 51.656\ b_{12.34} + 5.053\ b_{13.24} + 18.007 b_{14.23}$$

$$34.401 = 834\ a_{1.234} + 197.271\ b_{12.34} + 18.007\ b_{13.24} + 68.820 b_{14.23}$$

/$a_{1.234} = 0,31771$

$$a_{1.234} = 0{,}31771$$
$$b_{12.34} = 0{,}12394$$
$$b_{13.24} = 0{,}61978$$
$$b_{14.23} = -0{,}02363$$

$$\underline{X_1 = 0{,}31771 + 0{,}12394\ X_2 + 0{,}61978\ X_3 - 0{,}02363\ X_4}$$

Tercer paso: tabulación de una serie de productos necesarios para los cálculos (Cuadro 3 (c)

Cuarto paso: cómputo de las varianzas y coeficientes de correlación.

La varianza total:

$$n\ S^2_{X_1} = \sum X^2_1 - \bar{X}_1 \sum X_1$$
$$15\ S^2_{X_1} = 18.820 - 12.326{,}67 = 6.493{,}33$$
$$S^2_{X_1} = 432{,}89$$

Para $r_{1.2}$

$$n\ S^2_{X_{c_{1.2}}} = a_{1.2} \sum X_1 + b_{12} \sum X_1X_2 - \bar{X}_1 \sum X_1$$
$$15\ S^2_{X_{c_{1.2}}} = 23{,}64 + 17.545{,}04 - 12.326{,}67 - 5.242{,}01$$
$$S^2_{X_{c_{1.2}}} = 349{,}47$$

$$n\ S^2_{X_{c_{1.2}}} = \sum X^2_1 - (a_{1.2} \sum X_1 + b_{12} \sum X_1X_2)$$
$$15\ S^2_{X_{s_{1.2}}} = 18.820 - 23{,}64 - 17.545{,}04 - 1.251{,}32$$
$$S^2_{X_{s_{1.2}}} = 83{,}42$$

$$r^2_{1.2} = \frac{349{,}47}{432{,}89} = 0{,}8073$$
$$r_{1.2} = 0{,}898$$

Del mismo modo:

$$S^2_{X_{c_{1.3}}} = 320{,}77$$
$$S^2_{X_{s_{1.3}}} = 112{,}12$$
$$r_{1.3} = 0{,}861$$

$$S^2_{X_{c_{1.4}}} = 326{,}95$$
$$S^2_{X_{s_{1.4}}} = 105{,}94$$
$$r_{1.4} = 0{,}869$$

/Para $R_{1.23}$:

Para $R_{1.23}$:

$$n\, S^2_{X_{c1.23}} = a_{1.23} \sum X_1 + b_{12.3} \sum X_1X_2 + b_{13.2} \sum X_1X_3 - \bar{X} \sum X_1$$

$$15\, S^2_{X_{c1.23}} = 228,32 + 11.504,15 + 6.139,87 - 12.326,67 = 5.545,67$$

$$S^2_{X_{c1.23}} = 369,71$$

$$n\, S^2_{X_{s1.23}} = \sum X^2_1 - (a_{1.23} \sum X_1 + b_{12.3} \sum X_1X_2 + b_{13.2} \sum X_2X_3)$$

$$15\, S^2_{X_{s1.23}} = 18.820 - 228,32 - 11.504,15 - 6.139,87 = 947,66$$

$$S^2_{X_{s1.23}} = 63,18$$

$$R^2_{1.23} = \frac{369,71}{432,89} = 0,8541$$

$$R_{1.23} = 0,924$$

Del mismo modo:

$$S^2_{X_{c1.24}} = 361,09$$

$$S^2_{X_{s1.24}} = 95,06$$

$$R_{1.24} = 0,914$$

$$S^2_{X_{c1.34}} = 337,83$$

$$S^2_{X_{s1.34}} = 95,06$$

$$R_{1.34} = 0,883$$

Para $R_{1.234}$:

$$n\, S^2_{X_{c1.234}} = a_{1.234} \sum X_1 + b_{12.34} \sum X_1X_2 + b_{13.24} \sum X_1X_3 + b_{14.23} \sum X_1X_4 - \bar{X}_1 \sum X_1$$

$$15\, S^2_{X_{c1.234}} = 136,62 + 12.844,76 + 5.710,69 - 812,82 - 12.326,67 = 5.552,58$$

$$S^2_{X_{c1.234}} = 370,17$$

$$S^2_{X_{s1.234}} = 62,72$$

$$R^2_{1.234} = \frac{370,17}{432,89} = 0,8551$$

/Para $r_{12.34}$:

Para $r_{12.34}$:

$$r_{12.34} = \sqrt{\frac{s^2_{X_{c1.234}} - s^2_{X_{c1.34}}}{s^2_{X_{s1.34}}}}$$

$$r_{12.34} = \sqrt{\frac{370,17 - 337,83}{95,06}} = \sqrt{\frac{32,34}{95,06}} = \sqrt{0,3402}$$

$$\underline{r_{12.34} = 0,583}$$

Del mismo modo:

$$\underline{r_{13.24} = 0,356}$$

$$\underline{r_{14.23} = 0,027}$$

Finalmente, determinación de $X_{c1.234}$ (Cuadro 3(f) y comparación con los valores reales (última columna del Cuadro 3 (a).

Es interesante comparar las magnitudes relativas de los coeficientes de correlación simple $r_{1.3}$ y $r_{1.4}$ ($r_{1.4} > r_{1.3}$) y los correspondientes coeficientes de correlación parcial ($r_{13.24} > r_{14.23}$).

Esto ilustra los riesgos que ofrecen conclusiones basadas en la correlación simple, sin complementarlo con el cómputo de los coeficientes de correlación parcial.

El bajísimo coeficiente de correlación parcial $r_{14.23}$ se confirma también para la escasa diferencia de los dos valores de los coeficientes de correlación múltiple $R_{1.23}$ y $R_{1.234}$, que en la práctica llevarían a desechar la consideración de X_4 como variable independiente suficientemente importante para los efectos del análisis.

/América Latina.

América Latina. Volumen Físico de sus Exportaciones hacia Estados Unidos. Su relación con el ingreso real de éste y con los términos del intercambio.

(Indices, 1937 = 100)

Cuadro 1 (a)

Años	X_1	X_2	Y	X_1^2	X_1Y	X_2^2	X_2Y	Y_{c1}	Y_{c2}
1921	80,8	74,6	74,0	6,529	5.979	5.565	5.520	71,8	125,8
1922	75,2	79,0	97,5	5.655	7.332	6.241	7.702	66,1	122,9
1923	92,7	103,3	100,5	8.593	9.316	10.671	10.382	84,0	107,3
1924	93,7	111,3	96,0	8.780	8.995	12.388	10.685	85,0	102,1
1925	95,2	111,2	92,8	9.063	8.834	12.365	10.319	86,5	102,2
1926	99,9	106,0	100,1	9.980	10.000	11.236	10.611	91,3	105,5
1927	100,5	120,2	89,9	10.100	9.035	14.448	10.806	89,9	95,3
1928	100,3	129,2	84,0	10.060	8.425	16.693	10.853	91,7	90,5
1929	103,3	118,7	98,4	10.650	10.155	14.090	11.680	94,6	96,3
1930	100,2	94,9	86,3	10.060	8.656	9.006	8.190	91,7	112,7
1931	93,2	96,3	75,3	8.686	7.018	9.274	7.251	84,5	111,8
1932	82,0	96,7	55,2	6.724	4.526	9.351	5.338	73,1	111,5
1933	80,5	94,9	57,6	6.480	4.637	9.006	5.466	71,5	112,7
1934	87,4	98,7	59,5	7.639	5.200	9.742	5.873	78,6	110,2
1935	90,0	97,0	75,8	8.100	6.822	9.409	7.352	81,2	111,3
1936	97,9	97,3	80,5	9.584	7.881	9.467	7.833	89,3	111,1
1937	100,0	100,0	100,0	10.000	10.000	10.000	10.000	91,4	109,4
1938	98,6	84,9	81,0	9.722	7.987	7.208	6.877	90,0	119,1
1939	111,2	84,3	94,1	12.365	10.464	7.106	7.933	102,8	119,5
1940	122,0	73,4	118,4	14.884	14.445	5.388	8.691	113,8	126,5
1941	140,3	79,8	172,3	19.684	24.174	6.368	13.750	132,4	122,4
1942	163,0	80,3	134,0	26.569	21.842	6.448	10.760	155,5	122,1
1943	193,0	75,7	173,8	37.249	33.543	5.730	13.157	186,1	125,1
1944	205,5	65,5	206,6	42.230	42.456	4.303	13.532	198,8	131,6
1945	202,8	68,8	202,1	41.128	40.986	4.733	13.904	196,1	129,5
1946	173,7	101,9	168,9	30.172	29.338	10.384	17.211	166,4	108,1
1947	157,1	115,9	153,2	24.680	24.068	13.433	17.756	149,5	99,1
1948	161,6	120,4	154,5	26.115	24.967	14.496	18.602	154,1	96,2
1949	170,0	127,8	149,1	28.900	25.347	16.333	19.055	162,7	91,4
	3.471,6	2.808,0	3.231,4	460.381	432.428	280.882	307.089		

Y = Indice del volumen físico de las exportaciones latinoamericanas a EE.UU.

X_1 = Indice del ingreso real de los Estados Unidos

X_2 = Indice de la relación de precios del intercambio entre A.L. y EE.UU.

$Y = 1{,}018\ X_1 - 10.4$ $\qquad$ $Y = -\ 0{,}6453\ X_2 + 173.9$

/América Latina.

América Latina. Volumen Físico de sus Exportaciones hacia Estados Unidos. Su relación con el ingreso real de éste y con los términos del intercambio.

(Indices, 1937 = 100)

Cuadro 1 (b)

Años	$Y_i - \bar{Y}$	$Y_{c1} - \bar{Y}$	$Y_{c2} - \bar{Y}$	$Y_i - Y_{c1}$	$Y_i - Y_{c2}$
1921	- 37,4	- 39,6	- 14,4	2,2	- 51,8
1922	- 13,9	- 45,3	11,5	31,4	- 25,4
1923	- 10,9	- 27,4	- 4,1	16,5	- 6,8
1924	- 15,4	- 26,4	- 9,3	11,0	- 6,1
1925	- 18,6	- 24,9	- 9,4	6,3	- 9,4
1926	- 11,3	- 20,1	- 5,9	8,8	- 5,4
1927	- 21,5	- 21,5	- 16,1	0	- 5,4
1928	- 27,4	- 19,7	- 20,9	- 7,7	- 6,5
1929	- 13,0	- 16,8	- 15,1	3,8	2,1
1930	- 25,1	- 19,7	1,3	- 5,4	- 26,4
1931	- 36,1	- 26,9	0,4	- 9,2	- 36,5
1932	- 56,2	- 38,3	0,1	17,9	- 56,3
1933	- 53,8	- 39,9	1,3	- 13,9	- 55,1
1934	- 51,9	- 32,8	- 1,2	- 19,1	- 50,7
1935	- 35,6	- 30,2	- 0,1	5,4	- 35,5
1936	- 30,9	- 22,1	- 0,3	- 8,8	- 30,6
1937	- 11,4	- 20,0	- 2,0	9,6	- 9,4
1938	- 30,4	- 21,4	7,7	- 9,0	- 38,1
1939	- 17,3	- 8,6	8,1	- 8,7	- 25,4
1940	7,0	2,4	15,1	4,6	- 8,1
1941	60,9	21,0	11,0	39,9	49,9
1942	22,6	44,1	10,7	- 21,5	11,9
1943	62,4	74,7	13,7	- 12,3	48,7
1944	95,2	87,4	20,2	7,8	75,0
1945	90,7	84,7	18,1	6,0	72,6
1946	57,5	55,0	- 3,3	2,5	60,8
1947	41,8	38,1	- 12,3	3,7	54,1
1948	43,1	42,7	- 15,2	0,4	58,3
1949	37,7	51,3	- 20,0	- 13,6	57,7
	0	0	0		

/América Latina.

América Latina. Volumen Físico de sus Exportaciones hacia Estados Unidos. Su relación con el ingreso real de éste y con los términos del intercambio.

(Indices, 1937 = 100)

Cuadro 1 (c)

Años	$(Y_i - \bar{Y})^2$	$(Y_{c1} - \bar{Y})^2$	$(Y_{c2} - \bar{Y})^2$	$(Y_i - Y_{c1})^2$	$(Y_i - Y_{c2})^2$
1921	1.399	1.568	207	5	2.683
1922	193	2.052	132	986	645
1923	119	751	17	272	46
1924	237	697	86	121	37
1925	346	620	88	40	88
1926	128	404	35	77	29
1927	462	462	259	0	29
1928	751	388	437	59	42
1929	169	282	228	14	4
1930	630	388	2	29	697
1931	1.303	724	0	85	1.332
1932	3.158	1.467	0	320	3.170
1933	2.894	1.592	2	193	3.036
1934	2.694	1.076	1	365	2.570
1935	1.267	912	0	29	1.260
1936	955	488	0	77	936
1937	130	400	4	92	88
1938	924	458	59	81	1.452
1939	299	74	66	76	645
1940	49	6	228	21	66
1941	3.709	441	121	1.592	2.490
1942	511	1.945	114	462	142
1943	3.894	5.580	188	151	2.372
1944	9.063	7.639	408	61	5.625
1945	8.226	7.174	328	36	5.271
1946	3.306	3.025	11	6	3.697
1947	1.747	1.452	151	14	2.927
1948	1.858	1.823	231	0	3.399
1949	1.421	2.632	400	185	3.329
	51.842	46.520	3.803	5.449	48.107

$$s\,y^2 = \frac{51.842}{29} = 1.787.6 \qquad r^2_{Y,X_1} = \frac{1.604.1}{1.787.6} = 0,897$$

$$s^2\,y_{c1} = \frac{46.520}{29} = 1.604.1 \qquad r_{Y,X_1} = 0,947$$

$$s^2\,y_{s1} = \frac{5.449}{29} = 187.9 \qquad r^2_{Y,X_2} = \frac{131.1}{1.787.6} = 0,073$$

$$s\,y^2_{c2} = \frac{3.803}{29} = 131.1 \qquad r_{Y,X_2} = -0,27$$

$$s^2\,y_{s2} = \frac{48.107}{29} = 1.650.9$$

/Consumo de

Cuadro 2 (a)

Consumo de papel de diarios: su relación con el ingreso por capita

(año 1949)

Paises	Ingreso = X (Dólares por habitante)	Consumo = Y (Gramos por habitante)
Italia	235	1.600
Holanda	502	5.300
Noruegq	587	7.600
Suecia	780	16.000
Inglaterra	773	12.000
India	57	100
Japon	100	1.400
Turquía	125	600
Australia	679	16.000
Canada	870	22.000
Estados Unidos	1.453	34.000
Argentina	346	7.025
Bolivia	55	687
Brasil	112	1.567
Colombia	132	1.013
Costa Rica	125	1.796
Cuba	296	5.267
Chile	188	3.858
Ecuador	40	761
El Salvador	92	1.063
Guatemala	77	728
Haiti	40	62
Honduras	83	234
Mexico	121	2.385
Nicaragua	89	555
Panamá	183	3.204
Paraguay	84	205
Perú	100	1.208
Rep. Dominicana	75	688
Uruguay	331	7.239
Venezuela	322	1.823

/Consumo de

Cuadro 2 (b)

Consumo de papel de diarios: su relación con el ingreso per capita

(año 1949)

Paises	log X	log Y	log X • log Y	$(\log X_1)^2$	a log X	$\log Y_c = a \log X + b$
Italia	2,3710679	3,2041200	7,5971861	5,6219630	3,4100324	3,4447399
Holanda	2,7007037	3,7242759	10,0581657	7,2938005	3,8841094	3,9188169
Noruega	2,7686381	3,8808136	10,7445684	6,6653569	3,9818116	4,0165191
Suecia	2,8920946	4,2041200	11,8695033	8,3642112	4,1593648	4,1940723
Inglaterra	2,8881795	4,0791812	11,7814075	8,3415808	4,1537341	4,1884416
India	1,7558749	2,0000000	3,5117498	3,0830967	2,5252715	2,5599790
Japón	2,0000000	3,1461280	6,2922560	4,0000000	2,8763684	2,9110759
Turquía	2,0969100	2,7781513	5,8255332	4,3970315	3,0157428	3,0504503
Australia	2,8318698	4,2041200	11,9055205	8,0194866	4,0727504	4,1074579
Canada	2,9395193	4,3424227	12,7646353	8,6407737	4,2275702	4,2622777
EE.UU.	3,1622656	4,5514789	14,3297398	9,9999237	4,5479204	4,5826279
Argentina	2,5390761	3,8466463	9,7669277	6,4469074	3,6516591	3,6863666
Bolivia	1,7403627	2,8369557	4,9373536	3,0288625	2,5029621	2,5376696
Brasil	2,0492180	3,1950690	6,5473929	4,1992944	2,9471529	2,9818604
Colombia	2,1205739	3,0056094	6,3736168	4,4968337	3,0497759	3,0833834
Costa Rica	2,0969100	3,2543063	6,8239874	4,3970315	3,0157428	3,0504503
Cuba	2,4712917	3,7215633	9,1970685	6,1072827	3,5541727	3,5888802
Chile	2,2741578	3,5863622	8,1559536	5,1717937	3,2706578	3,3053653
Ecuador	1,6020600	2,8813847	4,6161512	2,5665962	2,3040574	2,3387649
El Salvador	1,9637878	3,0265333	5,9434692	3,8564625	2,8242886	2,8589961
Guatemala	1,8864907	2,8621314	5,3993843	3,5588472	2,7131211	2,7478286
Haiti	1,6020600	1,7923917	2,8715190	2,5665962	2,3040574	2,3387649
Honduras	1,9190781	2,3692159	4,5467103	3,6828608	2,7599878	2,7946953
Mexico	2,0827854	3,3774884	7,0345835	4,3379950	2,9954291	3,0301566
Nicaragua	1,9493900	2,7442930	5,3496973	3,8001214	2,8035819	2,8382894
Panamá	2,2624511	3,5056925	7,9314579	5,1186850	3,2538214	3,2885289
Paraguay	1,9242793	2,3117539	4,4484602	3,7028508	2,7674762	2,8021837
Perú	2,0000000	3,0820753	6,1641506	4,0000000	2,8763768	2,9110843
R.Dominicana	1,8750613	2,8375884	5,3206522	3,5158549	2,6966914	2,7313989
Uruguay	2,5198280	3,8596786	9,7257262	6,3495331	3,6239874	3,6586949
Venezuela	2,5078559	3,2607867	8,1775832	6,2893412	3,6067693	3,6414768
	69,7938412	101,4523302	236,3012838	162,6209746		

/Consumo de

Cuadro 2 (c)

Consumo de papel de diarios: su relación con el ingreso per capita
(año 1949)

Paises	$\log Y_i - \overline{\log Y}$	$\log Y_c - \overline{\log Y}$	$\log Y_i - \log Y_c$	$(\log Y_i - \overline{\log Y})^2$	$(\log Y_c - \overline{\log Y})^2$	$(\log Y_i - \log Y_c)^2$
Italia	-0,0685339	0,1720861	-0,2406199	0,0046968	0,0296136	0,0578979
Holanda	-0,4516221	0,6461631	-0,1945410	0,2039625	0,4175268	0,0378462
Noruega	0,6081598	0,7438653	-0,1357055	0,3698583	0,5533356	0,0184160
Suecia	-0,9314662	0,9214185	0,0100477	0,8676293	0,8490121	0,0001010
Inglaterra	0,8065274	0,9157878	-0,1092604	0,6504864	0,8386675	0,0119378
India	-1,2726538	-0,7126748	-0,5599790	1,6196477	0,5079054	0,3135765
Japón	-0,1265258	-0,3615779	0,2350521	0,0160088	0,1307386	0,0552495
Turquía	-0,4945025	-0,2222035	-0,2722990	0,2445327	0,0493744	0,0741467
Australia	0,9314662	0,8348041	0,0966621	0,8676293	0,6968979	0,0093435
Canada	1,0697589	0,9896239	0,0801450	1,1444055	0,9793555	0,0064232
EE.UU.	1,2588251	1,3099741	-0,0511490	1,5846406	1,7160321	0,0026162
Argentina	0,5739925	0,4137128	0,6027970	0,3294674	0,1711583	0,0256896
Bolivia	-0,4356971	0,0349842	0,2992871	0,1898520	0,5402018	0,0895728
Brasil	-0,0775848	-0,2907934	0,2132086	0,0060194	0,0845608	0,0454579
Colombia	-0,2670444	-0,1881704	-0,0788740	0,0713127	0,0354081	0,0062211
Costa Rica	-0,0183475	-0,2222035	0,2038560	0,0003366	0,0493744	0,0415573
Cuba	0,4489095	0,3162264	0,1326831	0,2015197	0,0999991	0,0176048
Chile	0,3137084	0,0327115	0,2809969	0,0984130	0,0010700	0,0789593
Ecuador	-0,3912691	-0,9338889	0,5426198	0,1530915	0,8721485	0,2944362
El Salvador	-0,2461205	-0,4136577	0,1675372	0,0605753	0,1711127	0,0280687
Guatemala	-0,4105224	-0,5248252	0,1143028	0,1685286	0,2754415	0,0130651
Haiti	-1,4802621	-0,9338889	-0,5463732	2,1911759	0,8721485	0,2985237
Honduras	-0,9034379	-0,4779585	-0,4254794	0,8162000	0,2284443	0,1810327
Mexico	0,1048346	-0,2425172	0,3473518	0,0109903	0,0588146	0,1206533
Nicaragua	-0,5283608	-0,4343644	-0,0939964	0,2791651	0,1886724	0,0088353
Panamá	0,2330387	0,0158751	0,2171636	0,0543070	0,0002520	0,0471600
Paraguay	-0,9608999	-0,4704782	-0,4904217	0,9233286	0,2213497	0,2405134
Perú	-0,1905869	-0,3615779	0,1709910	0,0363234	0,1307386	0,0292379
R.Dominicana	-1,3975945	-0,5412569	0,1061895	1,9532704	0,2929590	0,0112762
Uruguay	-0,7528278	0,3860391	0,2009837	0,5667497	0,1490262	0,0403944
Venezuela	-0,7647999	0,3688210	-0,3806901	0,5849189	0,1360289	0,1449250
				13,6981052	11,3473621	2,3507371

/América Latina:

Cuadro 3(a)

América Latina: Consumo de acero; su relación con el ingreso, las importaciones bienes de capital y el consumo de acero per capita.

	Consumo acero X_1	Ingreso X_2	Capitalización X_3	Consumo cemento X_4	Columna de verificación X	X_c	Variación %
1) Bolivia 1947-49	5	55	4	10	74	9	- 44,4
2) Ecuador 1947-49	8	40	7	17	72	9	- 11,1
3) Guatemala 1947-49	8	77	5	14	104	13	- 38,5
4) Perú 1947-49	10	100	7	36	153	16	- 37,5
5) México 1939	14	95	4	22	135	14	- 0
6) Brasil 1939	15	90	4	18	127	14	+ 7,1
7) Colombia 1947-49	16	132	12	42	202	23	- 30,4
8) Brasil 1950	24	112	7	34	177	18	+ 33,3
9) México 1950	28	121	10	55	214	20	+ 40,0
10) Cuba 1947-49	37	296	13	69	415	43	- 14,0
11) Chile 1940	38	170	13	80	301	28	+ 35,7
12) Uruguay 1947-49	38	331	25	120	514	54	- 29,6
13) Chile 1951	50	190	24	90	354	37	+ 35,1
14) Venezuela 1947-49	62	322	47	133	564	66	- 6,1
15) Argentina 1947-49	77	404	29	94	604	66	+ 16,7
	430	2.535	211	834	4.010		

/Cuadro 3(b)

Cuadro 3 (b)

Continuación

	X_1^2	$X_1 X_2$	$X_1 X_3$	$X_1 X_4$	Columna de chequeo $X_1 \cdot X$
1)	25	275	20	50	370
2)	64	320	56	136	576
3)	64	616	40	112	832
4)	100	1.000	70	360	1.530
5)	196	1.330	56	308	1.890
6)	225	1.350	60	270	1.905
7)	256	1.112	192	672	3.232
8)	576	2.686	168	816	4.248
9)	784	3.388	280	1.540	5.992
10)	1.369	10.952	481	2.553	15.355
11)	1.444	6.460	494	3.040	11.438
12)	1.444	12.578	950	4.560	19.532
13)	2.500	9.500	1.200	4.500	17.700
14)	3.844	19.964	2.914	8.246	34.968
15)	5.929	31.108	2.233	7.238	46.508
	18.820	103.641	9.214	34.401	166.076

Cuadro 3 (c)

Continuación

	X_2^2	$X_2 X_3$	$X_2 X_4$	Columna de chequeo $X_2 X$
1)	3.025	220	550	4.070
2)	1.600	280	680	2.880
3)	5.929	385	1.078	8.008
4)	10.000	700	3.600	15.300
5)	9.025	380	2.090	17.825
6)	8.100	360	1.620	11.430
7)	17.424	1.584	5.544	26.664
8)	12.544	784	3.808	39.950
9)	14.641	1.210	6.655	25.894
10)	87.616	3.848	20.424	122.840

/Cuadro 3 (c)

Cuadro 3 (c)
(continuación)

	X_2^2	$X_2\ X_3$	$X_2\ X_4$	Columna de chequeo $X_2\ X$
11)	28.900	2.210	13.600	51.170
12)	109.561	8.275	39.720	170.134
13)	36.100	4.570	17.100	67.270
14)	103.684	15.134	42.826	181.608
15)	163.216	11.716	37.976	244.016
	611.365	51.656	197.271	989.059

Cuadro 3 (d)
Continuación

	X_3^2	X_3X_4	Columna de chequeo $X_3\ X$	X_4^2	Columna de chequeo $X_4\ X$
1)	16	40	296	100	740
2)	49	119	504	289	1.224
3)	25	70	520	196	1.456
4)	49	252	1.071	1.296	5.508
5)	16	88	620	484	3.410
6)	16	72	508	324	2.286
7)	144	504	2.424	1.764	8.484
8)	49	238	1.645	1.156	7.990
9)	100	550	2.140	3.025	11.770
10)	169	897	5.395	4.761	28.635
11)	169	1.040	3.913	6.400	24.080
12)	625	3.000	12.850	14.400	61.680
13)	576	2.160	8.506	8.100	31.860
14)	2.209	6.251	26.508	17.689	75.012
15)	841	2.726	17.516	8.836	56.776
	5.053	18.007	84.416	68.820	320.911

/Cuadro 3 (e)

Cuadro 3 (e)
Continuación

$\bar{X}_1 \sum X_1$	=	28,6666	x	430	=	12.326,67
$a_{1.2} \sum X_1$	=	0,05497	x	430	=	23,64
$a_{1.3} \sum X_1$	=	7,3135	x	430	=	3.144,81
$a_{1.4} \sum X_1$	=	2,67933	x	430	=	1.152,11
$a_{1.23} \sum X_1$	=	0,530966	x	430	=	228,32
$a_{1.24} \sum X_1$	=	-0,5882	x	430	=	- 252,93
$a_{1.234} \sum X_1$	=	3,71533	x	430	=	1.597,59
$b_{1.2} \sum X_1X_2$	=	0,1692866903	x	103.641	=	17.545,04
$b_{12.3} \sum X_1X_2$	=	0,1110	x	103,641	=	11.504,15
$b_{12.4} \sum X_1X_2$	=	0,1134	x	103,641	=	11.752,89
$b_{12.34} \sum X_1X_2$	=	0,12393517	x	103,641	=	12.844,76
$b_{1.3} \sum X_1X_3$	=	1,518714011	x	9.214	=	13.993,43
$b_{13.2} \sum X_1X_3$	=	0,666362718	x	9.214	=	6.139,87
$b_{13.4} \sum X_1X_3$	=	0,702188295	x	9.214	=	6.469,96
$b_{13.24} \sum X_1X_3$	=	0,6197836048	x	9.214	=	5.710,69
$b_{1.4} \sum X_1X_4$	=	0,467394209	x	34.401	=	16.078,83
$b_{14.2} \sum X_1X_4$	=	0,18148	x	34.401	=	6.243,09
$b_{14.3} \sum X_1X_4$	=	0,271114641	x	34.401	=	9.326,61
$b_{14.23} \sum X_1X_4$	=	-0,023627878	x	34.401	=	- 812,82

Determinación de $X_{c_{1.234}}$

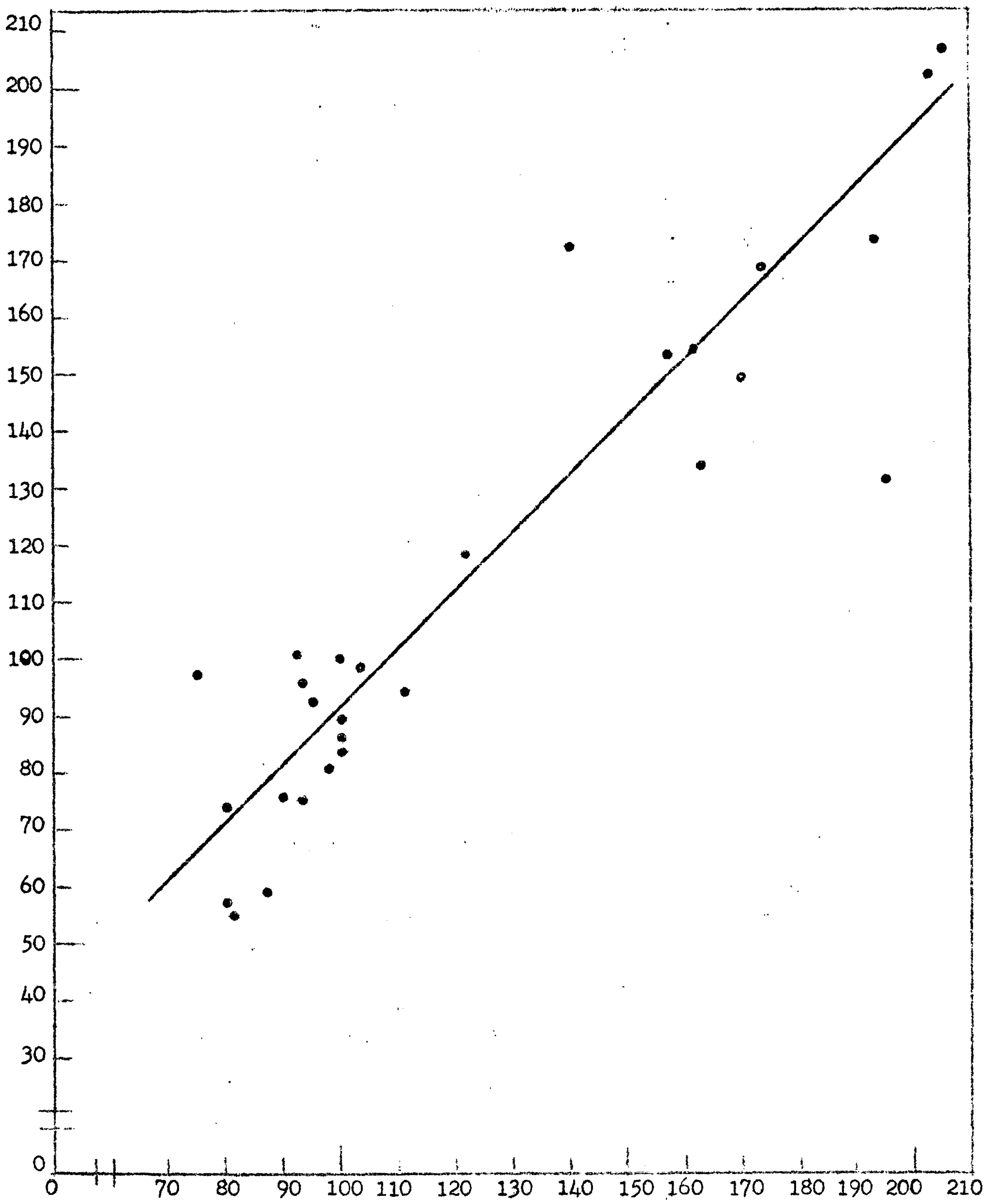
Cuadro 3 (f)
Continuación

A $0{,}12394X_2$	B $0{,}61978X_3$	C $-0{,}02363X_4$	A + B + C + 0,31771 $X_{c1.234}$
6,82	2,48	- 0,24	9.38
4,96	4,34	- 0,40	9.22
9,54	3,10	- 0,33	12.63
12,39	4,34	- 0,85	16.20
11,77	2,48	- 0,52	14.05
11,15	2,48	- 0,43	13.52
16,36	7,44	- 0,99	23.13
13,88	4,34	- 0,80	17.74
15,00	6,20	- 1,30	20.22
36,67	8,06	- 1,63	43.42
21,07	8,06	- 1,89	27.56
41,02	15,49	- 2,84	53.99
23,55	14,87	- 2,13	36.61
39,91	29,13	- 3,14	66.22
50,07	17,97	- 2,22	66.14

/Gráfico I

Gráfico I

RELACION ENTRE EL INDICE DEL INGRESO REAL EN LOS ESTADOS UNIDOS Y EL INDICE DE VOLUMEN FISICO DE LAS EXPORTACIONES DE AMERICA LATINA HACIA ESTADOS UNIDOS



Indice del Ingreso real de los Estados Unidos

/Gráfico II

Gráfico II

RELACION ENTRE EL INDICE DE LA RELACION DE PRECIOS DEL INTERCAMBIO ENTRE AMERICA LATINA Y ESTADOS UNIDOS Y EL INDICE DE VOLUMEN FISICO DE LAS EXPORTACIONES DE AMERICA LATINA HACIA ESTADOS UNIDOS