PRESENTACION

El presente material ofrece los lineamientos generales presentados en las sesiones: " Estadísticas Exploratorias " y " Estimación Econométrica " durante el " CURSO SOBRE EL PAPEL DE LA SOCIOECONOMIA EN LA GENERACION DE TECNOLOGIA AGRICOLA ", realizado en CIAT, entre Septiembre 16 y Octubre 11 de 1991.

Se pretende ofrecer a los usuarios una copia de los acetatos que se proyectan, con el fin de que puedan registrar en ella explicaciones adicionales, comentarios, ejemplos o cualquier elemento que contribuya a la claridad del tema considerado. De esta manera, al ser un material para uso paralelo a una explicación, permite a quien vuelva sobre sus páginas recordar un concepto o reconsiderar un ejemplo en forma directa.

Debido a la diferencia entre participantes en téminos de conocimientos estadísticos, este manual debe plantear un bosquejo del tema, permitiendo detenerse o avanzar rápidamente según el grupo lo requiera. Por estas razones, no recomendamos su uso para alguien que no haya estado presente en las exposiciones. Los enunciados de los conceptos son muy generales y al no existir interacción con el expositor, no hay la posibilidad de hacer correciones, ampliaciones, o de ofrecer mayor claridad sobre puntos específicos.

Esta constituye la primera versión del material y está sujeta a revisión. Si usted tiene sugerencias, nos sería de mucha utilidad conocerlas.

Deseamos agradecer la colaboración especial de Luis Roberto Sanint con respecto a la definición de temas, y a Marta Elena Carvajal en la transcripción.

- SIGNIFICADO DEL TERMINO LINEAL

- FORMAS FUNCIONALES MAS COMUNES

- MODELO DE REGRESION LINEAL CON 2 VARIABLES
  . Objetivo del análisis de regresión
  . Bases conceptuales del método
  . Precisión de los estimadores de cuadrado mínimos
  . Calidad del ajuste

- SUPUESTOS SOBRE LOS CUALES SE APOYA EL METODO DE REGRESION POR MINIMOS CUADRADOS

- VIOLACION DE LOS SUPUESTOS, DIAGNOSTICO, IMPLICACIONES Y MEDIDAS REMEDIALES

- SUPUESTO ADICIONAL : SUPUESTO DE NORMALIDAD

- **PRUEBAS DE HIPOTESIS**

- **PREDICCION MEDIA**

- **REGRESION LINEAL MULTIPLE**
  . Objetivos del análisis
  . Calidad del ajuste

- **SUPUESTOS**

- **VIOLACION DE LOS SUPUESTOS, DIAGNOSTICO, IMPLICACIONES Y MEDIDAS REMEDIALES**

¿QUE ES ECONOMETRIA?

Literalmente, *Econometría* significa "medición económica". Si bien es cierto que la medición es una parte importante de la econometría, el campo de acción de esta disciplina es mucho más amplio, como puede verse en las siguientes citas.

> La econometría, que es el resultado de cierta posición sobre el papel de la economía, consiste en la aplicación de la estadística matemática a datos económicos, para dar apoyo empírico a los modelos construidos por la economía matemática, y para obtener resultados numéricos[1].
>
> ... la econometría puede ser definida como el análisis cuantitativo de fenómenos económicos reales basados en los desarrollos simultáneos de la observación y la teoría, relacionados mediante métodos apropiados de inferencia[2].
>
> La econometría puede definirse como la ciencia social en la cual las herramientas de la teoría económica, las matemáticas y la inferencia estadística se aplican al análisis de los fenómenos económicos[3].
>
> La econometría se refiere a la determinación empírica de las leyes económicas[4].

[1] Gerhard Tintner, *Methodology of Mathematical Economics and Econometrics*, The University of Chicago Press, Chicago, 1968, p. 74.

[2] P. A. Samuelson, T. C. Koopmans, y J. R. N. Stone, "Report of the Evaluative Committee for *Econometrica*," *Econometrica*, vol. 22, No. 2, abril 1954, pp. 141-146.

[3] Arthur S. Goldberger, *Econometric Theory*, John Wiley & Sons, Inc., New York, 1964, p. 1.

[4] H. Theil, *Principles of Econometrics*, John Wiley & Sons, Inc., New York, 1971, p. 1.

METODOLOGIA DE LA ECONOMETRIA

(1) Teoría económica

↓

(2) Modelo econométrico de la teoría

↓

(3) Recolección de cifras apropiadas

↓

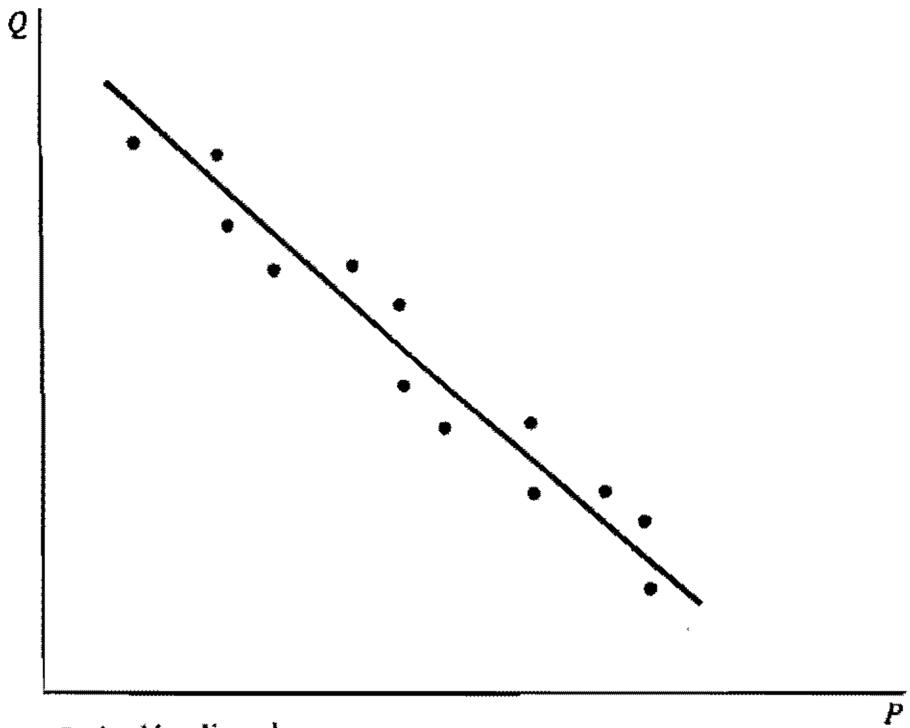
(4) Estimación de los parámetros del modelo

↓

(5) Inferencia estadística

↓	↓
Aceptar la teoría si las cifras (datos) son compatibles con la teoría	Rechazar la teoría si las cifras (datos) no son compatibles con la teoría
↓	↓
Predicción	Revisión de la teoría o teoría nueva
	↓
	Para la verificación, seguir los pasos (2) a (5)

SIGNIFICADO DEL TERMINO LINEAL



Relación lineal

Linealidad en las variables :

Y es una función lineal de X_i se refiere a que la relación tiene la representación geométrica de una recta.

Se dice que Y es una función lineal de X si X aparece con una potencia de 1 (o sea que términos como X^2, $\sqrt{X,}$ ln(X) no aparecen).

Además X no está multiplicada ni dividida por otra variable.

$Y = b_0 + b_1 X$

Linealidad en los parámetros :

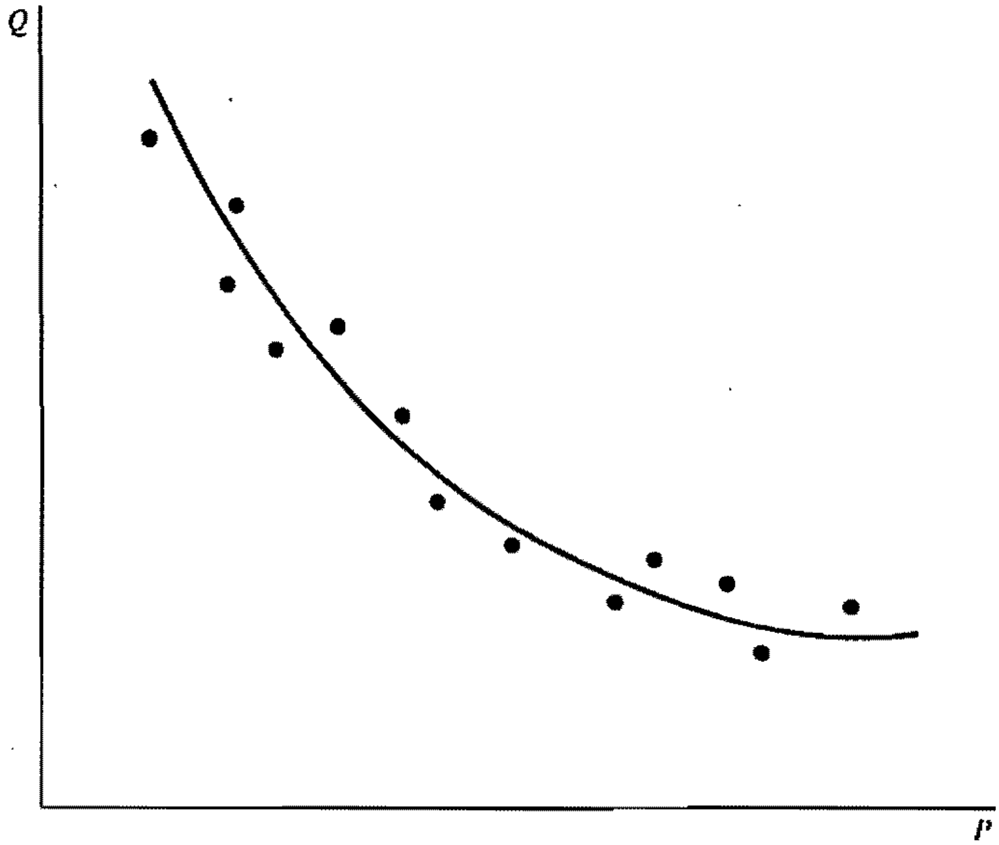
Y es una función lineal en los **parámetros,** es decir en los $\mathbf{B_i}$ pudiendo ser o no lineal en X

En este sentido

1. $$Y = B_0 + B_1 X$$

es lineal en las variables y en los parámetros.

2. $Y = B_o + B_1 X_i^2$



Relación curvilínea.

Es lineal en los parámetros y no lo es en la variable

3. $Y = B_o + \sqrt{B_1}\, X$

No es lineal en los parámetros, y si lo es en la variable.

CON LA EXPRESION " REGRESION LINEAL ",

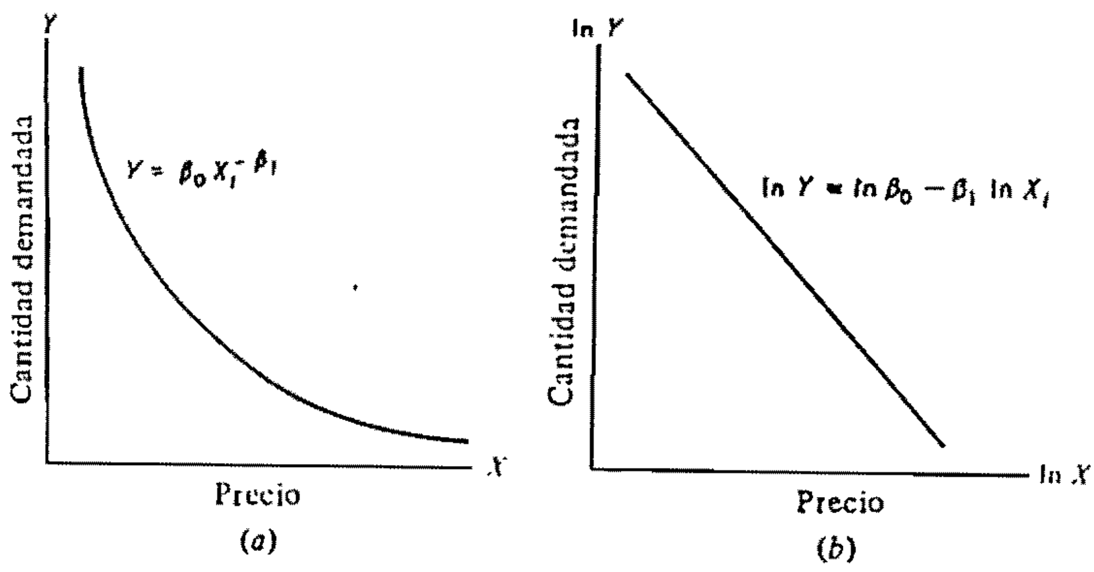
SE HACE REFERENCIA A LINEALIDAD EN LOS PARAMETROS.

FORMAS FUNCIONALES MAS COMUNES

Existen modelos lineales en los parámetros que no lo son en las variables, pero mediante algunas transformaciones pueden volverse lineales en ambos.

Algunos de ellos son:

- <u>Modelos log-log o modelos de elasticidad constante</u>



Modelos de elasticidad constante.

$Y_i = B_o X_i^B \exp(U_i)$

exp = base de los logaritmos naturales = 2.718

$$\ln(Y_i) = \ln B_o + B_1 \ln(X_i) + U_i$$
$$w_i = \alpha + B_1 Z_i + U_i$$

B_1 = elasticidad de Y con respecto a X

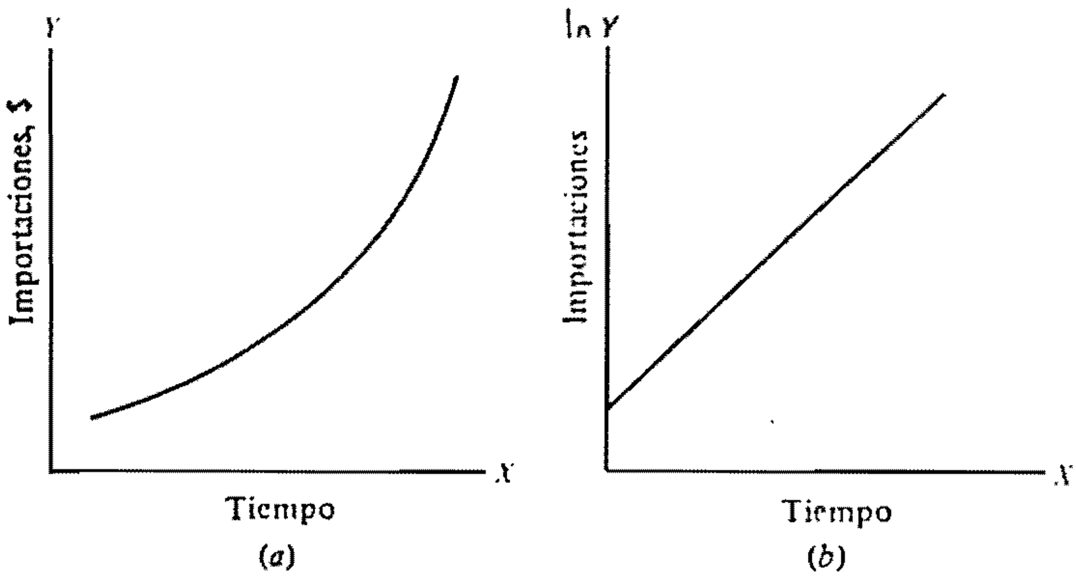
B_1 = cambio porcentual en Y asociado con un cambio porcentual en X

$B_1 = (dy / y) / (dx / x)$

Características: Se supone que la elasticidad no cambia

$B_o = \exp(\alpha)$ resulta sesgado

Modelos semilogarítmicos



Modelo de crecimiento porcentual constante

En ellos la variable X o la variable Y está expresada en términos logarítmicos.

CASO 1

$$\ln Y_i = \alpha_o + \alpha_1 X_i + U_i$$

$$w_i = \alpha_o + \alpha_1 X_i + U_i$$

correspondiente al modelo en variables originales:

$$Y_i = \exp(\alpha_o) \,.\, \exp(\alpha_1 X_i) \,.\, \exp(U_i)$$

$$Y_i = A \,.\, \exp(\alpha_1 X_i) \,.\, \exp(U_i) \qquad A = \exp(\alpha_0)$$

$$Y_i = A \,.\, B^{X_i} \,.\, e^{u} \qquad B = \exp(\alpha_1)$$

CARACTERISTICAS

α_1 = Cambio relativo en Y / cambio absoluto en X

recibe el nombre de modelo de crecimiento constante (por cada unidad X, crece un porcentaje constante).

Elasticidad de Y con respecto a X = $\alpha_1 \cdot \overline{X}$

CASO 2

$$Y_i = B_o + B_1 \ln X_i + U_i$$

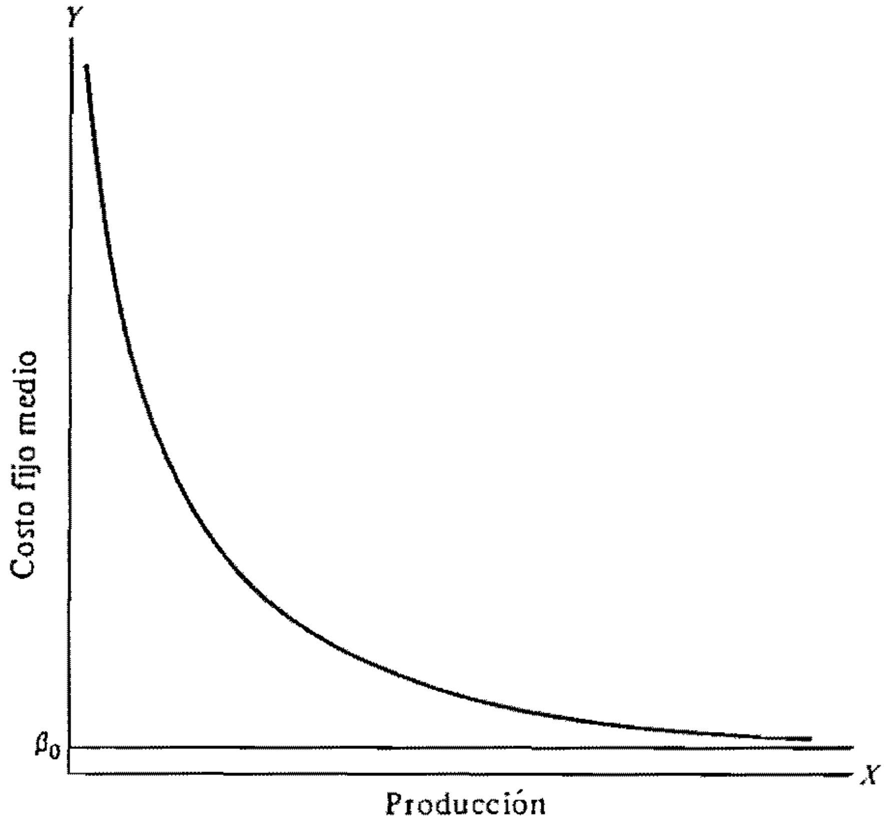
$$Y_i = B_o + B_1 Z_i + U_i$$

CARACTERISTICAS

B_1 = Cambio <u>absoluto</u> en Y / cambio <u>relativo</u> en X

elasticidad de Y con respecto a X = $B_1 / \overline{Y}$

MODELOS INVERSOS O RECIPROCOS



$$Y_i = B_o + B_1 X_i^{-1} + U_i$$

$$Y_i = B_o + B_1 W_i + U_i \qquad W_i = X_i^{-1}$$

CARACTERISTICAS

El valor B_o es la asíntota

elasticidad de Y con respecto a X $= -(\overline{X}\overline{Y})^{-1}$

ALGUNAS FUNCIONES DE USO COMUN Y SUS ELASTICIDADES

	Funciones	Elasticidades
Lineal	$y = a + bx$	$e = bx / y$
Cuadrática	$y = a + bx + cx^2$	$e = (b + 2cx)\, x / y$
Log-log	$\ln y = a + b[\ln x]$	$e = b$
Semilog	$y = a + b[\ln x]$	$e = b / y$
Log inverso	$\ln y = a + b / x$	$e = -b / x$
Inversa	$y = a + b / x$	$e = -b / xy$
Log cuad.	$\ln y = a + b[\ln x] + c[\ln x]^2$	$e = b + 2\, c[\ln x]$
Expon	$\ln y = a + bx$	$e = bx$

MODELOS DE REGRESION LINEAL CON DOS VARIABLES

Existen varios métodos de estimación, pero el más usado es el de mínimos cuadrados ordinarios por sus propiedades estadísticas y facilidad relativa de cálculo.

$$Y_i = B_0 + B_1 X_i + U_i$$

Y_i = valor de la variable de respuesta (dependiente). Valor observable

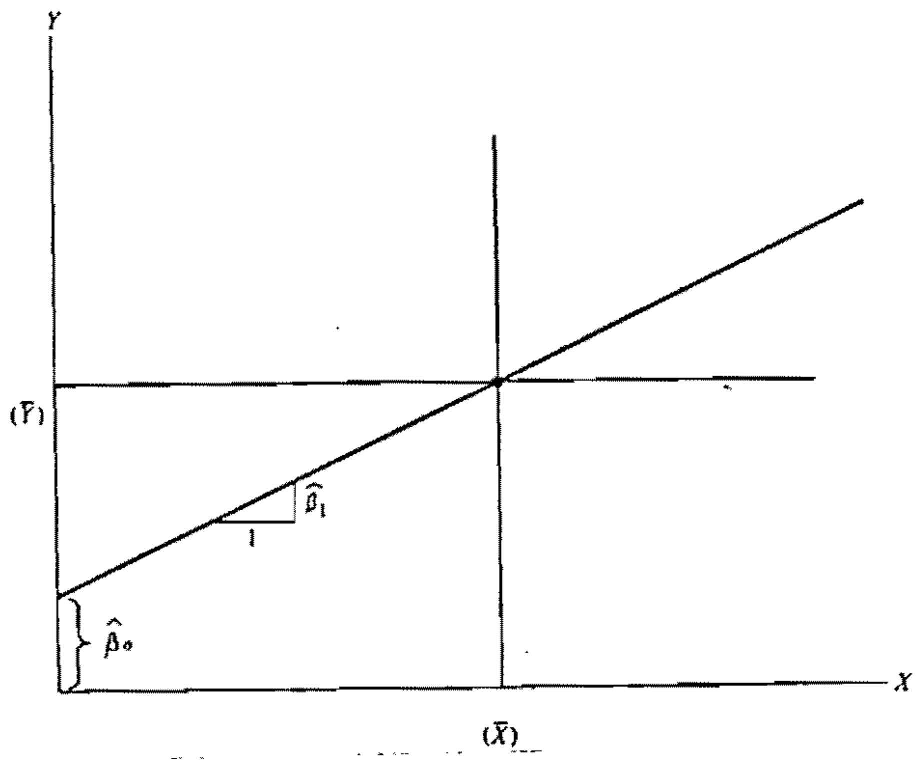
X_i = valor de la variable independiente. Valor observable

U_i = perturbación estocástica o término de error estadístico. Valor no observable
Variable que resume todas las variables omitidas

B_o , B_1 = parámetros desconocidos pero fijos llamados coeficientes de regresión

B_o = intercepto. Refleja el valor promedio de Y, cuando X = 0

B_1 = pendiente. Representa el cambio en Y por cada unidad de incremento en X



OBJETIVO DEL ANALISIS DE REGRESION

Estimar la función de regresión poblacional

$$Y_i = B_o + B_1 X_i + U_i$$

con base en la función de regresión muestral

$$Y_i = B_o + B_1 X_i + e_i$$

donde e_i = error o residuo en la evaluación muestral. Es análogo al concepto de U_i

BASES CONCEPTUALES DEL METODO

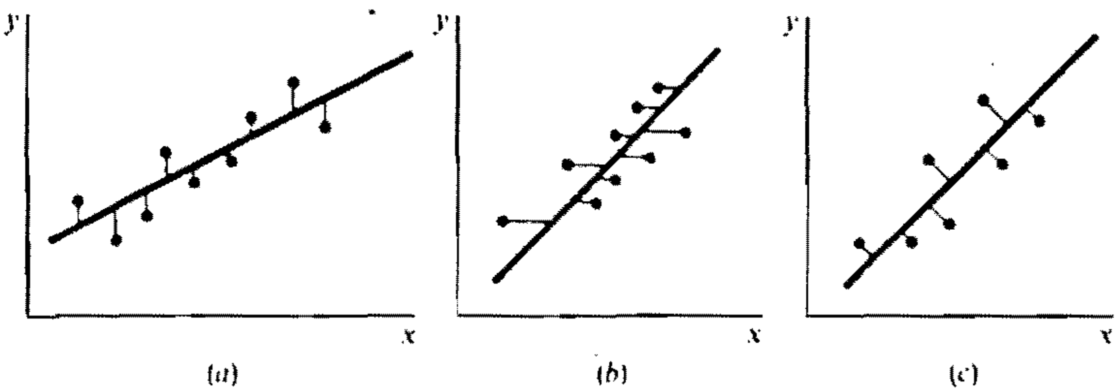
Hemos dicho que:

$$Y_i = \underbrace{\widehat{B}_o + \widehat{B}_1 X_1}_{} + e_i$$

$$Y_i = \widehat{Y}_i + e_i$$

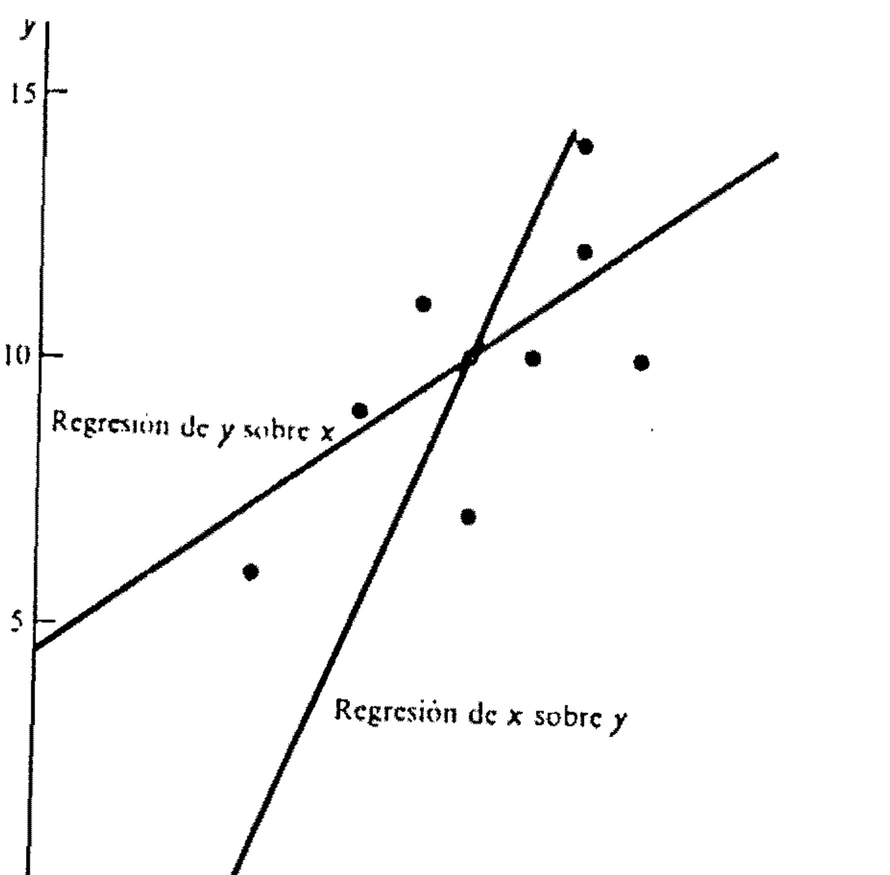
Valor verdadero = valor estimado + error o residuo

error = valor verdadero - valor estimado



(*a*) Regresión de *y* sobre *x*; (*b*) regresión de *x* sobre *y*; (*c*) regresión ortogonal.

DIFERENCIA ENTRE LA REGRESION DE Y SOBRE X
Y
DE X SOBRE Y



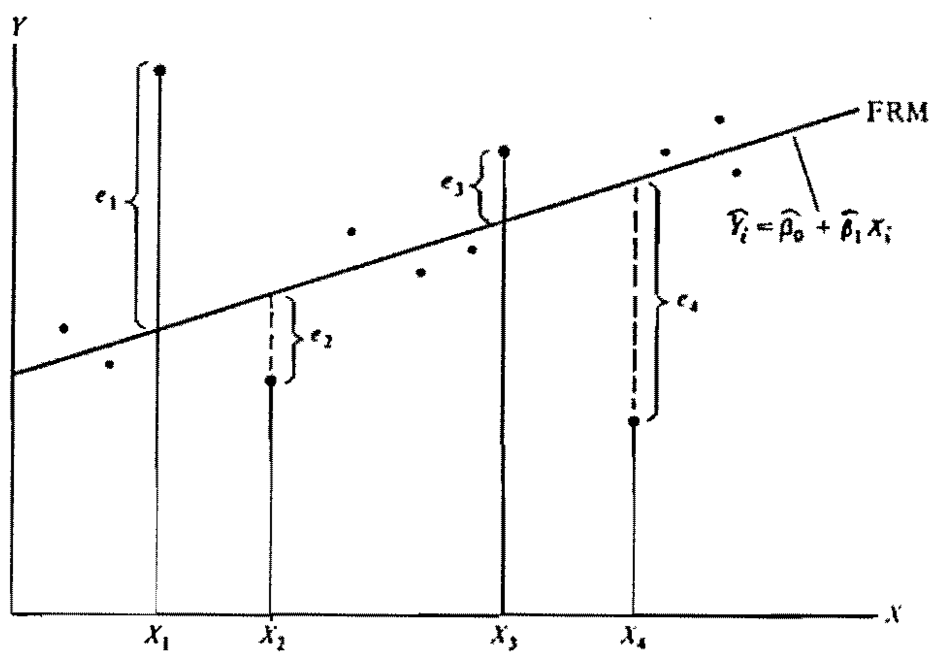
Rectas de regresión de y sobre x y x sobre y.

Objetivo : encontrar un " Valor estimado " que minimice la suma de cuadrados de los errores.

$$\widehat{B}_1 = \Sigma (X_i - \overline{X}) (Y_i - \overline{Y}) / \Sigma (X_i - \overline{X})^2$$

$$\widehat{B}_1 = \text{cov} (X, Y) / S^2_x$$

$$\widehat{B}_o = \overline{Y} - \widehat{B}_1\overline{X}$$



Criterio de los cuadrados mínimos

MODELO DE REGRESION A TRAVES DEL ORIGEN

Si $Y_i = B_1 X_i + U_i$

el intercepto no aparece

En este caso

$$\sigma^2 = \Sigma e_i^2 / (N - 1)$$
$$B_1 = \Sigma X_i Y_i / \Sigma X_i^2$$

Nota : Σe_i no necesariamente $= 0$

Una vez obtenidos los estimadores de los mínimos cuadrados, puede ajustarse la regresión que posee las siguientes propiedades :

1. La línea de regresión pasa por los valores $\overline{X}$, e , $\overline{Y}$

2. El valor promedio de Y estimado ($\overline{\hat{Y}}$) es igual al valor promedio de Y observado ($\overline{Y}$)

3. El promedio de los residuos es 0

4. Los residuos no están correlacionados con el valor predicho de Y_i ($\hat{Y}_i$)

5. Los residuos no están correlacionados con X_i

PRECISION DE LOS ESTIMADORES DE CUADRADOS MINIMOS

Estadísticamente, precisión está asociada con error estandar.

Si $Y = B_o + B_1 X$

$$\sigma^2 = \Sigma e_i^{\,2} / (N-2)$$

$$\Sigma e_i^{\,2} = \Sigma (Y_i - Y)^2 - B_1^2 \ \Sigma (X_i - X)^2$$

$$Var (B_o) = [\sigma^2 \Sigma_i^{\,2}] / (N \Sigma (X_i - X)^2)$$

$$Var (B_1) = \sigma^2 / \Sigma (X_i - X)^2$$

Si $Y = B_1 X$

$$\sigma^2 = \Sigma e_i^2 / (N - 1)$$

$$Var (B_1) = \sigma^2 / \Sigma X_i^2$$

Nota: Σe_i puede ser diferente de 0

MODELO GENERADOR DE LOS DATOS : 1500 + 200*X1 + 3*X2 + 300*U
DONDE U SE HA GENERADO CON RANNOR

OBS	X1	X2	U	Y1
1	20	60	-1.34286	5277.14
2	30	110	-0.52040	7673.88
3	40	80	0.39375	9858.12
4	50	100	0.51144	11953.43
5	60	70	-0.05042	13694.87
6	70	40	-0.02216	15613.35
7	80	80	-0.41586	17615.24
8	10	10	0.18234	3584.70
9	90	50	1.67943	20153.83
10	100	40	-2.47241	20878.28
11	120	30	-0.34063	25487.81
12	130	20	0.34550	27663.65
13	140	40	1.06885	29940.65
14	150	50	-2.34187	30947.44
15	160	10	-0.89694	33260.92
16	170	40	-0.49987	35470.04
17	180	30	-1.87542	37027.37
18	190	20	0.94159	39842.48
19	200	40	-1.46312	41181.06
20	210	60	1.02777	43988.33
21	220	80	0.46616	45879.85
22	230	100	0.72639	48017.92
23	240	90	0.88455	50035.36
24	250	80	0.42881	51868.64
25	260	70	-1.65398	53213.81
26	270	30	-1.51764	55134.71
27	280	50	-2.70201	56839.40
28	290	40	0.33685	59721.05
29	300	20	0.89743	61829.23

MODEL: MODEL1
DEPENDENT VARIABLE: Y1

SOURCE	DF	SQUARES	SQUARE	F VALUE	PROB>F
MODEL	1	8877976146.5	8877976146.5	59535.501	0.0001
ERROR	27	4026259.1395	149120.70887		
C TOTAL	28	8882002405.6			

ROOT MSE	386.16151	R-SQUARE	0.9995
DEP MEAN	32884.57189	ADJ R-SQ	0.9995
C.V.	1.17429		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|
INTERCEP	1	1623.889962	146.82076575	11.060	0.0001
X1	1	199.682770	0.81837544	243.999	0.0001

MODELO GENERADOR DE LOS DATOS : 1500 + 200*X1 + 3*X2 + 300*U
DONDE U SE HA GENERADO CON RANNOR

MODEL: MODEL2
DEPENDENT VARIABLE: Y1

ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	1	122022515.82	122022515.82	0.376	0.5448
ERROR	27	8759979889.8	324443699.62		
C TOTAL	28	8882002405.6			

ROOT MSE	18012.32077	R-SQUARE	0.0137
DEP MEAN	32884.57189	ADJ R-SQ	-0.0228
C.V.	54.77438		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|
INTERCEP	1	36820	7236.1669184	5.088	0.0001
X2	1	-74.103927	120.83448782	-0.613	0.5448

MODEL: MODEL3
DEPENDENT VARIABLE: Y1

ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	2	8878514361.6	4439257180.8	33090.375	0.0001
ERROR	26	3488044.0658	134155.54099		
C TOTAL	28	8882002405.6			

ROOT MSE	366.27250	R-SQUARE	0.9996
DEP MEAN	32884.57189	ADJ R-SQ	0.9996
C.V.	1.11381		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|
INTERCEP	1	1329.819710	202.35703936	6.572	0.0001
X1	1	199.878589	0.78235797	255.482	0.0001
X2	1	4.960401	2.47652740	2.003	0.0557

MODELO 1

PLOT OF R*X1. LEGEND: A = 1 OBS, B = 2 OBS, ETC.

```
         ---+---------+---------+---------+---------+---------+---------+---
    1000 +                                                                 +
         |                                                                 |
R        |                                                                 |
E        |                    A                       A   A A              |
S        |            A                 A         A     A     A         A  |
I        |          A   A                                             A    |
D      0 +    A   A       A A         A                                    +
U        |                          A         A                            |
A        |      A                           A       A           A A        |
L        |                                      A                          |
         |                      A         A                         A      |
         |                                                                 |
   -1000 +                                                                 +
         ---+---------+---------+---------+---------+---------+---------+---
            0        50       100       150       200       250       300
```

X1
INCLUIDA

PLOT OF R*X2. LEGEND: A = 1 OBS, B = 2 OBS, ETC.

```
         ---+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+---
    1000 +                                                                 +
         |                                                                 |
R        |                                                                 |
E        |                          A     A                 A     A        |
S        |        B           A                       B           A        |
I        |                    A                 A     A                    |
D      0 +  A     A           A                       A                 A  +
U        |              A     A                                            |
A        |  A           A     A           A     A                          |
L        |              A                                                  |
         |                    A     B                                      |
         |                                                                 |
   -1000 +                                                                 +
         ---+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+---
           10    20    30    40    50    60    70    80    90   100   110
```

X2
EXCLUIDA

MODELO 2

PLOT OF R*X1. LEGEND: A = 1 OBS, B = 2 OBS, ETC.

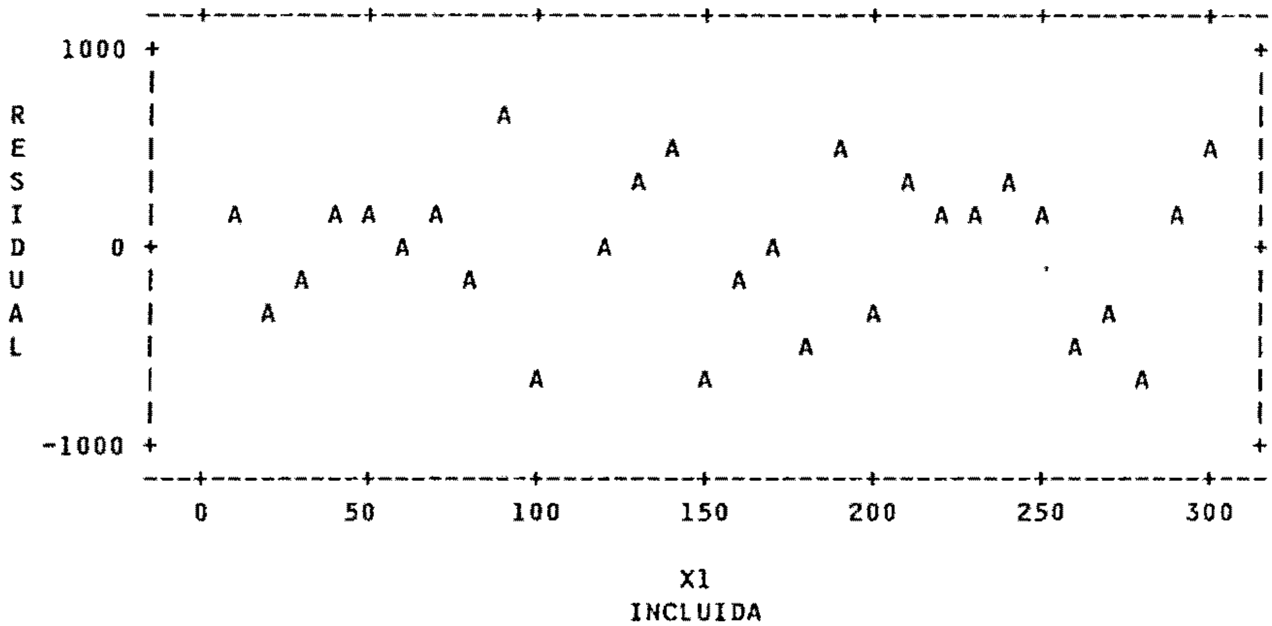
```
          ---+---------+---------+---------+---------+---------+---------+---
    50000 +                                                                 +
          |                                                                 |
 R        |                                                                 |
 E        |                                                    A A   A A A  |
 S        |                                              A A A     A        |
 I        |                                        A A A                    |
 D      0 +                              A A A A A                          +
 U        |                          A A                                    |
 A        |            A A A A A A                                          |
 L        |      A A A                                                      |
          |    A                                                            |
          |                                                                 |
   -50000 +                                                                 +
          ---+---------+---------+---------+---------+---------+---------+---
             0        50       100       150       200       250       300

                                         X1
                                      EXCLUIDA
```

PLOT OF R*X2. LEGEND: A = 1 OBS, B = 2 OBS, ETC.

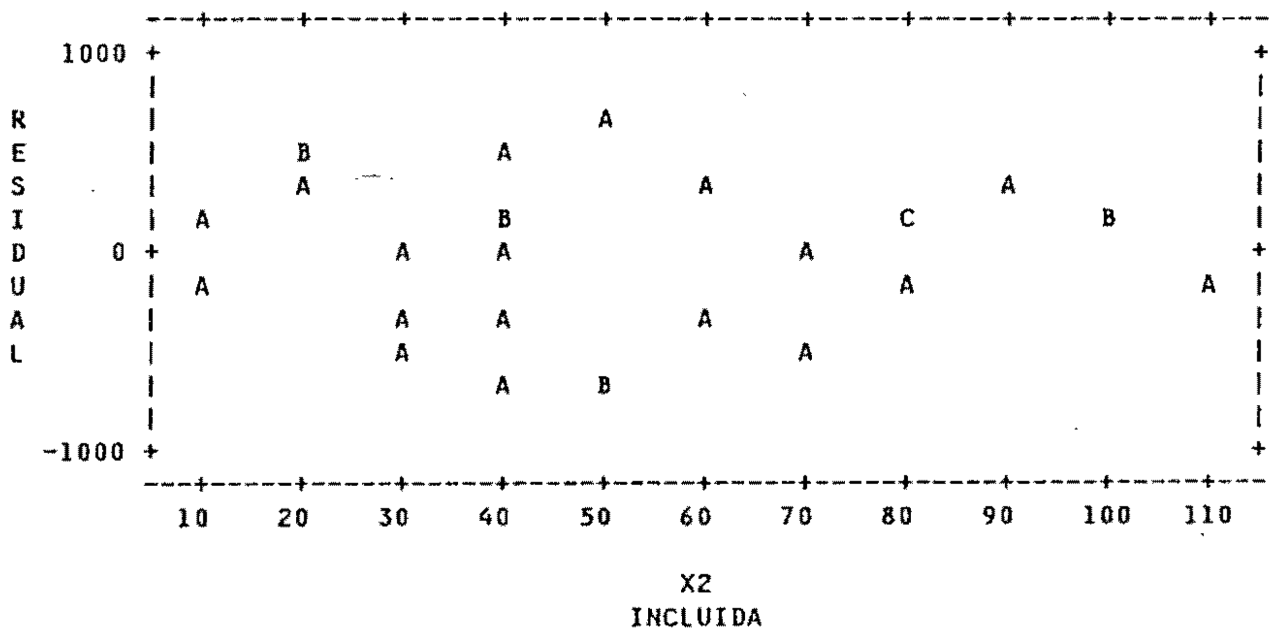
```
          ---+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+---
    50000 +                                                                 +
          |                                                                 |
 R        |                                                                 |
 E        |        A           A     A           A     A                    |
 S        |              A                             A     A     A        |
 I        |        A           A           A                                |
 D      0 +  A           A     B     A                                      +
 U        |        A     A                                                  |
 A        |                    B     A           A     A           A        |
 L        |                                A           A                 A  |
          |  A                                                              |
          |                                                                 |
   -50000 +                                                                 +
          ---+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+---
            10    20    30    40    50    60    70    80    90   100   110

                                         X2
                                      INCLUIDA
```

MODELO 3

PLOT OF R*X1. LEGEND: A = 1 OBS, B = 2 OBS, ETC.

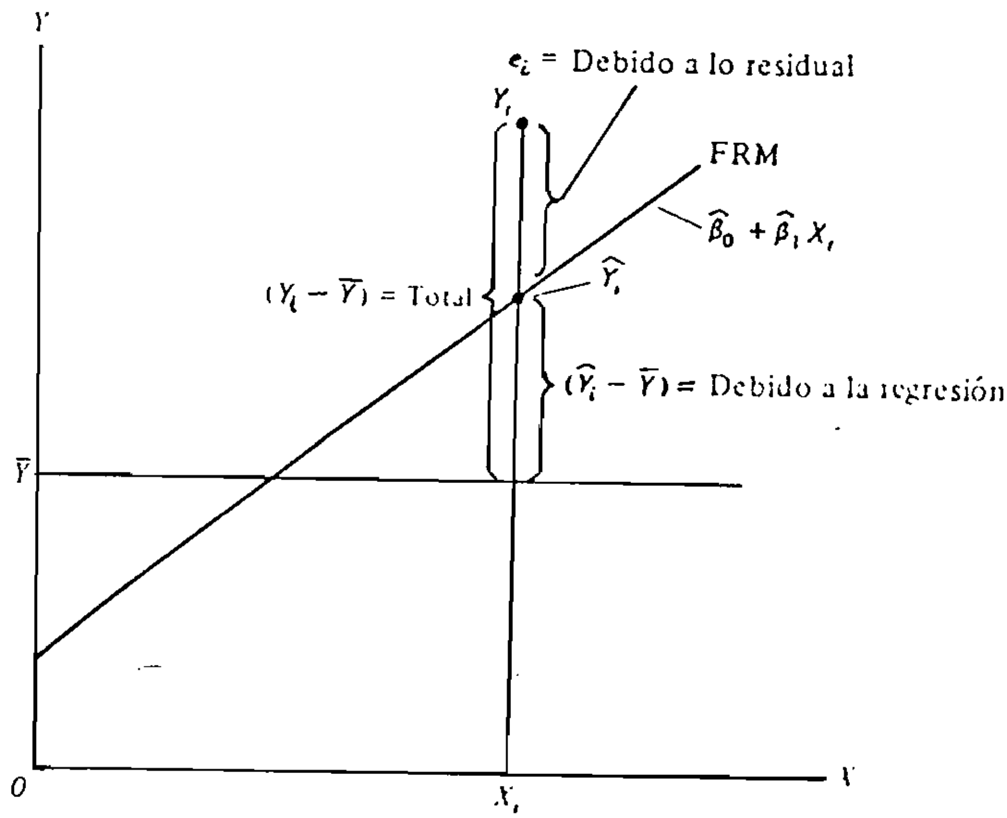


PLOT OF R*X2. LEGEND: A = 1 OBS, B = 2 OBS, ETC.



CALIDAD DEL AJUSTE

Se refiere a una medida que refleje en qué medida se ajusta la línea de regresión muestral a los datos.



Partición de los componentes de variación de Y_i.

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

error (no explicada) — regresión (explicada)

suma total de cuadrados = suma de cuadrados de los residuos + suma de cuadrados explicada por la regresión

Al porcentaje de la variación total explicado por la regresión se le considera un indicador de la bondad o calidad del ajuste.

Se le designa por R^2 y recibe el nombre de coeficiente de determinación.

Características

1. $0 \leq R^2 \leq 1$

   $R^2 = 1$: ajuste perfecto

   $R^2 = 0$: no hay relación entre la variable dependiente y las explicatorias

2. R^2 está numéricamente relacionada con r (coeficiente de correlación) pero conceptualmente son diferentes.

   R^2 = (coeficiente de correlación)^2

3. $R^2 = \hat{B}_1^2 \; S_x^2 / S^2_y$

   $R^2 = 1 - \Sigma\, e_i^2 / \Sigma\, (y_i - \overline{Y})^2$

Algunas consideraciones adicionales sobre r (coeficiente de correlación)

1. $-1 \leq r \leq 1$

2. Correlación (x, y) = correlación (y, x)

3. Independencia de origen y escala, esto es:

   $z = a + bx$

   $w = c + dy$

   $r(z, w) = r(x, y)$

   Si a, c son constantes y b, d > 0

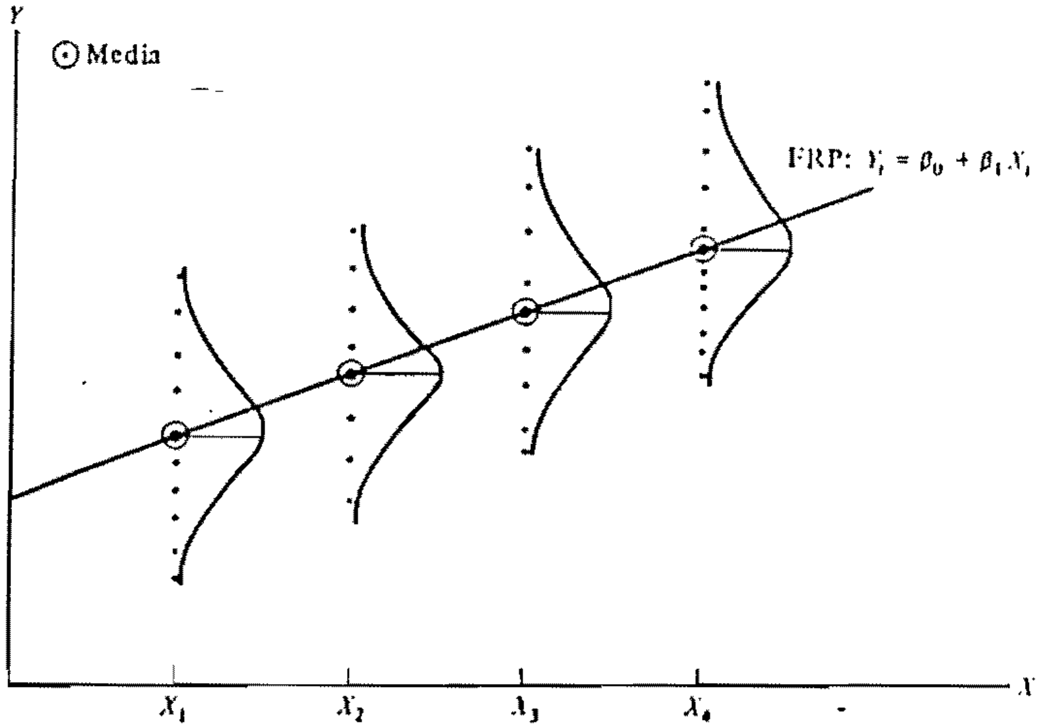
4. r es una medida de asociación lineal entre las variables x, y. No tiene sentido utilizarlo para medir relaciones no lineales.

5. Si x es independiente de y estadisticamente entonces $r(x, y) = 0$ pero si $r(x, y) = 0$ No implica independencia necesariamente, pues puede tratarse de una relación no lineal entre ellas.

6. No implica causalidad

SUPUESTOS SOBRE LOS CUALES SE APOYA EL METODO DE REGRESION POR MINIMOS CUADRADOS

SUPUESTO 1



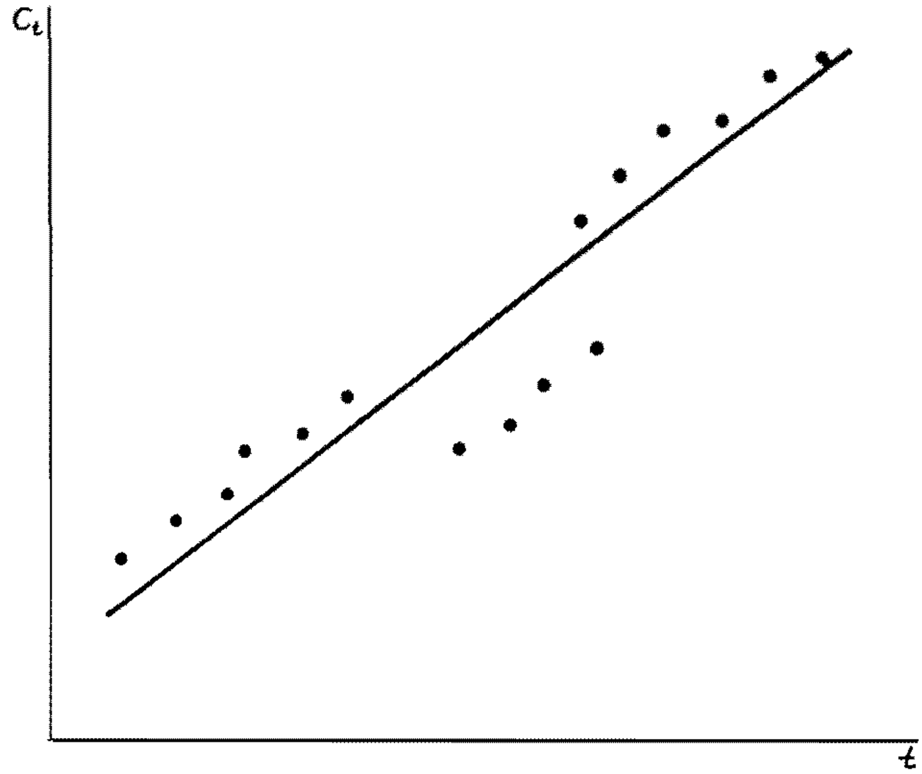
Distribución condicional de las perturbaciones u_i.

$E(U_i) = 0$

Para cada valor de X se presenta una población de Y distribuida al rededor de su valor medio.

Las distancias de estos valores de Y al promedio son los U_i . Por estar localizados encima y debajo de él, se espera que el promedio de los U_i sea = 0

SUPUESTO 2

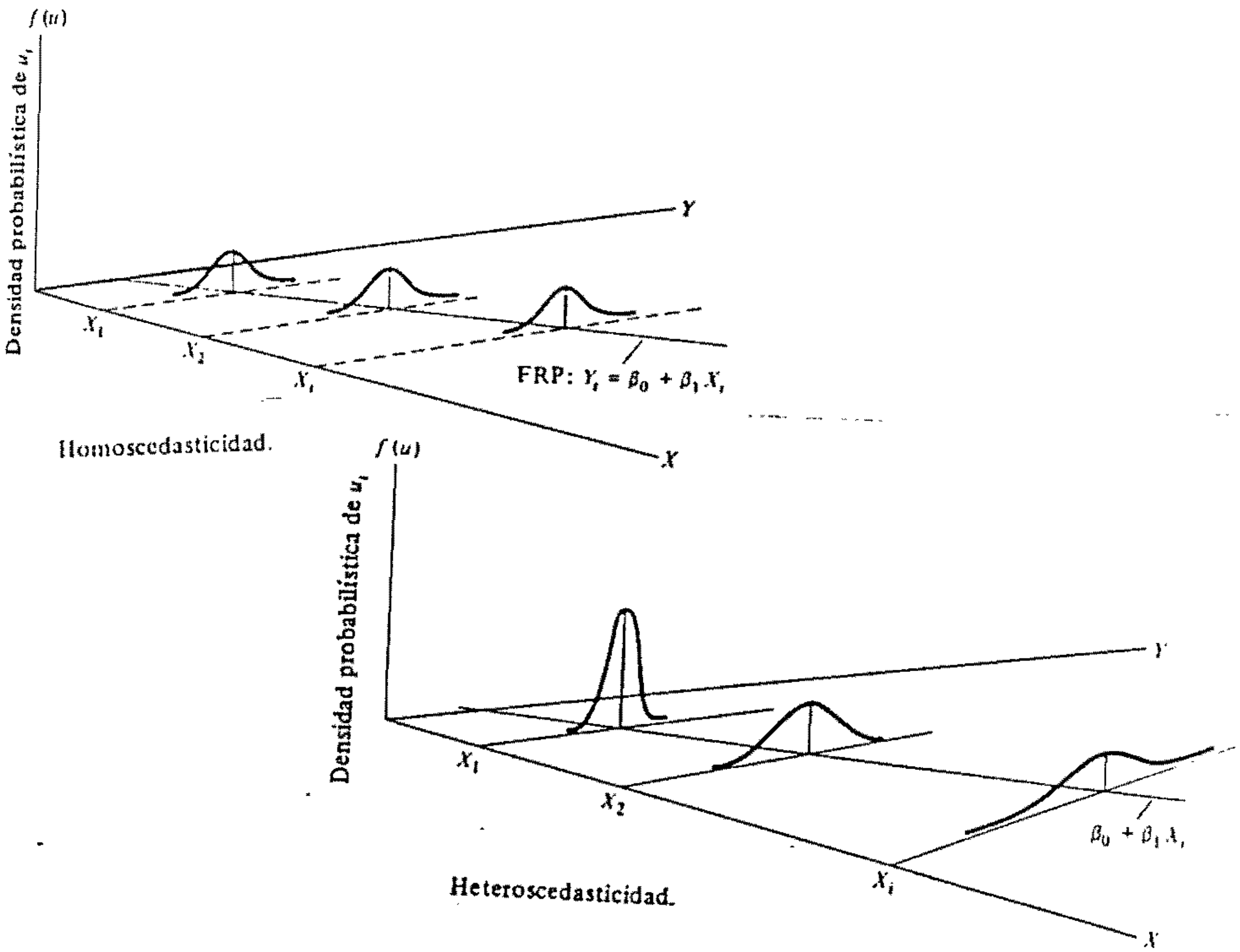


$Cov\ (U_i, U_j) = 0$ para $i \neq J$

No hay correlación entre las perturbaciones U_i, U_j

En caso de que las perturbaciones sigan un patrón sistemático hay violación del supuesto.

SUPUESTO 3



Para cada X se presentan perturbaciones que tienen promedio = 0 (supuesto 1), y una varianza que debe ser la misma para todos los valores de X (Supuesto 3)

Técnicamente, este supuesto recibe el nombre de Homoscedasticidad. Si no se cumple, se dice que hay Heteroscedasticidad.

SUPUESTO 4

$\mathrm{Cov}\,(U_i, X_i) = 0$

La perturbación U, y la variable X no están correlacionadas

Si lo están, no es posible establecer efecto individual sobre Y

Este supuesto se cumple automáticamente si X es una variable no aleatoria y se cumple el supuesto 1

Si X es aleatoria, la validez de este supuesto permite aplicar todos los resultados de la teoría de regresión.

TEOREMA DE GAUSS-MARKOV

Dados los supuestos del modelo de regresión lineal, los estimadores de cuadrados mínimos, en la clase de los estimadores lineales insesgados tienen varianza mínima. Es decir son los mejores estimadores lineales insesgados. (MELI)

VIOLACION DE SUPUESTOS

DIAGNOSTICO, IMPLICACIONES Y MEDIDAS REMEDIALES

VALIDACION DE SUPUESTOS

No encontrar suficiente evidencia para rechazarlos

METODOS

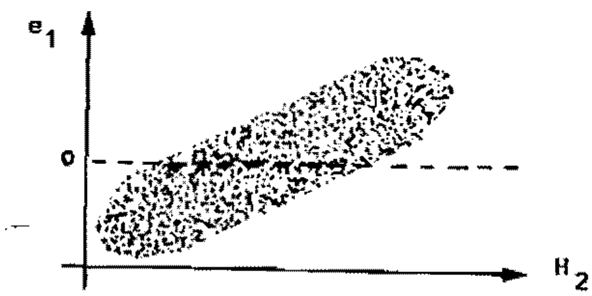
gráficos
análiticos

Los métodos gráficos sugieren formas particulares de incumplimiento del supuesto, los análiticos evalúan su importancia.

SUPUESTO 1

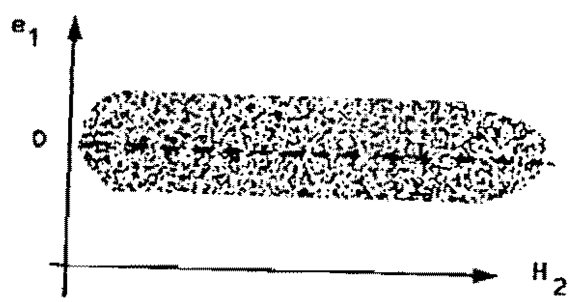
INCORRECTA ESPECIFICACION

CAUSA 1. Se excluyen variables relevantes



Variable no incluida en el modelo

Variable importante. La " pendiente " de la gráfica de la importancia de la variable excluida.



Variable no incluida en el modelo

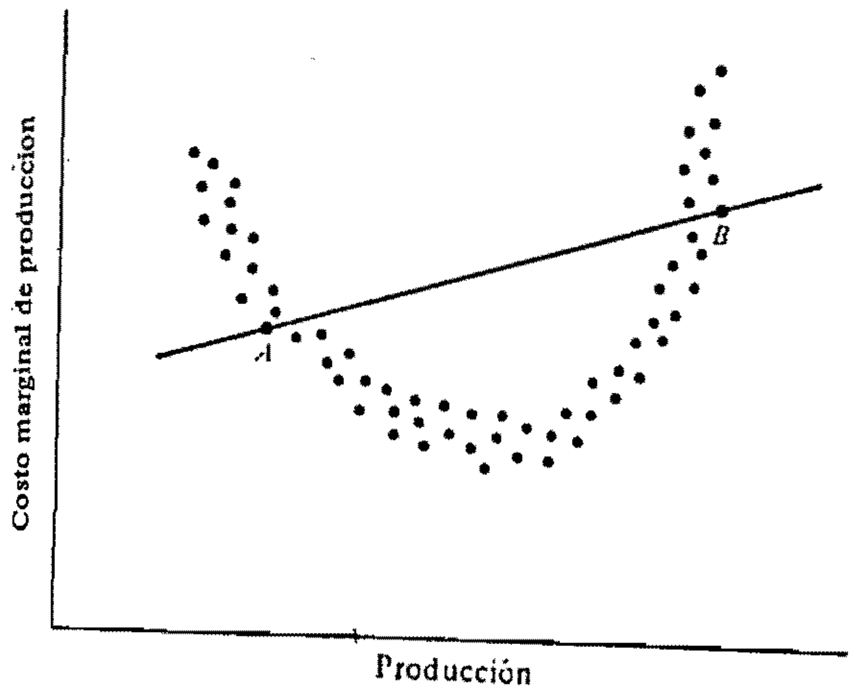
Variable no importante

Efecto: Hay sesgo en B_i proporcional a la importancia de las variables excluídas.

CAUSA 2. Se incluyen variables irrevelantes

Efecto : No hay sesgo, pero la varianza se incrementa (ineficiencia) por lo tanto la potencia de las pruebas de hipótesis se reduce o puede llegar a la incapacidad de aplicarlas.

CAUSA 3. Están las variables adecuadas, pero la forma funcional no lo es.



Sesgo de especificación, forma funcional incorrecta

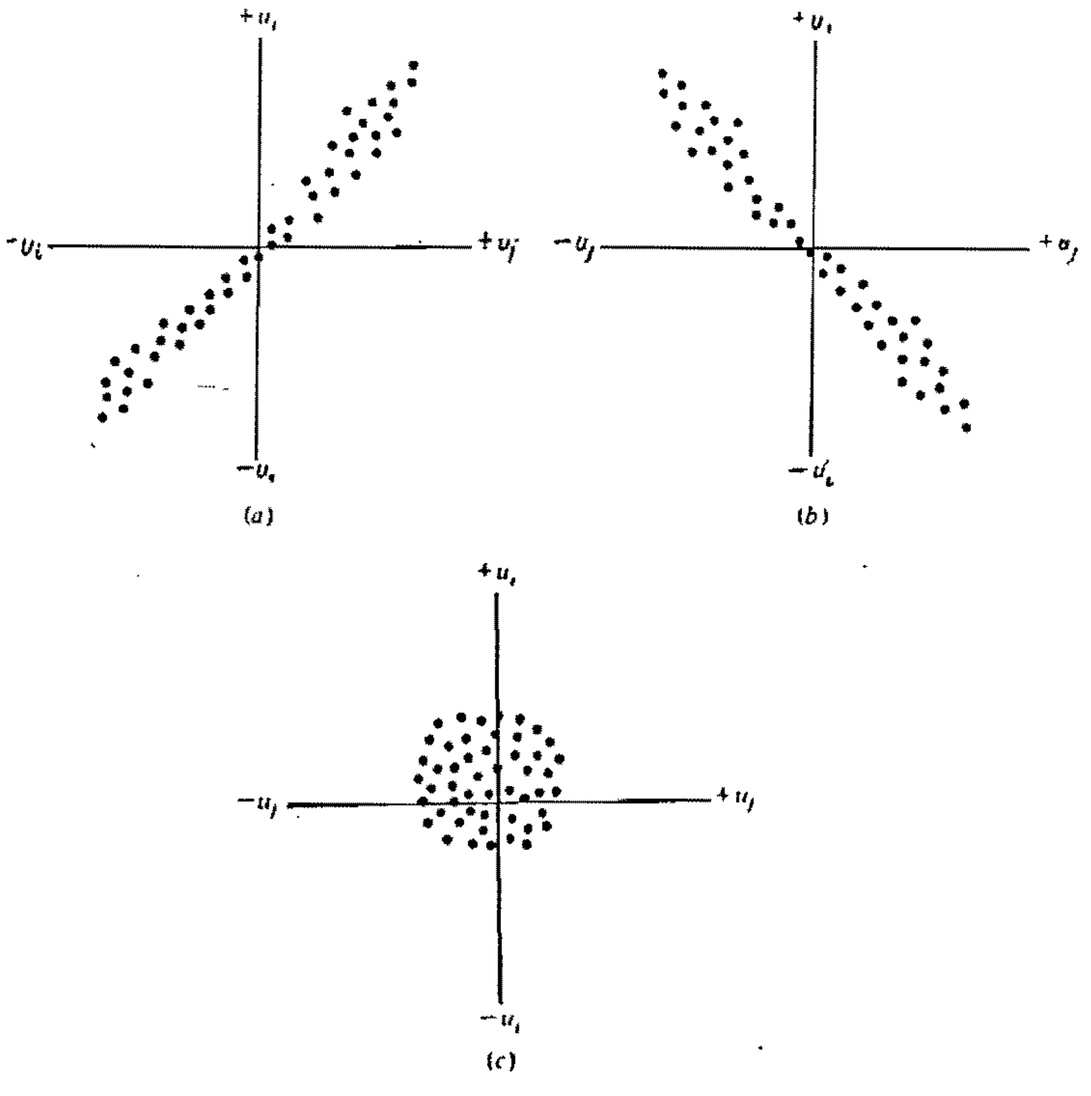
El patrón de los residuos orienta sobre la forma funcional adecuada.

MEDIDAS CORRECTIVAS

Busque mejorar el modelo o plantear otro.

SUPUESTO 2

La perturbación de una observación no debe afectar la de otra.



Patrones de correlación entre las perturbaciones. (*a*) Correlación serial positiva. (*b*) Correlación serial negativa. (*c*) Correlación cero.

CAUSAS DE AUTOCORRELACION

- Impulso o tendencias en los datos principalmente si se trata de series de tiempo.

- Costumbre que se están liberando o adoptando

- Violaciones al supuesto 1

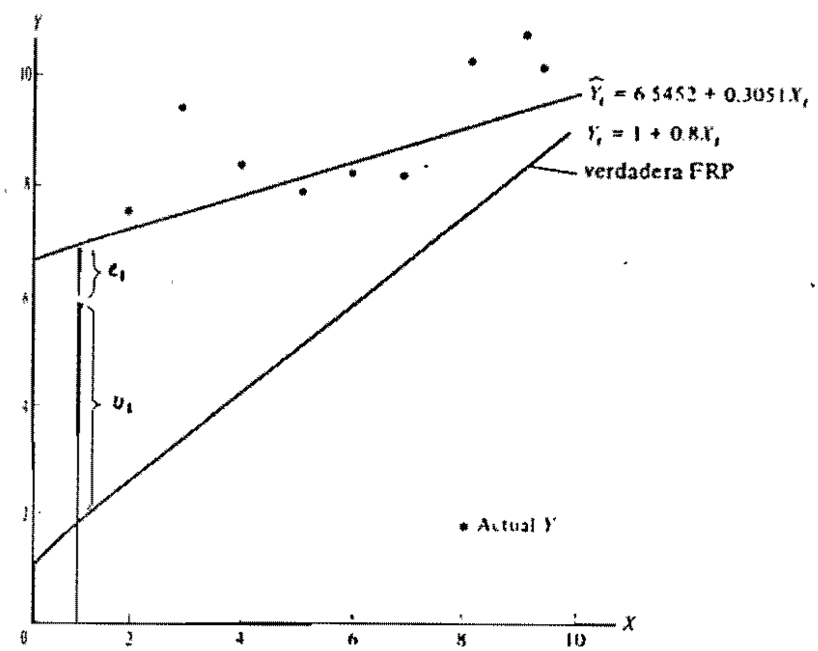
- Hacer regresiones con datos calculados

EFECTOS

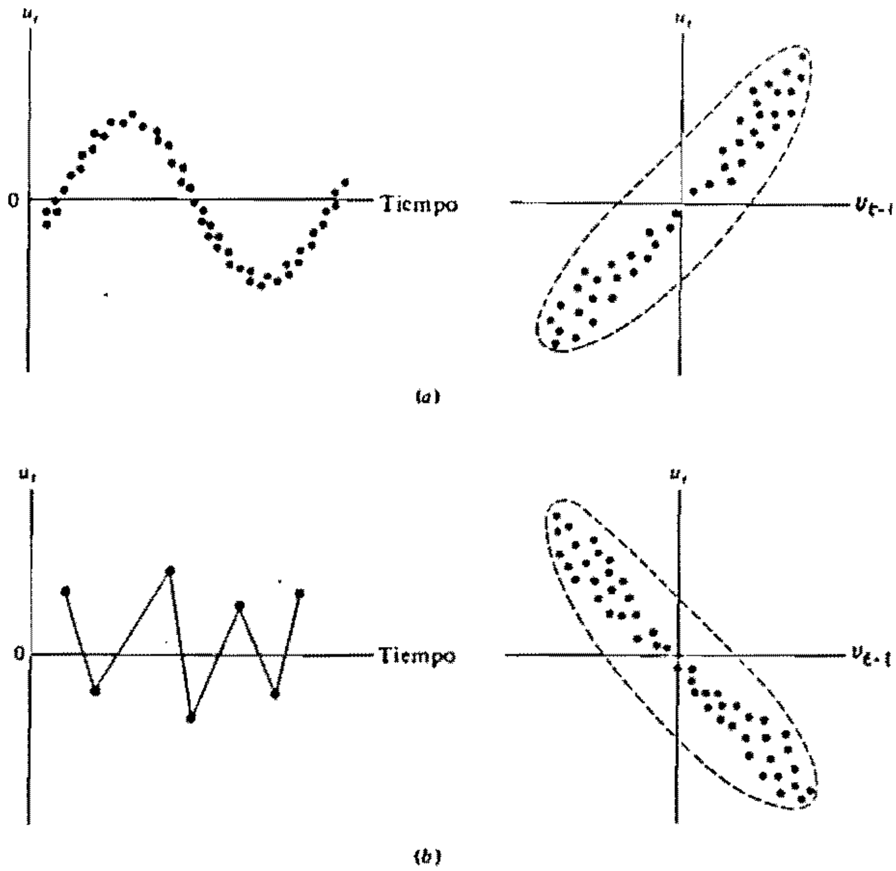
- Los estimadores son ineficientes, es decir ya no tienen varianza mínima.

- La varianza es subestimada, lo mismo que las varianzas de los estimadores, por lo tanto la inferencia que se haga puede ser no válida.

- Los estimadores son insesgados pero se vuelven muy sensibles a fluctuaciones muestrales.



METODO GRAFICO PARA EL DIAGNOSTICO



Autocorrelación (*a*) positiva y (*b*) negativa.

METODO ANALITICO

Prueba de Durbin-Watson : d

$$d = \Sigma\ (e_t\ -\ e_{t-1})^2\ /\ \Sigma\ e_t^2$$

exige :

- Mínimo número de observaciones : 15
- Considerar siempre el intercepto aunque la regresión deba pasar por el origen
- Los valores de X deben ser fijos en muestras repetidas o no ser estocásticas
- Parte del supuesto de que la estructura de autocorrelación es

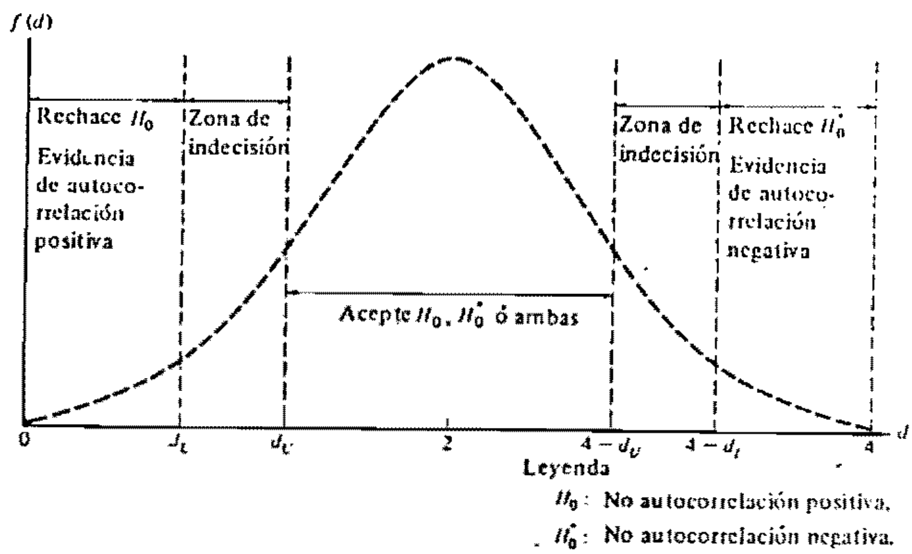
$$U_t = rU_{t-1} + e_t$$

o sea un esquema **AUTORREGRESIVO DE PRIMER ORDEN (AR1)**
- El modelo no incluye valores rezagados de Y como variable independiente.

CALCULO Y USO DE " d "

1. Haga el análisis de regresión y obtenga e_i

2. Calcule "d" y con K= número de observaciones calcular los valores críticos.



Estadístico Durbin-Watson.

PRUEBA DE RACHAS

1. Haga el análisis de regresión y obtenga e_t

2. Construya la siguiente tabla

	+ t	- t
+(t - 1)	n_1	n_2
-(t - 1)	n_3	n_4

n_1 : número de veces en que el residuo en t-1, y t era positivo

n_2 : número de veces en que el residuo en t-1 era positivo y en t negativo

n_3 : número de veces en que el residuo en t-1 era negativo y en t positivo

n_4 : número de veces en que el residuo en t-1 y t era negativo

Se hace una prueba X^2 de independencia para la tabla anterior.

MEDIDAS REMEDIALES

1. Método de las primeras diferencias

Calcular

$$Y_t = Y_t - Y_{t-1}$$

$$X_t = X_t - X_{t-1}$$

hacer la regresión:

$$Y_t = B_l X_t + e_t$$

la cual <u>no</u> tiene intercepto

SOLO es útil si r = 1

2. Ecuación de diferencias generalizadas

$$r = N^2 (1 - d / 2) + K^2 / (N^2 - K^2)$$

d = Durbin - Watson

K = número de coeficientes, incluyendo el intercepto

N = número total de observaciones

PASOS

1. Los primeros valores de X, Y se multiplican por $\sqrt{1 - r^2}$

2. Calcule

$$Y = Y_t - r\, Y_{t-1}$$
$$X = X_t - r\, X_{t-1}$$

$$Y = B_0 (1 - r) + B_1\, X + e_t$$

- Estimado de r por el método de Cochrane-Orcutt

Se supone AR1

$r = \Sigma\, e_t\, e_{t-1} \;/\; \Sigma\, e^2_{t-1}$

con este valor se hace la ecuación de deficiencia generalizada y se generan nuevos valores e. Con ellos se vuelve a calcular r y nuevamente la ecuación de diferencia generalizada.

El proceso iterativo termina cuando el valor de r se estabilice.

Nota 1

Si hay razones para pensar que la correlación no es de primer orden, sino de orden 2, 3....n, el proceso puede adaptarse al orden supuesto.

(Calcular DW_n hacer n-esimas diferencias)

Nota 2

Usualmente algunos casos dan ecuaciones con R^2 altos si se hacen las regresiones con valores originales y al hacerlos con las diferencias R^2 disminuye.

No olvidar que el primero está construido sobre una base falsa, y que el segundo puede ser más real.

En resumen:

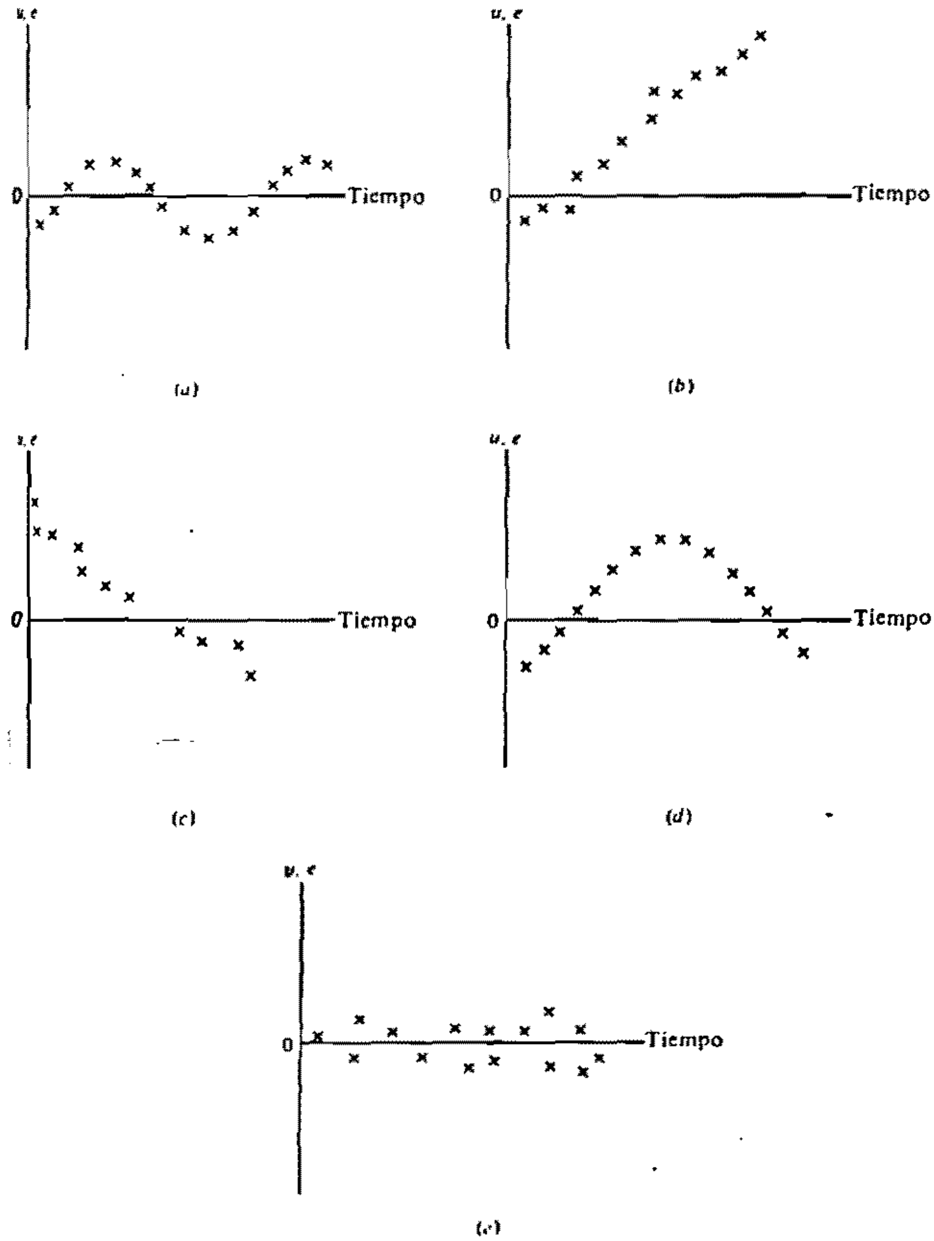
R^2 alto y $d = 0$ en valores originales

R^2 bajo y $d = 2$ en diferencias

=> R^2 en valores originales es espureo

OTROS PATRONES DE AUTOCORRELACION

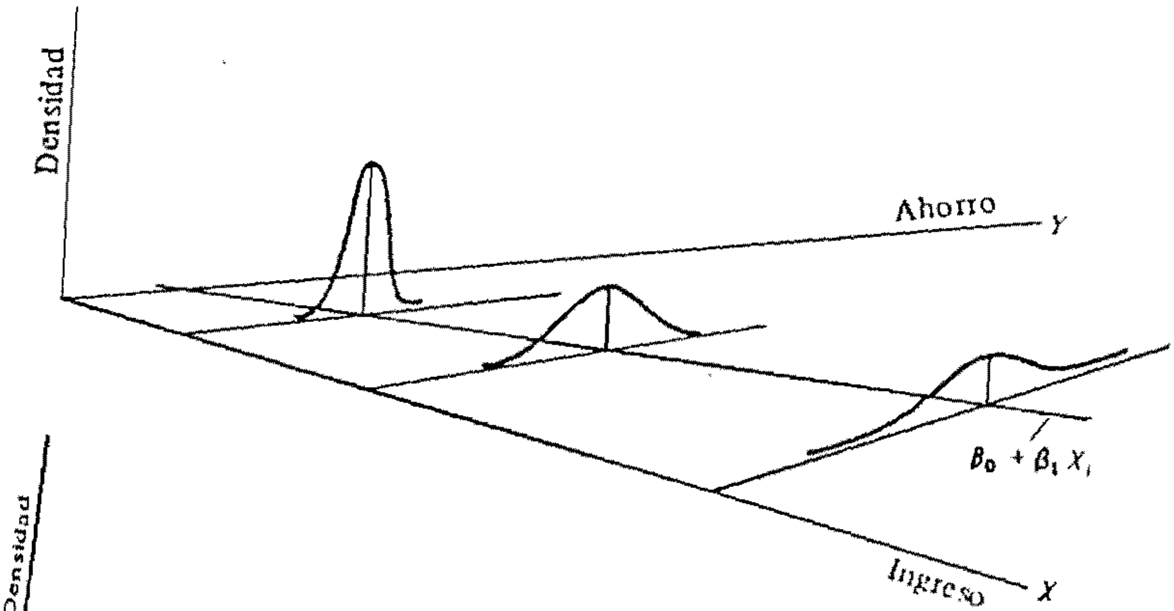
AUTOCORRELACION

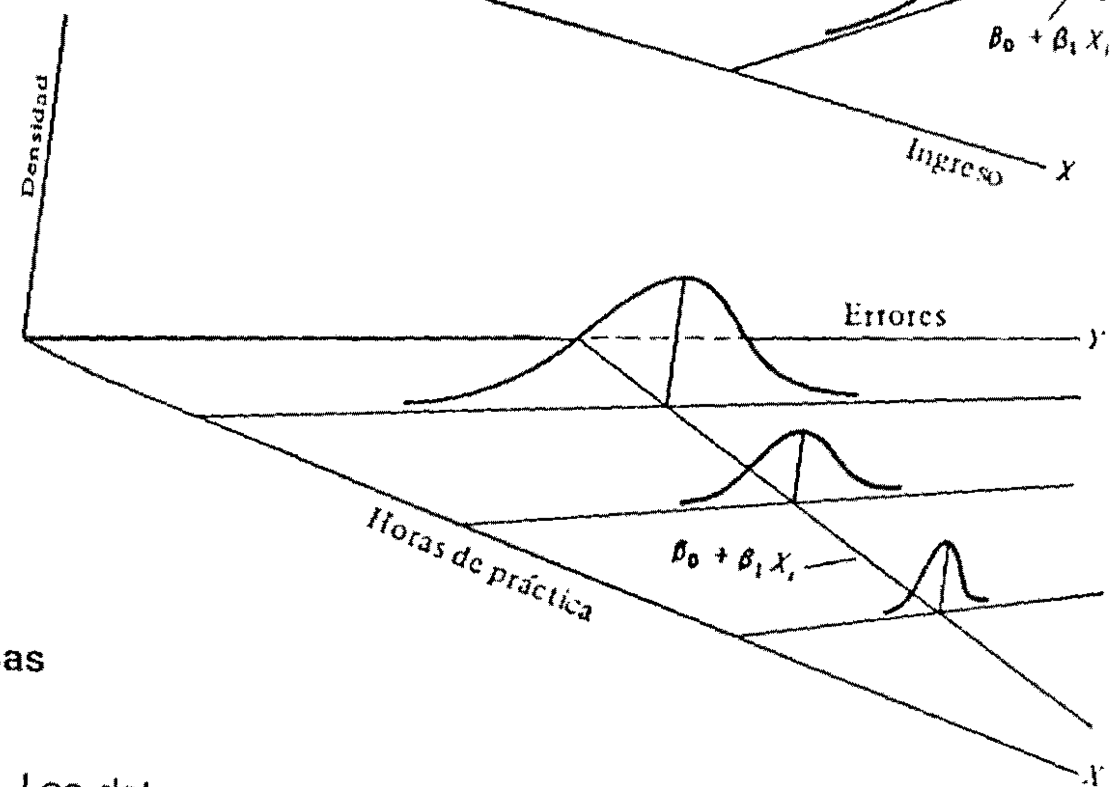


Patrones de autocorrelación

SUPUESTO 3

Heterogeneidad de varianza





Causas

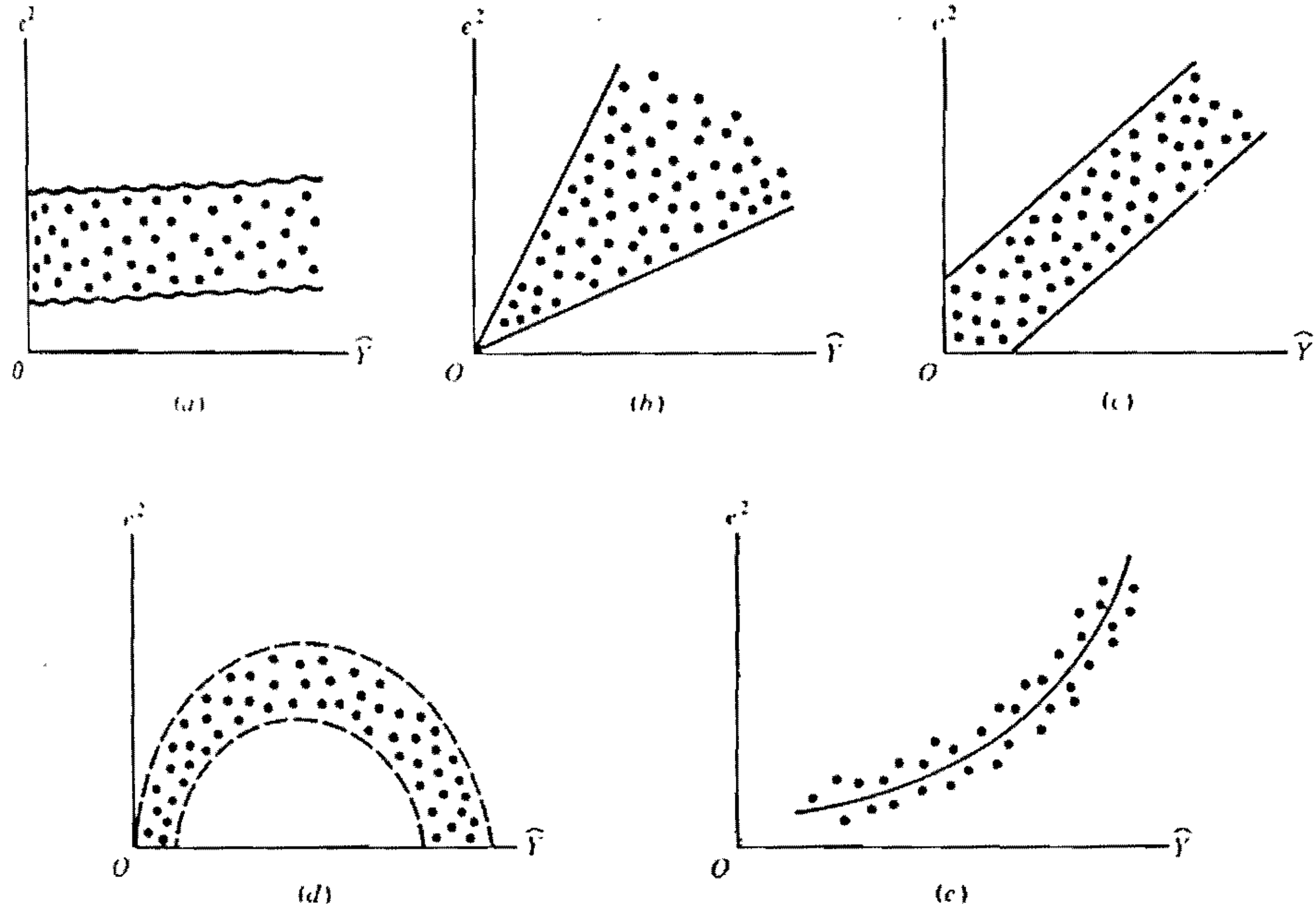
1. Los datos <u>son</u> así
2. Los procesos de medición o colección mejoran

EFECTO

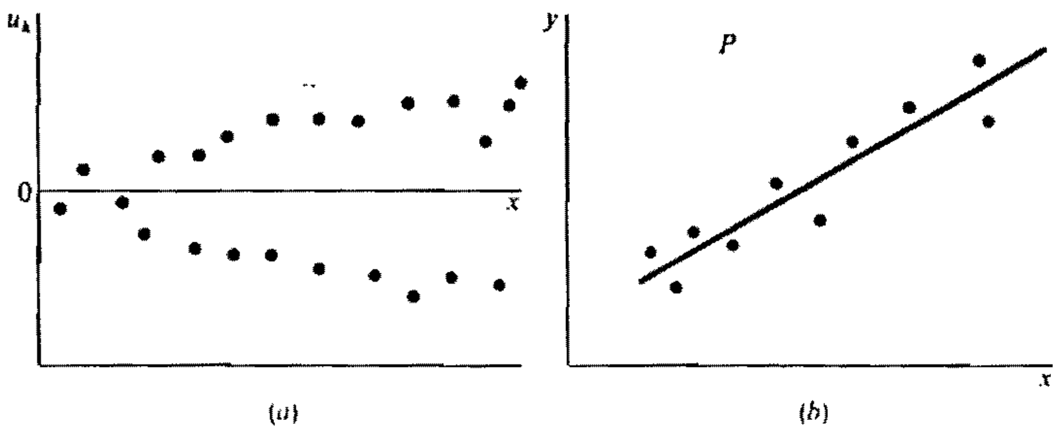
Los estimadores se mantienen insesgados, consistentes, pero aunque el tamaño de muestra aumente son ineficientes (varianza alta).

La estimación de σ^2 es falsa y por lo tanto la de B_t resulta subestimada. Esto significa " falsos significativos ", e intervalos de confianza equivocados.

METODOS GRAFICOS PARA EL DIAGNOSTICO



Patrones hipotéticos de residuos estimados al cuadrado.



(*a*) Residuos heterocedásticos; (*b*) un caso de observación atípica.

ANALITICOS

- PRUEBA DE PARK

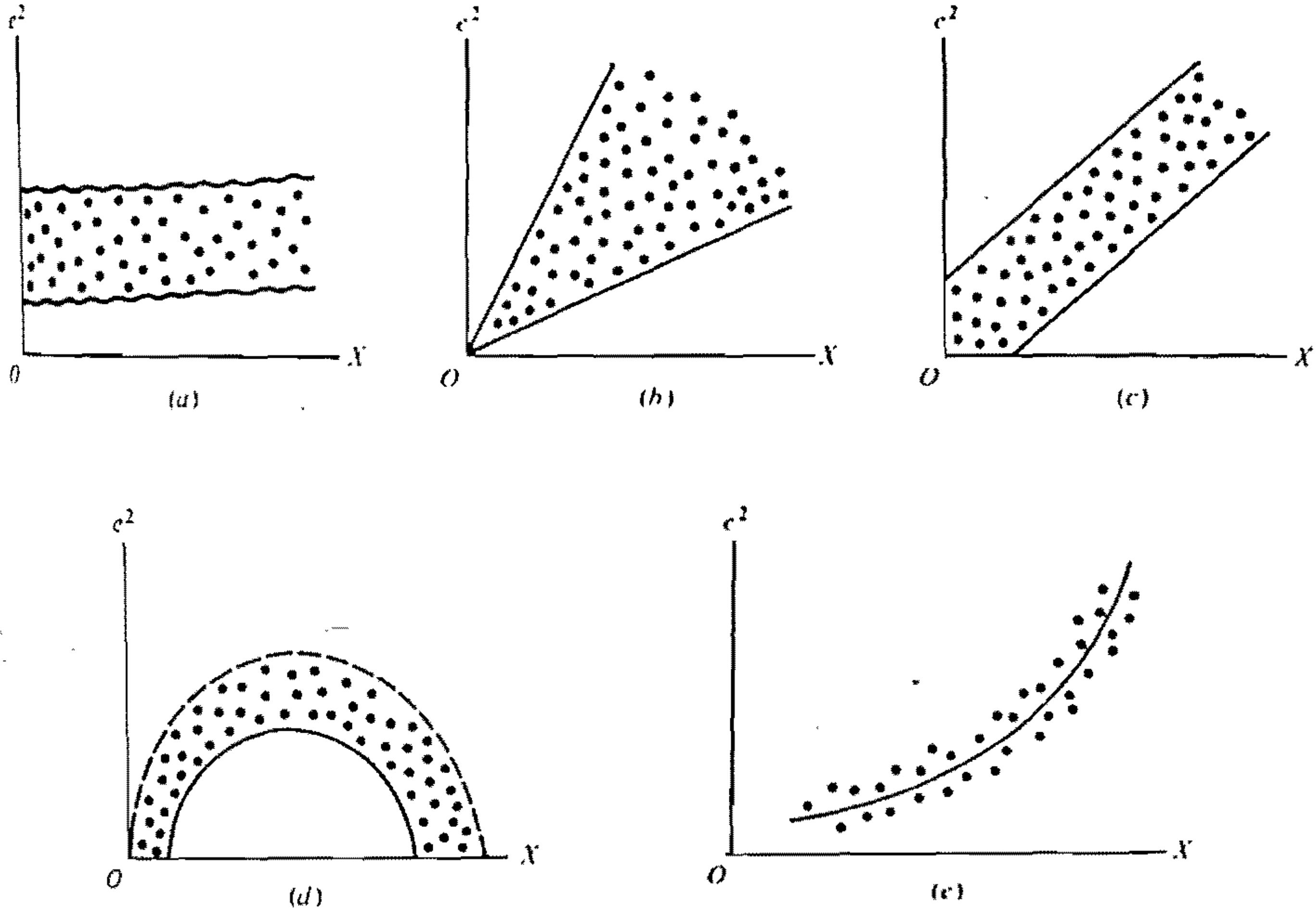


Diagrama de los residuos estimados al cuadrado, contra X.

$$\ln e^2 = \alpha + B \ln X_i + V_i$$

Si $B \neq 0$ hay heteroscedasticidad

- **Correlación de rango de Spearman**

$r_1 = 1 - 6 \text{ ABS } (\Sigma\, d_i^2 / N (N^2 - 1))$

d_i = rango X_i - rango e_i

N = número de individuos

Calcular $t_c = (r_s \sqrt{N - 2}) / \sqrt{1 - r_s^2}$

con N-2 grados de libertad

Si este valor de $t_c > t$ entonces hay heteroscedasticidad

- Pruebas de Bartlett, Cochran y Hartley

MEDIDAS REMEDIALES

- **Mínimos cuadrados ponderados**

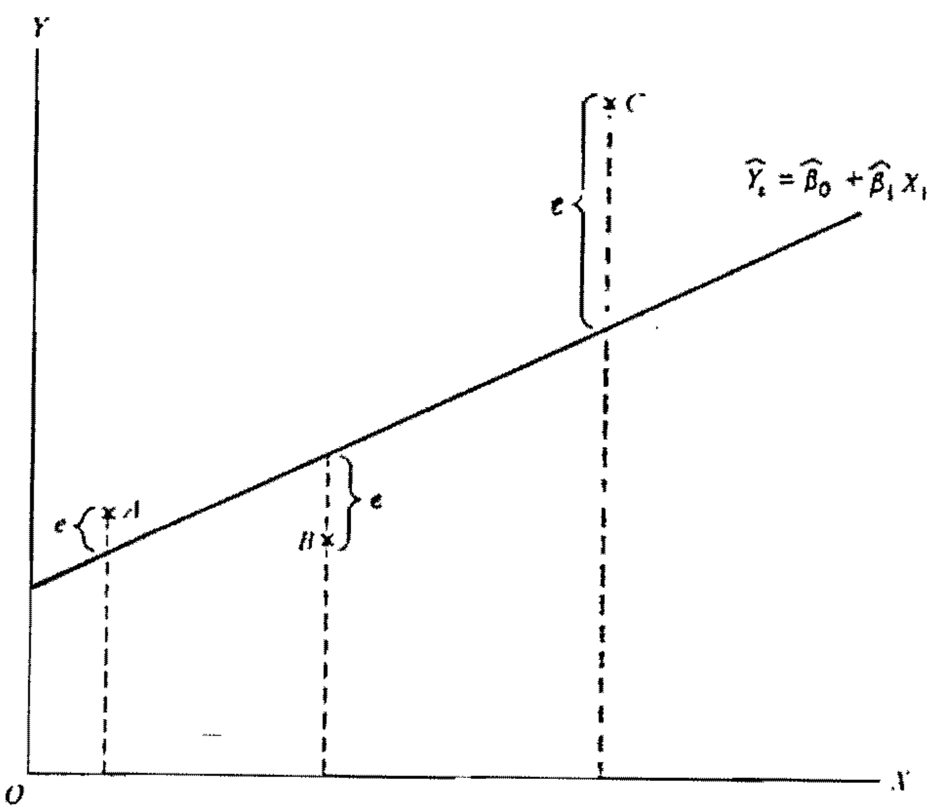


Diagrama hipotético.

Minimiza la importancia de las observaciones extremas ponderandolas por el enverso de su varianza

$$Y^* = \Sigma\ w_i\ y_i\ /\ \Sigma\ w_i$$

$$X^* = \Sigma\ w_i\ y_i\ /\ \Sigma\ w_i$$

$$B_1 = \Sigma\ w_i\ (\ y_i - y^*\ )\ (x_i - x^*\ )\ /\ \Sigma\ w_i\ (x_i - x^*\ )^2$$

$$B_o = y^* - B_1\ x^*$$

La precisión se calcula a partir de :

$$\mathrm{var}\ (\ B_1\ ) = \Sigma\ w_i\ /\ [(\Sigma\ w_i)(\Sigma\ w_i\ x_i^2) - (\Sigma\ w_i\ x_i)^2]$$

- **Transformacion logaritmica** (log-log)

  Reduce considerablemente el problema debido a los cambios de escala que realiza.

- **Uso de deflactores**
  Calcular variables, dividiendo las originales por alguna medida de " tamaño ".

  Puede ser convertir los valores en unidad de área, por persona, por unidad de producción, etc.

SUPUESTO 4

Falla frecuentemente en el caso de sistemas de ecuaciones produciendo estimadores sesgados e inconsistentes o sea, que el sesgo no desaparece aún cuando el tamaño de muestra crezca indefinidamente.

SUPUESTO ADICIONAL : SUPUESTO DE NORMALIDAD :

Antes :

$\bar{U} = 0$, varianza constante, perturbaciones no correlacionadas

Si el objetivo es estimación puntual : suficiente

Si lo es hacer inferencias y obtener intervalos de confianza entonces el supuesto adicional

$U_i \sim N(0, \sigma^2)$

es de gran importancia

En estas condiciones los estimadores de mínimos cuadrados son:

- insesgados
- tienen varianza mínima (eficientes), sean los estimadores lineales o no
- Si aumenta el tamaño de la muestra convergen al valor poblacional (consistentes)
- $Var(\hat{B}_o) = \sigma^2 (\Sigma X_i^2) / (N \Sigma (X_i - \bar{X})^2)$

  $\hat{B}_o \sim N(B_o, Var(\hat{B}_o))$

- $Var(\hat{B}_1) = \sigma^2 / \Sigma (X_i - \bar{X})^2$

  $\hat{B}_1 \sim N(B_1, Var(\hat{B}_1))$

y como siempre : $\hat{\sigma}^2 = \Sigma e_i^2 / (N - 2)$

Bajo el supuesto adicional de normalidad de los errores podemos establecer la formulación general siguiente:

$$\Pr\left(- t_{\alpha/2} \leq t \leq t_{\alpha/2} \right) = 1 - \alpha$$

$t_{\alpha/2}$ = valor de t para nivel de significancia $\alpha/2$ y N-2 grados de libertad

$$\Pr\left[-t_{\alpha/2} \leq \frac{\hat{B}_1 - B_1}{es\,(\hat{B}_1)} \leq t_{\alpha/2} \right] = 1 - \alpha$$

$$\Pr\left[\hat{B}_1 - t_{\alpha/2}\ es(\hat{B}_1) \leq B_1 \leq \hat{B}_1 + t_{\alpha/2}\ es(\hat{B}_1) \right) = 1 - \alpha$$

PRUEBAS DE HIPOTESIS DE REGRESION LINEAL SIMPLE

Bajo el supuesto de normalidad, la variable

$$t = (\hat{B}_1 - B_1) / Es (\hat{B}_1)$$

sigue una distribución t con N-2 grados de libertad

La anteior constituye una prueba de hipótesis sobre los coeficientes de regresión.

Si en una regresión lineal simple, $B_1 = 0$ esto es :

La variable explicatoria no tiene influencia lineal en Y y toda la variación de Y se explica por los errores aleatorios U_i , se espera que

$$\frac{\text{SC. debida a la regresión}}{\text{grados de libertad} = 1} = \frac{\text{S.C. residuos}}{\text{grados de libertad} = \text{N-2}}$$

$$= \hat{\sigma}^2$$

Además, si $B_1 \neq 0$, parte de la variación puede atribuirse a X, por lo tanto la igualdad no se cumple.

Por lo tanto, la comparación de estos valores habla de la significancia de la regresión.

$$F = \frac{\text{suma de cuadrados explicada por la regresión}}{\text{suma de cuadrados de residuos / (N-2)}}$$

sigue una distribución F con 1 y N-2 grados de libertad

NOTA : El cuadrado de t con N-K grados de libertad es un valor F con 1 y N-K grados de libertad. Por lo tanto, en una regresión lineal simple estos dos valores son un solo concepto.

PREDICCION MEDIA

Si a X se le asigna un valor, con la ayuda de los coeficientes de regresión podemos hacer una predicción puntual del valor de Y.

El valor medio de una predicción es

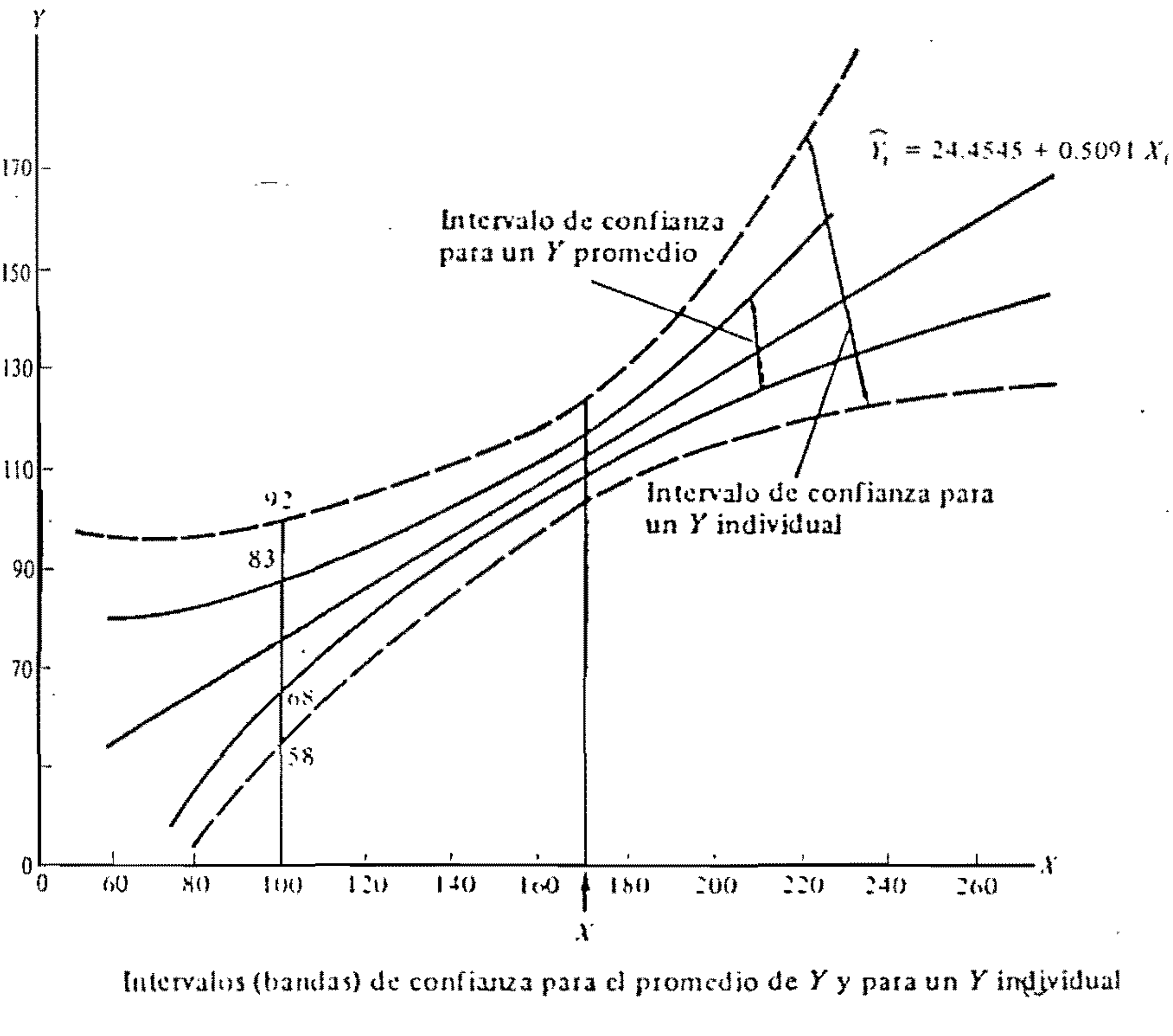
$$Y_o = B_o + B_1 X_o$$

y su varianza

$$V(Y_o) = \alpha^2 (N^{-1} + (X_o - \overline{X})^2 \ / \ \Sigma (X_i - \overline{X})^2)$$

regidos por una distribución " t " con N-2 grados de libertad.

Con estos elementos la construcción de intervalos de confianza es directa.



Intervalos (bandas) de confianza para el promedio de Y y para un Y individual

REGRESION LINEAL MULTIPLE

Generalizando lo visto anteriormente

$$Y_i = B_o + B_1 X_{1i} + B_2 X_{2i} + ... + B_n X_{ni} + U_i$$

B_o = intercepto

B_i = coeficientes parciales de regresión

mide el cambio en el valor de Y por cambio de una unidad en X_i manteniendo las demás constantes.

ESTIMACION

OBJETIVOS

Encontrar los valores de los parámetros desconocidos de tal manera que la suma de residuos al cuadrado sea tan pequeña como sea posible.

Bajo este enfoque, las fórmulas de cálculo y las propiedades de los estimadores son una extensión del caso de la regresión lineal simple.

CALIDAD DEL AJUSTE

Constituye básicamente una extensión del caso de la regresión lineal simple.

R^2 o coeficiente de determinación múltiple representa la proporción de la variabilidad de Y explicada conjuntamente por todas las X_i

R o coeficiente de correlación múltiple dará una medida de asociación lineal entre Y y el conjunto de las X_i

R es un concepto de poca importancia, no así R^2

OBSERVACION

R^2 es una función no decreciente del número de variables explicatorias, por lo tanto suele calcularse el valor

$\bar{R}^2$ = R^2 ajustado

$$\bar{R}^2 = 1 - [\ \Sigma\ e_i^2 / (N-K) / \Sigma\ (y_i - \bar{Y})^2 / (N-1)\ ]$$

K = número de parámetros incluyendo intercepto

N = número de datos

RELACION IMPORTANTE

$$\overline{R}^2 = 1 - (1 - R^2)(N - 1) / (N - K)$$

USO :

$\overline{R}^2$ es útil para comparaciones, en ese caso Y debe ser la misma.

NOTAS ADICIONALES SOBRE R^2

Si se requiere comparar R^2 de un modelo lineal (original) con el de uno linealizado (transformado) calcular así:

$$R^2_o = 1 - \Sigma (y - Y)^2 / \Sigma (y - Y)^2$$

para el transformado:

Generar los valores W y aplicar la transformación inversa para restituirlo a las unidades originales. Con estos valores calcular R^2_t

SUPUESTOS

Los mismos que en el caso de regresión lineal simple (1 a 4) y el supuesto de normalidad de U_i

SUPUESTO 5

No colinealidad

o sea:

no existe relación lineal exacta entre las variables explicatorias

No existen $L_1, L_2 \ldots . L_n \neq 0$

tales que

$$L_1 X_1 + L_2 X_2 + \ldots .. + L_n X_n = 0$$

Si la única forma posible es que $L_1 = L_2 = \ldots .. = L_n = 0$ entonces son linealmente independientes.

Nota: Se refiere solo a relación lineal

VIOLACION DEL SUPUESTO

Si la colinealidad es perfecta los estimadores son indeterminados (0/0) y sus varianzas indefinidas ($\alpha^2 / 0 = \infty$)

Si la colinealidad es alta, los estimadores pueden obtenerse, pero la varianza de ellos aumenta proporcionalmente al grado de colinealidad.

Esto causa dificultades para la estimación de intervalos de coanfianza y pruebas de hipótesis.

Los coeficientes de regresión y sus errores estándar se vuelven muy sensibles a pequeños cambios.

En presencia de multicolinealidad alta, se vuelve casi imposible separar los efectos de las variables.

Si el objeto de la regresión es la predicción, la multicolinearidad, no es problema serio, pues mientras más alto sea R^2 mejor es la predicción, siempre y cuando haya argumentos sólidos para pensar que esa situación persiste en el futuro.

Si el objeto es la estimación de parámetros confiables, entonces si es problema grave.

METODOS PARA DETECTARLA

1. Valor alto de R^2

Ninguno o muy pocos coeficientes significativos (según prueba de t)
Aunque valores bajos de correlación (X_i X_j)

NOTA : Si hay solo dos variables explicatorias, la correlación entre X_1 y X_2 si sirve como indicador de colinealidad.

Si hay más de dos variables explicatorias, la correlación no es una buena guia para detectar multicolinearidad.

MEDIDAS REMEDIALES

A veces se utiliza como solución, la eliminación de variables que estén causando problemas. Esta solución a veces es peor que la enfermedad, pues puede caerse en sesgo de espicificación y con ello en estimaciones sesgadas de los coeficientes.

Hay otras soluciones:

Regresión con componentes ortogonales y regresión de borde, que suelen mitigar el efecto del problema.

INFERENCIA ESTADISTICA EN REGRESION LINEAL MULTIPLE

Por extensión del caso de la regresión lineal simple pueden construirse intervalos de confianza y efectuarse pruebas de hipótesis.

El número de grados de libertad = N - K

La prueba F, mencionada anteriormente constituye una forma de verificar significancia global

$H_o : B_1 = B_2 = ... B_n = 0$

su nueva formulación es

F = (N - K) s. de c. explicada / (k-1) s. de c. de residuos

Y tiene una distribución F con K-1 y N-K grados de libertad

K = número de parámetros a estimar, incluido el intercepto

RELACION ENTRE F, R^2 y t

Si $R^2 = 0$ $F = 0$

Si $R^2 = 1$ F tiende a infinito

Si R^2 aumenta F aumenta

F prueba global de la regresión de una prueba para R^2

$\bar{R}^2$ aumenta al incluir una variable si ella tiene un valor de t (en valor absoluto) mayor que 1

R^2	B_i	
1. signifi.	todos signific.	Si los signos son correctos posiblemente ya terminó su trabajo.
2. signifi.	algunos signif.	Frecuentemente, sobre todo si el número de variables es grande. Se eliminan? Cuáles eliminar?
3. signifi.	ninguno	Frecuente. Posible multicolinearidad.
4. no	todos signif.	Escaso. Suelo presentarse si las correlaciones (Y_t, X_i) son bajas y correl $(X_i, X_j) < 0$
5. no	algunos	Buscar eliminar las variables no significativas y revisar qué puede faltar por incluir
6. no	ninguno	No insista

RESUMEN SOBRE VIOLACION DE SUPUESTOS

Violación de supuesto	Efecto
1. Incorrecta especificación Omisión Inclusión de variables irrelevantes	Sesgo Ineficiencia
2. Cov (U_i, U_j) = 0	Ineficiencia, inestabilidad Subestimación de las varianzas
3. Homoscedasticidad	Ineficiencia, estimación falsa de las varianzas
4. Cov (U_i, X_i) = 0	Sesgo Inconsistencia
5. Colinearidad	Varianza alta Sensibilidad

BEHAR, G.R., 1990, Métodos para la validación de los supuestos del modelo de regresión. En Memorias del I Simposio de Estadística-Análisis de Regresión. Mayo 31-Junio 2 de 1990. Universidad Nacional de Colombia.

DRAPER, N.R. And H. SMITH, 1981. "Applied regression Analysis" John Wiley and sons. New York.

FLURY B. And H. RIEDWYL, 1988, Multivariate Statistics, a practical approach. Chapman and Hall. London-New York. 126 pp.

GUJARATI, D., 1981. Econometría Básica Mc Graw-Hill Latinoamericana S.A. Bogotá, Colombia.

HABER A, R.P. RUNYON, 1972. Estadística General. Fondo Educativo Interamericano S.A. Bogotá. 371 pp.

MADDALA, G.S., 1985. Econometría Mc Graw-Hill Latinoamericana S.A. Bogotá, Colombia.

PINDYCK R.S. y D.L. RUBINFELD, 1976. Econometric Models and Economic Forecasts. Mc Graw-Hill Koga Kusha Ltd. Tokyo, Japón.

SHAO, S.P., 1975. Estadística para economistas y administradores de empresas. Henero Hermanos, sucs., S.A. Mexico. 786 pp.

STEEL R.G.D., J.H. TORRIE, 1985. Bioestadística : Principios y Procedimientos 2a Ed., 1a en español, Mc Graw-Hill Latinoamericana S.A. Bogotá, Colombia. 622 pp.

REFERENCIAS BIBLIOGRAFICAS

1. ABRAHAMSE, A.P.L. And A. S. LOUTER. (1971). On a New Test for autocorrelation in least squares regression. BIOMETRIKA, 58, 53-60.
2. ABRAHAMSE, A.P.S. And J. KOERTS. (1971). New estimators of disturbances in regression analysis. J. Am Statist. Assoc. 66, 71-74.
3. ALI, M. M. (1984). An approximation to the null distribution an power of the Durbin-Watson. Statistics. BIOMETRIKA, 71, 253-261.
4. AMEMIYA. T. (1977). A note on a heterocedastic model. Journal of Econometrics. 6, 365-370.
5. ANDREWS, D.F. GNANADESIKAN R. And WARNER, J. L. (1973). "Methods for assesing multivariate normality" In: P. R. KRISHNAIAH Ed. Multivariate analysis III. Academic Press. New York. 95-116.
6. ANSCOMBE, F.J. And J.W. TUKEY. (1967). Topics in investigation of linear relations fitted by the method of least squares. J. Roy. Statist Soc. B-29. 1-29. Discussion 29-52.
7. ANSCOMBE, F.J. And J.W. TUKEY. (1963). Examinations and analysis of residuals. Technometrics 5. 141-160.
8. ANSCOMBE, F.J. (1961). Examination of residuals. Proc. Fourth Berkley Symp. Math Statist. Prob. 1. 1-36.
9. ANSCOMBE, F.J. (1973). "Graphs in statistical analysis" Am. Statistician 27. 17-21.
10. BANDEMER, H. (1981). Methods for checking assumptions in regression models. BIOMETRY J. 23, 419-427.
11. BARLETT. (1937). "Proerties of sufficiency and statistical tests". Proc. R. Soc. A. 160, 268-282.
12. BARNETT. V. (1975). "Probability plotting methods and order statistics". Appl. Statist. 24. 95-108.
13. BECKMAN, R. J. And TRUSSEL (1974). The distribution of an

arbitrary studentized residual and the effects of updating in multiple regression. J. Amer. Statist. Assoc. 69, 199-201.
14. BEHNKEN, D.W. And N.R. DRAPER (1972). Residuals and their variance patterns. Technometrics 14, 101-111.
15. BENNETT, B.M. (1967). Use of Haldane-Smith test in examining randomness of residuals. Metron, 26 No. 3-4, 1-3.
16. BERA, A.K. And C.M. JARQUE (1982). Model Specification test: a simultaneous approach. Journal of econometrics. 20, 59-82.
17. BERENBLUT, I.I. And G.I. WEBB (1973). A new test for autocorrelation errors in the linear regression. Model J. Roy. Statist. Soc. B-35, 33-50.
18. BESLEY D.A.; E. KUH; R.E. WELSH (1980) "Regression Diagnostics" John Wiley and Sons. New York.
19. BICKEL, P.J. (1978). Using residuals robustly I: Test for heteroscedasticity, non lineanty. The annals of Statistics. 6. 266-291.
20. BIERENS, H. (1982). Test of model specification in the absence of alternative hypotheses paper presented at 1982 North American Summer Meeting of the Economemetric society.
21. BLOM, G. (1958). 'Statitical estimate and transformed Beta-Variables'. Wiley, New York.
22. BOX, G.E.P. (1953). Nonnormality and tests on variances. BIOMETRIKA 40 318-335.
23. BOX, G.E.P. And COX D.R. (1964). 'An analysis of transformations J. Roy Statist. Soc. B, 26.
24. BREUSCH, T.S. And A.R. PAGAN (1979). A simple test for heteroscedasticity and coefficient variation. Econometrica, 47, 1287-1294.
25. BREUSCH, T.S. And PAGAN A.R. (1980). "The Lagrange multipliers test and its applications to model specification in econometrics Review of economic studies. 47, 239-253.
26. BROWN, R.L.; DURBIN J. And EVANS J.M. (1975). "Techniques for testing the constancy of Regression Relationships over time" (with discussion). Journal of the Royal statistical society Ser. B. 37, 149-192.
27. BROWN, M.B. And FORSYTHE A.B. (1974). "Robust test for the equality of variances". J. Am. Stat. Assoc. 69, 364-367.
28. CHAMBERS, J.M. W.S. CLEVELAND, B. KLEINER And P.A. TUKEY (1983). Graphical Methods for data Analysis. Duxbury press. boston M.A.
29. CHAN, N. N. And T.K. MAK (1984). Heteroscedastic error in a linear functional relationship. BIOMETRIKA 71, 212-215.
30. CHATTERJEE, S. And B. PRICE (1977) "Regression analysis by example". John Wiley and sons. New York.
31. CLEVELAND, W.S, AND KLEINER (1975). 'A graphical Technique for enhancing scatterplots with proving statitics". Technometrics 17,447-454.
32. COCHRAN, W.G. (1941). "The distribution of the largest of a set of estimated variances as a fraction of their total". Ann Eugenics Lond. 11, 47-52
33. COOK, R.D. And WEISBERG (1983). diagnostics for heterocedasticity in regression. BIOMETRIKA, 70, 1-10.
34. COOK, R.D. And C.L. TSAI (1985). Residuals in nonlinear regression. BIOMETRIKA. 72, 23-29.
35. COX, D.R. (1968). Regression methods. notes on some aspects of regression analysis J.R. Statist Soc.Vol. 131 No. 3, 265-79

36. COX, D.R. And E.J. SNELL (1971). On test statistics calculated from residuals. BIOMETRIKA 58, 589-594.
37. COX, D.R. And E.J. SNELL (1968). A general definition of residuals. J. Roy Statistic Soc. B-30, 248-265. Discussion 265-275.
38. D'AGOSTINO, R.B. And PEARSON E.S. (1972). "Tests for departure from normality. Empirical results for the distributions of b_2 an $\sqrt{b_1}$ BIOMETRIKA 60, 613-622.
39. DANIEL, C. And WOOD (1980) "Fitting Equations to data" John Wiley and sons. New York.
40. DEMPSTER, A.P.; GASKO-GREEN M. (1981). "New Tools for residual analysis". The annals of statistics. Vol 9 No. 5, 945-959.
41. DIXON, J. And MASSEY F.J. Jr. (1969) "Introduction to statistical analysis 3th ed. McGraw-Hill. New York.
42. DRAPER, N.R. And W.E. LAWRENCE (1970). Anote on residuals in two dimensions. Technometrics, 19,394-398.
43. DRAPER, N.R. And H. SMITH (1981). "Applied regression analysis" John Wiley and sons. New York.
44. DURBIN, J. (1970). Testing for serial correlation in least square regression when some of regression are lagajed dependent variables. Econometrica, 38, 410-421.
45. DURBIN, J. And G.S. WATSON (1950). Testing for serial correlation in least squares regression, I. BIOMETRIKA, 37, 409-428.
46. DURBIN, J. (1970). An alternative to the bounds test for testing for serial correlation least squares regression. Econometrica 38, 422-429.
47. DURBIN, J. And G.S. WATSON (1951). Test for serial correlation in least squares regression II. BIOMETRIKA, 38, 159-178.
48. DURBIN, J. And G.S. WATSON (1971). Testing for serial correlation in least squares regression III. BIOMETRIKA, 58, 1-19.
49. DURBIN, J. (1969). Test for serial correlation in regression analysis based on the periodgram of least squares residuals. BIOMETRIKA 56, 1-15.
50. ELLENBERG, J.H. (1973). Tje joint distribution of the standardized least squares residuals from a general linear regression. J. Am statist Assoc 68, 941-943.
51. ENGLE (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation". Econometrica 50, 987-1007.
52. EPPS, T.W. And M.L. EPPS (1977). The robustness of some standard test for autocorrelation and heteroscedasticity when both problems are present. Econometrica, 47, 209-210.
53. FAREBROTHER, R.W. (1979). A grouping test for misspecifications. Econometrica, 47, 209-210.
54. FEDER, P.I. (1974). "Graphical techniques in statistical data analysis-tools for extracting information from data". Technometrics, 16, 287-299.
55. FILLIBEN, J.J. (1975). "The probability plot correlation coefficient tests for normality". Technometrics 17, 111-117.
56. FISCHER, R.A. (1930) "The moments of the distribution for normal samples of measures of departure from normality". Proc. Roy Soc. A, 130-136.

57. GALE, W.A. And PREGIBON, D. (1983). An expert system for regression analysis. Computer science and statistics: *Proc.* 14 th. Symposium on the interface edited by Heiner, Sacher, and Wilkinson. Springer Verlag. N.Y.

58. GALPIN, J.S. And D.M. HAWKINS (1984) The use of recursive residuals in checking model fit in linear regression. Ann statist 38, 94-105.

59. GARTSIDE, P.S. (1972). "A study of methods for comparing serveral variances" J. Am Stat. Assoc. 67, 342-346.

60. GIBBONS J.D. (1971). "Nomparametric statistics inference' McGraw-Hill. New York.

61. GNANADESIKAN, R. (1977). "Methods for statistical analysis of multivariate data" John Wiley. New York.

62. GODFREY, L.G. (1978). Testing for multiplicative heterocedasticity. Journal of Econometrics 8. 227-236.

63. GOLDFELD, S.M. And R.E. QUANDT. (1965). "Some tests for homoskedasticity' Jour. Am. Stat. Assoc. 60, 539-47.

64. GOLDFELD, S.M. And R.E. QUANDT. (1972). "Non linear methods of Econometrics" North-Holland Publisching Co. N.Y.

65. GOURIEROUX, C. HOLLY A. And MONFORT, A. (1982) "Likelihood ratio test, wald test, and KUHN-TUCKER Test in linear models with inequality constraints on the regression parameters". Econometrica, 50, 63-80.

66. GRAYBILL, F. (1976). "Theory and application of the linear model". Duxbury press. Massachussetts.

67. GREEN J.R. (1971). "Testing Departure from a Regression without using replication' Technometrics Vol. 13 No. 3. 609-615.

68. GUNST, R.F. And R.L. MASON (1980) "Regression analysis and its application". Marcel Dekker. New York.

69. HANNAN, E. (1957). Testing for serial correlation in least squares regression. BIOMETRIKA 44, 57-66.

70. HEDAYAT, A. And D.S. ROBSON (1970). Independent stepwise residuals for testing homocedasticity. J. Am Statist Assoc. 65, 1573-1581.

71. HOAGLIN D.C. And WELSCH R. (1978) "The Hat matrix in regression and ANOVA" The American Statistician 32, 17-22.

72. HOCKING, R.R. (1973). "Misspecification in Regression'. Am Stat. 28(1), 39-40.

73. HUANG, G.J. And BOLCH (1974). On the testing of regression disturbances for normality. J. Am statist Assoc. 69. 330-335.

74. IMAN, R.L. (1982). "Graphs for use with the Lilliefors test for Normal And exponential Distributions'. The Am Statistician. Vol. 36 No. 2. 109-112.

75. INTRILIGATOR M.D. (1978). "Econometric Models. Techniques and applicattions". Prentice Hall. New Jersey.

76. JOHNSON, R. And D. WICHERN (1982). 'Applied Multivariate statistical analysis" Prentice-Hall. New Jersey.

77. KELEJIAN, H.H. And OATES, E.E. (1972). "Introduction to econometrics, principles and applications". New York. Harper Row.

78. KING, M.L. (1982). A bounds test for heteroscedasticity. Monash University Working paper No. 5 /82

79. KOENKER, R. (1981). A note on studentizing a test for heterocedasticity. Journal of econometrics. 17, 107-112.

80. LARSEN, A.W. And MCLEARY. S. (1972). The use of partial

residual plots in regression analysis. Technometrics. Vol. 14, 3. 781-790.

81. LEAMER, E.E. (1983). "Let's take the con out of econometrics" American Economic review, 73, 31-43.

82. LILLIEFORS, H.W. (1969). "On the Kolmogorov-Smirnov test for normality with mean and variance unknown". Journal of the american statistical association 62, 399-402.

83. LOONEY, S.W. And GULLEDGE, T.R. (1985) "Use of the correlation coefficient with normal probability plots . The American statistician 39, 75-79.

84. LOYNES, R.M. (1969). On cox on Snell's general defition of residuals. J. Roy Statist Soc B-31, 103-106.

85. MAGE, D.T. (1982). "An objetive graphical methods for testing normality. Distributional assumptions, using probability plots". The Am. Stat. 36, 116-120.

86. MALKOVICH, J.F. And AFIFI, A.A. (1973). "On tests for multivariate normality" J. Am. Stat. Assoc. 68, 176-179.

87. MARDIA, K.V. (1980). Test for univariate and multivariate normality. Hand book of statistcs, vol 1. (P.R. Krishnaiah Ed). Amsterdam: North Holland, 279-320.

88. MARTINEZ, J. And B. IGLEWICZ (1981). A test for departure from normality based on the biweight estimator of scale. BIOMETRIKA, 68, 331-333.

89. MORAN, P.A.P. (1970) "On asymtotically optimal tests of composite hypotheses". BIOMETRIKA 57, 47-55.

90. NELSON, L.S. (1981). "A simple test for normality" Journal of Quality technology, Vol. 13, No. 1, 76-77.

91. OJA, H. (1983). New Tests for normality. BIOMETRIKA, 70, 297-299.

0½ 92. PAGAN Y HALL (1983). Residuals analysis. Communications statistics. Econometrics Reviens New York.

93. PAGAN AR, HALL A.D. And TRIVEDI P.K. (1981) "Assessing the variability of inflations" Australian National University Working papers in economics and econometrics No. 049.

94. PEARSON, E.S. (1930). "A further development of tests of normality" BIOMETRIKA 22, 239.

95. PEARSON, E.S. And HARTLEY, H.O. (1970). "Biometrika tables for statisticions" Vol. 1, 3rd. Ed. Cambridge university Press.

96. PHILLIPS, G.D.A. And HARVEY, A.C. (1974) "A simple test for serial correlation in Regression analysis" Journal of the American Statistical association 69, 935-939.

97. PREGIBON, DARYL (1983). "Comentarios al articulo de PAGAN Y HALL (1983) 235-339.

98. PUTTER, J. (1967) Orthonormal bases of error spaces and their use for investigating the normality and variance of residuals. J. Am Statist Assoc. 62, 1022-1036.

99. RAMSEY, J.B. (1969) Test for specification errors in classical linear least-squares regression analysis. J. Roy statist soc B-31, 350-371.

100. RAMSEY, J.B. (1983) "Comentarios al articulo de PAGAN Y HALL (1983).

101. RAMOLES, R.H. (1984). On tests applied to residuals. JASA, 79, 349-354.

102. RAO, C.R. (1973) . "Linear statistical inference and its applications" 2nd. Ed. Wiley, New York.

103. SCHEFFE, H. (1959). "The analysis of variance" John Wiley and sons. New York.

104. SEARLE, S.R. (1971). "Linear models" John Wiley and sons. New York.
105. SEBER G.A.F (1977). "Linear Regression analysis". John Wiley and sons. New York.
106. SEN, P.K. (1982). Invariance principles for recursive residuals. Ann Statist. 10, 307-312.
107. SHAPIRO, S.S. And FRANCIA, R.S. (1972). "An approximate analysis of variance test for normality". J. Am Stat. Assoc. 67, 215-216.
108. SHAPIRO, S.S. And WILK, M.B. (1965). "An analysis of variance test for normality (complete samples)". BIOMETRIKA 52, 591-611.
109. SHAPIRO, S.S. And WILK, M.B. AND CHEN, H.J. (1968). "A comparative study of various tests for normality". Journal of the american statistical association, 63, 1343-1372.
110. SHILLINGTON, E.R. (1979). Testing lack of fit in regression without replication. Can J. Statist 7, 137-146.
111. SNEE, R.D. (1977). Validation of regression models: Methods and examples. Technometrics. 19, 415-428. Technometrics, 13, 430-437.
112. SNEE, R.D. (1971). A note on the use of residuals for examining the asumptions of covariance analysis.
113. STEPHENS, MA.A. (1974). "EDF Statistics for goodsness of fit and some comparisons". Journal of american statistical association". 69, 730-737.
114. SZOETER. J. (1979). A class of parametric test for heteroscedasticity in linear economic models. Econometrica, 46, 1311 1327.
115. THEIL, H. (1965). "The analysis of Disturbances in Regression analysis" Journal of the american statistical Association. 60, 1067-1079.
116. THEIL, H. (1971). Principles of Econometrics" John Wiley and Sons. New York.
117. THEIL, H. And NAGAR (1961). "Testing the independency of regression disturbances" J. Am Stat. assoc. 56, 793-806.
118. THURSBY, J.G. (1982). Misspecification, heteroscedasticity and the chow and Goldfeld-Quantdt test. The review of economics and statistics, LXIV, 314-321.
119. VAN DOBBEN DE BRUYN C.S. (1968). "Acumulative Sum test: Theory and practice" Griffin's statistical monographs and courses".
120. WEISBERG, S. And BINGHAN C. (1975). "An approximate analysis of variance test for non-normality suitable for machine calculation". Technometrics Vol. 17 No. 1 133-134.
121. WESOLOWSKY, G.O. (1976) "Multiple regression and analysis of variance". John Wiley and sons. New York.
122. WHITE, H. And G.M. MCDONALD (1980) Some sample test for non normality in the linear regression model. Journal of the American Statistical association 75, 16-28.
123. WILK, M.B. And GNANADESIKAN (1968). Probability plotting methods for the analysis of data. BIOMETRIKA 55, 1-17.
124. WOOD, F.S. (1973). The use of individual effects and residuals in fitting equations to data. Technometrics, 15, 677-695.
125. ZYSKIND G. And P.A. JOHNSTON (1973) On a zero residuals sum in regression. Am Statist 27, 43-44.