

MANUAL DE ECONOMETRIA

Universidad Autónoma Agraria
"ANTONIO NARRO"

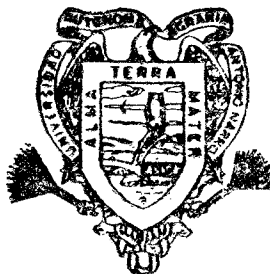


BIBLIOTECA

MARIO CANTU SIFUENTES

T E S I S

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS
EN ESTADISTICA EXPERIMENTAL



Universidad Autónoma Agraria
Antonio Narro

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coah.

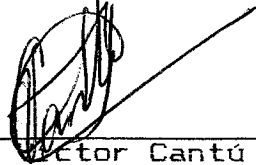
NOVIEMBRE DE 1989

Tesis elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial, para obtener el grado de

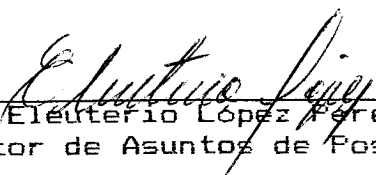
MAESTRO EN CIENCIAS
EN ESTADISTICA EXPERIMENTAL

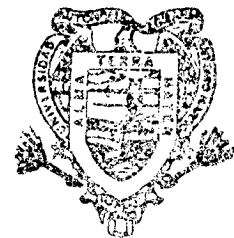
COMITE PARTICULAR

Asesor Principal: 
Ing. M.C. Félix de Jesús Sánchez Pérez.

Asesor: 
Ing. M.C. Hector Cantú Hernández.

Asesor: 
Q.F.B. M.C. Clara Rosa Leyva Moreno.


Dr. Eleuterio López Pérez.
Subdirector de Asuntos de Postgrado.



BIBLIOTECA
EGIDIO G. REBONATO
BANCO DE TESIS
U.A.A.A.N.

Buenavista, Saltillo, Coahuila. Noviembre de 1989.

AGRADECIMIENTO

A Clara Rosa porque sin su desinteresada solidaridad no hubiera sido posible la realización del presente trabajo. De hecho, la mayor parte del crédito es de ella.

DEDICATORIA

Para:

Diana

Alejandro

Daniel

Mariana

A mis Papás:

Juan e Irene

A mis hermanos.

INDICE DEL CONTENIDO

	Página
INDICE DE CUADROS.....	viii
INDICE DE FIGURAS.....	ix
INTRODUCCION.....	1
MANUAL DE ECONOMETRIA	
INTRODUCCION.....	4
GENERALIDADES.....	4
ALCANCES.....	5
REGRESION LINEAL SIMPLE.....	7
ESPECIFICACIONES LINEALES.....	9
ESTIMACION DE LOS PARAMETROS.....	11
PROPIEDADES DE LOS ESTIMADORES.....	24
COEFICIENTE DE CORRELACION.....	33
ESTIMACION POR INTERVALO.....	40
PRUEBA DE HIPOTESIS.....	45
ANALISIS DE VARIANZA.....	48
PREDICCIÓN DEL MODELO DE MINIMOS - CUADRADOS.....	50
ESPECIFICACIONES LINEALES.....	53
EL MODELO LOG-LOG.....	53
EL MODELO SEMI-LOG.....	55
EL MODELO RECIPROCO.....	57
EL MODELO LOGARITMO RECIPRO	

CO.....	58
REGRESION MULTIPLE.....	61
EL MODELO DE REGRESION MULTIPLE....	61
ESTIMACION PUNTUAL.....	63
LOS COEFICIENTES DE CORRELACION PAR	
CIAL Y MULTIPLE.....	70
INTERVALOS DE CONFIANZA Y PRUEBAS -	
DE HIPOTESIS.....	73
ESPECIFICACIONES LINEALES DE VARIAS	
VARIABLES.....	75
VIOLACION DE LOS SUPUESTOS DEL MODELO CLASI	
CO.....	77
NORMALIDAD Y MEDIA CERO.....	78
VARIABLES REGRESORAS NO ESTOCASTI -	
CAS.....	79
MULTICOLINEALIDAD.....	79
MULTICOLINEALIDAD EXACTA...	80
CASIMULTICOLINEALIDAD.....	82
EFECTOS.....	82
DETECCION.....	84
SOLUCIONES.....	86
HETEROCEDASTICIDAD.....	91
DETECCION.....	92
CONSECUENCIAS.....	94
CORRECCION.....	95
AUTOCORRELACION.....	99
DETECCION.....	100
EFECTOS.....	105

CORRECCION.....	105
LITERATURA CITADA.....	108

INDICE DE CUADROS

	Página
CUADRO 2.1 ANALISIS DE VARIANZA.....	49
CUADRO 3.1 ANALISIS DE VARIANZA MATRICIAL.....	75

INDICE DE FIGURAS

	Página
2.1 DIAGRAMAS DE DISPERSION.....	8
2.2 DIAGRAMA DE DISPERSION.....	13
2.3 FUNCION DE DENSIDAD DE PROBABILIDAD CONDICIO NAL.....	16
2.4 RELACION DE REGRESION.....	17
2.5 DIAGRAMA DE DISPERSION PARA UNA MUESTRA DE - TAMANO n.....	18
2.6 REPRESENTACION GRAFICA DE LA FORMA DE DESVIA CION DE LA RECTA DE MINIMOS CUADRADOS.....	22
2.7 DIAGRAMAS DE DISPERSION.....	36
2.8 INTERPRETACION GRAFICA DE UN INTERVALO DE - CONFIANZA.....	43
2.9 INTERVALO DE CONFIANZA DEL 100(1- α) PORCIEN- TO PARA χ^2 CON N-2 GRADOS DE LIBERTAD.....	45
2.10 RELACIONES NO LINEALES DE LA FORMA $Z = \alpha_0 W^{\beta_0}$	54
2.11 EL MODELO SEMILOG.....	56
2.12 EL MODELO RECIPROCO.....	58
2.13 LA CURVA LOGISTICA.....	59
2.14 EL MODELO LOGARITMICO-RECIPROCO.....	60
4.1 GRAFICA DE RESIDUOS ESTANDARIZADA.....	78
4.2 PERTURBACIONES HETEROCEDASTICAS.....	92
4.3 PRESENCIA DE HETEROCEDASTICIDAD.....	93
4.4 DIAGRAMAS DE DISPERSION.....	101

MANUAL

DE

ECONOMETRIA

INTRODUCCION

La ciencia económica tiene como objetivo la distribución de recursos limitados entre fines que compiten entre sí. Para esto, es necesario desarrollar políticas que conduzcan a minimizar problemas y maximizar beneficios.

En el estudio cuantitativo de los sistemas económicos, es necesario conjuntar la teoría económica, la matemática y las técnicas estadísticas en una parte del conocimiento humano llamada Econometría.

La Universidad Autónoma Agraria "Antonio Narro" ha contemplado en su programa de graduados la formación de Planificadores Agropecuarios, para los cuales se incluye en su plan de estudios, la materia de Econometría.

Generalmente, no es posible encontrar bibliografía del tema en español con un nivel adecuado a los propósitos de dicha materia. Además de esto, la notación en algunos textos es obsoleta y en otros no es uniforme.

Ante tal inconveniente se propone el presente trabajo cuyo objetivo es presentar de manera accesible la teoría sobre la que se sustentan los diferentes métodos econométricos.

CAPITULO I

INTRODUCCION

Generalidades

Econometría significa literalmente "medida de la economía". Sin embargo, su definición no ha podido ser delimitada en forma precisa, y es por esto que no se ha logrado la aceptación general de una sola definición. Muchas son las que existen actualmente, pero parece que todas ellas concuerdan en dos puntos fundamentales: expresar en forma matemática las relaciones de los fenómenos económicos y medir estadísticamente esas relaciones.

El desarrollo lógico de la teoría económica conduce al establecimiento de una función o conjunto de funciones las cuales constituyen el modelo económico . Estas relaciones pueden ser expresadas en forma cualitativa o cuantitativa. Evidentemente la primera forma, aunque muy importante, no concierne a la econometría. Son las relaciones cuantitativas las que le interesan.

Las relaciones que se investigan en la economía son de tres tipos:

1. Uniecuacionales.
2. Multiecuacionales.
3. Ecuaciones simultáneas.

En las relaciones uniecuacionales existe una variable

dependiente o explicada, que viene determinada por una o más independientes o explicatorias; por ejemplo, cuando se dice que el consumo C depende del ingreso Y y la tasa de interés r , escrito como:

$$C = f(Y, w, r)$$

C es la variable dependiente y Y, w y r las independientes.

En las relaciones multiecuacionales, se parte de un conjunto de ecuaciones. Por ejemplo, sean C_A, C_D, C_{ND} y C_S los gastos de consumo en automóviles, bienes duraderos, no duraderos y de servicios respectivamente. Cada uno de ellos podría ser función del ingreso y la riqueza; pero la forma en la que dependen puede ser distinta. Así, en vez de estudiar el comportamiento del consumo total como función del ingreso y la riqueza, se adquiere mayor conocimiento tratando cada una de esas relaciones por separado.

En las relaciones de ecuaciones simultáneas, dos o más variables vienen determinadas "simultáneamente" por un cierto número de variables explicativas; por ejemplo, la función de oferta y demanda para un bien determinado, están "explicadas" por el precio del bien.

En todo modelo económico aparecen variables interrelacionadas, y parámetros.

Las variables que intervienen en todo modelo económico son:

1. Continuas.
2. Discretas.
3. Endógenas.
4. Exógenas.

Una variable discreta toma solo valores particulares.

Una variable continua puede tomar todos los valores posibles dentro de un intervalo.

En los modelos económicos aparecen por lo general variables discretas, pero frecuentemente se aplican a ellos las propiedades de las variables continuas dado que son matemáticamente más fáciles de manejar.

Las variables endógenas son aquellas que se interrelacionan dentro del modelo económico, esto es, son variables determinadas dentro del propio sistema económico.

Las variables exógenas son magnitudes determinadas fuera del sistema; sin embargo, estas influyen sobre las variables endógenas sin que estas influyan sobre aquellas.

Las constantes o parámetros son números que miden las relaciones entre las variables existentes dentro de un modelo. Están relacionados estrechamente con propensiones, elasticidades etc..

Gran parte del trabajo econométrico consiste en estimar los parámetros que intervienen en los modelos. Para esto, es necesario contar con conjuntos de datos.

Dado que la economía no es ciencia experimental, no existe forma de diseñar y ejecutar experimentos controlados; así los datos con los que se cuentan son obtenidos, a saber, en dos formas:

1. De sección cruzada.
2. De series de tiempo.

En los datos de sección cruzada se tienen observaciones sobre unidades individuales en un momento del

tiempo; por ejemplo, ingresos y gastos en alimentos para un conjunto de familias. Estos datos son recogidos normalmente mediante encuestas muestrales. Cuando se trata de series de tiempo, se dispone de observaciones sobre un cierto período de tiempo, por ejemplo, datos trimestrales del PNB.

Alcances

Este trabajo no pretende ser, un elemental compendio de econometría, es más bien un esfuerzo encaminado a ayudar al desarrollo de una comprensión básica de ella. La mayor parte de estos principios son poco más que refinamientos del pensamiento cotidiano. Aún así, no cabe duda de que para poder dominar este tema, el lector tendrá que leer y releer algunas secciones.

Aunque en este trabajo se acentúa la aplicación de métodos estadísticos a la investigación económica, debe tenerse en cuenta que la mayor parte de lo que se expone puede usarse sin inconvenientes en otras disciplinas, pues se basan en principios fundamentales idénticos.

La novedad que pueda presentarse quizá, es la simplificación en la notación y aclaración de conceptos que en los textos es confusa; sobre todo para estudiantes que tienen un primer contacto con esta disciplina.

En el capítulo II se muestran los conceptos básicos desarrollando el modelo económico de dos variables; de ahí en adelante, se abordará directamente el caso de k variables. Se trata además conceptualmente, la violación de supuestos del modelo clásico. La notación matricial se usa

siempre y cuando la exposición no esté fuera de la comprensión del lector que ha tomado solo cursos elementales de álgebra lineal.

CAPITULO II

REGRESION LINEAL SIMPLE

El análisis de regresión es una rama de la teoría estadística cuyo uso está muy difundido en casi todas las disciplinas científicas.

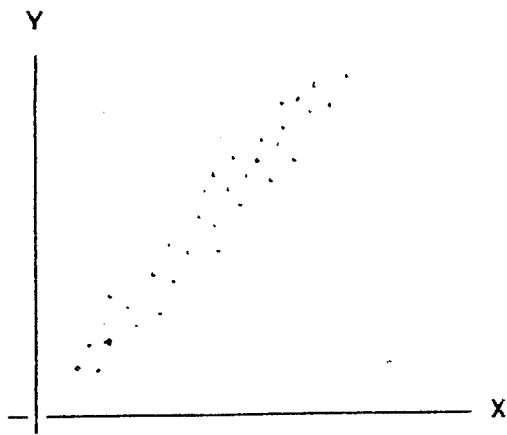
En economía, es la técnica básica para medir o estimar las relaciones entre variables económicas, que constituyen la esencia de la teoría y vida económica. En este capítulo, se presentan las técnicas de regresión simple, cuyo propósito fundamental es estimar la relación que existe entre dos variables X e Y.

Se puede suponer por ejemplo, que se desea estimar la cantidad de trigo que se compraría a diferentes precios de producción. Esta información podría resultar muy útil a nuestro gobierno entre otras cosas, para fijar los precios de garantía de tal cereal. Para tal efecto, se podría decidir la estimación de una curva de demanda a partir de una serie de observaciones.

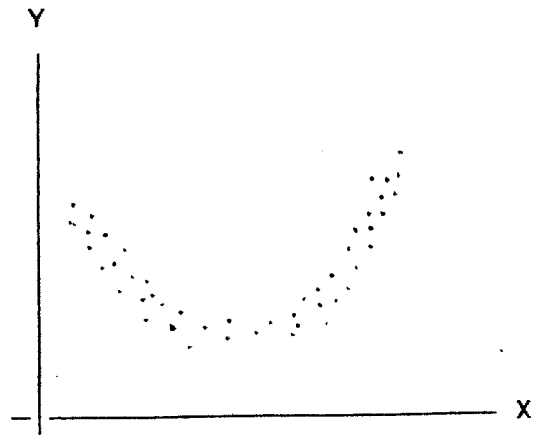
Dado que el ajuste de la relación entre las dos variables se hace mediante un conjunto de observaciones

$$\{X_i, Y_i\}_{i=1}^n$$

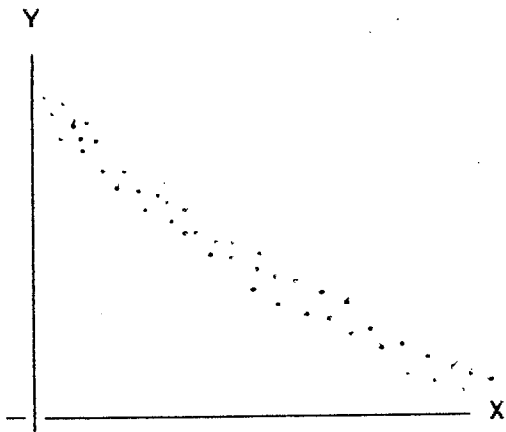
un primer paso para tomar la decisión de la forma de la función a ajustar, es graficar los datos en un diagrama de dispersión como los que se muestran en la figura 2.1.



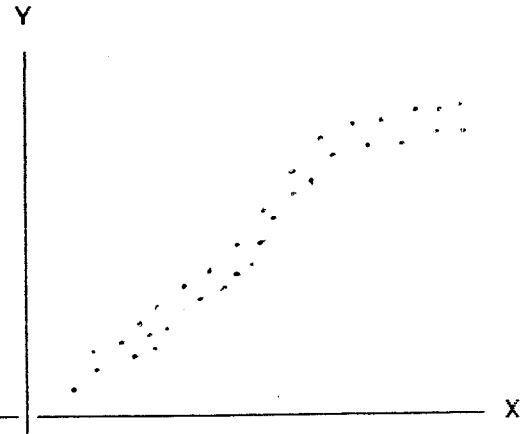
(2.1.a)



(2.1.b)



(2.1.c)



(2.1.d)

Figura 2.1. Diagramas de dispersión.

Para el caso del diagrama 2.1.a , salvo un error que puede deberse a fluctuaciones causadas por variables no consideradas, ó a fallas al observar, se encuentran alrededor de una línea recta, a la cual se pueden ajustar.

En la gráfica 2.1.b parece que la relación verdadera es una parábola.

En la gráfica 2.1.c la evidencia gráfica es de que el mejor ajuste es una curva exponencial.

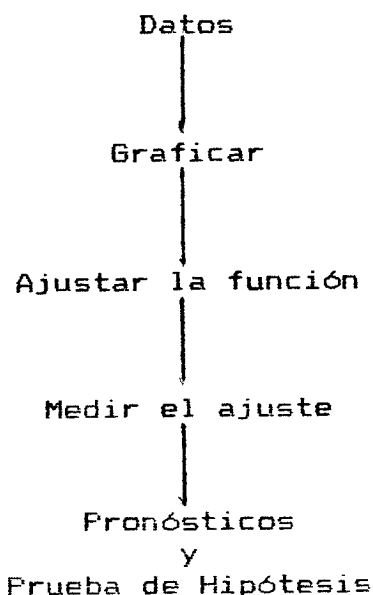
Finalmente, la gráfica 2.1.d hace pensar que la curva logística proporciona un ajuste más adecuado.

Una vez que se ha decidido que curva representa mejor los datos, se procede a ajustar la función correspondiente.

Luego que se ha obtenido una curva de ajuste es conveniente medir que tan bueno es éste para tener idea de que tanto se puede confiar en las inferencias hechas.

Para una mejor comprensión del análisis de regresión, es necesario tener conocimientos acerca de las variables aleatorias, muestras, parámetros, estimadores, funciones de distribución, media y varianza de las variables aleatorias, covarianza entre variables y prueba de hipótesis.

En resumen, en forma esquemática, el proceso de ajuste y manejo de datos, podría ser el siguiente:



Especificaciones Lineales.

Se principia por postular una relación entre dos

variables de la forma:

$$Y = f(X)$$

donde Y indica la variable dependiente (explicada) y X la variable independiente (explicativa), ambas endógenas al modelo. Es posible de acuerdo al sistema económico que se esté tratando, hacer expectativas teóricas acerca del signo de $f'(X)$, o acerca del rango de valores en el cual esta relación esté definida.

En teoría económica, son de gran importancia las especificaciones lineales.

Una especificación lineal puede ser interpretada de dos maneras:

1. Linealidad en las variables. Existe linealidad en las variables si Y, o alguna transformación de Y, puede ser expresada como una función lineal de X. En este sentido:

$$Y = \alpha + \beta X$$

$$Y = \alpha X^\beta$$

$$Y = \exp \left\{ \alpha + \beta \frac{1}{X} \right\}$$

son especificaciones lineales. La primera especificación es directa, la segunda especificación puede ponerse tomando logaritmos como:

$$\ln Y = \ln \alpha + \beta \ln X,$$

la cual es lineal en $\ln Y$ y $\ln X$, mientras que la última especificación toma la forma :

$$\ln Y = \alpha + \beta \left[\frac{1}{X} \right]$$

Sin embargo la función:

$$Y = \alpha + \beta X + \delta X^2$$

siendo lineal en Y , X y en X^2 , no es lineal en las variables X e Y .

2. Linealidad en los parámetros. Un segundo sentido es la linealidad en los parámetros, pudiendo o no, haber linealidad en las variables.

De ésta forma:

$$Y = \alpha + \beta X + X^2$$

$$Y = \alpha + \beta X^3$$

son lineales en los parámetros, mientras que

$$Y = \alpha + \sqrt{\alpha} X + \beta X^2$$

no es lineal en el parámetro α . En éste trabajo se trata solo con funciones lineales en los parámetros.

Estimación de los Parámetros

Las funciones lineales son de gran importancia en la economía dado que, son de fácil manejo y pueden utilizarse con frecuencia para aproximar funciones no lineales. En general, si X e Y denotan datos transformados apropiadamente, la forma algebraica general de una función lineal es:

$$Y = \alpha + \beta X \quad (2.1)$$

donde α y β son los parámetros de la función. La constante α se llama ordenada al origen y β , la pendiente de la función; de tal forma que si la función (2.1) representara la relación entre ingreso (X) y consumo (Y), α

representaría el consumo autónomo y β la propensión marginal a consumir.

Es muy remoto el caso de que una relación como (2.1) describa con exactitud los datos con los que se está trabajando, sin embargo, si se supone que los datos observados fueron generados por una ecuación lineal estocástica tal como

$$Y = \alpha + \beta X + U \quad (2.2)$$

donde el término U es una perturbación aleatoria, la ecuación permite que Y sea mayor o menor que $\alpha + \beta X$, lo cual depende de que U sea positivo ó negativo.

Lo ideal sería, que el término U fuera pequeño y que no estuviera correlacionado con X , a fin de poder cambiar a X sin que se modifique U y determinar así que le ocurrirá en promedio a Y . Se utilizará a (2.2) como ecuación básica para el modelo de regresión desarrollado en éste capítulo. En general, los datos de insumo para el análisis de regresión simple son un conjunto de pares ordenados de números tales como.

$$X_1, X_2, \dots, X_i, \dots, X_n$$

$$Y_1, Y_2, \dots, Y_i, \dots, Y_n$$

donde el par (X_i, Y_i) es la i -ésima observación de las variables X e Y .

Para cada observación (X_i, Y_i) existe una perturbación U_i .

Si cada par de datos se grafican se obtiene un diagrama de dispersión tal como el mostrado en la figura (2.1).

Si se conocieran los parámetros α y β , estaría determinada la función $\alpha + \beta X$. Las perturbaciones estocásticas están representadas por la distancia vertical entre los puntos observados (X_i, Y_i) y los puntos correspondientes en la línea $(X, \alpha + \beta X)$.

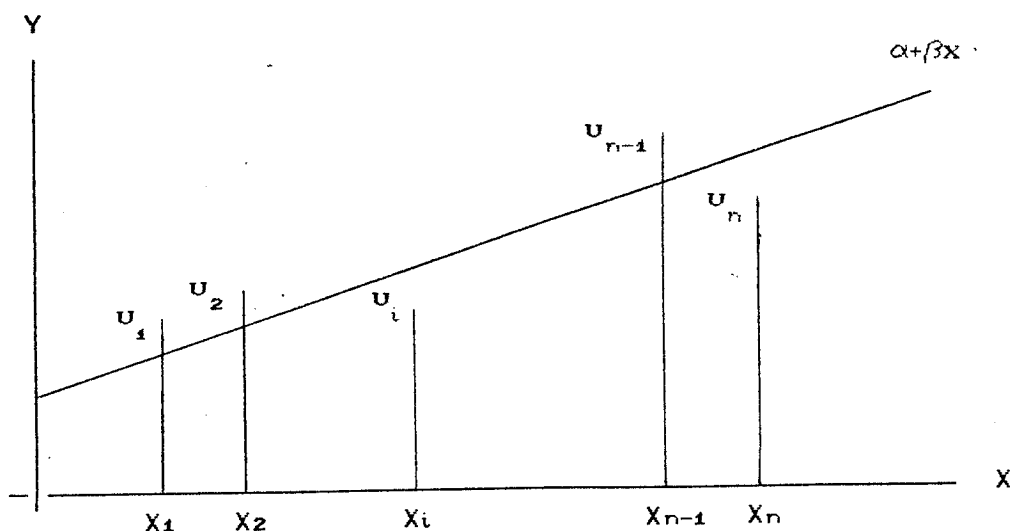


Figura 2.2. Diagrama de Dispersión.

El siguiente paso es desarrollar un modelo básico de regresión, el cual consiste en un conjunto de supuestos sobre la distribución de los términos de error y las relaciones entre las variables X e Y . Primero se supone que las perturbaciones U son variables aleatorias independientes con valores esperados iguales a cero y varianzas σ^2 para todo valor de X . Las dos últimas suposiciones acerca de los términos U se expresan en forma matemática como:

$$E(U_i) = 0 \quad \text{para toda } i \quad (2.3)$$

$$\text{Var}(U_i) = \sigma^2 \quad \text{para toda } i \quad (2.4)$$

Por otra parte, la suposición de independencia

requiere que la covarianza entre U_i y U_j ; $i \neq j$ sea cero, esto es:

$$\begin{aligned} \text{Cov}(U_i, U_j) &= E \{ [U_i - E(U_i)][U_j - E(U_j)] \} \\ &= E \{ U_i U_j \} = 0 \quad \text{para } i \neq j \quad (2.5) \end{aligned}$$

además

$$\begin{aligned} \text{Cov}(U_i, U_j) &= E \{ U_i, U_i \} \quad \text{para } i = j \\ &= E \{ U_i^2 \} \\ &= \text{Var}(U_i) \quad (2.6) \\ &= \sigma^2 \end{aligned}$$

Es necesario, para poder construir intervalos de confianza y efectuar pruebas de hipótesis, considerar una distribución para las perturbaciones estocásticas, lo que dará como consecuencia una distribución para las Y .

Ciertamente, no es posible predecir, el valor específico de U en una observación, sin embargo, se pueden hacer proposiciones acerca de las principales características de su distribución de probabilidad. Antes que nada es claro que las U 's pueden tomar valores positivos y negativos, ahora bien, dado que U es la suma algebraica de los efectos positivos y negativos se esperan, con mayor frecuencia valores numéricos pequeños para U , de tal manera que la función deberá ser unimodal alrededor de algún valor pequeño para U . Si se añade además la suposición de simetría, entonces el valor unimodal coincide con el valor esperado de cero.

Si se desea ajustar un modelo probabilístico alrededor de cero y simétrico, se puede tomar a la distribución normal o bien a la t , sin embargo, la más

usada es la primera, lo cual se justifica mediante el teorema del límite central que en su caso general indica que: "si se tiene una suma de variables aleatorias independientes, con medias y varianzas finitas, y con la condición de Lindberg ($\sum_{i=1}^{\infty} \sigma_i^3 < \infty$) que implica que cada varianza por si sola es de poca importancia, la suma tiene distribución normal"

En resumen, la forma matemática del modelo es:

$$Y_i = \alpha + \beta X_i + U_i \quad i = 1, 2, \dots, n.$$

$$E(U_i) = 0 \quad \text{para toda } i$$

$$E(U_i U_j) = \begin{cases} 0 & i \neq j \\ \sigma^2 & i = j \end{cases}$$

$$U_i \sim N(0, \sigma^2) \quad \text{para toda } i$$

Ahora bien, dado que en observaciones sucesivas la variable independiente X es fija y que la variable dependiente Y es una función lineal de X , el valor esperado de Y para un valor dado de X se obtiene de la función de regresión $\alpha + \beta X$; es decir,

$$\begin{aligned} E(Y/X) &= E(\alpha + \beta X + U) \\ &= \alpha + \beta X + E(U) \\ &= \alpha + \beta X \end{aligned} \quad (2.7)$$

y la varianza de Y para un valor dado de X es igual a

$$\text{Var}(Y/X) = \text{Var}(\alpha + \beta X + U) = \text{Var}(U) = \sigma^2 \quad (2.8)$$

La suposición de que $\text{Var}(Y/X)$ es igual para todos los valores de X es útil, dado que permite usar en la estimación de σ^2 todas las observaciones de X e Y . Sin embargo, cuando se trabaja con datos económicos, tal

suposición no es justificable totalmente, ya que si se trata de ajustar el gasto en consumo (Y) al ingreso (X), es obvio que la varianza en el consumo no será la misma para todos los niveles de ingreso. Las familias ricas consumen en promedio más que las pobres, y también las primeras tienen mayor variación en el consumo. Este caso se analiza en el Capítulo IV.

Del supuesto de normalidad en U, y de las fórmulas (2.7) y (2.8) se sigue que para un valor dado de X, Y se distribuye como $N(\alpha + \beta X, \sigma^2)$. En la figura 3 se muestra una función de distribución de este tipo.

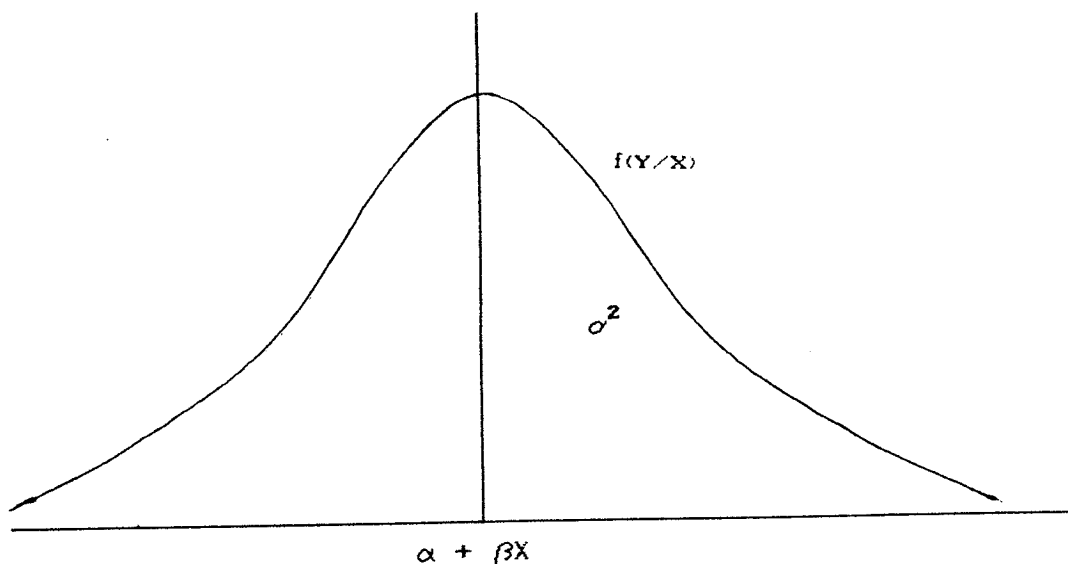


Figura 2.3. Función de densidad de probabilidad condicional.

La posición de $f(Y/X)$ cambia con cada valor de X. Esto se muestra geoméricamente en la figura 4.

Cuando se conoce toda la población de valores (X_i, Y_i) es posible calcular los valores exactos de los parámetros de regresión.

Sin embargo, dado que se trabaja generalmente con

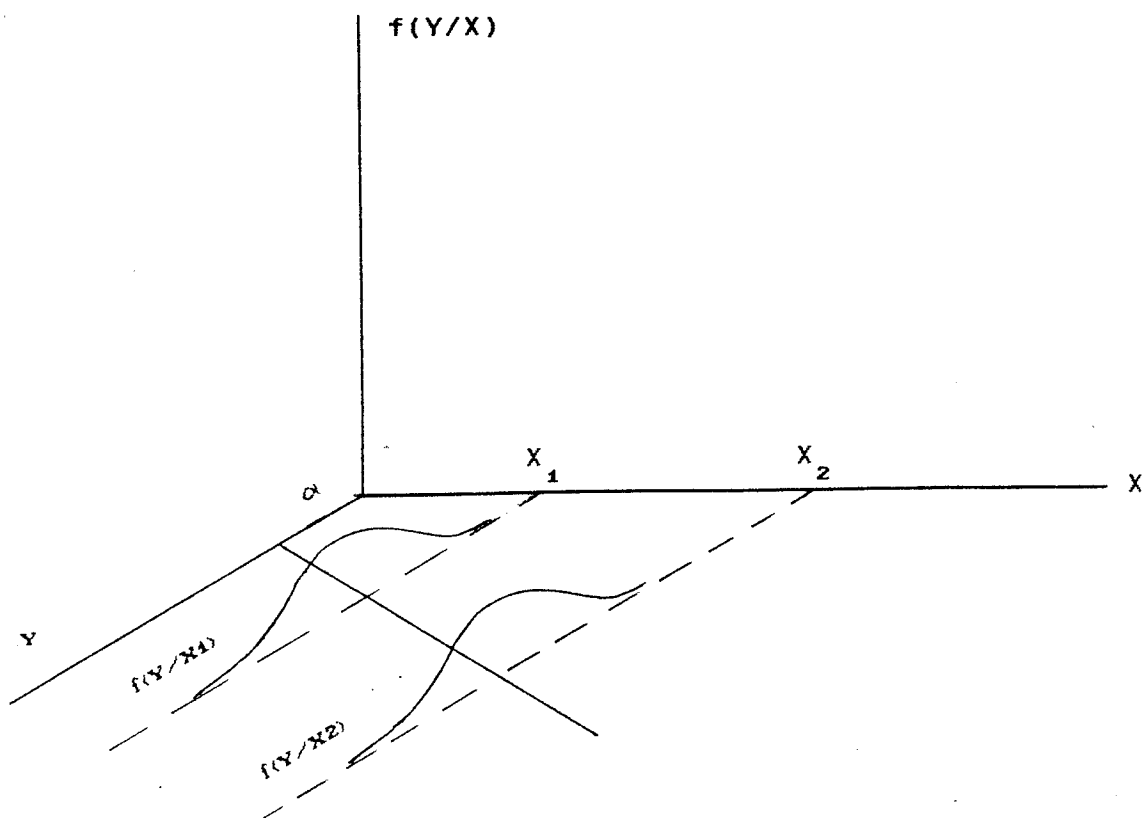


Figura 2.4. Relación de regresión.

muestras, el problema desde el punto de vista estadístico consiste en estimar de la mejor manera dichos parámetros.

En el proceso de estimación, se comienza primero con una muestra de tamaño n , cuyo diagrama de dispersión se muestra en la figura 2.5.

Los puntos de dispersión se grafican, en el cuadrante positivo, siendo esto convencional dado que las variables económicas en general, solo tienen sentido cuando son no negativas. Una línea recta através del diagrama de dispersión puede tomarse como una estimación de la relación hipótetica:

$$Y = \alpha + \beta X + U$$

Supongamos que (a, b) son los estimadores de (α, β) ,

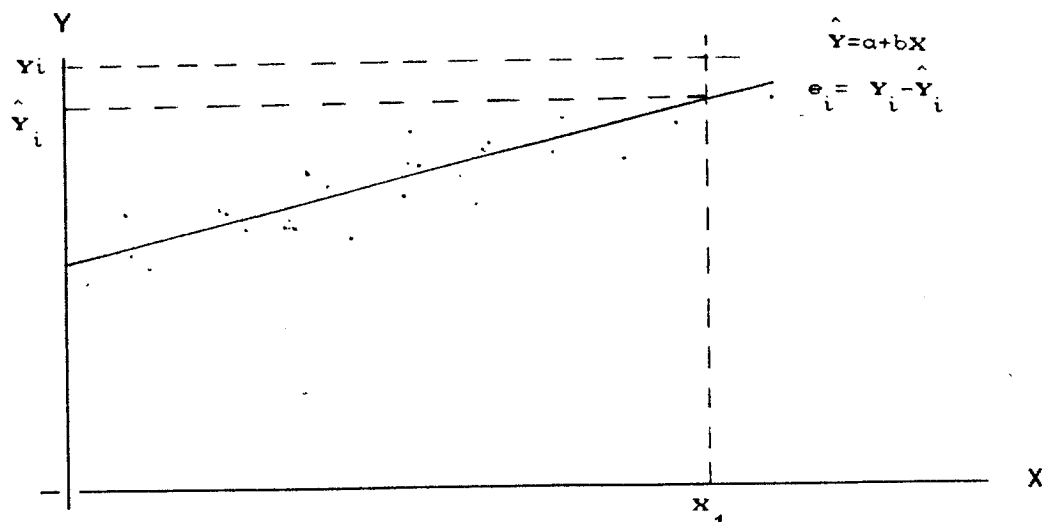


Figura 2.5. Diagrama de dispersión para una muestra de tamaño n .

y que definimos los residuos para cada par (X_i, Y_i) , como:

$$\begin{aligned} e_i &= Y_i - \hat{Y} \\ &= Y_i - a - bX_i \quad i = 1, 2, \dots, n \end{aligned} \quad (2.9)$$

El siguiente paso consiste en escoger el criterio que haga los residuos tan pequeños como sea posible. Uno de ellos podría ser:

seleccionar a , b tales que:

$$\sum_{i=1}^n e_i = 0$$

esto es:

$$\sum_{i=1}^n e_i = \sum_{i=1}^n (Y_i - a - bX_i) = 0$$

dividiendo entre n :

$$\bar{Y} = a + b\bar{X} \quad (2.10)$$

La ecuación (2.10) exige que a y b sean tales que la recta de estimación pase por el punto de coordenadas (\bar{X}, \bar{Y}) y así, la suma de los residuos será nula. Esta exigencia es

demasiado ambigua ya que existe un número infinito de rectas de estimación que cumplen con tal requisito.

Existe un criterio más fuerte, el de mínimos cuadrados, que pide:

seleccionar a, b tales que:

$$\sum_{i=1}^n e_i^2$$

sea mínimo.

Así pues, es necesario que la función:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - a - bX_i)^2 = f(a, b) \quad (2.11)$$

sea minimizada.

Para minimizar (2.11), se deriva parcialmente con respecto a a y b . Las derivadas parciales se igualan con cero y se resuelve el sistema resultante, cumpliéndose así la condición necesaria para obtener un mínimo en $f(a, b)$. De ésta forma, se obtiene en forma sucesiva

$$\frac{\delta \sum e_i^2}{\delta a} = -2 \sum (Y_i - a - bX_i) = -2 \sum e_i = 0 \quad (2.12)$$

$$\frac{\delta \sum e_i^2}{\delta b} = -2 \sum (Y_i - a - bX_i)X_i = -2 \sum e_i X_i = 0 \quad (2.13)$$

Simplificando tales expresiones se obtiene:

$$\sum_{i=1}^n Y_i = n a + b \sum_{i=1}^n X_i \quad (2.14.a)$$

$$\sum_{i=1}^n Y_i X_i = a \sum_{i=1}^n X_i + b \sum_{i=1}^n X_i^2 \quad (2.14.b)$$

Las ecuaciones (2.14), son llamadas las ecuaciones normales de la regresión por mínimos cuadrados. Usando

matrices se obtienen las expresiones para los estimadores a y b como sigue:

siendo

$$\begin{bmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma XY \end{bmatrix} \quad (2.15)$$

entonces:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma Y \\ \Sigma XY \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{n\Sigma X^2 - (\Sigma X)^2} \begin{bmatrix} \Sigma X^2 - \Sigma X \\ -\Sigma X & n \end{bmatrix} \begin{bmatrix} \Sigma Y \\ \Sigma XY \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{n\Sigma X^2 - (\Sigma X)^2} \begin{bmatrix} \Sigma X \Sigma Y - \Sigma X \Sigma XY \\ n \Sigma XY - \Sigma X \Sigma Y \end{bmatrix}$$

es decir:

$$a = \frac{\Sigma X^2 \Sigma Y - \Sigma X \Sigma XY}{n \Sigma X^2 - (\Sigma X)^2} \quad (2.16)$$

$$b = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2} \quad (2.17)$$

La condición de suficiencia para que el punto (a,b) de f(a,b) sea mínimo, es difícil de obtener si no se usa notación matricial, por lo que la derivación de tal condición se deja para el Capítulo III.

Mientras tanto, intuitivamente se puede ver que si f(a,b) es cuadrática en a y b con coeficientes positivos, entonces f(a,b) es geoméricamente, un paraboloides en revolución, abriéndose hacia arriba por lo que (a,b) corresponde a un mínimo de f(a,b).

La regresión lineal tiene ciertas propiedades

importantes:

1. La línea de regresión pasa por (\bar{X}, \bar{Y}) ($\sum e_i = 0$)
2. Los residuos tienen covarianza cero con respecto a los valores muestrales X y también con los valores estimados \hat{Y} . Esto es:

$$\begin{aligned} \text{Cov}(X, e) &= \frac{1}{n} \sum (X - E(X)) (e - E(e)) \\ &= \frac{1}{n} \sum (X - E(X)) e \\ &= \frac{1}{n} \sum eX - \frac{1}{n} E(X) \sum e = 0 \end{aligned}$$

Por otra parte:

$$\begin{aligned} \text{Cov}(\hat{Y}, e) &= \frac{1}{n} \sum (\hat{Y} - E(\hat{Y})) (e - E(e)) \\ &= \frac{1}{n} \sum (\hat{Y} e - E(\hat{Y}) e) \\ &= \frac{1}{n} \hat{Y} \sum e - E(\hat{Y}) \sum e \\ &= 0 \end{aligned}$$

3. Los estimadores de regresión pueden ponerse (y calcularse) en forma de desviación, trasladando el origen a (\bar{X}, \bar{Y}) de la forma siguiente como se muestra en la figura 2.6.

Originalmente la línea de regresión estimada es:

$$\hat{Y} = a + bX$$

Si se hace el cambio de coordenadas se tiene que la ecuación de regresión es:

$$\hat{y} = bx \quad (2.18)$$

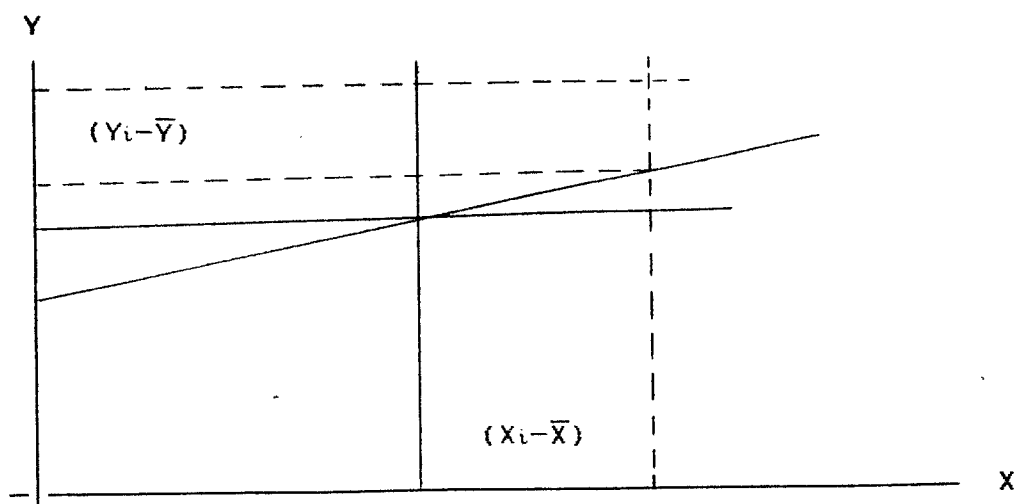


Figura 2.6. Representación gráfica de la forma desviación de la recta de mínimos cuadrados.

usando el criterio de mínimos cuadrados se tiene ahora:

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - bx_i)^2$$

usando la condición necesaria:

$$\frac{\delta \sum e_i^2}{\delta b} = -2 \sum (y_i - bx_i)x_i = 0$$

con lo que:

$$b = \frac{\sum x_i y_i}{\sum x_i^2} \quad (2.19)$$

y

$$a = \bar{Y} - b\bar{X}$$

donde: $x_i = (X_i - \bar{X})$; $y_i = (Y_i - \bar{Y})$

4. Descomposición en suma de cuadrados.

La variación total en Y puede ser explicada como

la suma de dos componentes:

- a. La variación explicada por la regresión lineal.
- b. La variación no explicada por la regresión. De la propiedad 3.

$$y_i = \hat{y}_i + e_i = bx_i + e_i$$

elevando al cuadrado ambos miembros y sumando para n , se obtiene:

$$\begin{aligned} \sum y_i^2 &= \sum \hat{y}_i^2 + \sum e_i^2 + 2 \sum \hat{y}_i e_i \\ &= b^2 \sum x_i^2 + \sum e_i^2 + 2b \sum x_i e_i \\ &= b^2 \sum x_i^2 + \sum e_i^2 \end{aligned} \quad (2.20)$$

Las cantidades claves en (2.20) son:

$\sum y_i^2$: Suma total de cuadrados en la variable dependiente medida con respecto a su media (STC).

$\sum e_i^2$: Suma de residuales al cuadrado (SRC) (no explicados).

$\sum \hat{y}_i^2 = b^2 \sum x_i^2 = b \sum xy = \frac{(\sum xy)^2}{\sum x^2}$: Suma de cuadrados explicada (SEC).

Así, la ecuación (2.20) se puede expresar en forma alterna como:

$$STC = SEC + SRC.$$

Falta, para terminar con el problema de la estimación de parámetros, obtener un estimador para σ^2 . Desafortunadamente se hacen necesarios resultados que se encuentran en la derivación de las propiedades de los estimadores a y b . Por lo pronto se propone, sin derivación

que:

$$S^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} \quad (2.21)$$

es un estimador insesgado de σ^2 .

Propiedades de los Estimadores

El método de mínimos cuadrados es solo uno de los posibles para estimar los parámetros de regresión. Se pueden definir fácilmente otros estimadores; por ejemplo, se puede ajustar una línea recta tal que pase por el primer y último punto en el diagrama de dispersión, o bien, obtener un promedio de los dos primeros y los dos últimos y hacer que pase una línea a través de las dos medias.

Se puede, de igual forma tomar promedios de las $n/2$ primeros puntos y otro de los $n/2$ restantes. La pregunta ahora es: que principio de estimación es el mejor?

La respuesta dada por la estadística clásica es escoger el principio de estimación en base a ciertas propiedades importantes de los parámetros que resulten. Esas propiedades se refieren, al comportamiento de los estimadores en muestreo repetido.

Para efectuar la comparación entre estimadores, se comienza suponiendo que los datos son generados por el modelo lineal.

$$Y = \alpha + \beta X + U$$

Siendo que las X 's están fijas, la única fuente de variación en la ecuación generadora es el término U .

Si se extraen 10,000 muestras de tamaño n , la

aplicación de cualquier principio de estimación da 10,000 valores para los estimadores a, b, σ^2 . En particular 10,000 pares de estimadores (a, b) ; los cuales pueden arreglarse en una distribución bivariada. Si el número de muestras es "grande" la distribución será aproximadamente continua. De esta manera se obtendrá una función conjunta.

$$F(a, b)$$

de la cual pueden obtenerse las funciones marginales $f(a)$ y $f(b)$.

A fin de desarrollar conceptos, se toma primero a $f(b)$ donde

$$f(b) = \int_{-\infty}^{\infty} f(a, b) da$$

Existen, a saber, tres características de la distribución muestral (marginal) cruciales en la evaluación de un estimador. Estas son, la media, la varianza y el error cuadrático medio. El valor medio.

$$E(b) = \int_{-\infty}^{\infty} b f(b) db$$

da el promedio que puede ser generado para el estimador en aplicaciones sucesivas. El sesgo del estimador con respecto al parámetro real se define como:

$$\text{Sesgo } (b) = E(b) - \beta \quad (2.22)$$

Si el sesgo es cero, el estimador se llama insesgado.

La varianza de la distribución.

$$\text{Var}(b) = \sigma^2 = E [b - E(b)]^2$$

$$\text{Var}(b) = \int_{-\infty}^{\infty} [b - E(b)]^2 f(b) db.$$

mide la dispersión del estimador alrededor de la media.

Si con éstas dos características se trata de comparar dos estimadores ambos insesgados, resulta obvio tomar el de varianza más pequeña. Al estimador así escogido se le llama mejor estimador insesgado. Sin embargo, la dificultad radica en la comparación eficiente de dos estimadores sesgados y con distinta varianza. De tal forma que se crea la necesidad de definir el error cuadrático medio como:

$$\begin{aligned} \text{ECM} &= E \{ (b - \beta)^2 \} \\ &= \int_{-\infty}^{\infty} (b - \beta)^2 f(b) db. \\ &= E \{ (b - E(b)) + (E(b) - \beta) \}^2 \\ &= E(b - E(b))^2 + E(E(b) - \beta)^2 \\ &\quad + 2 E \{ (b - E(b)) (E(b) - \beta) \} \\ &= \text{Var}(b) + \{ \text{sesgo}(b) \}^2 \end{aligned} \quad (2.23)$$

Así usando el criterio del ECM se puede encontrar un estimador sesgado que sea preferible a uno insesgado, siempre y cuando la varianza del primero sea lo suficientemente pequeña para compensar su ECM con el estimador insesgado (que por supuesto tendrá mayor varianza).

Con estos conceptos introductorios, a continuación se procede a obtener las medias y varianzas de los estimadores a y b; principiando con b de la ecuación (2.19) se tiene:

$$\begin{aligned}
 b &= \frac{\sum x_i y_i}{\sum x_i^2} \\
 &= \frac{\sum x_i (Y_i - \bar{Y})}{\sum x_i^2} \\
 &= \frac{\sum x_i Y_i}{\sum x_i^2} - \frac{b \sum x_i}{\sum x_i^2} \\
 &= \frac{\sum x_i Y_i}{\sum x_i^2}
 \end{aligned}$$

o bien:

$$b = \sum w_i Y_i; \quad w_i = \frac{x_i}{\sum x_i^2} \quad (2.24)$$

Lo que muestra que b es un estimador lineal, dado que es función lineal de Y . De la ecuación (2.24) se sigue que:

$$\begin{aligned}
 \sum w_i &= \frac{\sum x_i}{\sum x_i^2} = 0; \\
 \sum w_i x_i &= \frac{\sum (x_i - \bar{x}) x_i}{\sum x_i^2} \\
 &= \frac{\sum (x_i^2 - \frac{1}{n} (\sum x_i)^2)}{\sum x_i^2} \\
 &= \frac{\sum x_i^2}{\sum x_i^2} \\
 &= 1 \\
 \sum w_i^2 &= \frac{\sum x_i^2}{(\sum x_i^2)^2} \\
 &= \frac{1}{\sum x_i^2} \quad (2.25)
 \end{aligned}$$

Siendo que:

$$b = \sum w_i Y_i = \sum w_i (\alpha + \beta x_i + U_i)$$

con los resultados (2.25):

$$b = \beta + \sum w_i U_i$$

(note que $(b - \beta) = \sum w_i U_i$)

tomando esperanzas se tiene:

$$\begin{aligned} E(b) &= \beta + \sum w_i E(U_i) \\ &= \beta \end{aligned} \quad (2.26)$$

Lo cual muestra que b es un estimador insesgado de β .

Por otra parte, la expresión para la varianza de b se obtiene como sigue:

por definición:

$$\begin{aligned} \text{Var}(b) &= E\{b - E(b)\}^2 \\ &= E\{b - \beta\}^2 \\ &= E\left\{\sum w_i U_i\right\}^2 \\ &= E\left\{\sum w_i^2 U_i^2 + 2 \sum_{i < j} w_i w_j U_i U_j\right\} \\ &= \sum w_i^2 E(U_i)^2 + \sum_{i < j} w_i w_j E(U_i U_j) \end{aligned}$$

con lo que

$$\text{Var}(b) = \frac{\sigma^2}{\sum x_i^2} \quad (2.27)$$

De forma similar para a , se obtienen la media y la varianza como sigue:

siendo

$$\begin{aligned} a &= \bar{Y} - b \bar{X} \\ &= \alpha + \beta \bar{X} + \bar{U} - b \bar{X} \end{aligned}$$

(note que $(a - \alpha) = (\beta \bar{X} + \bar{U} - b \bar{X}) = \bar{U} - (b - \beta) \bar{X}$).

Tomando esperanzas

$$E(a) = \alpha + E(\bar{U})$$

$$E(a) = \alpha \frac{1}{n} \sum_{i=1}^n E(U_i)$$

$$E(a) = \alpha \quad (2.28)$$

Con lo que también se muestra que a es un estimador insesgado de α . Continuando con a , se tiene que:

$$\begin{aligned} \text{Var}(a) &= E(a - E(a))^2 \\ &= E(a - \alpha)^2 \\ &= E\{\bar{U} - (b - \beta)\bar{X}\}^2 \\ &= E(\bar{U}^2) + \bar{X}^2 E(b - \beta)^2 - 2\bar{U}\bar{X}E(b - \beta) \\ &= \frac{\sigma^2}{n} + \frac{\bar{X}^2 \sigma^2}{\sum x_i^2} \end{aligned}$$

asi:

$$\text{Var}(a) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} \right] \quad (2.29)$$

además:

$$\begin{aligned} \text{Cov}(a,b) &= E\{(a - \alpha)(b - \beta)\} \\ &= E\{[\bar{U} - (b - \beta)\bar{X}](b - \beta)\} \\ &= -\bar{X} E(b - \beta)^2 \end{aligned}$$

dado que $E\{(b - \beta)\bar{U}\} = 0$

$$\text{Cov}(a,b) = -\bar{X} \frac{\sigma^2}{\sum x_i^2} \quad (2.30)$$

En este momento se está ya en condición de mostrar que:

$$S^2 = \frac{\sum e^2}{n - 2}$$

es un estimador insesgado de σ^2 . Además el hecho de mostrarlo en este momento se justifica, dado que en las

expresiones para $\text{Var}(a)$ y $\text{Var}(b)$ se necesita conocer a σ^2 , el cual no es posible obtener si se trabaja con datos muestrales. Así pues, considerando al i -ésimo residual.

$$\begin{aligned} e_i &= Y_i - \hat{Y}_i \\ &= \alpha + \beta X_i + U_i - a - bX_i \\ &= U_i - (b - \beta)X_i - (a - \alpha) \\ &= U_i - (b - \beta)X_i - (\bar{U} - (b - \beta)\bar{X}) \\ &= (U_i - \bar{U}) - (b - \beta)x_i \end{aligned}$$

elevando el cuadrado y sumando se tiene

$$\begin{aligned} \sum e_i^2 &= \sum [(U_i - \bar{U})^2 + (b - \beta)^2 x_i^2 - 2(U_i - \bar{U})(b - \beta)x_i] \\ &= \sum U_i^2 - n\bar{U}^2 + (b - \beta)^2 \sum x_i^2 - 2(b - \beta) \sum U_i x_i \end{aligned}$$

notando que:

$$\begin{aligned} E\left[\sum U_i^2\right] &= n \sigma^2 \\ E(\bar{U}^2) &= \text{Var}(\bar{U}) \\ (\text{Var}(\bar{U})) &= E(\bar{U}^2) - [E(\bar{U})]^2 \\ &= \frac{\sigma^2}{n} \\ E(b - \beta)^2 &= \frac{\sigma^2}{\sum x_i^2} \end{aligned}$$

y:

$$\begin{aligned} E\left\{(b - \beta) \sum U_i x_i\right\} &= E\left\{\sum W_i U_i\right\} \left\{\sum U_i x_i\right\} \\ &= E\left\{(W_{11} U_1 + W_{22} U_2 + \dots + W_{nn} U_n) (U_1 x_1 + U_2 x_2 + \dots + U_n x_n)\right\} \\ &= E\left\{W_{11} U_1^2 x_1 + W_{22} U_2^2 x_2 + \dots + W_{nn} U_n^2 x_n + \dots + W_{11} U_1 U_2 x_2 + \dots + W_{11} U_1 U_n x_n + \dots\right\} \\ &= \sigma^2 \sum W_i x_i \end{aligned}$$

$$E\left\{(b - \beta) \sum U_i x_i\right\} = 0$$

Así

$$\begin{aligned} E(\sum e_i^2) &= n\sigma_i^2 - \sigma_i^2 + \sigma_i^2 - 2\sigma_i^2 \\ &= (n - 2)\sigma^2 \end{aligned}$$

de donde

$$E\left[\frac{\sum e_i^2}{n - 2}\right] = \sigma^2 \quad (2.31)$$

con lo cual se muestra que el estimador propuesto para σ^2 es insesgado.

El paso siguiente es mostrar, mediante el teorema de Gauss-Markov, que los estimadores de mínimos cuadrados ordinarios son los de mínima varianza en la clase de los estimadores lineales insesgados.

Se comienza con mostrarlo para b .

Según se ha visto:

$$b = \sum W_i Y_i; \quad W_i = \frac{x_i}{\sum x_i^2}$$

por lo que b es un estimador de β lineal e insesgado, con varianza

$$\frac{\sigma^2}{\sum x_i^2}$$

Ahora, se define un estimador lineal general de β como.

$$b^* = \sum c_i Y_i \quad (2.32)$$

donde c_i ($i = 1, \dots, n$) es un conjunto de ponderaciones. Suponiendo ahora que los datos son generados por el modelo lineal

$$Y_i = \alpha + \beta X_i + U_i$$

se tiene

$$b^* = \alpha_i \sum c_i + \beta \sum c_i x_i + \sum c_i U_i$$

Tomando esperanzas:

$$E(b^*) = \alpha \sum c_i + \beta \sum c_i X_i$$

Si b^* ha de ser un estimador insesgado de β ; se requiere que

$$\sum c_i = 0 ;$$

$$\begin{aligned} \sum c_i X_i &= \sum c_i (x_i + \bar{X}) \\ &= \sum c_i x_i + \bar{X} \sum c_i \end{aligned}$$

$$\sum c_i X_i = \sum c_i x_i = 0$$

de esta forma, la varianza de b^* es:

$$\begin{aligned} \text{Var}(b^*) &= E\{[b^* - E(b^*)]^2\} \\ &= E\{(b^* - \beta)^2\} \\ &= E(b^{*2}) + \beta^2 - 2\beta E(b^*) \\ &= E(b^{*2}) - \beta^2 \end{aligned}$$

donde:

$$\begin{aligned} E(b^{*2}) &= E\{(\alpha \sum c_i + \beta \sum c_i x_i + \sum c_i U_i)^2\} \\ &= E\{(\beta + \sum c_i U_i)^2\} \\ &= E\{\beta^2 + (\sum c_i U_i)^2 + 2\beta \sum c_i U_i\} \\ &= \beta^2 + E\{(\sum c_i U_i)^2\} + 2\beta E\{\sum c_i U_i\} \\ &= \beta^2 + E\{\sum c_i^2 U_i^2 + \sum_{i < j} c_i c_j U_i U_j\} \\ &= \beta^2 + \sum c_i^2 E(U_i^2) + \sum c_i c_j E(U_i U_j) \\ &= \beta^2 + \sigma^2 \sum c_i^2 \end{aligned}$$

Así:

$$\text{Var}(b^*) = \sigma^2 \sum c_i^2 \quad (2.33)$$

para comparar la var (b^*) con var (b) se pone:

$$c_i = w_i + (c_i - w_i)$$

así:

$$\begin{aligned} \Sigma c_i^2 &= \Sigma w_i^2 + \Sigma (c_i - w_i)^2 + 2 \Sigma w_i (c_i - w_i) \\ &= \Sigma w_i^2 + \Sigma (c_i - w_i)^2 + 2 [\Sigma w_i c_i - \Sigma w_i^2] \\ &= \Sigma w_i^2 + \Sigma (c_i - w_i)^2 + 2 \left[\Sigma \frac{c_i x_i}{\Sigma x^2} - \frac{1}{\Sigma x^2} \right] \end{aligned}$$

de tal manera que:

$$\sigma^2 \Sigma c_i^2 = \sigma^2 \Sigma w_i^2 + \sigma^2 \Sigma (c_i - w_i)^2$$

o bien:

$$\text{Var}(b^*) = \text{Var}(b) + S^2 \Sigma (c_i - w_i)^2$$

donde se ha usado el estimador S^2 para σ^2 , donde

$$\Sigma (c_i - w_i) \geq 0$$

por lo que:

$$\text{Var}(b^*) \leq \text{Var}(b) \quad (2.34)$$

Este resultado, establece que b es el estimador de mínima varianza dentro de la clase de los estimadores lineales insesgados. Esto se puede escribir como:

b es el mejor estimador lineal insesgado de β
de manera semejante, se puede mostrar que:

a es el mejor estimador lineal insesgado de α

Coeficiente de Correlación

Los desarrollos mostrados hasta ahora se han hecho en base a ciertos supuestos, uno de ellos es la relación lineal entre X e Y . Como se recordará, para ajustar una función de regresión se sugirió como un primer paso, que los datos disponibles se graficaran en un diagrama de

dispersión, sin embargo, es obvio que una inspección visual no indica de manera cuantitativa la "bondad de ajuste" a una línea recta.

Existe un concepto estadístico que indica el grado de asociación lineal entre variables, tal es el coeficiente de correlación lineal. El hecho de que tal concepto se trate después de haber obtenido estimadores, descomposición del modelo, etc., obedece a que el coeficiente de correlación está estrechamente relacionado con b , el estimador de la pendiente de la línea de regresión, y puede ponerse en términos de la SCR y STC.

En el trabajo práctico, antes de obtener los estimadores, es recomendable analizar los datos a ajustar usando el coeficiente de correlación y en base a su valor, decidir si el ajuste lineal es el más adecuado para relacionar las variables en cuestión.

Cuando se trata de ver que tan asociadas están dos variables aleatorias, digamos X e Y , con una distribución bivariada, se utiliza el coeficiente de correlación poblacional, definido por

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} \quad (2.35)$$

donde:

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

$$E(X) = \mu_x; E(Y) = \mu_y; \text{Var}(X) = \sigma_x^2; \text{Var}(Y) = \sigma_y^2$$

Cuando se trabaja con muestras aleatorias de una distribución bivariada normal, el estimador de máxima verosimilitud de ρ es:

$$\hat{\rho} = r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}}$$

$$\hat{\rho} = r = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2][n \sum Y_i^2 - (\sum Y_i)^2]}}$$

El estimador r se denomina "coeficiente de correlación de la muestra". En otras palabras, dicho coeficiente es igual al estimador de $\text{Cov}(X, Y)$ entre el producto de los estimadores de σ_x y σ_y . Otra manera conveniente de expresar a r es:

$$r = \frac{\sum xy}{n S_x S_y} \quad (2.36)$$

donde:

$$S_x = \sqrt{\frac{\sum x^2}{n}} \quad ; \quad S_y = \sqrt{\frac{\sum y^2}{n}}$$

Son las desviaciones standard muestrales de X e Y .

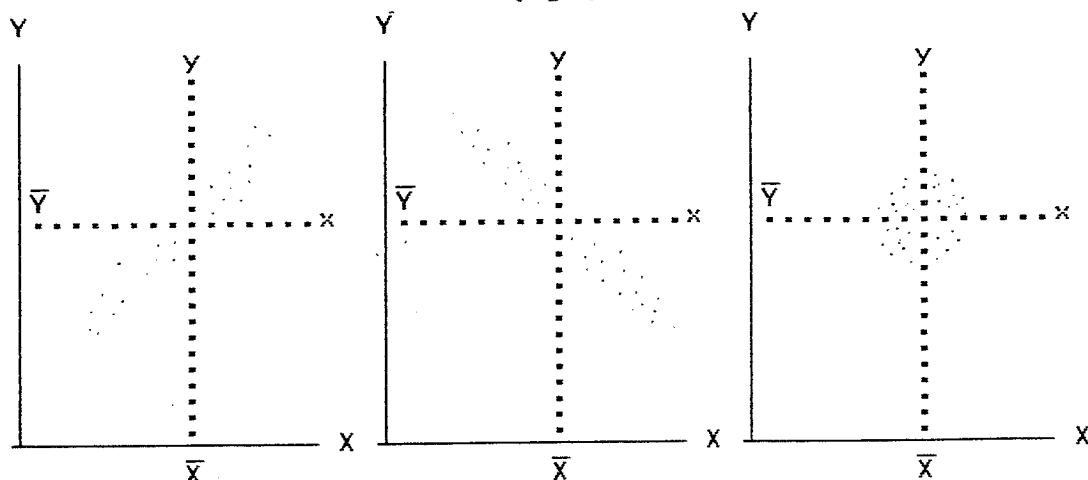
En el modelo de regresión desarrollado, se establece explícitamente que la variable X no es aleatoria, por lo que un coeficiente de correlación muestral calculado en base a los datos generados por un modelo de regresión simple, no mide exactamente lo mismo que un coeficiente de correlación muestral basado en datos extraídos de una distribución bivariada.

La relación entre el análisis de regresión y el de correlación puede entenderse como sigue: en base a los diagramas de dispersión presentados en la figura 2.7. y en la expresión para r dada en la ecuación (2.36) el signo

dependerá solo del numerador.

El signo del producto $x_i y_i$ en el "plano de dispersión", con origen en (\bar{X}, \bar{Y}) es positivo en los cuadrantes I y III, y negativo en II y IV. Por tanto, si la mayor parte de las observaciones muestrales se encuentran en los cuadrantes I y III, la suma de los productos

$$\sum_{i=1}^n X_i Y_i$$



(2.7.a)

(2.7.b)

(2.7.c)

Figura 2.7. Diagramas de dispersión.

tenderá a ser positiva (figura 2.7.a); mientras que si caen en II y IV (figura.2.7.b) la suma de los productos tenderá a ser negativa.

Por otra parte, si el diagrama de dispersión es como el mostrado en la figura 2.7.c, la suma $\sum_{i=1}^n x_i y_i$ tenderá a anularse.

La cantidad $\sum_{i=1}^n x_i y_i$ podría servir como una medida de asociación entre X e Y, si no fuera por dos defectos:

1. Su valor numérico aumenta con el tamaño de la muestra. Esto puede corregirse parcialmente dividiendo $\sum_{i=1}^n x_i y_i$ entre el tamaño de la muestra,

obteniéndose así la covarianza entre X e Y:

$$\text{Cov}(x,y) = \frac{\sum xy}{n}$$

2. La suma $\sum x_i y_i$ por ende la covarianza depende de las unidades de X e Y, de tal forma, que el valor de la covarianza se ve afectado si X e Y se miden en pesos o centavos. Existe una forma de estandarizar la covarianza, esta es, dividir cada desviación (de X y Y) con respecto a su media entre las desviaciones estándar de cada variable. Esto da el coeficiente de correlación.

Existe, según se ha mencionado, una relación entre el coeficiente de correlación y el estimador b.

Tomando el coeficiente de correlación:

$$r = \frac{\sum xy}{\sqrt{\sum x^2} \sqrt{\sum y^2}}$$

se arregla de la forma:

$$r = \frac{\sum xy}{\sum x^2} \left\{ \frac{\sqrt{\sum x^2}}{\sqrt{\sum y^2}} \right\}$$

usando ahora la expresión (2.19) se obtiene:

$$r = b \frac{S_x}{S_y}$$

o bien

$$b = r \frac{S_y}{S_x} \quad (2.37)$$

La expresión (2.37) muestra la relación entre la pendiente de la línea de regresión y el coeficiente de correlación. Además, como también ya se mencionó, el coeficiente de correlación se escribe, implícitamente, en términos de SRC y STC, según se muestra a continuación.

Elevando al cuadrado r se obtiene

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2} = \frac{b \sum xy}{\sum y^2}$$

o bien:

$$r^2 = \frac{SEC}{STC}$$

ahora, dado que $STC = SEC + SRC$; se tiene:

$$\frac{SEC}{STC} = 1 - \frac{SRC}{STC}$$

$$r^2 = 1 - \frac{SRC}{STC} \quad (2.38 \text{ a})$$

$$= 1 - \frac{\sum e^2/n}{\sum y^2/n} \quad (2.38 \text{ b})$$

De acuerdo a las ecuaciones (2.38) se tiene que

$$r^2 \geq 0$$

y como

$$0 \leq SRC \leq STC$$

$$0 \leq r^2 \leq 1 \quad (2.39)$$

r^2 se llama el coeficiente de determinación.

La condición (2.39) muestra que

$$-1 \leq r \leq 1 \quad (2.40)$$

r puede calcularse a partir de r^2 .

Falta solo de aclarar, Qué valor debe tener r para que se considere que el ajuste es bueno?. Para contestar se supone que existe un estimador b tal que

$$y_i = -bx_i \quad \text{para } i = 1, 2, \dots, n$$

o bien:

$$y_i + bx_i = 0$$

Elevando al cuadrado y sumando de uno a n , se tiene:

$$\sum (y_i + bx_i)^2 = 0$$

desarrollando:

$$\sum (y_i^2 + 2b y_i x_i + b^2 x_i^2) = 0$$

$$\sum y_i^2 + 2b \sum x_i y_i + b^2 \sum x_i^2 = 0$$

la cual es una ecuación cuadrática en b ; y para que se cumpla la proposición de que el valor de b sea único, es necesario que el discriminante sea nulo, esto es:

$$(2 \sum xy)^2 - 4 \sum x^2 \sum y^2 = 0$$

o bien

$$(\sum xy)^2 = \sum x^2 \sum y^2$$

Esto ocurre si y solo si $r^2 = 1$, con lo que $r = \pm 1$. Así pues, si $r = \pm 1$ el ajuste lineal será perfecto; es decir todos los puntos de dispersión caen en línea recta, de pendiente positiva o negativa según sea el signo de r , (el cual obviamente coincidirá con el signo del estimador de la pendiente).

En general, se puede pensar que entre más cercano esté r de ± 1 , el ajuste será mejor, mientras que un valor para r de cero indicará un mal ajuste.

Conviene aclarar, que el hecho de que dos variables presenten un alto grado de correlación lineal, no implica que exista una relación causa efecto de una sobre la otra, ambas pueden estar influenciadas por otras variables, de manera que den una relación puramente matemática. A manera de ejemplos:

a) En Alemania se encontró que el número de nidos de cigüeña y la producción de acero presentaban $r = 0.99$

b) En E.U.A. el salario de profesores y el consumo de licor $r = 0.98$

En la práctica, r se calcula a partir de r^2 ,

asociándosele el signo de b , y el éxito en el manejo del coeficiente de correlación estará determinado por el conocimiento que se tenga del campo que se esté trabajando.

Estimación por Intervalos.

La estimación por intervalos de los parámetros de regresión, difiere de la estimación puntual, en que no proporciona un valor estimado para cada parámetro, si no que proporciona un intervalo dentro del cual, con una probabilidad deseada, se encuentra el verdadero valor del parámetro en cuestión.

En forma más específica, la idea de construir intervalos de confianza consiste en encontrar dos números positivos δ y α que estén entre 0 y 1, de manera que la probabilidad de que el intervalo $(\theta - \delta, \theta + \delta)$ contenga al valor verdadero Ξ sea $1 - \alpha$, donde θ es un estimador de Θ .

Simbólicamente:

$$P(\theta - \delta \leq \Theta \leq \theta + \delta) = 1 - \alpha \quad (2.41)$$

tal intervalo, en caso de existir, se conoce como intervalo de confianza. $(1 - \alpha)$ se llama coeficiente de confianza y α ($0 \leq \alpha \leq 1$) se conoce como nivel de significancia. Los límites del intervalo de confianza se conocen como límites de confianza o bien como valores críticos.

Para poder construir intervalos de confianza es necesario conocer la distribución de probabilidades de los estimadores. Esto se resuelve usando la suposición de normalidad del modelo de regresión, que hasta este momento

no se había usado.

Dado que a y b son funciones lineales de las e 's, tanto a como b se distribuyen normalmente, y conociendo ya su media y varianza la distribución está completamente determinada. Con esto:

$$a \sim N \left[\alpha, \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma x^2} \right] \right] \quad (2.42)$$

$$b \sim N \left[\beta, \frac{\sigma^2}{\Sigma x^2} \right] \quad (2.43)$$

en donde, por supuesto, en lugar de poner la varianza, se usa su estimador:

$$S^2 = \frac{\Sigma e^2}{n - 2}$$

además, se puede demostrar que

$$\frac{\Sigma e^2}{\sigma^2} \sim \chi^2_{(n-2)} \quad (2.44)$$

y que Σe^2 se distribuye independiente de $f(a,b)$ (la función conjunta de a y b).

Se comienza por construir intervalos de confianza para β .

Dado que la distribución t está dada por la razón de una normal estándar y la raíz cuadrada de una χ^2 dividida entre sus grados de libertad, y como:

$$\frac{b - \beta}{\sigma / \sqrt{\Sigma x^2}} \sim N(0,1) \quad (2.45)$$

entonces:

U.A.A.A.N.

$$\frac{\frac{(b - \beta) \sqrt{\sum x^2}}{\sigma}}{\frac{\sqrt{\sum e^2}}{\sigma \sqrt{n-2}}} = \frac{b - \beta}{S / \sqrt{\sum x^2}} \sim t_{(n-2)} \quad (2.46)$$

si la muestra es grande ($n \geq 30$) se considera que:

$$\frac{(b - \beta) \sqrt{\sum x^2}}{S} \sim N(0,1) \quad (2.47)$$

Así, el intervalo de confianza para β se construye de la siguiente manera:

$$P \left[-t_{\alpha/2} \leq t \leq t_{\alpha/2} \right] = 1 - \alpha \quad (2.48)$$

donde t está dada por (2.46). $t_{\alpha/2}$ es el valor de t obtenido de la distribución t para el nivel de significancia $\alpha/2$ y $n-2$ grados de libertad.

Reemplazando (2.46) en (2.48) se tiene:

$$P \left[-t_{\alpha/2} \leq \frac{b - \beta}{S} \sqrt{\sum x^2} \leq t_{\alpha/2} \right] = 1 - \alpha$$

reordenando, se llega a:

$$P \left[b - \frac{t_{\alpha/2} S}{\sqrt{\sum x^2}} \leq \beta \leq b + \frac{t_{\alpha/2} S}{\sqrt{\sum x^2}} \right] = 1 - \alpha \quad (2.49)$$

La ecuación (2.49) proporciona el intervalo de confianza de $100(1-\alpha)$ por ciento para β .

Según se ha establecido, el valor de b es aleatorio, por lo que el intervalo de confianza también lo es.

La interpretación de un intervalo de confianza es que dado un nivel de confianza $(1-\alpha)$ (o bien, un nivel de significancia α) en un muestreo sucesivo a largo plazo, $100(1-\alpha)$ por ciento de los intervalos construidos deberán

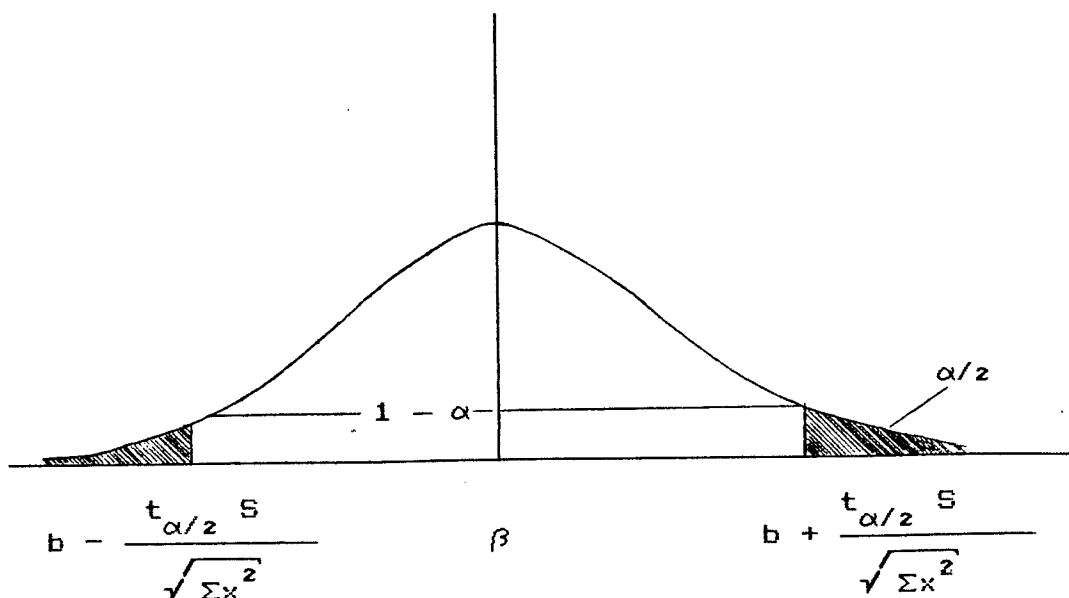


Figura 2.8. Interpretación gráfica de un intervalo de confianza para β .
contener al valor β .

Gráficamente, un intervalo de confianza para β usando la distribución t se representa según la figura 8.

De igual forma, para construir un intervalo de confianza para a .

Si

$$a \sim N \left[\alpha, \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma x^2} \right] \right]$$

entonces:

$$\frac{a - \alpha}{\sigma \sqrt{\frac{1}{n} + \left[\bar{X}^2 / \Sigma x^2 \right]}} \sim N(0,1)$$

y

$$\frac{\frac{a - \alpha}{\sigma \sqrt{\frac{1}{n} + \left[\bar{X}^2 / \Sigma x^2 \right]}}}{\frac{\sqrt{\Sigma e^2}}{\sigma \sqrt{n - 2}}} = \frac{a - \alpha}{S \sqrt{\frac{1}{n} + \left[\bar{X}^2 / \Sigma x^2 \right]}} \sim t_{\alpha/2}$$

así:

$$P \left[-t_{\alpha/2} \leq \frac{a - \alpha}{S \sqrt{\frac{1}{n} + \left[\bar{X}^2 / \Sigma x^2 \right]}} \leq t_{\alpha/2} \right] = 1 - \alpha$$

reordenando se tiene:

$$P \left[a - t_{\alpha/2} S \sqrt{\frac{1}{n} + \left[\bar{X}^2 / \Sigma x^2 \right]} \leq \alpha \leq a + t_{\alpha/2} S \sqrt{\frac{1}{n} + \left[\bar{X}^2 / \Sigma x^2 \right]} \right] = 1 - \alpha \quad (2.50)$$

la expresión (2.50) define un intervalo de confianza para α . Su interpretación y representación gráfica son análogas a las de β .

Finalmente, se procede a construir un intervalo de confianza para σ^2 . De acuerdo a la ecuación (2.44):

$$\frac{\Sigma e^2}{\sigma^2} = \frac{S^2 (n - 2)}{\sigma^2} \sim \chi^2_{(n-2)}$$

Se puede establecer el intervalo de confianza para σ^2 usando la distribución χ^2 . La idea para la construcción del intervalo de confianza para σ^2 es la misma que la usada para a y b .

$$P \left[\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2} \right] = 1 - \alpha \quad (2.52)$$

donde el valor de χ^2 es el dado en la ecuación (2.51) con

$n-2$ grados de libertad. $\chi^2_{1-\alpha/2}$ y $\chi^2_{\alpha/2}$ dan los valores de χ^2 obtenidos en la misma tabla, de tal manera que en forma gráfica las probabilidades (áreas bajo la curva de χ^2) queden repartidas según se muestra en la figura 9.

Reemplazando el valor de χ^2 en (2.52) y ordenando términos, se obtiene:

$$P \left[(n-2) \frac{S^2}{\chi^2_{1-\alpha/2}} \leq \sigma^2 \leq (n-2) \frac{S^2}{\chi^2_{\alpha/2}} \right] = 1 - \alpha \quad (2.53)$$

la ecuación (2.53) define un intervalo de confianza de $100(1-\alpha)$ por ciento de confianza para σ^2

Pruebas de hipótesis.

El problema de la prueba de hipótesis radica en como justificar de manera significativa si el valor de un estimador da evidencia de que el verdadero valor del parámetro en cuestión sea un número dado.

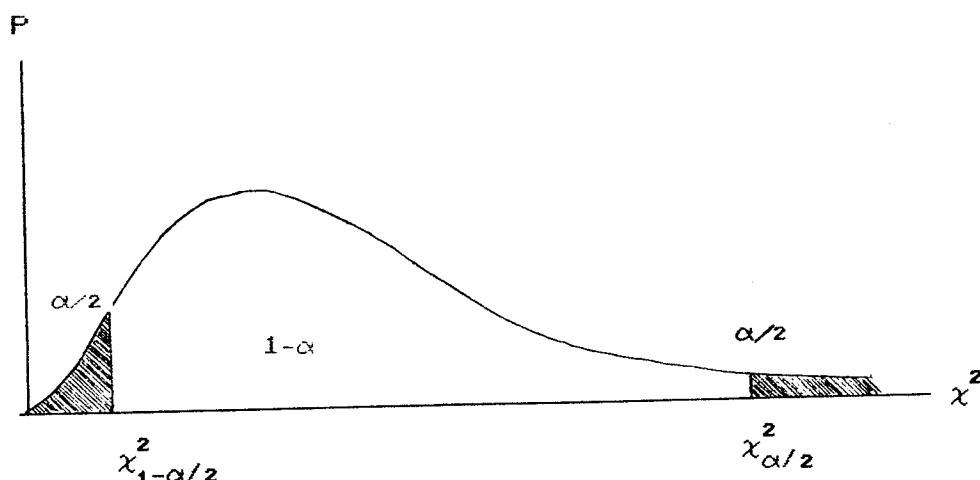


Figura 2.9. Intervalo de confianza del $100(1-\alpha)$ por ciento para χ^2 con $n-2$ g.l.

En otras palabras, el problema es decir que tan "cerca" debe estar el valor de un estimador, del valor

hipotético del parámetro, de tal forma que permita aceptar tal valor.

En lenguaje estadístico, la hipótesis propuesta o hipótesis sometida al análisis se conoce como hipótesis nula y se designa como H_0 ; la cual se compara con una hipótesis alternativa, designada H_1 .

La hipótesis alternativa puede ser simple o compuesta, por ejemplo: $H_1: \theta = 3$ es simple, mientras que $H_1: \theta \neq 3$ es compuesta.

La teoría de las pruebas de hipótesis se ocupa de proporcionar reglas que permitan decir si una hipótesis nula se acepta o rechaza.

En el presente trabajo, se muestra el enfoque más usado que es el de la prueba de significancia.

En general, una prueba de significancia es un procedimiento mediante el cual, se usan los resultados de la muestra para probar la veracidad de una hipótesis nula. La idea es, hacer tales pruebas usando un estadístico de prueba y es la distribución de tal estadístico bajo la hipótesis nula.

A fin de ilustración, supóngase la siguiente prueba de hipótesis.

$$H_0: \beta = \beta_0$$

$$H_1: \beta \neq \beta_0$$

donde H_0 es una hipótesis simple, mientras que H_1 es un hipótesis compuesta.

El problema es decidir si b , el estimador de β , es lo suficientemente cercano a β_0 para aceptar H_0 . Bajo la

hipótesis nula la ecuación (2.48) se escribe:

$$P \left[-t_{\alpha/2} \leq \frac{b - \beta_0}{S} \sqrt{\Sigma x^2} \leq t_{\alpha/2} \right] = 1 - \alpha$$

o bien:

$$P \left[\beta_0 - \frac{t_{\alpha/2} S}{\sqrt{\Sigma x^2}} \leq b \leq \beta_0 + \frac{t_{\alpha/2} S}{\sqrt{\Sigma x^2}} \right] = 1 - \alpha \quad (2.54)$$

Así la expresión (2.54) proporciona un intervalo cuyos extremos son los valores críticos para b .

En este caso se usa como estadístico de prueba el valor calculado para b . Si este valor está contenido en el intervalo, H_0 se acepta, en caso contrario se rechaza.

De una manera análoga se pueden construir los marcos de referencia para probar hipótesis para α y σ^2 los cuales son:

a) Para α , probar:

$$H_0: \alpha = \alpha_0$$

$$H_1: \alpha \neq \alpha_0$$

Marco de referencia:

$$P \left[\alpha_0 - t_{\alpha/2} S \sqrt{\frac{1}{n} + \left[\bar{X}^2 / \Sigma x^2 \right]} \leq \alpha \leq \alpha_0 + t_{\alpha/2} S \sqrt{\frac{1}{n} + \left[\bar{X}^2 / \Sigma x^2 \right]} \right]$$

Estadístico de prueba: a .

Criterio de prueba:

Aceptar H_0 si el valor de a cae dentro del intervalo de confianza. Rechazar en caso contrario.

b) Para σ^2 , probar:

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 \neq \sigma_0^2$$

Marco de referencia:

$$P \left[\frac{\sigma_0^2 \chi_{1-\alpha/2}}{(n-2)} \leq S^2 \leq \frac{\sigma_0^2 \chi_{\alpha/2}}{(n-2)} \right] = 1 - \alpha$$

Estadístico de prueba: S^2 .

Criterio de prueba:

Aceptar H_0 si el valor de S^2 cae dentro del intervalo de confianza. Rechazar en caso contrario.

En la práctica, lo que se hace es tomar, por ejemplo para b:

$$P \left[-t_{\alpha/2} \leq t \leq t_{\alpha/2} \right] = 1 - \alpha$$

se establece un nivel de significancia, encontrándose con este los valores " de tablas " $-t_{\alpha/2}$ y $t_{\alpha/2}$, los cuales definen el intervalo $(-t_{\alpha/2}, t_{\alpha/2})$.

A continuación se obtiene un t "calculado" bajo H_0 siendo éste:

$$\frac{b - \beta_0}{S} \sqrt{\sum x^2}$$

Si t calculado cae en la "región de aceptación" H_0 se acepta, caso contrario se rechaza. El mismo procedimiento se usa para probar hipótesis para α y σ^2 .

Análisis de Varianza.

En el análisis de regresión es importante probar la hipótesis $H_0: \beta = 0$ contra $H_1: \beta \neq 0$, dado que es la pendiente de la recta la que permite ver si existe a no una relación entre X e Y.

La prueba de significancia para $H_0: \beta = 0$, derivada en la sección anterior puede llevarse a cabo usando un análisis de varianza, el cual se describe a continuación:

Dado que:

$$\frac{b - \beta}{\sigma / \sqrt{\sum x^2}} \sim N(0,1)$$

Por definición de la variable χ^2

$$\frac{(b - \beta)^2}{\sigma^2 / \sum x^2} \sim \chi^2_{(1)}$$

y como:

$$\frac{\sum e^2}{\sigma^2} \sim \chi^2_{(n-2)}$$

independiente de b, se construye una distribución F dada por:

$$F = \frac{(b-\beta)^2 \sum x^2}{\sum e^2 / (n-2)} \sim F_{(1, n-2)}$$

la cual bajo H_0 resulta:

$$F = \frac{b^2 \sum x^2 / 1}{\sum e^2 / (n-2)} \sim F_{(1, n-2)}$$

o bien, usando la descomposición de suma de cuadrados

$$F = \frac{SEC/1}{SCR/(n-2)}$$

De acuerdo a esta aproximación los datos se dan en un análisis de varianza según se muestra en el cuadro (2.1):

f fuente de variación	suma de cuadrados	g.l.	suma del cuadrado* de los promedios
X	SCE = $b^2 \sum x^2$	1	SCE/1
residuos	SCR = $\sum e^2$	(n-2)	SCR/(n-2)
total	SCT = $\sum y^2$	(n-1)	

* Se obtiene dividiendo SC por g.l.

Cuadro 2.1. Análisis de varianza.

El criterio para probar:

$$H_0: \beta = 0$$

rechace $H_0: \beta = 0$ al $\alpha(100)$ por ciento si:

$$F = \frac{SEC/1}{SCR/(n-2)} > F_{(1-\alpha)(1, n-2)}$$

Predicción en el Modelo de Mínimos Cuadrados.

La ecuación de regresión $Y = a + bx$ puede usarse para predecir los valores de Y para algún valor dado de X . Se pueden hacer dos tipos de predicción:

1. Predicción puntual.
2. Predicción por intervalos.

En la práctica, la estimación puntual se usa indicando la precisión de la predicción. El análisis de predicción puntual se desarrolla de la manera siguiente:

Siendo X_F un valor futuro X , el correspondiente valor de Y estimado es:

$$\hat{Y}_F = a + b X_F \quad (2.55)$$

el valor real de Y_F viene dado por:

$$Y_F = \alpha + \beta X_F + U_F$$

por lo que el error de predicción puede definirse como:

$$e_F = (Y_F - \hat{Y}_F) = U_F - (a - \alpha) - (b - \beta)X_F \quad (2.56)$$

tomando esperanzas, se tiene que:

$$\begin{aligned} E(e_F) &= E\{U_F - (a - \alpha) - (b - \beta)X_F\} \\ &= 0 \end{aligned}$$

o bien:

$$E(Y_F - \hat{Y}_F) = 0$$

con lo que $E(\hat{Y}_F) = Y_F$, es decir el predictor \hat{Y}_F es insesgado. La varianza del error de predicción se encuentra como sigue:

dado que

$$\begin{aligned}
 \text{Var}(e_F) &= E(e^2) - E(e)^2 \\
 &= E \{ U_F^2 + (a-\alpha)^2 + (b-\beta)^2 X_F^2 - 2U_F(a-\alpha) - 2U_F(b-\beta)X_F \\
 &\quad + 2(a-\alpha)(b-\beta)X_F \} \\
 &= E(U_F^2) + E(a-\alpha)^2 + X_F^2 E(b-\beta)^2 + 2X_F E(a-\alpha)(b-\beta) \\
 &= \sigma^2 + \text{Var}(a) + X_F^2 \text{Var}(b) + 2X_F \text{Cov}(a,b) \\
 &= \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma X^2} \right] + X_F^2 \left[\frac{\sigma^2}{\Sigma X^2} \right] + 2X_F \left[\frac{-\bar{X} \sigma^2}{\Sigma X^2} \right]
 \end{aligned}$$

o bien en forma abreviada:

$$\text{Var}(e_F) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\Sigma X^2} \right] \quad (2.57)$$

de la última expresión, se puede ver que la varianza aumenta a medida que X_F se aleja de la media. De hecho cuando $X_F = \bar{X}$, $\text{Var}(e_F)$ toma su valor mínimo y se incrementa en forma no lineal.

Por otra parte, dado que e_F es una función de variables distribuidas normalmente, entonces:

$$e_F \sim N \left[0, \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\Sigma X^2} \right] \right]$$

o bien, estandarizando:

$$\frac{e_F}{\sigma \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\Sigma X^2} \right]^{1/2}} \sim N(0,1)$$

más aún, remplazando σ por su estimador:

$$\frac{Y_F - \hat{Y}_F}{S \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\Sigma x^2} \right]^{1/2}} \sim t_{(n-2)} \quad (2.58)$$

La expresión (2.58). permite construir intervalos de confianza al $(1-\alpha)100$ por ciento para Y_F según se muestra a continuación .

Se desea un intervalo de confianza tal que:

$$P \left[-t_{\alpha/2} \leq t \leq t_{\alpha/2} \right] = 1 - \alpha$$

$$P \left[-t_{\alpha/2} \leq \frac{Y_F - \hat{Y}_F}{S \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\Sigma x^2} \right]^{1/2}} \leq t_{\alpha/2} \right] = 1 - \alpha$$

o en forma equivalente:

$$P \left[(a+bX_F) - t_{\alpha/2} S \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\Sigma x^2} \right]^{1/2} \leq Y_F \leq (a+bX_F) + t_{\alpha/2} S \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\Sigma x^2} \right]^{1/2} \right] = 1 - \alpha$$

es decir, el intervalo de confianza al $(1-\alpha)100$ por ciento está definido por:

$$\left[(a+bX_F) - t_{\alpha/2} S \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\Sigma x^2} \right]^{1/2}, \right. \\ \left. (a+bX_F) + t_{\alpha/2} S \left[1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\Sigma x^2} \right]^{1/2} \right] \quad (2.59)$$

Especificaciones Lineales

A veces, la teoría económica o los hechos empíricos indican que para describir la relación entre dos variables debe utilizarse una forma no lineal. En tales casos, a menudo es posible hallar una transformación, tal que una vez aplicada se obtenga una especificación lineal por lo que pueden aplicarse entonces las técnicas de regresión lineal.

El modelo log-log.

Existen relaciones que pueden representarse con una función de la forma

$$Z = \alpha_0 W^{\beta_0} \quad (2.60)$$

donde α_0 y β_0 son constantes. Si X es igual a uno, Y es igual a α_0 . De la ecuación (2.60):

$$\frac{dZ}{dW} = \beta_0 \alpha_0 W^{\beta_0-1}$$

así, si β_0 es positiva, la pendiente es siempre positiva y Z tiende al infinito a medida que W tiende al infinito.

Si β_0 es mayor que uno, la pendiente crece a medida

que W crece. Si $0 < \beta_0 < 1$ la pendiente decrece continuamente, aunque siempre permanezca positiva. Cuando β_0 es negativa, la pendiente es siempre negativa.

Estas consideraciones dan las formas posibles representadas en la figura 2.10.

No todos los puntos de un diagrama de dispersión producidos por una relación de la forma (2.60) caen exactamente sobre la curva. Una descripción más exacta de los datos requiere de un término de perturbación.

Si se postula un término de perturbación multiplicativo (en lugar de aditivo) no negativo V_i , la relación toma la forma:

$$Z_i = \alpha_0 W_i^{\beta_0} V_i \quad (2.61)$$

El logaritmo de la función es:

$$\log Z_i = \log \alpha_0 + \beta_0 \log W_i + \log V_i \quad (2.62)$$

poniendo

$Y_i = \log Z_i$; $\alpha = \log \alpha_0$, $\beta = \beta_0$, $X_i = \log W_i$, y $U_i = \log V_i$

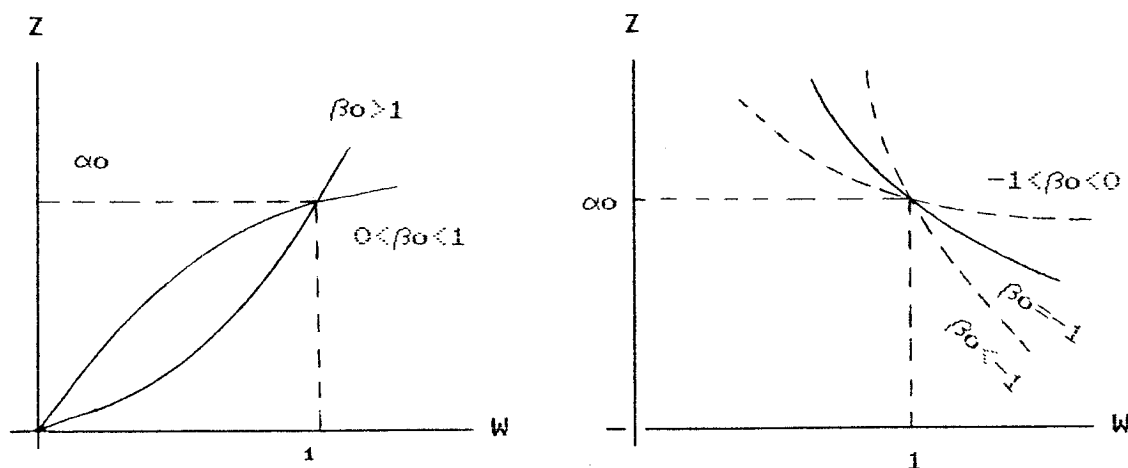


Figura 2.10. Relaciones no lineales de la forma $Z = \alpha_0 W^{\beta_0}$

la función (2.62) puede escribirse de la forma:

$$Y_i = \alpha + \beta X_i + U_i$$

la cual es la forma estándar para la aplicación de los mínimos cuadrados en la regresión lineal simple.

La relación (2.61) llamada modelo log-log tiene propiedades importantes para la teoría económica. Es una función de elasticidad constante e igual a β_0 . En efecto, por definición, la elasticidad es:

$$\eta = \frac{dZ}{dW} \frac{W}{Z} \quad (2.63)$$

así:

$$\eta = \alpha_0 \beta_0 W^{\beta_0-1} \frac{W}{\alpha_0 W^{\beta_0}} = \beta_0$$

En particular, cuando $\beta_0 = -1$ la ecuación:

$$ZW = \alpha_0 \quad (2.64)$$

es una hipérbola equilátera. Si Z representa la cantidad demandada y W el precio unitario de algún bien; la ecuación (2.64) representa una curva de demanda con elasticidad constante igual a -1 y un gasto total constante e igual a α_0 sin importar el precio del bien.

Por otra parte, cuando $\beta_0 = 1$ se tiene

$$Z = \alpha_0 W \quad (2.65)$$

si Z representa la cantidad ofrecida y W el precio unitario de un bien, la ecuación (2.65) representa una curva de oferta de elasticidad unitaria para todo precio.

El modelo semi-log.

Cuando se producen diagramas de dispersión como los mostrados en las figuras 2.11.a y 2.11.b:

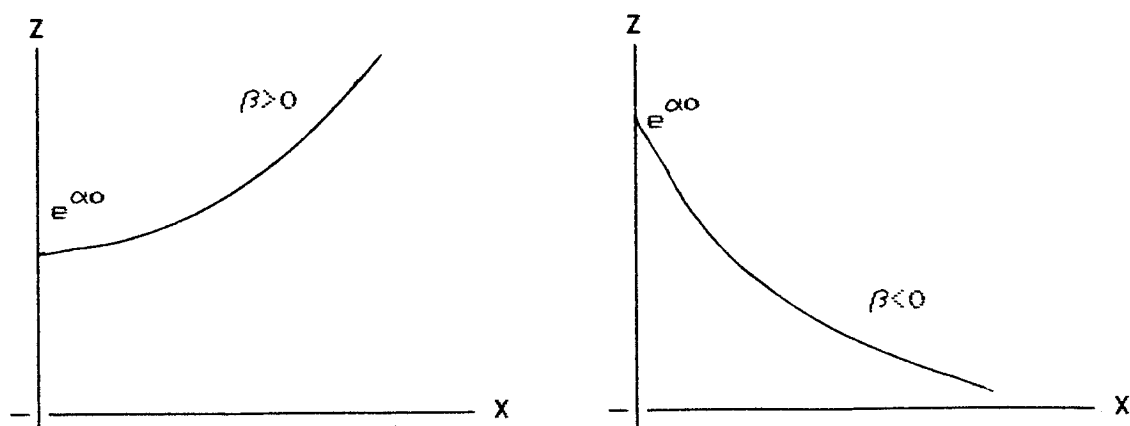


Figura 2.11. El modelo semilog.

Puede usarse como función de ajuste a

$$Z = e^{(\alpha_0 + \beta_0 X)} \quad (2.66)$$

la cual es una especificación lineal, pues tomando logaritmos naturales se obtiene

$$\ln Z = \alpha_0 + \beta_0 X$$

poniendo $Y = \ln Z$:

$$Y = \alpha_0 + \beta_0 X \quad (2.67)$$

Es claro, que la perturbación aleatoria ha de suponerse nuevamente multiplicativa y no negativa.

Una característica de la relación (2.66) es que la razón proporcional de cambio en Y por cambio unitario en X es constante e igual a β_0 . Esto es:

$$\frac{1}{Y} \frac{dY}{dX} = \frac{1}{Y} e^{(\alpha_0 + \beta_0 X)} \beta_0$$

Otra característica de la función es que su rango está compuesto solo por valores positivos. Además, el intercepto está dado por e^{α_0} y la pendiente es positiva o negativa, dependiendo del signo de β_0 .

Un caso especial en la ecuación (2.66) ocurre cuando

X denota al tiempo y la función describe una variable Y la cual muestra una razón proporcional de cambio constante en su crecimiento ($\beta > 0$) o decrecimiento ($\beta < 0$).

El modelo recíproco.

Otra especificación lineal de gran utilidad en la teoría económica está definida por el modelo recíproco dado por:

$$Y = \alpha + \beta \left[\frac{1}{X} \right] + U \quad (2.68)$$

La pendiente es:

$$\frac{dY}{dX} = -\beta/X^2$$

Así, si β es positivo, la curva es siempre decreciente. Cuando X tiende a infinito Y tiende a α .

Además, como:

$$\frac{d^2Y}{dX^2} = \beta/X^3$$

para valores positivos de X, la función es cóncava hacia arriba.

Ahora bien, si β es negativo y X toma solo valores positivos la función es creciente y cóncava hacia abajo, interseca al eje X en $-\beta/\alpha$ y tiene como asíntota a $y=\alpha$. La forma de ésta curva se muestra en la figura 2.12.

Si se piensa que Y representa la tasa de aumento de los sueldos monetarios y X la tasa de desempleo, la figura 12.a indica la forma típica de la curva de Phillips.

La convexidad de la curva se justifica por las siguientes razones: cuando el desempleo disminuye en cantidades constantes, los salarios aumentarán a una razón

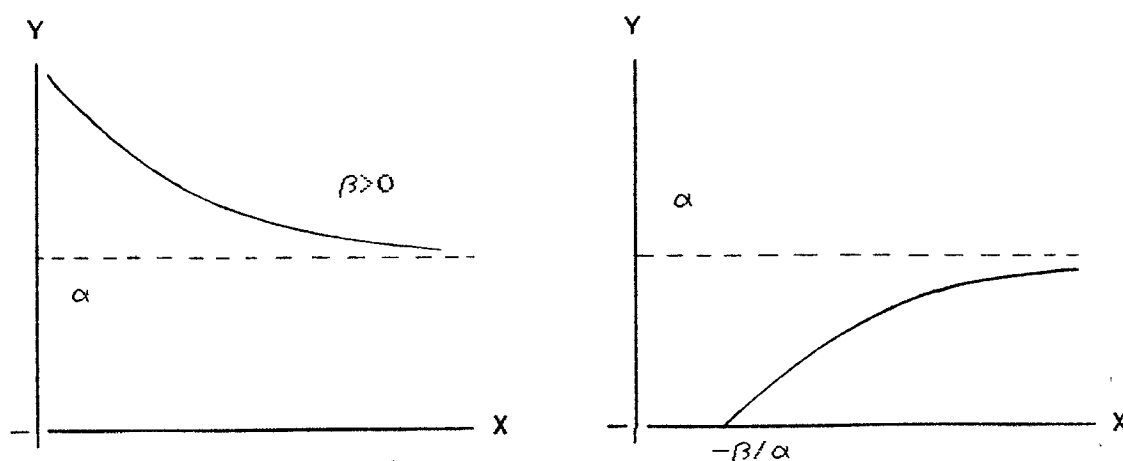


Figura 2.12. El modelo recíproco.

creciente tendiendo a infinito, a medida que el desempleo tiende a cero.

Por otra parte, debe haber cierto límite institucional inferior por debajo del cual Y no pueda descender.

La figura 2.12.b se usa con frecuencia para representar funciones de gasto, donde Y representa el gasto en algún bien o servicio específico y X representa el ingreso total.

Esta aplicación en particular tiene solo sentido cuando α es positivo y β negativo. α indica el nivel asintótico de gasto, y $-\beta/\alpha$ es el monto en el ingreso que hace que no exista gasto en el bien o servicio.

El Modelo Logarítmico Recíproco.

Existen algunas relaciones económicas, tales como el uso de bienes de servicio (autos, televisión, etc.) con respecto al tiempo que pueden ser descritas mediante una curva logística definida por:

$$Y = \frac{k}{1 + be^{-at}} \quad (2.68)$$

donde k , a y b son parámetros a determinarse, y cuya forma general se muestra en la figura 2.13.

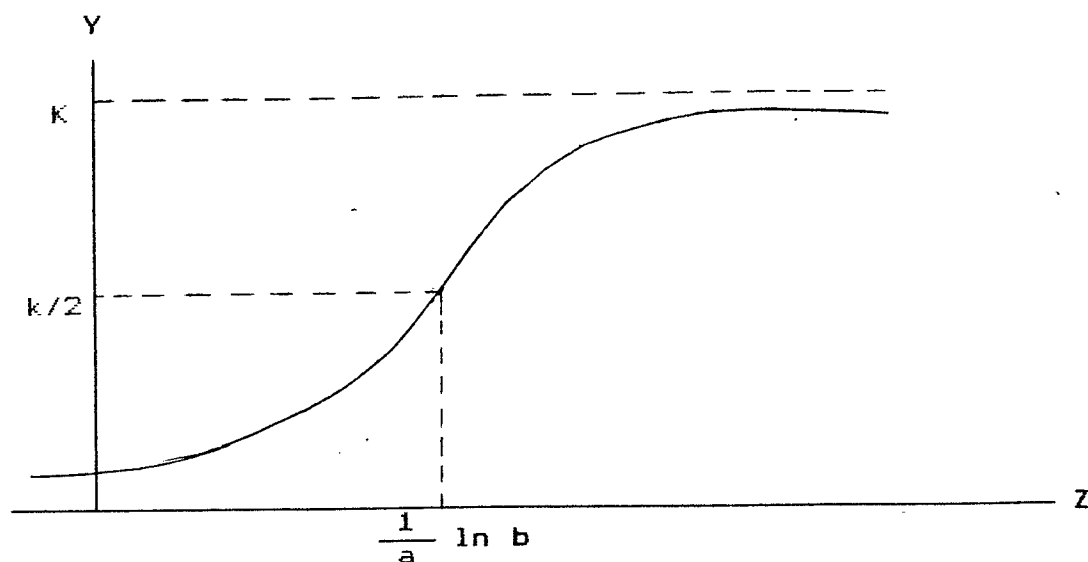


Figura 2.13. La curva logística.

La curva logística no es una especificación lineal, por lo que su tratamiento cae fuera de los propósitos del presente trabajo. Sin embargo, existe una especificación lineal con características similares llamada modelo logaritmo recíproco, definida por:

$$Y = e^{\alpha - \beta/X} \quad (2.69)$$

cuya forma lineal es:

$$\ln Y = \alpha - \beta(1/X) \quad (2.70)$$

Cuando $X \rightarrow 0$ y $Y \rightarrow 0$. Además, cuando $X \rightarrow \infty$ y $Y \rightarrow e^{\alpha}$

Por otra parte, dado que:

$$\frac{dY}{dX} = \left[\frac{\beta}{X^2} \right] e^{\alpha - \beta(1/X)} \quad (2.71)$$

para $\beta > 0$ la curva es creciente y presenta un punto de inflexión cuando:

$$\frac{d^2Y}{dX^2} = \left[\frac{\beta^2}{X^4} - \frac{2\beta}{X^3} \right] e^{\alpha - \beta(1/X)} = 0$$

Esto ocurre para $X = \beta/2$ al nivel

$$Y = e^{\alpha - 2} = 0.135 e^{\alpha}$$

Gráficamente, el modelo logarítmico recíproco se ve según se muestra en la figura 2.14.

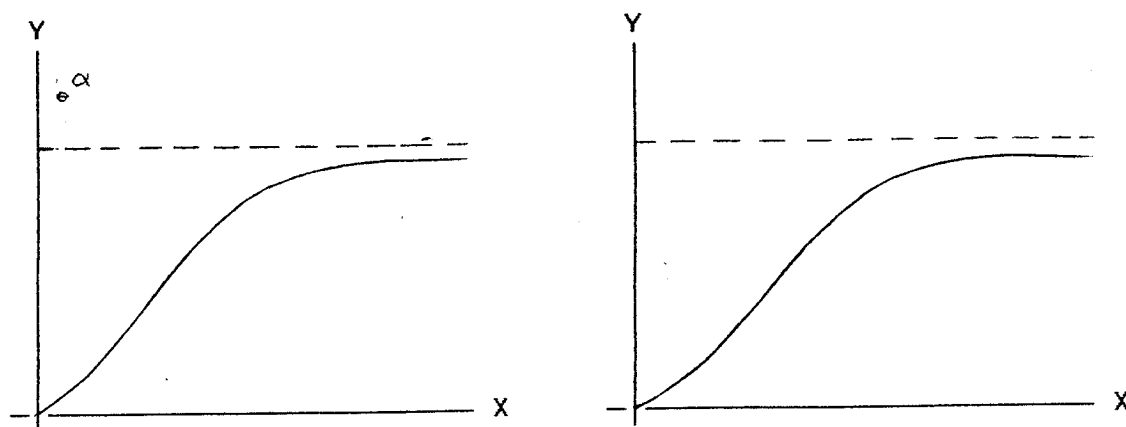


Figura 2.14. El modelo logarítmico-recíproco.

CAPITULO III

REGRESION MULTIPLE

En el capítulo anterior se presentaron los conceptos lógicos que forman la base del análisis de regresión. En este capítulo se muestra la forma en la que se hacen extensivas tales ideas fundamentales, a ecuaciones de regresión que abarcan más de dos variables.

Dada la dificultad en el manejo algebraico de las expresiones resultantes, se usará la notación matricial.

El Modelo de Regresión Múltiple.

En este modelo, se supone que en un sistema económico existe una relación lineal tanto en los parámetros como en las variables; para una de ellas dependiente Y y k independientes X_1, X_2, \dots, X_k es decir, existe una relación de la forma:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (3.1)$$

De igual manera que en el modelo de regresión simple, la ecuación 3.1 no explica las variaciones que se puedan dar. Por esto es que en términos mas generales, se supone que los datos a recolectar y analizar son generados por una ecuación estocástica tal como:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + U \quad (3.2)$$

El problema ahora consiste en: dado un modelo de regresión y un conjunto de observaciones:

$$\left\{ X_{1i}, X_{2i}, \dots, X_{ki}, Y_i \right\}_{i=1}^n$$

estimar los parámetros de regresión β_j para $j = 1, \dots, k$.

Antes de construir el modelo de regresión, es conveniente precisar algunas ideas.

Los datos pueden presentarse como un conjunto de ecuaciones lineales

$$Y = \beta_1 X_{11} + \beta_2 X_{21} + \dots + \beta_k X_{k1} + U_1$$

$$Y = \beta_1 X_{12} + \beta_2 X_{22} + \dots + \beta_k X_{k2} + U_2$$

⋮

$$Y = \beta_1 X_{1n} + \beta_2 X_{2n} + \dots + \beta_k X_{kn} + U_n$$

o simplemente

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i \quad i = 1, \dots, n \quad (3.3)$$

donde los coeficientes β_j , $j=1, \dots, k$ son los parámetros a estimar. Utilizando notación matricial, las n ecuaciones (3.3) se escriben como

$$Y = X\beta + U$$

donde

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \quad U = \begin{bmatrix} U_1 \\ \vdots \\ U_n \end{bmatrix}$$

$$X = \begin{bmatrix} X_{11} & X_{21} & X_{31} & \dots & X_{k1} \\ X_{12} & X_{22} & X_{32} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ X_{1n} & X_{2n} & X_{3n} & \dots & X_{kn} \end{bmatrix}$$

La generalización de los supuestos del modelo de regresión se presentan a continuación.

1. Los valores esperados de los términos de perturbación son nulos. Es decir $E(U_i) = 0$ para toda i .

En notación matricial:

$$E(U) = \begin{bmatrix} E(U_1) \\ \vdots \\ E(U_n) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

2. Se supone que los elementos de perturbación no están mutuamente correlacionados y que poseen varianza constante. Es decir:

$$E(UU') = \begin{bmatrix} E(U_1U_1)E(U_1U_2)\dots\dots E(U_1U_n) \\ E(U_2U_1)E(U_2U_2)\dots\dots E(U_2U_n) \\ \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ E(U_nU_1)E(U_nU_2)\dots\dots E(U_nU_n) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 I_n$$

3. Los valores $X_{1i}, X_{2i}, \dots, X_{ki}$ $i=1, \dots, n$ no son aleatorios. En otras palabras la matriz X no es estocástica.

4. Los términos de perturbación se distribuyen normalmente con media cero y varianza σ^2 . En términos matriciales, el vector U tiene una distribución normal multivariada. Es decir:

$$U \sim N(0, \sigma^2 I)$$

5. Las variables explicatorias no forman un conjunto linealmente dependiente. Es decir no existen relaciones de la forma $X_r = C_1 + C_2 X_m$. En términos de matrices se pide que X sea una matriz de rango $k < n$.

Estimación Puntual

Cuando se satisfacen las cinco condiciones

precedentes, el criterio de mínimos cuadrados ofrece los estimadores lineales insesgados de mínima varianza. El análisis de la estimación puntual es idéntico al del Capítulo II, salvo que ahora se trabaja con k variables independientes y notación matricial.

Se indica con B el vector de estimadores. Es decir:

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

En la deducción del estimador por mínimos cuadrados, se utilizan los residuos de la función de regresión ajustada, los cuales son iguales a:

$$\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} - \begin{bmatrix} X_{11} & X_{21} & X_{31} & \dots & X_{k1} \\ X_{12} & X_{22} & X_{32} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ X_{1n} & X_{2n} & X_{3n} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

que en forma matricial se escriben como:

$$e = Y - XB$$

Entonces, la suma del cuadrado de los residuos está dada por:

$$e'e = (Y - XB)'(Y - XB) = S(b)$$

El criterio de mínimos cuadrados exige pues, un vector B que minimice la expresión $S(b)$.

La función a minimizar puede escribirse como:

$$S(b) = Y'[I - X(X'X)^{-1}X']Y + [B - (X'X)^{-1}X'Y]'X'X[B - (X'X)^{-1}X'Y] \quad (3.4)$$

El primer término no contiene a B ; luego, analizando el segundo, se tiene que éste es una forma cuadrática en la

matriz $(X'X)$, definida positiva.¹ Por tanto, está acotada inferiormente por cero. Esto ocurre cuando se toma a B como

$$B = (X'X)^{-1} X'Y \quad (3.5)$$

Siendo éste el estimador de β por mínimos cuadrados.

Si se desea estimar los parámetros de regresión simple, se toma $k = 2$ con $X_{1i} = 1$ para toda i .

Consecuentemente, la ecuación normal (3.5) toma la siguiente forma:

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \end{bmatrix} \begin{bmatrix} 1 & X_{21} \\ 1 & X_{22} \\ \vdots & \vdots \\ 1 & X_{2n} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

o bien

$$\begin{bmatrix} n & \sum X_{2i} \\ \sum X_{2i} & \sum X_{2i}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{2i} \end{bmatrix}$$

y

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} n & \sum X_{2i} \\ \sum X_{2i} & \sum X_{2i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{2i} \end{bmatrix}$$

finalmente:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \frac{1}{n \sum X_{2i}^2 - (\sum X_{2i})^2} \begin{bmatrix} \sum Y_i \sum X_{2i}^2 - \sum X_{2i} \sum Y_i X_{2i} \\ n \sum Y_i X_{2i} - \sum Y_i \sum X_{2i} \end{bmatrix} \quad (3.6)$$

La ecuación (3.6) es idéntica a las ecuaciones (2.16) y (2.17), salvo en la notación. Las propiedades de B como

¹ Una matriz A de orden m se llama semidefinida positiva si y solo si para todo vector X

$$X'AX \geq 0$$

y se llama definida positiva si para todo $X \neq 0$

$$X'AX > 0$$

estimador de β se obtienen como sigue:

De la ecuación (3.5) se tiene:

$$\begin{aligned} B &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + U) \\ &= \beta + (X'X)^{-1}X'U \end{aligned} \quad (3.7)$$

tomando esperanzas:

$$\begin{aligned} E(B) &= E(\beta) + (X'X)^{-1}X'E(U) \\ E(B) &= \beta \end{aligned} \quad (3.8)$$

Luego, B resulta ser un estimador insesgado de β .

Por otra parte:

$$\begin{aligned} \text{Var}(B) &= E \{ (B - \beta) (B - \beta)' \} \\ &= E \{ (B - \beta) (B' - \beta') \} \\ &= E \{ (X'X)^{-1}X'UU'X(X'X)^{-1} \} \end{aligned}$$

donde se ha usado el hecho de que $(X'X)$ es una matriz simétrica y como:

$$((X)^{-1})' = (X')^{-1}$$

entonces

$$((X'X)^{-1})' = (X'X)^{-1}.$$

Así

$$\begin{aligned} \text{Var}(B)' &= (X'X)^{-1}X' E(UU')X(X'X)^{-1} \\ \text{Var}(B) &= \sigma^2(X'X)^{-1} \end{aligned} \quad (3.9)$$

La expresión (3.9) debe interpretarse como una matriz de varianzas y covarianzas. Esto es, por ejemplo, en el caso de regresión simple:

$$\text{Var}(B) = \begin{bmatrix} \text{Var}(b_1) & \text{Cov}(b_1b_2) \\ \text{Cov}(b_1b_2) & \text{Var}(b_2) \end{bmatrix}$$

o bien, según la notación para los estimadores usada en el Capítulo II:

$$\text{Var}(B) = \begin{bmatrix} \text{Var}(a) & \text{Cov}(a,b) \\ \text{Cov}(a,b) & \text{Var}(b) \end{bmatrix}$$

Para el mismo caso, la matriz de varianzas y covarianzas es:

$$\begin{aligned} \text{Var}(B) &= \sigma^2 (X'X)^{-1} \\ &= \frac{\sigma^2}{n \Sigma X^2 - (\Sigma X)^2} \begin{bmatrix} \Sigma X^2 & -\Sigma X \\ -\Sigma X & n \end{bmatrix} \\ &= \frac{\sigma^2}{n \Sigma X^2} \begin{bmatrix} \Sigma X^2 & -\Sigma X \\ -\Sigma X & n \end{bmatrix} \end{aligned}$$

donde según se ve:

$$\begin{aligned} \text{Var}(b) &= \frac{\sigma^2}{\Sigma x^2} \\ \text{Var}(a) &= \frac{\sigma^2 \Sigma X^2}{n \Sigma x^2} = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma x^2} \right] \end{aligned}$$

Así mismo

$$\text{Cov}(a,b) = = \frac{-\sigma^2}{n \Sigma x^2} = - \frac{\sigma^2 \bar{X}}{\Sigma x^2}$$

que son las expresiones obtenidas para las varianzas y covarianzas en el capítulo anterior.

Al igual que para el caso de dos variables, los estimadores de los coeficientes β , son estimadores lineales insesgados y poseen una varianza más pequeña que cualquier otro estimador lineal insesgado. Para mostrar esto, nótese que según la ecuación (3.5), B es un estimador lineal (por ser función lineal de Y). Se toma un estimador lineal insesgado B_1 y se muestra que $\text{Var}(B) \leq \text{Var}(B_1)$. En el sentido de que la matriz $\text{Var}(B_1) - \text{Var}(B)$ es una matriz

semidefinida positiva. En particular, que todos los elementos de su diagonal principal son no negativos.

Sea $B_1 = [(X'X)^{-1}X' + C]Y$ cualquier estimador lineal insesgado de β . Así:

$$\begin{aligned} B_1 &= (X'X)^{-1}X'Y + CY \\ &= B + CY \\ &= \beta + (X'X)^{-1}X'U + C(X\beta + U) \\ &= \beta + CX\beta + [(X'X)^{-1}X' + C]U \end{aligned}$$

siendo que B es insesgado

$$\begin{aligned} E(B_1) &= E(\beta) + E(CX\beta) + [(X'X)^{-1}X' + C] E(U) \\ &= \beta \end{aligned}$$

por lo que se requiere que $CX = 0$.

De esta forma:

$$\begin{aligned} \text{Var}(B_1) &= E \{ (B_1 - \beta) (B_1 - \beta)' \} \\ &= E \{ [(X'X)^{-1}X' + C]UU' [(X'X)^{-1}X' + C]' \} \\ &= [(X'X)^{-1}X' + C] E(UU') [(X'X)^{-1}X' + C]' \\ &= \sigma^2 [(X'X)^{-1} + CC'] \end{aligned}$$

donde se ha usado la condición $CX = 0$, con lo que:

$$\begin{aligned} \text{Var}(B_1) &= \sigma^2 (X'X)^{-1} + \sigma^2 CC' \\ &= \text{Var}(B) + \sigma^2 CC' \end{aligned}$$

por tanto $\text{Var}(B) \leq \text{Var}(B_1)$.

Para concluir el estudio de la estimación puntual es necesario obtener un estimador de σ^2 , al cual se le llamará S^2 . Como en el caso de la regresión simple, éste se basa en la suma del cuadrado de los residuos. Dado que

$$Y = X\beta + U \text{ y } \hat{Y} = XB = X[(X'X)^{-1}X'Y]$$

los residuos pueden escribirse como:

$$e = Y - \hat{Y} = X\beta + U - X[(X'X)^{-1}X'(X\beta + U)]$$

$$\begin{aligned}
 &= U - X(X'X)^{-1}X'U \\
 &= (I - X(X'X)^{-1}X')U \\
 e &= MU
 \end{aligned}$$

donde $M = [I - X(X'X)^{-1}X']$.

La matriz M posee dos propiedades muy útiles:

1. Es una matriz simétrica, es decir:

$$M' = I - X(X'X)^{-1}X' = M$$

2. Es una matriz idempotente es decir:

$$\begin{aligned}
 M^2 &= [I - X(X'X)^{-1}X'] [I - X(X'X)^{-1}X'] \\
 &= I - 2X(X'X)^{-1}X' + X(X'X)^{-1}(X'X)(X'X)^{-1}X' \\
 &= I - X(X'X)^{-1}X' = M
 \end{aligned}$$

Con esto la suma de los cuadrados de los residuos es pues

$$e'e = U'M'MU = U'MU$$

dado que $M'M = M$.

Desarrollando detalladamente $e'e$ se obtiene

$$\begin{aligned}
 U'MU &= [U_1, U_2, \dots, U_n] \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & m_{2n} \\ \vdots & \vdots & \dots & \vdots \\ m_{n1} & m_{n2} & \dots & m_{nn} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \\
 &= \sum_i \sum_j m_{ij} u_i u_j
 \end{aligned}$$

$$\text{así, } E(e'e) = E(U'MU) = \sum_i \sum_j m_{ij} E(u_i u_j) = \sigma^2 \sum_i m_{ii}$$

dado que todos los términos con $i \neq j$ son cero debido a $E(u_i u_j) = 0$.

Queda solo por evaluar $\sum m_{ii}$ que representa la suma de los elementos de la diagonal principal de M . Esto es $\sum m_{ii}$

representa la traza de M . De esta forma, y de acuerdo a las propiedades de la traza¹:

$$\begin{aligned}\sum_i m_{ii} &= \text{tr } M \\ &= \text{tr } [D - X(X'X)^{-1}X'] \\ &= \text{tr } D - \text{tr } [X(X'X)^{-1}X'] \\ &= \text{tr } D - \text{tr } (X'X)(X'X)^{-1} \\ &= n - k\end{aligned}$$

Así puede obtenerse un estimador insesgado para σ^2 poniendo

$$S^2 = \frac{(e'e)}{n - k} = \frac{\sum e_i^2}{n - k} \quad (3.10)$$

dado que $E(S^2) = \frac{E(e'e)}{n - k} = \frac{\sigma^2(n-k)}{n - k} = \sigma^2$

Este resultado, tomando $k = 2$ con $X_1 = 1$, produce:

$$S^2 = \frac{\sum e_i^2}{n - 2}$$

que es la ecuación (2.21) de mínimos cuadrados.

Los coeficientes de correlación parcial y múltiple.

Una vez estimada la ecuación de regresión múltiple, se hace necesario determinar la exactitud del ajuste. Esto se hace en forma similar al modelo de regresión simple, solo que involucrando k variables regresoras. De acuerdo al Capítulo II:

$$R^2 = \frac{SEC}{STC}$$

donde R^2 : es el coeficiente de determinación múltiple.

¹ $\text{tr } (A + B) = \text{tr}(A) + \text{tr}(B)$. Si A y B son dos matrices cualesquiera, tales que AB y BA existen, $\text{tr}(AB) = \text{tr}(BA)$.

SEC: suma explicada de cuadrados.

STC: suma total de cuadrados.

Además

$$STC = SEC + SRC$$

donde SRC: suma de los residuales al cuadrado.

A fin de derivar una expresión para R^2 se nota que

$$\begin{aligned} SRC &= e'e = (Y - XB)'(Y - XB) \\ &= (Y' - B'X')(Y - XB) \\ &= Y'Y - B'X'Y - Y'XB + B'X'XB \\ &= Y'Y - 2B'X'Y + B'X'XB \\ &= Y'Y - 2B'X'Y + B'X'X(X'X)^{-1}X'Y \\ &= Y'Y - B'X'Y \end{aligned} \quad (3.11)$$

además,

$$STC = \sum y_i^2 = \sum (Y - \bar{Y})^2 = \sum Y^2 - n\bar{Y}^2$$

o bien, en notación matricial

$$STC = Y'Y - n\bar{Y}^2 \quad (3.12)$$

de esta forma,

$$\begin{aligned} SEC &= STC - SRC \\ &= Y'Y - n\bar{Y}^2 - (Y'Y - B'X'Y) \\ &= B'X'Y - n\bar{Y}^2 \end{aligned} \quad (3.13)$$

Estos resultados conducen a

$$R^2 = \frac{B'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} \quad (3.14)$$

Por supuesto, el coeficiente de correlación múltiple está dado por

$$R = \sqrt{R^2} \quad (3.15)$$

El coeficiente de correlación múltiple mide el ajuste

tomando en cuenta todas las variables, sin embargo, no proporciona información acerca del efecto individual o parcial de una variable independiente. Esto lo hace el coeficiente de correlación parcial. Las correlaciones parciales se definen como correlaciones entre dos variables cuando las demás están fijas. El símbolo $R_{Y1,23}$ se usa para la correlación muestral entre Y y X_1 cuando X_2 y X_3 son constantes. El desarrollo de las ecuaciones para obtener las correlaciones parciales se omite. Para calcular los coeficientes de correlación parcial se define primero R como la matriz simétrica de correlaciones simples entre el conjunto de variables Y, X_1, X_2, \dots, X_k ; ninguna de ellas señaladas como independiente. En general R está dada por

$$R = \begin{matrix} & Y & X_1 & X_2 & \dots & X_k \\ \begin{matrix} Y \\ X_1 \\ X_2 \\ \vdots \\ X_k \end{matrix} & \begin{bmatrix} 1 & r_{Y1} & r_{Y2} & \dots & r_{Yk} \\ r_{1Y} & 1 & r_{12} & \dots & r_{1k} \\ r_{2Y} & r_{21} & 1 & \dots & r_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{kY} & r_{k1} & r_{k2} & \dots & 1 \end{bmatrix} \end{matrix} \quad (3.16)$$

la inversa de R es también simétrica, y puede escribirse como

$$C = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1, k+1} \\ C_{21} & C_{22} & \dots & C_{2, k+1} \\ \vdots & \vdots & \ddots & \vdots \\ C_{k+1,1} & C_{k+1,2} & \dots & C_{k+1, k+1} \end{bmatrix} \quad (3.17)$$

Ahora para obtener la correlación parcial entre la variable i y la j se usa la relación:

$$r_{ij, \dots, i-1, j-1, j+1, \dots, k+1} = \frac{-c_{ij}}{\sqrt{c_{ii} c_{jj}}} \quad (3.18)$$

Intervalos de Confianza y Pruebas de Hipótesis

Hasta ahora, no se ha usado la suposición de la distribución del vector U . Pero si se desea construir intervalos de confianza y pruebas de hipótesis se hace necesaria.

Si

$$U \sim N(0, \sigma^2 I)$$

y X es no estocástica con rango k , y además

$$B = \beta + (X'X)^{-1}X'U$$

entonces

$$B \sim N(\beta, \sigma^2 (X'X)^{-1})$$

Por otra parte, se puede demostrar que B se distribuye independientemente de S^2 , con lo que:

$$b_i \sim N(\beta_i, \sigma^2 a_{ii})$$

donde a_{ii} denota el i -ésimo elemento sobre la diagonal principal de la matriz $(X'X)^{-1}$. De esa forma

$$\frac{b_i - \beta_i}{\sigma \sqrt{a_{ii}}} \sim N(0, 1)$$

Ahora bien

$$\frac{(n - k) S^2}{\sigma^2} \sim \chi^2_{(n-k)}$$

independiente de b_i . Con esto:

$$t = \frac{b_i - \beta_i}{\sigma \sqrt{a_{ii}}} \cdot \frac{\sigma \sqrt{(n-k)}}{S \sqrt{(n-k)}}$$

o bien:

$$t = \frac{b_i - \beta_i}{S \sqrt{a_{ii}}} \sim t_{(n-k)}$$

Este último resultado puede usarse para probar hipótesis y construir intervalos de confianza para β_i , usando los conceptos desarrollados en el capítulo anterior.

Cuando se desea efectuar la prueba global para determinar si la relación de regresión es o no significativa, la hipótesis que se prueba es:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

El estadístico de prueba apropiado se construye de la siguiente manera:

es posible probar que $B'X'Y/\sigma^2$ se distribuye como una χ^2 con k grados de libertad y es independiente de $e'e/\sigma^2$ que se distribuye como χ^2 con $n-k$ grados de libertad. Por tanto la razón:

$$F = \frac{B'X'Y/\sigma^2 k}{e'e^2 / \sigma(n-k)} = \frac{(n-k)B'X'Y}{k(e'e)}$$

tiene una distribución F con k y $(n-k)$ grados de libertad.

De esta forma, la tabla para el análisis de varianza es:

Cuadro 3.1 Análisis de varianza matricial.

Fuente de variación	Suma de cuadrados	Grados de libertad	Suma del promedio de los cuadrados
debido a las X 's	$B'X'Y$	k	$B'X'Y/k$
debido a residuos	$e'e$	$(n - k)$	$e'e/(n - k)$
total	$e'e + B'e'Y$	n	

La regla de decisión dice:

rechace

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

al $\alpha(100)$ por ciento si:

$$F = \frac{(n - k)B'X'Y}{k(e'e)} > F_{(1-\alpha)(k, n-k)}$$

Especificaciones Lineales de Varias Variables

La limitación de la linealidad aplicada a los parámetros da lugar a manejar modelos en apariencia más complicados, como por ejemplo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + U$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + U$$

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + U$$

los cuales son fáciles de redefinir a la forma estándar del modelo.

Así, si no es posible describir adecuadamente una tendencia por medio de una recta, el investigador puede decidir el ajuste de los datos usando una parábola. En tal caso, la ecuación polinómica de segundo grado

$$\hat{Y} = b_1 + b_2 X + b_3 X^2 \quad (3.19)$$

es la de uso más frecuente.

La estimación de los parámetros se efectúa como

sigue:

de acuerdo al modelo lineal múltiple

$$\hat{Y} = b_1X_1 + b_2X_2 + b_3X_3 \quad (3.20)$$

póngase $X_1 = 1$, $X_2 = X$, y $X_3 = X^2$ para obtener:

$$\hat{Y} = b_1 + b_2X + b_3X^2$$

si se dispone de n observaciones, entonces:

$$X = \begin{bmatrix} 1 & X_{21} & X_{31} \\ 1 & X_{22} & X_{32} \\ \vdots & & \\ \vdots & & \\ 1 & X_{2n} & X_{3n} \end{bmatrix}_{n \times 3}; \quad X' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ X_{31} & X_{32} & \dots & X_{3n} \end{bmatrix}_{3 \times n}; \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}$$

Es decir para el caso cuadrático:

$$X = \begin{bmatrix} 1 & X_1 & X_1^2 \\ 1 & X_2 & X_2^2 \\ \vdots & \vdots & \vdots \\ 1 & X_n & X_n^2 \end{bmatrix}_{n \times 3}; \quad X' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \\ X_1^2 & X_2^2 & \dots & X_n^2 \end{bmatrix}_{3 \times n}; \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}$$

matrices que permiten estimar los parámetros b_1 , b_2 , y b_3 .

CAPITULO IV

Violación De Los Supuestos Del Modelo Clásico.

Los desarrollos hechos en los capítulos precedentes están basados en varios supuestos simplificadores. Estos son:

1. El valor promedio de la perturbación estocástica U_i es nulo.
2. La varianza de $U_i = \sigma^2$ para toda i .
3. Las perturbaciones estocásticas no están autocorrelacionadas .
4. Las variables regresoras son no estocásticas.
5. Las variables regresoras son independientes.
6. Las U_i 's están normalmente distribuidas.

Matemáticamente el modelo es:

$$1. Y = \beta X + U \quad (4.1)$$

$$2. U \sim N(0, \sigma^2 I) \quad (4.2)$$

$$3. \text{Cov}(U_i, U_j) = E(U_i U_j) = 0, \text{ para } i \neq j \quad (4.3)$$

$$4. X \text{ es no estocástica de rango completo} \quad (4.4)$$

Con tales suposiciones, se mostró que los estimadores de los coeficientes de regresión por mínimos cuadrados son los mejores estimadores lineales insesgados y están normalmente distribuidos.

En este capítulo se analizan con mayor detenimiento estos supuestos, y se ve que ocurre si uno o mas de ellos no se cumplen.

Normalidad y Media Cero

Para comprobar ésta suposición, puede usarse una gráfica de los residuos estandarizados, ésto es e_i/S donde:

$$S^2 = \frac{\sum(e_i - \bar{e})^2}{n - k} = \frac{\sum e_i^2}{n - k} \quad (4.5)$$

Ya que si:

$$e_i/\sigma \sim N(0,1)$$

aproximadamente el 95 por ciento del total de residuos e_i/S debe de caer en una banda con límites en $(-2, +2)$ si n es "grande" .

La gráfica debe verse aproximadamente como la mostrada en la figura 4.1.

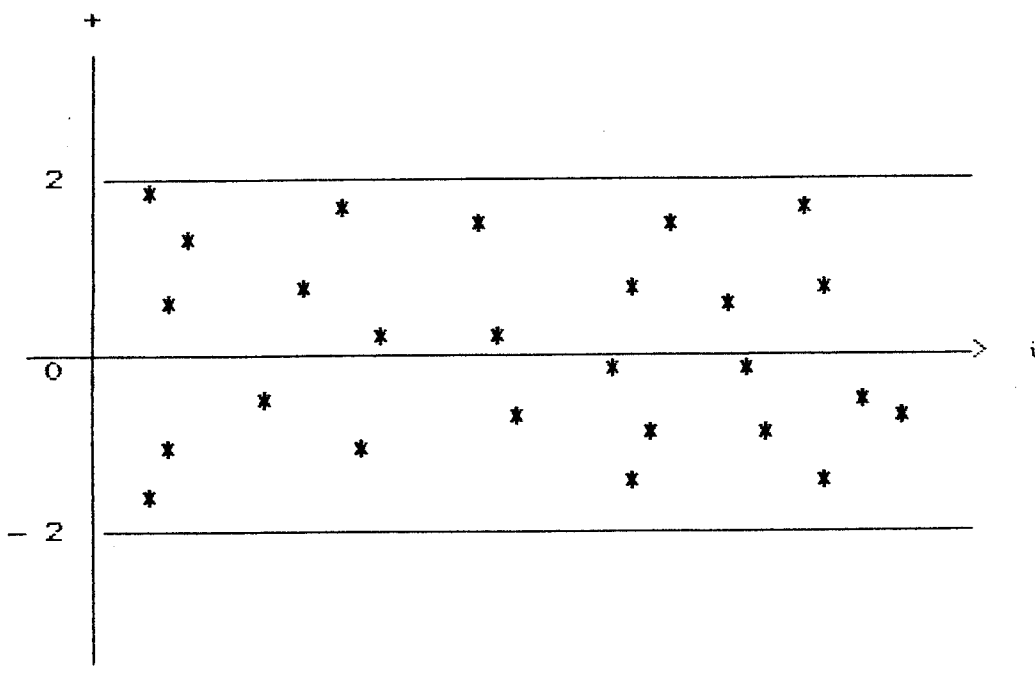


Figura 4.1. Gráfica de residuos estandarizada.

Donde el mayor porcentaje de los residuos se concentra en la banda entre -2 y 2 unidades del origen.

Esta gráfica es también indicativa de una media cero.

La violación de estos supuestos no es muy crítica desde el punto de vista práctico, ya que el supuesto de

normalidad no es esencial si el objetivo es solo la estimación de los parámetros de la regresión. Además, si los U_i no están normalmente distribuidas, puede demostrarse que los estimadores tienden a estarlo a medida que el tamaño de la muestra crece.

Por otra parte, suponiendo que $E(U_i)$ no es cero si no es igual a una constante r si tiene:

$$Y = X\beta + U$$

y

$$E(Y) = X\beta + R.$$

donde:

$$R = \begin{bmatrix} r \\ \vdots \\ r \end{bmatrix}$$

Por consiguiente, si el supuesto no se cumple no es posible estimar la ordenada original. Sin embargo, éste término es en general poco importante, por lo que no se le presta mucha atención.

Variables Regresoras no Estocásticas.

Este supuesto, no se viola tan fácilmente dado que los datos explicatorios son en economía, en general, no estocásticos. Por lo que no se harán más comentarios al respecto.

Multicolinealidad

Un problema que se presenta con frecuencia en los análisis de regresión es la multicolinealidad. Esto es, una alta correlación entre una o más de las variables independientes.

En la práctica, es raro que se presenten relaciones lineales exactas, sin embargo, la interdependencia general de los fenómenos económicos suele conducir a relaciones lineales aproximadas entre las variables. Por ejemplo, la posibilidad de incluir en una función de demanda, como variables independientes, la población y el producto nacional bruto. En la mayoría de los países, una y otra tienden a aumentar con el tiempo, y es frecuente que aparenten tener una relación lineal.

Multicolinealidad Exacta.

Para la regresión con k variables independientes X_1, \dots, X_k , se dice que existe una relación lineal exacta si se satisface la relación:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{bmatrix}' \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}' \quad (4.6)$$

Donde las λ 's son constantes, no todas simultáneamente nulas. Esto es, se dice que existe multicolinealidad exacta cuando hay dependencia lineal en las variables. Esto conduce a que el rango de $X'X$ es menor que k y $(X'X) = 0$, por lo que las ecuaciones normales:

$$(X'X)B = X'Y$$

no tienen solución única, y no es posible obtener un solo estimador por mínimos cuadrados ordinarios.

Si existe multicolinealidad perfecta, una manera de obtener el estimador B es la reparametrización.

Supóngase que $\text{rango}(X'X) = r < k$. Hay entonces un conjunto de r columnas linealmente independientes en X .

Tales columnas pueden arreglarse en las primeras r de una matriz particionada, definida como sigue:

$$X = [X_r, X_s] \quad (4.7)$$

donde: X_r es una matriz $n \times r$ de rango r .

X_s es una matriz de $n \times (r-s)$.

Cada vector columna en X_s puede expresarse como una combinación lineal de las columnas de X_r . Así, es posible escribir:

$$X_s = X_r W \quad (4.8)$$

donde W es una matriz $r \times s$ en la cual cada columna da los coeficientes de la combinación lineal para el correspondiente vector en X_s .

Combinando las ecuaciones (4.7) y (4.8) se tiene:

$$\begin{aligned} X &= [X_r, X_r W] \\ &= X_r [I_r, W] \\ &= X_r Z \end{aligned} \quad (4.9)$$

donde $Z = [I_r, W]$.

El modelo lineal puede ahora ser escrito como:

$$\begin{aligned} Y &= X\beta + U \\ &= X_r Z \beta + U \\ &= X_r \beta_r + U \end{aligned}$$

donde $\beta_r = Z \beta$.

Nótese que β_r indica un vector de r combinaciones lineales en las β 's originales. Los elementos de β_r pueden estimarse usando el principio de mínimos cuadrados, dado que X_r es una matriz de rango completo.

El estimador es pues:

$$B_r = (X_r' X_r)^{-1} X_r' Y \quad (4.10)$$

además,

$$\hat{Y} = Xr \quad (4.11)$$

Las estimaciones en las ecuaciones tienen las propiedades usuales de los estimadores por mínimos cuadrados.

El procedimiento operacional es encontrar la mayor submatriz en X de rango completo y reemplazarla en la ecuación (4.10). Si existe más de una, \hat{Y} será invariante a la matriz escogida.

Casimulticolinealidad.

Un caso menos extremo, pero también importante, surge cuando la hipótesis solo se satisface en parte, es decir, cuando algunas o todas las variables regresoras no son perfectamente colineales, pero lo son en alto grado. Es entonces necesario analizar tres casos.

1. Los efectos que produce la multicolinealidad.
2. Detección de la multicolinealidad.
3. Que acciones tomar para remediarla.

Efectos.

En el análisis de las consecuencias de la multicolinealidad, se toma un modelo simple, en forma de desviación, dado por:

$$Y_i = b_2 x_{2i} + b_3 x_{3i} + e_i$$

Supóngase que existe casimulticolinealidad entre X_2 y X_3 ; esto es:

$$x_{3i} = x_{2i} + u_i \quad (4.12)$$

donde u_i es un término estocástico de error.

Dado que las x son desviaciones:

$$\sum x_{2i} = \sum x_{3i} = 0$$

Las x son variables, así que para simplificar el análisis se pone:

$$\sum x_{2i}^2 = \sum x_{3i}^2 = 1 \quad (4.13.a)$$

$$\sum u_i = 0 \quad (4.13.b)$$

$$\sum u_i x_{2i} = 0 \quad (4.13.c)$$

Bajo tales condiciones:

$$\begin{aligned} r_{23} &= \frac{\sum x_{2i} \cdot x_{3i}}{\sqrt{\sum x_{2i}^2} \sqrt{\sum x_{3i}^2}} \\ &= \frac{\sum x_{2i} (\alpha x_{2i} + u_i)}{\sqrt{\sum x_{2i}^2} \sqrt{\sum x_{3i}^2}} \\ &= \alpha \end{aligned}$$

y además:

$$\begin{aligned} (X'X) &= \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \\ (X'X)^{-1} &= \frac{1}{1-\alpha^2} \begin{bmatrix} 1 & -\alpha \\ -\alpha & 1 \end{bmatrix} \end{aligned}$$

de esta manera:

$$\text{Var} (B) = \frac{\sigma^2}{1-\alpha^2} \quad (4.14)$$

$$\text{Cov} (b_2 \ b_3) = \frac{-\alpha \sigma^2}{1-\alpha^2} \quad (4.15)$$

La ecuación (4.14) muestra que cuando aumenta la multicolinealidad, medida en este caso por α , las varianzas muestrales de los estimadores crecen.

De hecho cuando:

$$\alpha = r_{23} \longrightarrow 1$$

$$\text{Var}(B) \longrightarrow \infty.$$

Cuando $\alpha > 0$, $\text{Cov}(b_2, b_3) < 0$, y esta crece con un aumento en la multicolinealidad.

Ciertamente, la multicolinealidad no destruye la propiedad de varianza mínima, pero esto no quiere decir que la varianza de todo estimador por cuadrados mínimos sea necesariamente pequeña. Por lo tanto, el hecho de que los estimadores por cuadrados mínimos sean los lineales de mínima varianza es, en la práctica de poco valor. Cuando existe casicolinealidad las consecuencias son:

1. La varianza de los estimadores tiende a ser mayor a medida de que aumenta el grado de colinealidad. Consecuentemente, los intervalos de confianza para los parámetros poblacionales crecen.
2. Es posible obtener un R^2 alto aunque con pocos o casi ningún coeficiente estimado estadísticamente significativo.

Detección.

Existen varias formas o más bien, una combinación de formas para detectar la multicolinealidad.

Se sospecha la presencia de la multicolinealidad cuando las varianzas y covarianzas de los estimadores son grandes.

Por otra parte, también puede existir cuando R^2 es alto y las r_{ij} son altas, y a la vez pocos o ninguno de los coeficientes de regresión, son individualmente significativos, en base a la prueba t convencional. No se prueba con

la F dado que si R^2 es alto la prueba F en el análisis de varianza rechazará:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

independientemente de la prueba t.

Cuando se tienen dos variables explicativas, es suficiente el análisis del coeficiente de correlación simple. Sin embargo, cuando hay más de dos variables explicativas se deben de examinar los coeficientes de correlación parcial y de orden cero (esto es entre X_1 y X_2 , entre X_1 y X_3 , etc.).

Ahora bien, dado que la multicolinealidad se presenta porque una o más de las regresoras son una " casi combinación lineal " de las restantes, el coeficiente de determinación múltiple, R_i^2 entre cada X_i y las restantes $(k-1)$ variables de X , es un indicador de cual de ellas está relacionada con el resto.

La F ya definida se calcula para cada R_i^2 reemplazando k por $(k-1)$, dado que se ha excluido Y y solo se contemplan las relaciones entre las X . Así, se tiene:

$$F_i = \frac{R_i^2 / (k-2)}{(1-R_i^2) / (n-k+1)} \quad (i = 2, \dots, k) \quad (4.16)$$

Si F calculado excede al F_i crítico para el nivel de significancia, escogido, se interpreta que X_i en particular es colineal con las otras X .

Caso contrario, la X_i no es colineal con el resto, en tal caso la variable se puede retener en el modelo.

Los criterios comentados para detectar la multicolinealidad deben tomarse con la reserva del problema

que se esté tratando, ya que el grado de multicolinealidad es también un fenómeno muestral, y un análisis económico o bien a priori puede "rechazar" la hipótesis de relación entre las variables regresoras del modelo.

Soluciones.

Una vez detectada la multicolinealidad, el problema radica en qué hacer cuando esta existe.

Algunas de las soluciones sugeridas son:

1. Análisis de información previa.
2. Regresión en cordillera.
3. Omisión de variables.

Análisis de información previa. Considérese el siguiente modelo con tres variables regresoras, donde $X_{1i} = 1$.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$$

Donde, supóngase que:

$$Y_i = \text{Consumo}$$

$$X_2 = \text{Ingreso}$$

$$X_3 = \text{Riqueza}$$

De la teoría económica se postula que el ingreso es una función de la riqueza. Esto es, son altamente colineales. Supóngase ahora que se conoce de estudios previos que:

$$\beta_3 = a \beta_2$$

es decir, la razón de variación del consumo respecto a la riqueza, es a veces la razón de variación del consumo al ingreso.

Bajo tales circunstancias, se puede obtener la siguiente regresión:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + a \beta_3 X_{3i} + U_i \\ &= \beta_1 + \beta_2 X_i + U_i \end{aligned}$$

donde

$$X_i = X_{2i} + a X_{3i}$$

Una vez que se obtenga $\hat{\beta}_2$ se puede proceder a obtener $\hat{\beta}_3$ a partir de la relación postulada entre β_2 y β_3 .

Regresión en Cordillera. Como una alternativa para atacar el problema de multicolinealidad se ha propuesto el método de regresión en cordillera. Este método, sugiere un estimador de β dado por:

$$bc = (X'X + cI)^{-1} X'Y \quad (4.17)$$

Así, dado que:

$$b = (X'X)^{-1} X'Y$$

se tiene:

$$bc = (X'X + cI)^{-1} (X'X) b$$

además:

$$E(bc) = (X'X + cI)^{-1} (X'X) b \quad (4.18)$$

lo cual implica que bc es un estimador sesgado, y:

$$\text{Var}(bc) = \sigma^2 (X'X + cI)^{-1} X'X (X'X + cI)^{-1} \quad (4.19)$$

de tal forma que el error cuadrático medio es:

$$\begin{aligned} \text{ECM}(bc) &= \sigma^2 (X'X + cI)^{-1} X'X (X'X + cI)^{-1} + \\ &\quad ((X'X + cI)^{-1} (X'X) b - \beta)^2 \quad (4.20) \end{aligned}$$

Esta expresión permite la posibilidad de que bc pueda tener un error cuadrático medio menor que b , el estimador por mínimos cuadrados.

La principal dificultad radica en la selección del escalar C . Se han publicado muchos artículos en relación a este tema y con la determinación del valor de C , bajo condiciones especiales, sin embargo, el procedimiento más usado es aún el descrito, donde usando ciertas reglas generales, el problema se resuelve iterativamente para valores de C a partir de cero, y los estimadores son graficados, $br(c)$ vs c .

De esta gráfica se selecciona el valor óptimo de C , como el mínimo valor, en el cual las curvas de todos los estimadores se estabilizan. La subjetividad del procedimiento es obvia. Además, otro inconveniente de éste método es que constituye una solución meramente estadística y por tanto no puede resultar atractiva para muchos economistas.

Omisión de Variables. Otra forma de enfrentar el problema de la multicolinealidad, es omitir variables cuyo coeficiente sea de menor interés en la investigación.

En tal caso, se pueden obtener estimadores para los parámetros de interés, con errores cuadráticos medios menores que los estimadores de mínimos cuadrados. Como una ilustración considérese el modelo de tres variables en forma de desviación:

$$Y = \beta_2 x_2 + \beta_3 x_3 + u \quad (4.21)$$

el problema es que x_2 y x_3 están altamente correlacionadas.

Es posible obtener por mínimos cuadrados al estimador B , donde:

$$B = \begin{bmatrix} b_2 \\ b_3 \end{bmatrix}$$

este estimador es insesgado y con varianza muestral

$$\begin{aligned} \text{Var}(B) &= \text{Var} \begin{bmatrix} b_2 \\ b_3 \end{bmatrix} \\ &= \sigma^2 (X'X)^{-1} \\ &= \frac{\sigma^2}{\Sigma x_2^2 \Sigma x_3^2 - (\Sigma x_2 x_3)^2} \begin{bmatrix} \Sigma x_3^2 & -\Sigma x_2 x_3 \\ -\Sigma x_2 x_3 & \Sigma x_2^2 \end{bmatrix} \end{aligned}$$

con lo cual

$$\begin{aligned} \text{Var}(b_2) &= \frac{\sigma^2 \Sigma x_3^2}{\Sigma x_2^2 \Sigma x_3^2 - (\Sigma x_2 x_3)^2} \\ &= \frac{\sigma^2}{\Sigma x_2^2 (1 - r_{23}^2)} \end{aligned} \quad (4.22)$$

y

$$\begin{aligned} \text{Var}(b_3) &= \frac{\sigma^2 \Sigma x_2^2}{\Sigma x_2^2 \Sigma x_3^2 - (\Sigma x_2 x_3)^2} \\ &= \frac{\sigma^2}{\Sigma x_3^2 (1 - r_{23}^2)} \end{aligned} \quad (4.23)$$

donde r_{23} es el coeficiente de correlación simple entre X_2 y X_3 .

Es claro que a medida de que r_{23}^2 se acerca a uno ambas varianzas se incrementan.

Considérese ahora la regresión simple de y sobre x_2 .

Denotando por b_{2y} al coeficiente de regresión, se

tiene:

$$b_{2y} = \frac{\Sigma y x_2}{\Sigma x_2^2} \quad (4.24)$$

reemplazando en la ecuación (4.21) se obtiene:

$$b_{2y} = \beta_2 + b_{32} \beta_3 + \frac{\sum x_2 u}{\sum x_2^2} \quad (4.25)$$

donde:

$$b_{32} = \frac{\sum x_2 x_3}{\sum x_2^2}$$

representa el coeficiente de regresión x_3 sobre x_2 (es decir x_2 de regresor de x_3). De la ecuación (4.25) se sigue que:

$$E(b_{2y}) = \beta_2 + b_{32} \beta_3 \quad (4.26)$$

y

$$\begin{aligned} \text{Var}(b_{2y}) &= E \left\{ \left[b_{2y} - E(b_{2y}) \right]^2 \right\} \\ &= E \left\{ \left[b_{2y} - \beta_2 - b_{32} \beta_3 \right]^2 \right\} \\ &= E \left\{ \left[\frac{\sum x_2 u}{\sum x_2^2} \right]^2 \right\} \\ &= \frac{\sigma^2}{(\sum x_2^2)^2} \end{aligned} \quad (4.27)$$

Es decir b_{2y} es un estimador sesgado de β_2 a menos que x_2 y x_3 sean ortogonales, de tal forma que $b_{23} = 0$,

Nótese ahora que:

$$\text{Var}(b_{2y}) = \frac{\sigma^2}{(\sum x_2^2)} < \text{Var}(b_2) = \frac{\sigma^2}{\sum x_2^2 (1 - r_{23}^2)}$$

Falta solo ver bajo que condiciones b_{2y} puede tener un menor error cuadrático medio que b_2 .

Para esto, dado que:

$$\text{ECM}(b_{2y}) = \frac{\sigma^2}{(\sum x_2^2)} + b_{32}^2 \beta_3^2$$

$$ECM(b_2) = \frac{\sigma^2}{\sum x_2^2 (1 - r_{23}^2)}$$

entonces:

$$\frac{ECM(b_{2y})}{ECM(b_2)} = 1 + r_{23}^2 (\tau^2 - 1) \quad (4.28)$$

donde:

$$\tau^2 = \frac{\beta_3^2}{\sigma^2 / \sum x_3^2 (1 - r_{23}^2)} = \frac{\beta_3^2}{\text{Var}(b_3)} \quad (4.29)$$

El estadístico τ^2 es la razón del cuadrado del valor real (pero desconocido) de β_3 a la varianza real (no estimada) de b_3 .

Si $\tau^2 < 1$:

$$ECM(b_{2y}) < ECM(b_2)$$

De ésta forma, si se está interesado principalmente en obtener una estimación tan buena como sea posible de β_2 , y si es posible confiar que τ^2 sea significativamente menor que uno, simplemente se omite x_3 de la regresión y se lleva a cabo una regresión simple de y contra x_2 .

Sin embargo, τ^2 no se conoce, pero se puede estimar con:

$$t^2 = \frac{b_3^2}{S^2 / \sum x_3^2 (1 - r_{23}^2)} \quad (4.30)$$

y se puede definir un estimador condicional de variable omitida de β_2 como:

$$b_{ECVO} = \begin{cases} b_{2y} & \text{Si } t^2 < 1 \\ b_2 & \text{Si } t^2 \geq 1 \end{cases}$$

Heterocedasticidad.

El supuesto a examinar es el que mantiene a los

residuos con una varianza constante σ^2 . Esto se conoce como homocedasticidad y la violación de este supuesto como heterocedasticidad.

Este problema de varianza no constante no es raro en los trabajos económicos, Gráficamente la heterocedasticidad se muestra en la figura 4.2.

Así como el ejemplo anterior, se puede pensar en muchas más relaciones económicas; tales como la regresión entre ingresos y ahorro, producción y tiempo de trabajo de un obrero. etc.

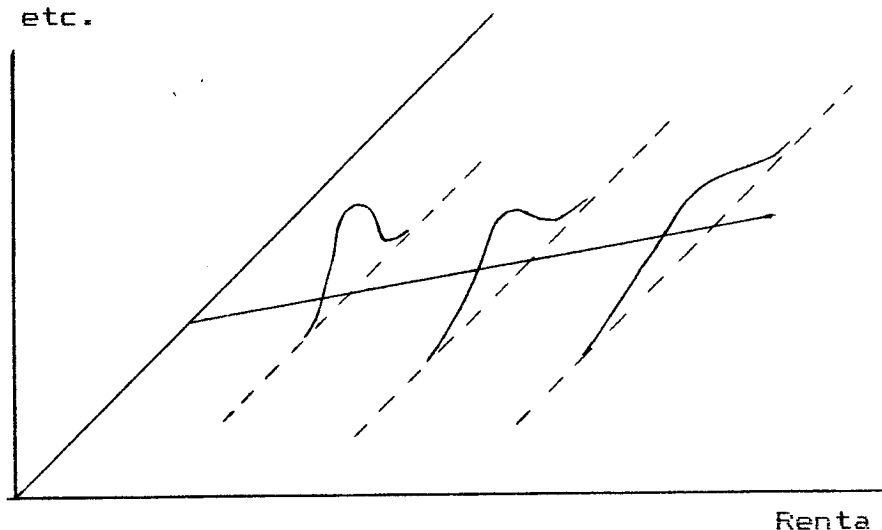


Figura 4.2. Perturbaciones heterocedásticas.

Detección.

La presencia de heterocedasticidad puede ser detectada de varias formas, algunas formales y otras informales.

Un primer indicio de heterocedasticidad lo sugiere la naturaleza del problema y la información previa.

En la práctica, cuando no existe información empírica, se puede hacer el análisis de regresión, y

efectuar un exámen gráfico posterior de los residuos estimados, para ver si se presenta algún patrón sistemático. Los residuos estimados (e_i) pueden graficarse contra x_i (o contra \hat{y}_i en el caso de regresión múltiple o no lineal).

La presencia de varianza no constante se mostrará en estas gráficas, en forma de una tendencia de la banda de residuos a ensancharse o a reducirse, según se muestra en la figura 4.3.

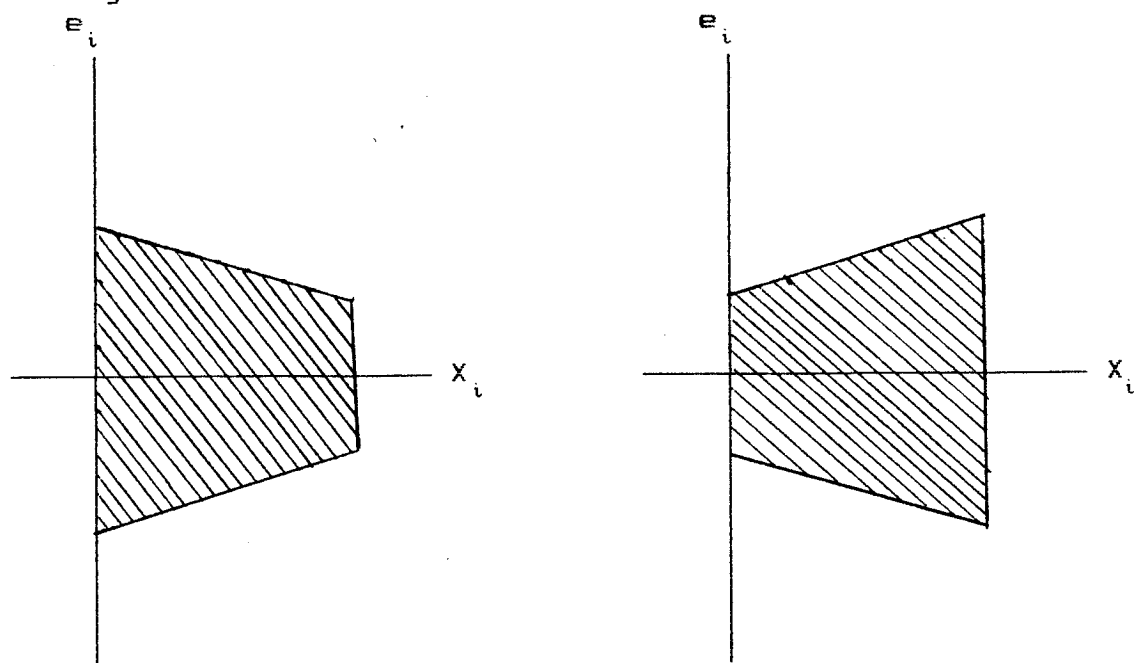


Figura 4.3. Presencia de heterocedasticidad.

Existe una formalización del método gráfico en la cual se propone que σ_i^2 es una función de la variable regresora X_i . La forma funcional es:

$$\sigma_i^2 = \sigma^2 X_i^\beta e^{v_i}$$

o bien:

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \quad (4.31)$$

donde v_i es el término estocástico de perturbación.

Siendo que σ_i^2 es en general desconocida, se propone que se use e_i^2 como aproximación, y que se realice la siguiente regresión:

$$\begin{aligned} \ln e_i^2 &= \ln \sigma^2 + \beta \ln X_i + v_i \\ \ln e_i^2 &= \alpha + \beta \ln X_i + v_i \end{aligned} \quad (4.32)$$

Si β resulta estadísticamente significativa, esto sugiere la existencia de heterocedasticidad.

Consecuencias.

Cuando existe la heterocedasticidad, los estimadores por mínimos cuadrados siguen siendo insesgados, sin embargo, sus varianzas ya no son las más pequeñas. Además, las estimaciones de las varianzas son sesgadas.

Para ver esto, considérese un modelo simple en su forma de desviación.

$$y_i = \beta x_i + u_i \quad \text{Var}(u_i) = \sigma_i^2 \quad (4.33)$$

El estimador por mínimos cuadrados es:

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \beta + \frac{\sum x_i u_i}{\sum x_i^2} \quad (4.34)$$

si se satisfacen los demás supuestos se tiene que:

$$E(b) = \beta$$

es decir b es aún insesgado.

Por otra parte:

$$\text{Var}(b) = \text{Var} \left\{ \frac{\sum x_i u_i}{\sum x_i^2} = \frac{x_1^2 u_1}{\sum x_i^2} + \dots + \frac{x_n^2 u_n}{\sum x_i^2} \right\}$$

$$= \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \quad (4.35)$$

La $\text{Var}(b)$ se compara en el siguiente apartado con la varianza de b^* el estimador por mínimos cuadrados ponderados, y resulta ser mayor.

Corrección.

Existen dos enfoques para corregir la heterocedasticidad:

1. Cuando σ_i^2 es conocido
2. Cuando σ_i^2 es desconocido

Cuando σ_i^2 es conocido, el método más sencillo de tratar la heterocedasticidad es el de cuadrados mínimos ponderados. La aplicación de este método se ilustra a continuación usando el modelo de dos variables.

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Cuando se usa el método de cuadrados mínimos ordinario. Cada e_i^2 tiene la misma ponderación en la estimación de los parámetros de regresión.

El método de mínimos cuadrados ponderados "pesa" cada una de las desviaciones. Este criterio exige:

Encontrar a^* , b^* tales que:

$$\sum w_i e_i^2 = \sum w_i (Y_i - a^* - b^* X_i)^2 \quad (4.36)$$

sea mínima.

Los w_i son constantes (no estocásticos) y a^* , b^* son los estimadores por cuadrados mínimos ponderados. Los w_i se escogen de tal forma que las observaciones extremas reciban menor ponderación. Si σ_i^2 se conoce, se puede poner:

$$w_i = \frac{1}{\sigma_i^2} \quad (4.37)$$

esto es, ponderar cada observación de manera inversamente proporcional a σ_i^2 .

El proceso de minimizar (4.36) es el usual y proporciona los estimadores:

$$\begin{aligned} a^* &= \frac{\sum w_i Y_i}{\sum w_i} - b^* \frac{\sum w_i X_i}{\sum w_i} \\ &= \bar{Y}^* - b^* \bar{X}^* \end{aligned}$$

donde \bar{Y}^* y \bar{X}^* son medias muestrales ponderadas con w_i como ponderación y:

$$b^* = \frac{\sum w_i y_i^* x_i^*}{\sum w_i x_i^{*2}} \quad (4.38)$$

nótese que si $w_1 = w_2 = \dots = w_n = W$ los estimadores se reducen a los de mínimos cuadrados ordinarios.

Ahora bien:

$$\begin{aligned} b^* &= \frac{\sum w_i y_i^* x_i^*}{\sum w_i x_i^{*2}} \\ &= \frac{\sum w_i x_i^* (\beta x_i^* + u_i^*)}{\sum w_i x_i^{*2}} \\ &= \beta + \frac{\sum w_i x_i^* u_i^*}{\sum w_i x_i^{*2}} \end{aligned}$$

y

$$E(b^*) = \beta$$

resultando que b^* es también insesgado.

Continuando:

$$\begin{aligned} \text{Var}(b^*) &= \text{Var} \left[\frac{w_1 x_1^* u_1^*}{\sum w_i x_i^{*2}} + \frac{w_2 x_2^* u_2^*}{\sum w_i x_i^{*2}} + \dots + \frac{w_n x_n^* u_n^*}{\sum w_i x_i^{*2}} \right] \\ &= \frac{1}{(\sum w_i x_i^{*2})^2} \text{Var} \left[w_1 x_1^* u_1^* + w_2 x_2^* u_2^* + \dots + w_n x_n^* u_n^* \right] \\ \text{Var}(b^*) &= \frac{\sum (w_i x_i^* \sigma_i)^2}{(\sum w_i x_i^{*2})^2} \end{aligned} \quad (4.39)$$

Nótese que bajo el supuesto de homocedasticidad, la varianza de b^* se reduce a la de los mínimos cuadrados ordinarios.

Comparando las ecuaciones (4.35) y (4.39) se tiene:

$$\frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} > \frac{\sum (w_i x_i^* \sigma_i)^2}{(\sum w_i x_i^{*2})^2} \quad (4.40)$$

Por tanto $\text{Var}(b) > \text{Var}(b^*)$. De lo cual se desprende que el criterio de mínimos cuadrados ponderados produce mejores estimadores insesgados que los que proporciona el criterio de cuadrados mínimos ordinarios.

El conocimiento previo de σ_i^2 es muy poco común, por lo que el método de cuadrados mínimos antes descrito no puede usarse fácilmente. En todo caso, el problema de la heterocedasticidad puede resolverse usando transformaciones llamadas estabilizadoras de varianza. La función de tales transformaciones es ajustar el modelo de regresión original, de tal forma que satisfaga el supuesto de homocedasticidad.

Sin una transformación tal, el problema de heterocedasticidad se torna prácticamente insoluble.

Mediante el análisis gráfico, supóngase que se determina que el error tiene la siguiente relación con X en

una regresión simple:

$$\text{Var}(e_i) = k^2 X_i^2 \quad (4.41)$$

esto es, que la desviación estándar de los residuos es directamente proporcional a X . Así:

$$\text{Var} \left[\frac{e_i}{X_i} \right] = \frac{\text{Var}(e_i)}{X_i^2} = \frac{k^2 X_i^2}{X_i^2} = k^2$$

y con esto, la varianza se estabiliza. Esto da la base para transformar el modelo original a:

$$\frac{Y_i}{X_i} = \beta + \frac{\alpha}{X_i} + \frac{U_i}{X_i}$$

o bien

$$Y'_i = \alpha' + \beta' X'_i + U'_i \quad (4.42)$$

donde el modelo con las variables transformadas cumple la suposición de la varianza constante.

Si se cree que la varianza es proporcional a X_i , entonces se tiene:

$$\text{Var}(e_i) = k X_i$$

y

$$\text{Var} \left[\frac{e_i}{\sqrt{X_i}} \right] = \frac{1}{X_i} k X_i = k$$

La varianza ahora se estabiliza transformando el modelo original a:

$$\frac{Y_i}{\sqrt{X_i}} = \frac{\alpha}{\sqrt{X_i}} + \beta \sqrt{X_i} + \frac{U_i}{\sqrt{X_i}}$$

o bien:

$$Y'_i = \alpha X'_i + \beta Z'_i + U'_i \quad (4.43)$$

En esta situación homocedástica se deberá hacer la regresión de:

$$\frac{Y_i}{\sqrt{X_i}} \text{ contra } \frac{1}{\sqrt{X_i}} \text{ y } \sqrt{X_i}$$

una vez realizada la regresión se puede volver el modelo original multiplicando por $\sqrt{X_i}$.

Otras estabilizadoras de varianza frecuentemente usadas son:

$$\text{Var}(e_i) = k^2 X^3 ; \text{Var}(e_i) = k^2 \ln^2 X_i ; \text{Var}(e_i) = k^2 \frac{1}{X_i^2}$$

En general las transformaciones que estabilizan la varianza en los casos a y b mostrados en la figura 4.3 son X^2 ó X^3 para a, y $\ln x$ ó $1/X$ para B. Aún más, si en lugar de efectuar la regresión:

$$Y_i = \alpha + \beta X_i + u_i$$

se efectúa:

$$\ln Y_i = \alpha + \beta \ln X_i + u_i \quad (4.44)$$

la heterocedasticidad se reduce. Esto es porque la transformación logarítmica reduce las escalas en que están medidas las variables.

Debe tomarse en cuenta que al efectuar transformaciones en modelos de más de dos variables, se corre el riesgo de correlacionarlas con la transformación.

Esto debe poner alerta al lector contra los problemas asociados con las transformaciones.

Autocorrelación.

El último supuesto del modelo clásico de regresión lineal que falta de analizar, es el que postula la independencia entre los términos de perturbación estocástica esto es:

$$E(U_i U_j) = 0 \quad \text{para toda } i \neq j$$

o en forma equivalente:

$$E(U_i U_{i+s}) = 0 \quad \text{para toda } i \text{ y toda } s \neq 0.$$

Sin embargo, en muchos casos, los residuos de la regresión se comportan de una manera intuitivamente incompatible con tal hipótesis. Por ejemplo, no parece aceptable que los tres primeros residuos sean positivos, los tres siguientes negativos etc. Se puede sospechar la omisión de alguna variable importante del conjunto de variables explicativas, o bien, que la estructura probabilística del proceso de los errores no pertenece a la clase de los formados por variables independientes e idénticamente distribuídas.

En esta sección se considera el caso en que los términos de error muestran una dependencia serial o autocorrelación.

Detección.

Existen varias formas prácticas para detectar la autocorrelación. Si bien, los supuestos del modelo clásico de no autocorrelación se refieren a las perturbaciones poblacionales no observables, se dispone de sus aproximaciones, los residuos obtenidos mediante el método de cuadrados mínimos ordinarios.

Para confirmar la suposición de independencia es necesario guardar el orden en el cual se obtuvieron las observaciones, y graficar los residuos contra este, digamos t .

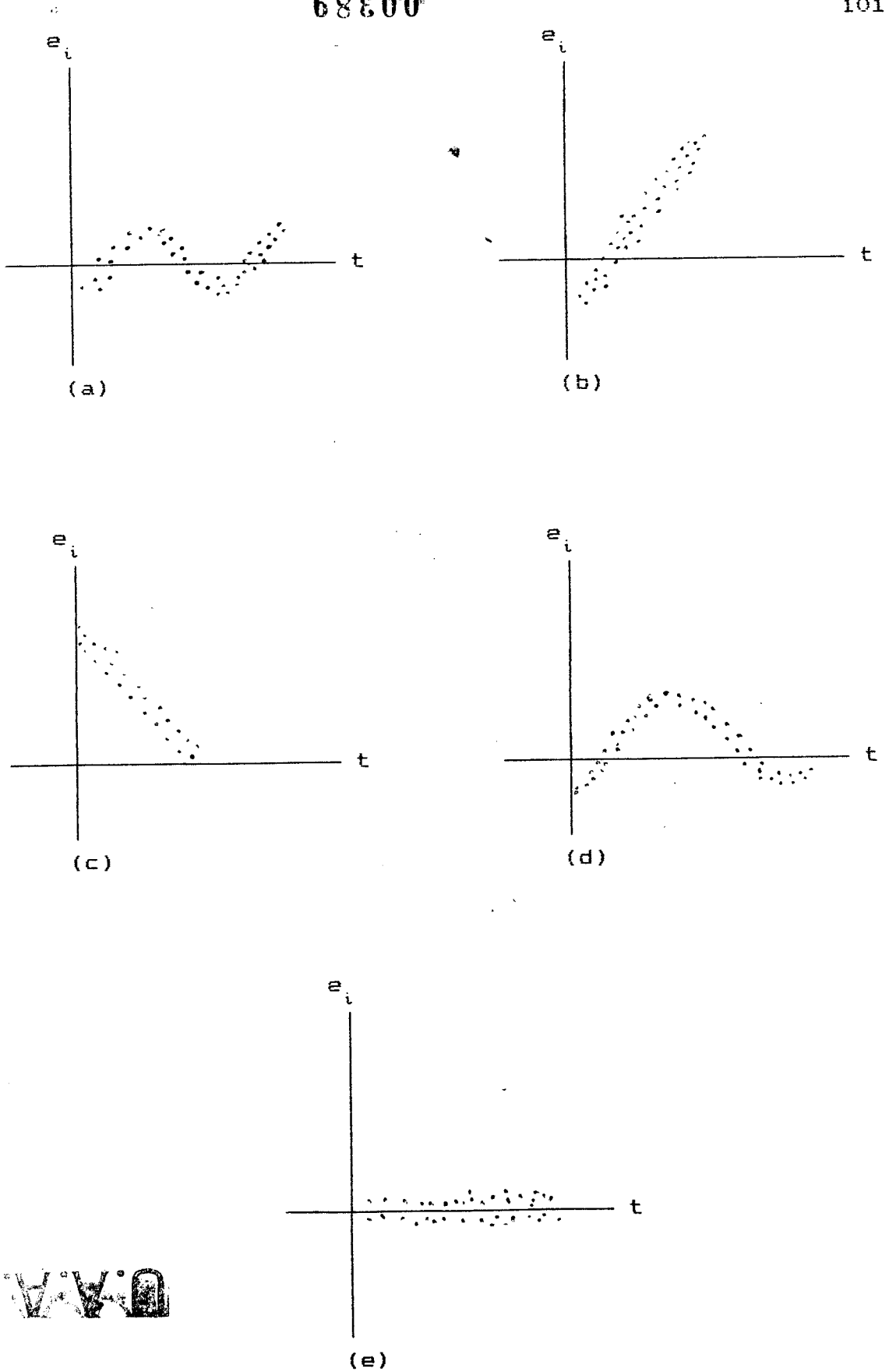


Figura 4.4. Diagramas de dispersión de e_i .

La falta de independencia se mostrará como una tendencia en el tiempo, la cual presentará distintos patrones de comportamiento. Figuras 4.4. a, b, c, d.

Por otra parte, la existencia de independencia se identifica como un patrón aleatorio alrededor de cero. Figura 4.4.e.

Existen otros comportamientos posibles. Pero en general una tendencia definida implica una cierta estructura de autocorrelación de los residuos.

Antes de mostrar un método estadístico para detectar la autocorrelación es necesario desarrollar algunos conceptos previos.

Supóngase el modelo de regresión múltiple.

$$Y = X \beta + U$$

y que las perturbaciones se generan de acuerdo al modelo.

$$U_t = P U_{t-1} + E_t \quad |P| < 1 \quad (4.45)$$

$$E(E_t) = 0 \quad (4.46)$$

$$\text{Var } E_t = \sigma_e^2 \quad (4.47)$$

$$\text{Cov}(E_t, E_{t+s}) = 0 \quad s \neq 0. \quad (4.48)$$

Donde P se llama coeficiente de autocovarianza y E_t es una perturbación estocástica que satisface los supuestos de mínimos cuadrados.

Este modelo corresponde al llamado autorregresivo de primer orden. Se llama autorregresivo porque es una regresión de U_t contra sí mismo, y de primer orden porque

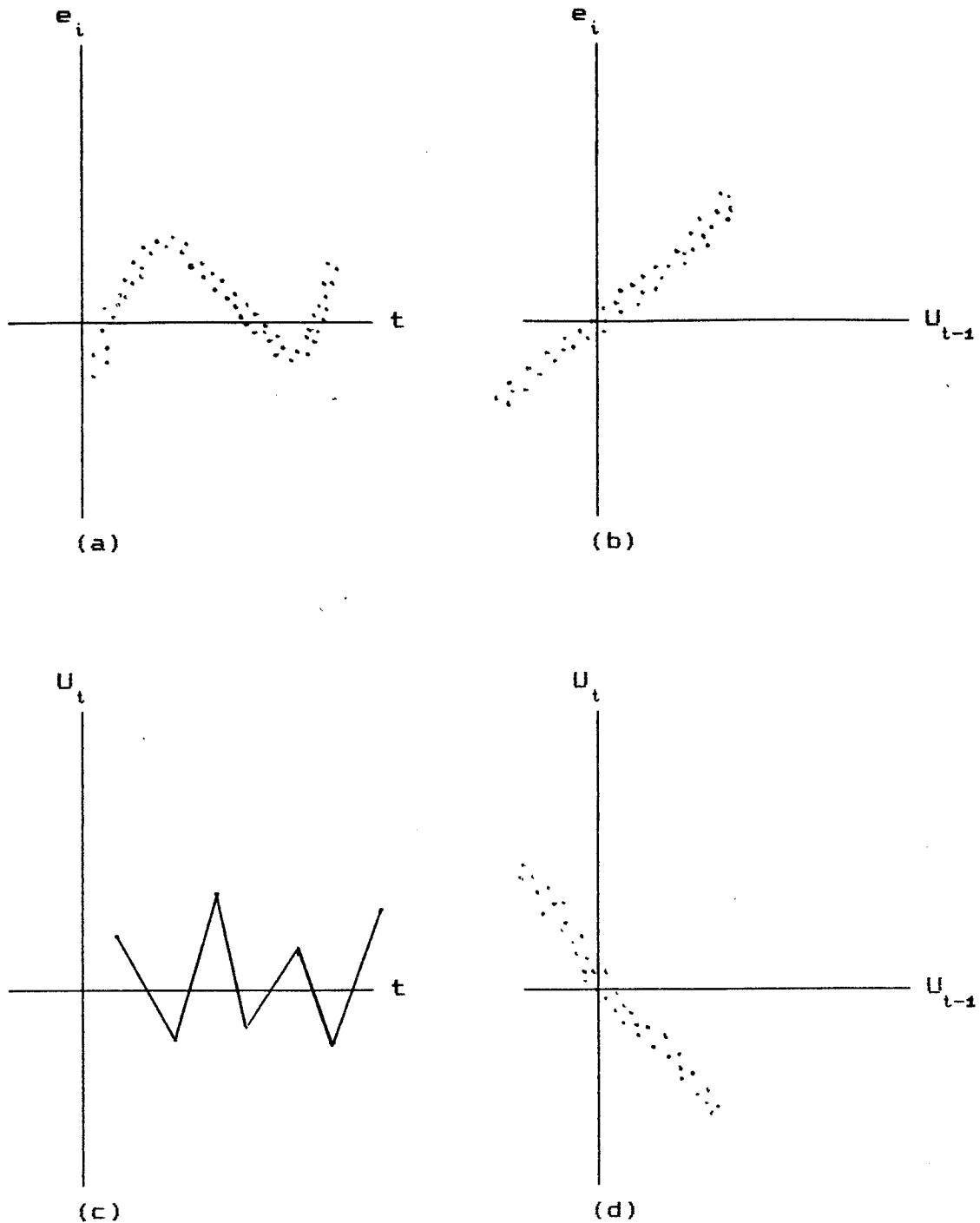


Figura 4.5. Dispersiones de U_t .

va retrasado un período.

Si el modelo fuese tal que $U_t = \rho U_{t-2} + E_t$, sería de segundo orden, y así sucesivamente.

La autorrelación puede ser positiva o negativa. La gráfica de los residuos para ambas se muestran en la figura

4.5.a, 4.5.b.

En economía se presenta con más frecuencia la positiva, dado que la mayoría de las series económicas se mueven hacia arriba o hacia abajo y no con movimientos ascendentes - descendentes, según se muestra en la figura 4.5.c, 4.5.d.

El modelo autoregresivo de primer orden puede detectarse calculando el estadístico de Durbin-Watson (D-W), definido como:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n (e_t)^2} \quad (4.45)$$

el cual tiene los siguientes supuestos subyacentes.

1. El modelo de regresión incluye la ordenada al origen.
2. Las variables regresoras son no estocásticas.
3. Las perturbaciones se generan mediante un modelo de autorregresión del primer orden.

La distribución del estadístico d se encuentra tabulada de tal forma, que pueden encontrarse acotaciones inferiores (d_i) y superiores (d_s), para distintas combinaciones de valores de n, k , y α .

La prueba de D-W para detección de autocorrelación serial se conduce de la siguiente manera.

Hipótesis:

H_0 : No hay correlación serial positiva.

H_1 : Hay correlación serial positiva.

Se calcula d y se compara con d_i . Si existe correlación serial positiva ($0 < P < 1$) d tiende a ser

pequeño, y si $d < d_i$ se rechaza H_0 y se acepta H_1 . Si $d > d_s$ se acepta H_0 . Si $d_i < d < d_s$ la prueba no es concluyente por lo que se requieren nuevos datos.

Para probar correlación serial negativa se usa $d' = 4 - d$ en lugar de d . Si $d' < d$ se rechaza H_0 , es decir, no hay correlación serial negativa. Si $d > d_s$ ó $d' > d_s$ se dice que d' no es significativo y se acepta H_0 . En los demás casos, la prueba D-W no es concluyente.

Efectos.

Los efectos de aplicar el criterio de mínimos cuadrados ordinarios, a una relación con perturbaciones autocorrelacionadas son similares cualitativamente, a los provocados por la heterocedasticidad. Esto es, la estimación es insegura e ineficiente. Además los procedimientos de inferencia no son válidos.

Para probar tal afirmación es conveniente desarrollar un modelo de regresión generalizado, lo cual está fuera de los propósitos del presente trabajo.

Corrección.

Las medidas correctivas de la autocorrelación, dependen del conocimiento que se tenga de la naturaleza de la independencia entre las perturbaciones.

En la práctica, se supone generalmente que U_t sigue un esquema autorregresivo de primer orden.

Si P , el coeficiente de correlación se conoce, el problema se resuelve de la siguiente forma.

Considérese sin pérdida de generalidad, el modelo de dos variables dado que la autocorrelación es una propiedad de las U's:

$$Y_t = \alpha + \beta X_t + u_t \quad (4.46)$$

Si el modelo es válido t, también es para t-1, esto es:

$$Y_{t-1} = \alpha + \beta X_{t-1} + u_{t-1} \quad (4.47)$$

multiplicando (4.47) por P y restando a (4.46) se obtiene:

$$(Y_t - PY_{t-1}) = \alpha(1-P) + \beta(X_t - PX_{t-1}) + E_t \quad (4.48)$$

Dado que E_t satisface todos los supuestos de cuadrados mínimos este criterio se puede aplicar, y obtener los mejores estimadores para α y β .

La regresión (4.48), se conoce como la ecuación de diferencias generalizada. Obteniendo las diferencias se pierde una información, dado que la primera no tiene un antecesor, para evitar esto la primera observación se transforma:

$$Y_1\sqrt{1-P^2} ; X_1\sqrt{1-P^2}$$

La regresión (4.48) es difícil de correr en la práctica, dado que P se conoce solo en muy contadas ocasiones, y por tal motivo, se recurre con frecuencia a procedimientos iterativos.

Para ilustrar uno de ellos, considérese la ecuación (4.48) escrita como:

$$Y'_t = \alpha' + \beta X'_t + E_t \text{ para } i = 1, 2, \dots, n \quad (4.49)$$

donde:

$$Y'_t = (Y_t - PY_{t-1})$$

$$\alpha' = \alpha(1-P)$$

$$X'_t = (X_t - PX_{t-1}) \quad (4.50)$$

Si las X'_t y las Y'_t fuesen conocidas, la regresión por mínimos cuadrados suministraría los mejores estimadores lineales insesgados para α' y β .

Para estimar X'_t y Y'_t , usando los valores originales de X_t y Y_t se obtienen primero estimadores simples para α y β , y estas se emplean para calcular los residuos.

$$e_t = Y_t - a - bX_t$$

Estos se emplean, posteriormente, para estimar la autocorrelación P del esquema autorregresivo de primer orden de la forma:

$$U_t = PU_{t-1} + E_t$$

donde el estimador de P por mínimos cuadrados es:

$$r = \frac{\sum_{t=2}^n (e_t - e_{t-1})}{\sum_{t=1}^n (e_{t-1})^2} \quad (4.51)$$

Después se usa r para calcular las variables transformadas X'_t y Y'_t . Luego, se emplean mínimos cuadrados ordinarios para estimar α' y β en (4.49). Si los residuos obtenidos en la nueva relación tienen una autocorrelación significativa pueden usarse para obtener una nueva estimación de P , este a su vez se empleará en la obtención de un nuevo conjunto de variables transformadas, sobre la base de las observaciones originales. Este procedimiento iterativo se continúa hasta eliminar la autocorrelación.

El sistema puede hacerse extensivo al caso de más de una variable independiente.

El método descrito es llamado de primeras diferencias.

LITERATURA CITADA

- Dagum C., y E.M.B. de Dagum. 1972. Introducción a la Econometría. 1a. Edición. Editores S.A. México. p. 6-36.
- Dhrymes P., J. 1978. Econometría. 1a. Edición. Editorial AC. España. p. 1-99.
- Gujarati D. 1978. Econometría Básica. 1a. Edición. Mc Graw Hill. México. p. 12-241.
- García P., A. 1968. Apuntes de Econometría. Manuales Universitarios UNAM. México. p. 8-10.
- García P., J. 1989. Análisis de Regresión con Aplicaciones. CINVESTAV. México. p. 9-66.
- Johnston S. 1984. Econometrics Methods. 3a. Edición. Mc. Graw Hill. República de Singapore. p. 1-74, 161-181.
- Maddala G.S. 1985. Econometría. 1a. Edición. Mc. Graw Hill. México. p. 76-218.
- Martínez G., A. 1982. Métodos Econométricos. 2a. Edición. Colegio de Posgraduados (Chapingo, México). México. p. 45-151.
- Merrill W., K. Fox. 1972. Introducción a la Estadística Económica. 1a. Edición. Amorrortu Editores. Argentina. p. 334-464
- Mood-Graybill and Boes. 1985. Introduction to the Theory of Statistics. 3a. Edición. República de Singapore. p. 482- 498.
- Ros B.J. 1984. Economía Mexicana. CIDE. México. p. 9-11.
- Steel R.G.D., J.H. Torrie. 1985. Bioestadística. Principios y Procedimientos. 1a. Edición. Mc. Graw Hill. México. p. 316-317.