ANÁLISIS DE REGRESIÓN

- **Bibliografía**

1.- Montgomery, D. C. (1992)
Introduction to Linear Regression Analysis.
New York: John Wiley and Sons.

2.- Gujarati, D. N. (2004)
Econometría (4a. edición)
México: McGraw-Hill Interamericana Editores S.A. de C.V.

3.- Johnston, J. y Dinardo, J. (1996)
Econometric Methods (4th edition)
New York: McGraw-Hill.

4.- Pyndick, R.S. y Rubinfeld, D. L. (1998)
Econometric Models and Economic Forecasts (4th edition)
Singapore: McGraw-Hill.

5.- Weisberg, S. (1985)
Applied Linear Regression.
New York: John Wiley and Sons.

- **Paquete estadístico recomendado**

E-Views Versión 7 (o más actual)

ANÁLISIS DE REGRESIÓN

TEMARIO

1. INTRODUCCIÓN A LOS MODELOS LINEALES.

2. MODELO DE REGRESIÓN LINEAL SIMPLE.

1.- Estimación por Mínimos Cuadrados Ordinarios (MCO).
2.- Medidas de bondad de ajuste.
3.- Propiedades de los estimadores.
4.- Estimación de la varianza del error.
5.- Pronóstico.
6.- Estimación por Máxima Verosimilitud (MV).
7.- Inferencia estadística.
8.- Análisis de varianza.
9.- Análisis de residuos.

3. MODELO DE REGRESIÓN LINEAL MÚLTIPLE.

1.- Notación matricial.
2.- Estimación por MCO.
3.- Pronóstico.
4.- Estimación por MV e inferencia estadística.
5.- Análisis de varianza.
6.- Prueba de una hipótesis lineal general.
7.- Mínimos Cuadrados Generalizados.
8.- Autocorrelación.
9.- Heteroscedasticidad.
10.- No-linealidad y no-normalidad.
11.- Multicolinealidad.

1. INTRODUCCIÓN A LOS MODELOS LINEALES

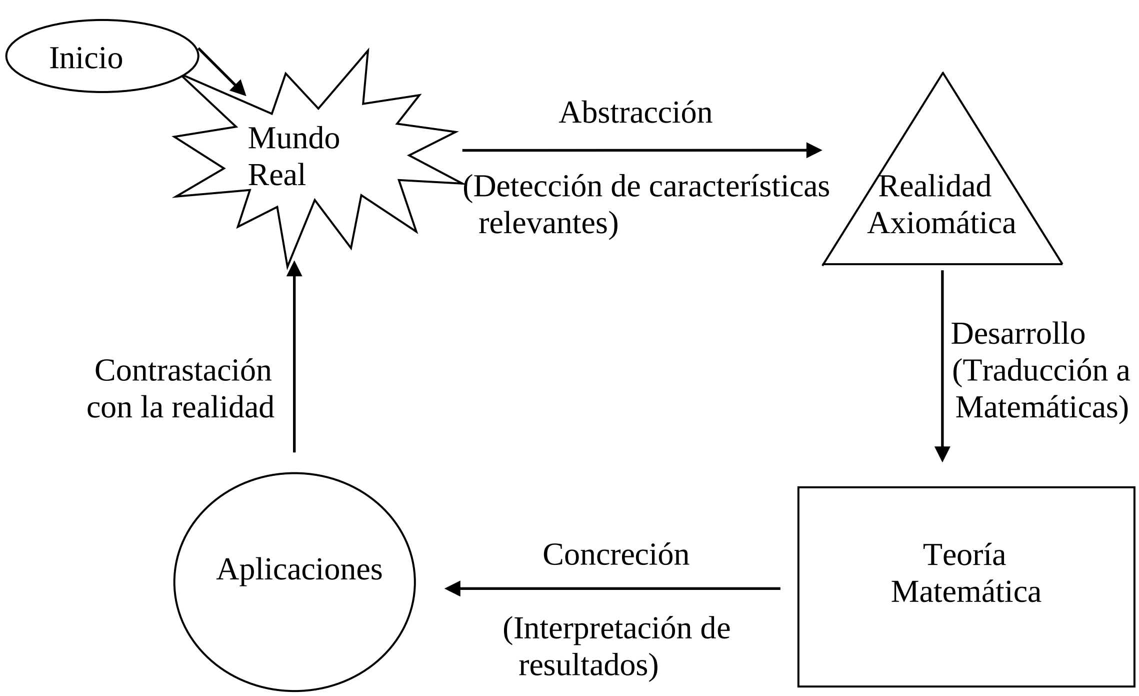
Los modelos son representaciones simplificadas de fenómenos reales. En consecuencia, **todo modelo es erróneo**, puesto que al simpificar se dejan fuera de consideración algunos elementos, supuestamente irrelevantes.

Los modelos a considerar aquí son aquellos que se escriben en términos matemáticos y que incorporan variables, parámetros y supuestos. Debe subrayarse que **los resultados de un modelo dependen de los supuestos que se hayan hecho**, en particular acerca de las variables (tanto de las incluídas, como de las omitidas del modelo).

Los supuestos comúnmente son acerca del tipo de relación funcional, de la independencia o de las posibles asociaciones de causa-efecto que pudieran existir entre las variables del modelo.

La construcción de un modelo matemático podría verse esquemáticamente como sigue.

MODELO MATEMÁTICO



Este es un proceso iterativo que continúa hasta que se obtienen resultados que ya se consideran útiles para algún fin en particular. Los fines que típicamente se busca conseguir al construir un modelo matemático son:

(i) explicar por qué la realidad se comporta de tal o cual manera,

(ii) predecir lo que ocurrirá con alguna variable de interés, como función de las otras variables incluídas en el modelo, y

(iii) simular lo que podría llegar a pasar con una cierta variable si las otras variables del modelo, que pretenden explicar su comportamiento, llegasen a tomar determinados valores.

Los modelos matemáticos pueden ser de tipo determinístico o de tipo no-determinístico (mejor conocidos como estadísticos). Los modelos determinísticos permiten predecir el valor de la variable de interés con exactitud, mientras que **los modelos estadísticos incorporan la idea de incertidumbre y de error**, al hacer predicciones o al obtener algún otro tipo de conclusión a partir del modelo.

Un ejemplo de modelo estadístico es

$$Y = \beta_1 + \beta_2 X_1^{X_2} + \beta_3\beta_4 X_3 + \beta_5^2 X_4 + \varepsilon \qquad (1.1)$$

donde $X_1, \ldots, X_4$ son **variables independientes**, explicativas, exógenas o factores (no necesariamente aleatorias), $\beta_1, \ldots, \beta_5$ son parámetros (constantes, comúnmente desconocidas), Y es la **variable dependiente**, por explicar, endógena o respuesta (variable aleatoria) y ε es un elemento aleatorio, conocido como error del modelo (variable aleatoria con distribución de probabilidad por determinar).

En la práctica se busca construir modelos del tipo

$$\eta = F(X_1, \ldots, X_k, \beta_1, \ldots, \beta_p) \quad \text{y} \quad Y = \eta + \varepsilon \qquad (1.2)$$

donde $F(\cdot)$ es una función que surge de la teoría del fenómeno que se estudia, k es el número de variables explicativas y p es el número de parámetros, de manera que se obtiene la ecuación

$$Y = F(X_1, \ldots, X_k, \beta_1, \ldots, \beta_p) + \varepsilon \tag{1.3}$$

que se interpreta como: Valor observado = Valor real + Error.

Una clasificación importante de los modelos estadísticos es la que distingue entre modelos lineales y no-lineales, en relación con la forma en que aparecen los parámetros en la relación funcional. Un ejemplo de modelo no-lineal es el de la ecuación (1.1), mientras que un modelo lineal es aquel que resulta ser **lineal en los parámetros**, como se muestra a continuación

$$Y = \beta_0 + \beta_1 X_1^{X_2} + \beta_2 X_3 + \beta_3 \log(X_4) + \varepsilon \tag{1.4}$$

En este caso, los parámetros β_0, β_1, β_2 y β_3, entran en la ecuación del modelo como combinación lineal de las variables 1, $X_1^{X_2}$, X_3 y $\log(X_4)$, con log(.) la función logaritmo natural, que no necesariamente son las variables originalmente observadas, o sea, 1, X_1, X_2, X_3 y X_4.

Una manera de verificar que un modelo es lineal (en los parámetros) es mediante la toma de derivadas de la ecuación que describe al modelo, respecto a cada uno de sus parámetros, ya que se esperaría que no hubiera interacción entre ellos, por lo cual la derivada con respecto a cualquiera de ellos no debería depender de ningún parámetro. Por ejemplo, en el caso de la ecuación (1.1) se tiene

$$\frac{\partial Y}{\partial \beta_3} = \beta_4 X_3 \quad \text{y} \quad \frac{\partial Y}{\partial \beta_5} = 2\beta_5 X_4 \tag{1.5}$$

por lo cual se concluye que dicho modelo no es lineal. En cambio, en el caso del modelo (1.4) ninguna derivada parcial de Y con respecto a β_0, β_1, β_2 y β_3 involucra alguno de los parámetros.

2.- MODELO DE REGRESIÓN LINEAL SIMPLE.

El modelo de Regresión Lineal Simple establece en principio que el modelo es lineal en los parámetros y debido a esto es que lleva dicho nombre el modelo.

La ecuación en la cual se basa el modelo es la siguiente

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (2.1)$$

en donde Y es la variable dependiente, X es la variable independiente, ε es el error aleatorio y β_0, β_1 son parámetros desconocidos.

En esta ecuación se puede observar que Y es una variable aleatoria porque está igualada con la suma de una variable aleatoria (ε) y otra variable que también podría serlo (X). Si la variable X es aleatoria, se supondrá que sus observaciones tienen ya sus **valores fijos**.

La idea del Análisis de Regresión Lineal es usar un conjunto de datos muestrales de tal manera que se puedan estimar los valores de β_0 y β_1, para **ajustar una recta a los datos muestrales** y así, entre otras cosas, poder explicar el comportamiento de los datos disponibles de Y.

El Cuadro 2.1 ilustra la forma en que se presentan los datos, de manera esquemática. Cabe notar que se presupone la existencia de n parejas de datos Y_i, X_i, con i = 1, …, n.

i	Y_i	X_i
1	Y_1	X_1
2	Y_2	X_2
...	...	...
n	Y_n	X_n

Cuadro2.1. Datos a usar en el modelo de Regresión Lineal Simple.

Por otro lado, las gráficas de las figuras 2.1, 2.2 y 2.3 resumen la manera en que los datos pueden visualizarse. Tales gráficas reciben el nombre de **Diagramas de Dispersión**.

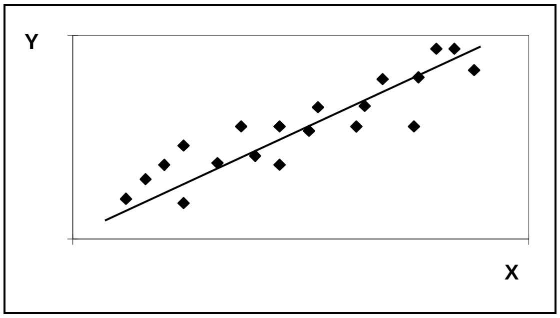


Figura 2.1. Tendencia lineal positiva

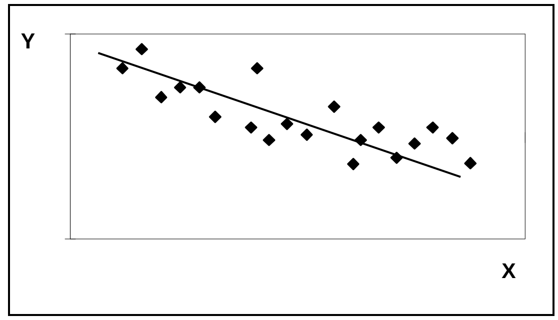


Figura 2.2. Tendencia lineal negativa

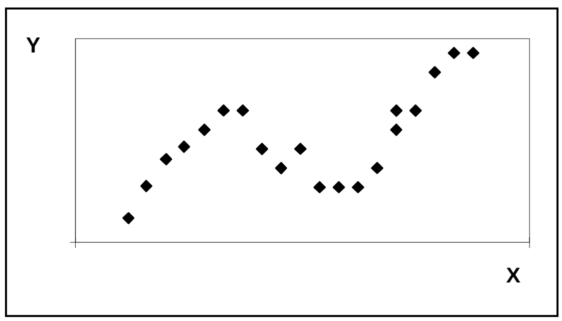


Figura 2.3. Tendencia no-lineal

Como se puede observar, en las figuras 2.1 y 2.2 existe una cierta tendencia, del tipo de línea recta, en los datos; es decir, los puntos siguen un movimiento genérico de tipo lineal. En cambio, en la Figura 2.3 los datos no parecen seguir una tendencia lineal, pero al separar esta gráfica en intervalos se puede ver claramente que si existe una tendencia lineal, aunque sea por intervalos. Nótese que ahora se hace referencia a una **relación lineal en las variables**, además de que el modelo sea lineal en los parámetros. Esto es porque la variable X entra en el modelo como combinación lineal de los parámetros. Estos es, al multiplicar el vector (1, X) por el vector $(\beta_{0,}\ \beta_1)$' se obtiene la relación funcional del modelo.

Una vez planteado el modelo, lo que resta es asignar valores a sus parámetros; para ello se usará un método ampliamente usado en la práctica debido a que sus resultados son óptimos en un sentido estadístico. El criterio que se utilizará para estimar los valores de β_0 y β_1 será el de Mínimos Cuadrados Ordinarios.

2.1 Estimación por Mínimos Cuadrados Ordinarios (MCO)

La idea del método de MCO es la de estimar los valores de β_0 y β_1 para obtener un valor estimado óptimo de la variable por explicar. Esto se hace de tal forma que la suma de diferencias al cuadrado, de los valores observados menos los valores estimados, sea mínima.

Para cada elección de valores de $\hat{\beta}_0$ y $\hat{\beta}_1$ se obtiene un conjunto de valores estimados $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$, para i = 1, …, n, con los cuales se calculan los **residuos**, definidos como

$$e_i = Y_i - \hat{Y}_i . \tag{2.2}$$

Se puede calcular entonces la Suma de Cuadrados Residual (SCR), que está definida como

$$SCR = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \tag{2.3}$$

De hecho, los valores estimados, $\hat{Y}_i$, producen la recta estimada que se muestra en la gráfica siguiente de la Figura 2.4

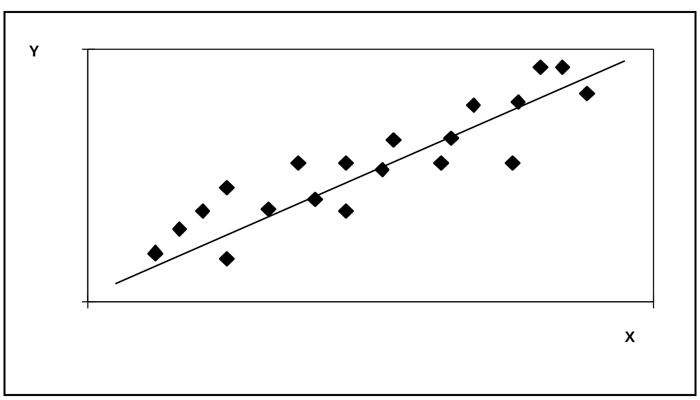


Figura 2.4. Valores observados y recta estimada.

Así que, para cada valor de Y_i observado existe su correspondiente valor estimado $\hat{Y}_i$ y, por consiguiente, su residuo respectivo e_i .

A continuación se elegirán los valores $\hat{\beta}_0$ y $\hat{\beta}_1$ de tal forma que la SCR sea mínima. Los respectivos **estimadores mínimo cuadráticos** serán denotados por b_0 y b_1.

Para obtener los valores b_0 y b_1 se deben resolver las siguientes ecuaciones

$$\frac{\partial\, SCR}{\partial\, \hat{\beta}_0}\Bigg|_{\substack{\hat{\beta}_0 = b_0 \\ \hat{\beta}_1 = b_1}} = 0 \quad \text{y} \quad \frac{\partial\, SCR}{\partial\, \hat{\beta}_1}\Bigg|_{\substack{\hat{\beta}_0 = b_0 \\ \hat{\beta}_1 = b_1}} = 0 \tag{2.4}$$

De la primera ecuación (2.3) se sigue que

$$\sum_{i=1}^{n} Y_i - b_0 n - b_1 \sum_{i=1}^{n} X_i = 0 \tag{2.5}$$

mientras que de la segunda se obtiene

$$\sum_{i=1}^{n} X_i Y_i - b_0 \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0 \tag{2.6}$$

Las dos ecuaciones anteriores forman un sistema lineal de dos ecuaciones con dos incógnitas, al cual se conoce como sistema de **Ecuaciones Normales**.

Al resolver las Ecuaciones Normales se obtiene lo siguiente (nótese que en las sumas se suprime el recorrido del índice i, por ser evidente y para simplificar la escritura)

$$b_1 = \frac{\begin{vmatrix} n & \sum Y_i \\ \sum X_i & \sum X_i Y_i \end{vmatrix}}{\begin{vmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{vmatrix}} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 + \left(\sum X_i\right)^2} = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum\left(X_i - \overline{X}\right)^2} \tag{2.7}$$

mientras que la ecuación (2.4) conduce a

$$b_0 = \overline{Y} - b_1\overline{X} \ . \tag{2.8}$$

Para obtener el valor b_1 se hizo uso de las siguientes relaciones

$$\sum(X_i - \overline{X})^2 = \sum(X_i^2 - 2X_i\overline{X} + \overline{X}^2) = \sum X_i^2 - 2\overline{X}\sum X_i + n\overline{X}^2 = \sum X_i^2 - (\sum X_i)^2/n$$

$$\sum(X_i - \overline{X})(Y_i - \overline{Y}) = \sum(X_iY_i - X_i\overline{Y} - \overline{X}Y_i + \overline{X}\overline{Y})$$

$$= \sum X_iY_i - \overline{Y}\sum X_i - \overline{X}\sum Y_i + n\overline{X}\overline{Y}$$

$$= \sum X_iY_i - \sum X_i\sum Y_i \big/ n$$

Debido a la forma que tiene b_0, se deduce que

$$\hat{Y}_i = b_0 + b_1X_i = \overline{Y} - b_1\overline{X} + b_1X_i = \overline{Y} + b_1(X_i - \overline{X})\,. \tag{2.9}$$

Por otro lado, la expresión para b_1 puede también escribirse como

$$b_1 = \frac{S_{XY}}{S_X^2} \tag{2.10}$$

en donde se hace uso de la siguiente notación:

$$S_{XY} = \frac{\sum X_iY_i - \sum X_i\sum Y_i/n}{n-1} = \text{Covarianza muestral de X y Y.} \tag{2.11}$$

$$S_X^2 = \frac{\sum X_i^2 - (\sum X_i)^2/n}{n-1} = \text{Varianza muestral de X.} \tag{2.12}$$

De manera similar se define

$$S_Y^2 = \frac{\sum Y_i^2 - (\sum Y_i)^2/n}{n-1} = \text{Varianza muestral de Y} \qquad (2.13)$$

2.2 Medidas de bondad de ajuste

Para determinar qué tan bueno es el ajuste logrado con la recta de regresión, se usa el coeficiente de correlación muestral entre X y Y, el cual toma la forma

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}} . \qquad (2.14)$$

- El coeficiente de correlación satisface la relación

$$-1 \le r_{XY} \le 1 .$$

Una demostración de este hecho se sigue de

$$0 \le \frac{1}{2}\sum_{i=1}^{n}\left[\frac{X_i - \overline{X}}{\sqrt{\sum (X_j - \overline{X})^2}} \pm \frac{Y_i - \overline{Y}}{\sqrt{\sum (Y_j - \overline{Y})^2}}\right]^2$$

$$= \frac{1}{2}\sum_{i=1}^{n}\left[\frac{(X_i - \overline{X})^2}{\sum (X_j - \overline{X})^2} \pm 2\frac{(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_j - \overline{X})^2 \sum (Y_j - \overline{Y})^2}} + \frac{(Y_i - \overline{Y})^2}{\sum (Y_j - \overline{Y})^2}\right]$$

$$= \frac{1}{2}\left[\frac{\sum (X_i - \overline{X})^2}{\sum (X_j - \overline{X})^2} \pm 2\, r_{XY} + \frac{\sum (Y_i - \overline{Y})^2}{\sum (Y_j - \overline{Y})^2}\right]$$

$$= \frac{1}{2}(2 \pm 2\, r_{XY})$$

$= 1 \pm r_{XY}$ de donde se obtiene el resultado.

Si $r_{XY} = 0$, no hay asociación lineal entre las variables X y Y. Por lo tanto, se puede concluir que r_{XY} **mide el grado de asociación lineal** entre las variables X y Y.

De hecho, ya que

$$b_1 = \sqrt{\frac{\Sigma\left(Y_i - \overline{Y}\right)^2}{\Sigma\left(X_i - \overline{X}\right)^2}} \frac{\Sigma\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\Sigma\left(X_i - \overline{X}\right)^2\left(Y_i - \overline{Y}\right)^2}} = \frac{S_Y}{S_X} r_{XY} \qquad (2.15)$$

se observa que si el coeficiente de correlación muestral es positivo, la pendiente de la recta es positiva; de igual manera se deduce el caso de pendiente negativa. Por consiguiente, **el signo de** r_{XY} **indica la dirección** de la relación entre X y Y.

Nótese en las gráficas de la Figura 2.5 que, cuando la mayoría de los puntos caen dentro de los cuadrantes positivos del plano coordenado $\left((X - \overline{X}), (Y - \overline{Y})\right)$, entonces la correlación es positiva y en caso contrario r_{XY} es negativo.

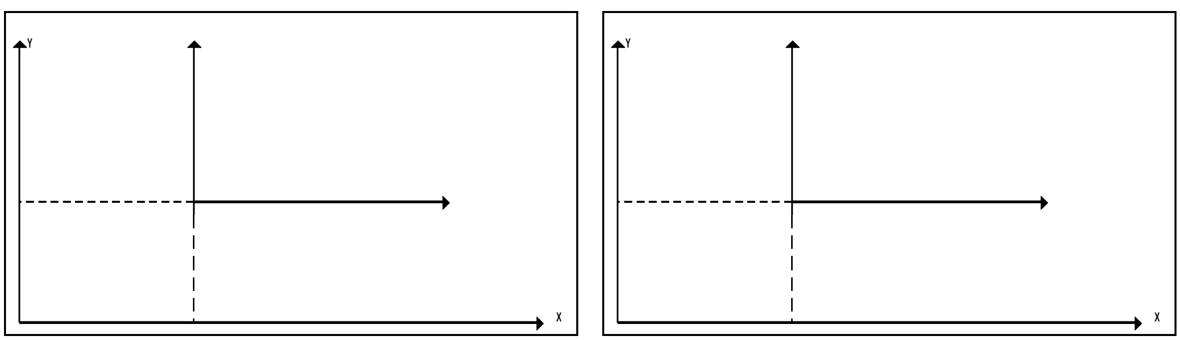


Figura 2.5. Datos con asociación lineal entre X y Y.

De igual manera es importante apreciar el papel que juega r_{XY}, mediante el grupo de gráficas que aparece como Figura 2.6, donde se muestra que la misma recta estimada ($\hat{Y} = 2 + X$) se ajusta a distintos grupos de datos, lo cual ilustra el hecho de que **la misma recta puede producir distintos grados de asociación lineal** (r_{XY} = 1, 0.8 y 0.2).

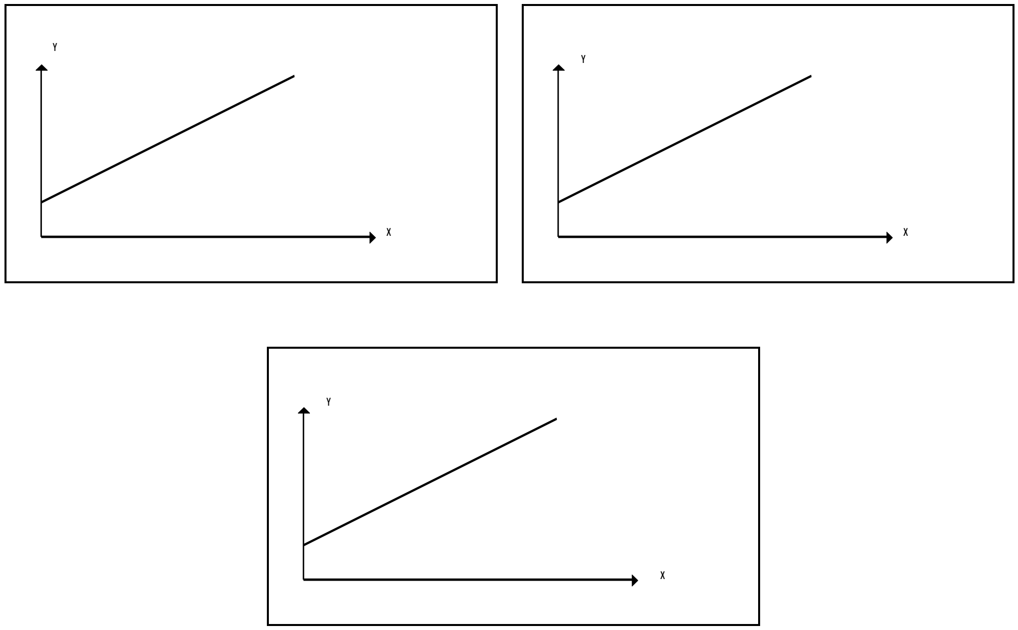


Figura 2.6. Misma recta ajustada a datos con diferentes grados de asociación lineal.

A partir de los cálculos básicos presentados en el Cuadro 2.2 se pueden calcular S_X^2, S_Y^2 y S_{XY} de manera sencilla, para después obtener los valores numéricos de b_1, b_0 y r_{XY}

i	Y_i	X_i	X_iY_i	X_i^2	Y_i^2
1	Y_1	X_1	X_1Y_1	X_1^2	Y_1^2
...	...	...	...	...	...
n	Y_n	X_n	X_nY_n	X_n^2	Y_n^2
Sumas	ΣY_i	ΣX_i	$\Sigma X_i Y_i$	ΣX_i^2	ΣY_i^2

Cuadro 2.2. Cálculos básicos para obtener b_0 y b_1.

A continuación se muestra un ejemplo numérico (hipotético) para comprender de mejor manera lo descrito anteriormente.

Ejemplo. Considérese el siguiente conjunto de n = 5 datos hipotéticos y sus cálculos básicos

i	X_i	Y_i	$X_i Y_i$	X_i^2	Y_i^2
1	2	4	8	4	16
2	3	7	21	9	49
3	1	3	3	1	9
4	5	9	45	25	81
5	9	17	153	81	289
Suma	20	40	230	120	444

Entonces $\sum X_i = 20$ y $\sum Y_i = 40$, de forma que $\overline{X} = 4$ y $\overline{Y} = 8$, por lo cual se obtiene

$$b_1 = \frac{5(230) - 20(40)}{5(120) - (20)^2} = \frac{1150 - 800}{600 - 400} = \frac{350}{200} = 1.75 \quad \text{y} \quad b_0 = 8\text{-}1.75(4)\text{=}1$$

o sea que la regresión de Y sobre X es

$$\hat{Y} = 1 + 1.75X$$

A partir de esta ecuación se pueden obtener los valores estimados por el modelo, correspondientes a las observaciones de X, es decir

$$\hat{Y}_1 = 1 + 1.75X_1 = 1 + 1.75(2) = 4.5$$

$$\hat{Y}_2 = 1 + 1.75X_2 = 1 + 1.75(3) = 6.25$$

$$\hat{Y}_3 = \ldots \qquad \ldots \qquad = 2.75$$

$$\hat{Y}_4 = \ldots \qquad \ldots \qquad = 9.75$$

$$\hat{Y}_5 = \ldots \qquad \ldots \qquad = 16.75$$

y, de esta forma, los residuos vienen a ser

$$e_1 = Y_1 - \hat{Y}_1 = 4 - 4.5 = -0.5,\ \ e_2 = 7 - 6.25 = 0.75,\ \ e_3 = 0.25,\ \ e_4 = -0.75\ \ \text{y}\ \ e_5 = 0.25$$

Nótese que la media aritmética de los residuos es igual a cero, esto siempre debe ocurrir cuando se estimen los parámetros por MCO y exista en el modelo ordenada al origen. En cambio, no necesariamente ocurre esto en un modelo de la forma $Y = \beta_1 X + \varepsilon$.

Asimismo, la SCR mínima correspondiente, que **constituye una medida de la precisión de la estimación**, en este ejemplo resulta ser

$$SCR = \sum e_i^2 = 1.5$$

Una vez visto el ejemplo numérico, conviene conocer lo siguiente:

- b_1 mide el cambio en Y debido a un cambio unitario en X.
- b_0 es la ordenada al origen o valor autónomo de Y; es decir, con un valor de X = 0 , la variable Y tomaría el valor b_0.
- r_{XY} no cambia si se aplican transformaciones lineales a las variables.
- Correlación no implica asociación causal, ni causalidad implica que haya correlación (a menos que la causalidad fuera estrictamente lineal).

Por otro lado, en muchas ocasiones conviene trabajar con los datos expresados en forma de desviaciones respecto a su media, esencialmente por facilidad de cálculo. Para ello, se definen las variables en **desviaciones respecto a su media**

$$x_i = X_i - \overline{X} \quad \text{y} \quad y_i = Y_i - \overline{Y} \tag{2.16}$$

de manera que ahora se postula el modelo

$$y_i = \beta_0' + \beta_1' x_i + \varepsilon_i \tag{2.17}$$

para el cual se obtienen los estimadores de MCO

$$b_1' = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum\left(X_i - \bar{X}\right)^2} = b_1 \quad \text{y} \quad b_0' = \bar{y} - b_1' \bar{x} = 0 \tag{2.18}$$

debido a que $\bar{x}$=0 y $\bar{y}$=0.

La Figura 2.7 muestra el hecho de que al trabajar con variables en desviaciones respecto a su media, la pendiente no cambia respecto a la de las variables originales. Además, nótese que el valor estimado de la ordenada al origen toma el valor cero, lo cual no implica que no exista ordenada al origen, sino que **los datos fueron trasladados al origen**.

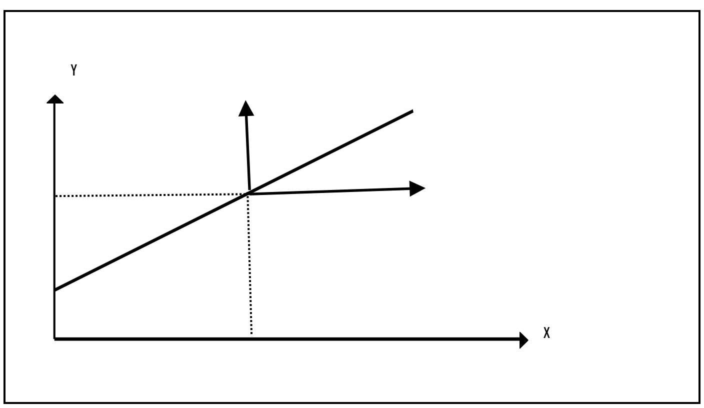


Figura 2.7. Efecto del cambio a variables en desviaciones respecto a su media.

Adicionalmente, se obtiene la siguiente expresión, que es muy simple y útil para calcular el valor de la pendiente de la recta, en el modelo estimado por MCO

$$b_1 = \frac{\sum x_i y_i}{\sum x_i^2}. \tag{2.19}$$

2.3 Propiedades de los estimadores

Los **estimadores mínimo cuadráticos son de fácil cálculo** y tienen además otras propiedades importantes, como es la de ser insesgados para los respectivos parámetros de la regresión. Para ver esto, conviene expresarlos como combinación lineal de las observaciones de Y. En principio, con las variables expresadas en desviaciones respecto a su media se obtiene

$$b_1 = \sum \frac{\left(X_i - \overline{X}\right)}{\sum\left(X_i - \overline{X}\right)^2}\left(Y_i - \overline{Y}\right) = \sum \frac{x_i}{\sum x_i^2} y_i = \sum w_i y_i \qquad (2.20)$$

donde

$$w_i = \frac{x_i}{\sum x_i^2} \qquad (2.21)$$

es tal que $\sum w_i x_i = 1, \sum w_i = 0$ y $\sum w_i^2 = \dfrac{1}{\sum x_i^2}$.

Las propiedades de los estimadores dependen de los supuestos que respalden al modelo. Conviene entonces suponer que:

1) Existe una relación lineal entre las variables X y Y, o sea

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{para } i = 1,...,n, \qquad (2.22)$$

2) $E(\varepsilon_i) = 0$ para toda i,

3) $Var(\varepsilon_i) = \sigma^2$ constante para toda i ,

4) $Cov(\varepsilon_i,\varepsilon_j) = 0$ para toda $i \neq j$.

A partir de tales supuestos, y si X está fija, se tiene

$$E(Y_i) = \beta_0 + \beta_1 X_i \tag{2.23}$$

con

$$Var(Y_i) = Var(\varepsilon_i) = \sigma^2. \tag{2.24}$$

Si, en cambio, X fuese una variable aleatoria, se debería considerar entonces que la esperanza y la varianza deben ser condicionales en los valores observados de X.

Por consiguiente, se tiene que

$$E(b_1) = E\left(\sum w_i y_i\right) = \sum w_i E(y_i) = \sum w_i (\beta_1 x_i) = \beta_1 \tag{2.25}$$

de donde se sigue que

$$E(b_0) = E\left(\overline{Y} - b_1\overline{X}\right) = E\left(\overline{Y}\right) - E(b_1)\overline{X} = \left(\beta_0 + \beta_1\overline{X}\right) - \beta_1\overline{X} = \beta_0 \,. \tag{2.26}$$

Por otro lado, como

$$\Sigma w_i \overline{Y} = 0 \tag{2.27}$$

se puede escribir b_1 de la forma

$$b_1 = \Sigma w_i Y_i \ , \tag{2.28}$$

por lo cual se deduce que

$$\mathrm{Var}(b_1) \;=\; \sum w_i^2 \;\mathrm{Var}(Y_i) \;=\; \frac{\sigma^2}{\sum x_i^2}\;. \tag{2.29}$$

De igual forma, para obtener la varianza de b_0 se tiene que

$$\mathrm{Var}(b_0) = \mathrm{Var}(\overline{Y} - b_1\overline{X})$$

$$= \mathrm{Var}(\overline{Y}) - 2\mathrm{Cov}(\overline{Y}, b_1) + \mathrm{Var}(b_1\overline{X})$$

$$= \sigma^2/n - 2\mathrm{Cov}(\overline{Y}, b_1) + \overline{X}^2\mathrm{Var}(b_1)$$

$$= \sigma^2\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum x_i^2}\right) \qquad \text{porque} \quad \mathrm{Cov}(\overline{Y}, b_1) = 0$$

$$= \sigma^2\left(\frac{\sum x_i^2 + n\overline{X}^2}{n\sum x_i^2}\right)$$

$$= \sigma^2\left(\frac{\sum X_i^2}{n\sum x_i^2}\right)\;. \tag{2.30}$$

Falta demostrar que $\mathrm{Cov}(\overline{Y}, b_1) \;=\; 0$, para ello nótese que

$$E(Y_i) = \beta_0 + \beta_1 X_i \qquad \text{y} \qquad Y_i - E(Y_i) = \varepsilon_i \tag{2.31}$$

por lo tanto

$$\mathrm{Cov}(\overline{Y}, b_1) = E\left\{\left[\frac{1}{n}\sum Y_i - \frac{1}{n}\sum E(Y_i)\right]\left[\sum w_i Y_i - \sum w_i E(Y_i)\right]\right\}$$

$$= E\left\{\frac{1}{n}\sum [Y_i - E(Y_i)] \sum w_i [Y_i - E(Y_i)]\right\}$$

$$= E\left[\left(\frac{1}{n}\sum \varepsilon_i\right)\left(\sum w_i \varepsilon_i\right)\right]$$

$$= E\left(\frac{1}{n}\sum w_i \varepsilon_i^2 + \sum_{i \neq j} w_i \varepsilon_i \varepsilon_j\right)$$

$$= \frac{\sigma^2}{n}\sum w_i \quad \text{porque} \quad E\left(\varepsilon_i \varepsilon_j\right) = 0 \quad \text{para} \;\; i \neq j$$

$$= 0 \tag{2.32}$$

De hecho, b_1 y b_0 son insesgados y tienen las varianzas Var(b_1) y Var(b_0) descritas anteriormente. Sin embargo, lo más importante es saber que estos estimadores son de varianza mínima y, por ende, los más eficientes. Eso es lo que establece el teorema siguiente.

Teorema de Gauss-Markov. Dentro de la clase de estimadores lineales e insesgados para β_0 y β_1 (en el modelo de regresión lineal simple) los estimadores mínimo cuadráticos poseen la menor varianza, es decir, son los MELI (Mejores Estimadores Linealmente Insesgados).

Esto se ilustra en la Figura 2.8 que sigue

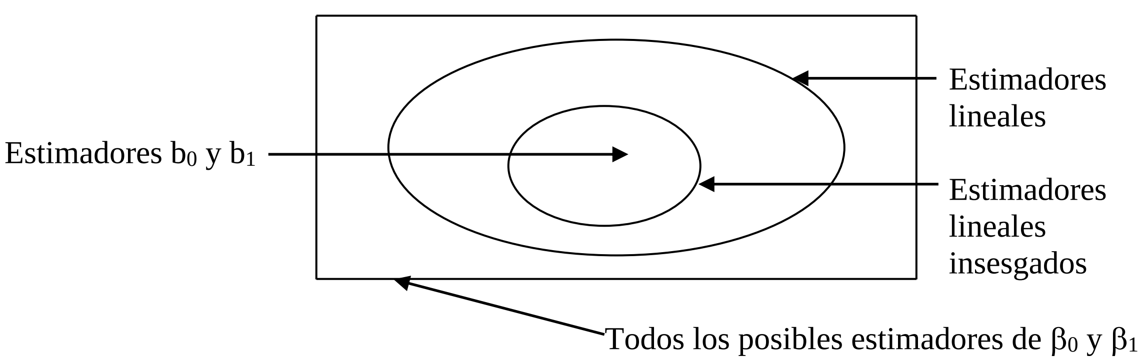


Figura 2.8. Ubicación de los estimadores de MCO.

2.4 Estimación de la varianza del error

Conviene ahora hacer notar que el modelo de regresión involucra otro parámetro, además de β_0 y β_1, el cual está asociado con la varianza del error. Por lo tanto, se requiere también estimar σ^2. Para ello, como

$$\sigma^2 = \text{Var}(\varepsilon_i) = E(\varepsilon_i^2) - [E(\varepsilon_i)]^2 = E(\varepsilon_i^2) \tag{2.33}$$

se propone usar un promedio de los residuos al cuadrado, ya que los errores no son observables. De hecho se usará la expresión

$$\frac{\text{SCR}}{\text{g.l.}} \tag{2.34}$$

con g.l. = grados de libertad = número de observaciones - número de parámetros estimados.

Dicho estimador es insesgado si $E\left(\frac{\text{SCR}}{\text{g.l.}}\right) = \sigma^2$. De hecho, se usa g.l. en lugar de n, para que $\frac{\text{SCR}}{\text{g.l.}}$ resulte ser insesgado para σ^2.

Para empezar, nótese que el residuo de la estimación mínimo-cuadrática está dado por

$$e_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_i \tag{2.35}$$

y, debido a que $b_0 = \bar{Y} - b_1\bar{X}$, se llega a

$$e_i = (Y_i - \bar{Y}) - b_1(X_i - \bar{X}) \tag{2.36}$$

así que

$$\sum e_i = \sum\left(Y_i - \overline{Y}\right) - b_1\sum\left(X_i - \overline{X}\right) = 0. \tag{2.37}$$

Asimismo, de la segunda Ecuación Normal se deduce que

$$\begin{aligned}\sum e_i X_i &= \sum\left(Y_i - \hat{Y}_i\right)X_i \\ &= \sum Y_i X_i - \sum\left(b_0 + b_1 X_i\right) \\ &= \sum X_i Y_i - b_0\sum X_i - b_1\sum X_i^2 \\ &= 0 \end{aligned} \tag{2.38}$$

y, de igual manera

$$\sum e_i \hat{Y}_i = \sum e_i\left(b_0 + b_1 X_i\right) = b_0\sum e_i + b_1\sum e_i X_i = 0. \tag{2.39}$$

Por otro lado, si se escribe el error como

$$\varepsilon_i = Y_i - \left(\beta_0 + \beta_1 X_i\right) = e_i + \left(\hat{Y}_i - \beta_0 - \beta_1 X_i\right) \tag{2.40}$$

se sigue que

$$\begin{aligned}E\left(\sum \varepsilon_i^2\right) &= E\left(\sum e_i^{\,2}\right) + E\left[\sum\left(\hat{Y}_i - \beta_0 - \beta_1 X_i\right)^2\right] + 2E\left[\sum e_i\left(\hat{Y}_i - \beta_0 - \beta_1 X_i\right)\right] \\ &= E\left(\sum e_i^2\right) + E\left\{\sum\left[\left(b_0 - \beta_0\right) + \left(b_1 - \beta_1\right)X_i\right]^2\right\} + 2E\left(\sum e_i \hat{Y}_i\right) - 2E\left[\sum e_i\left(\beta_0 + \beta_1 X_i\right)\right] \\ &= E\left(\sum e_i^2\right) + E\left\{\sum\left[\left(b_0 - \beta_0\right) + \left(b_1 - \beta_1\right)X_i\right]^2\right\} \\ &= E\left(\sum e_i^2\right) + \sum\left[\mathrm{Var}\left(b_0\right) + \mathrm{Var}\left(b_1\right)X_i^2 + 2X_i \mathrm{Cov}\left(b_0, b_1\right)\right] \end{aligned} \tag{2.41}$$

entonces, al sustituir los valores $\mathrm{Var}(b_0) = \left[\dfrac{\sum X_i^2}{n\sum(X_i - \overline{X})^2}\right]\sigma^2$, $\mathrm{Var}(b_1) = \left[\dfrac{1}{\sum(X_i - \overline{X})^2}\right]\sigma^2$ y

$$\mathrm{Cov}(b_0, b_1) = \mathrm{Cov}(\overline{Y} - b_1\overline{X}, b_1)$$

$$= \mathrm{Cov}(\overline{Y}, b_1) - \mathrm{Cov}(b_1\overline{X}, b_1)$$

$$= -\overline{X}\,\mathrm{Var}(b_1)$$

$$= \left[\frac{-\sum X_i}{n\sum(X_i - \overline{X})^2}\right]\sigma^2 \tag{2.42}$$

se sigue que la esperanza de la suma de los errores al cuadrado de la expresión (2.41) es igual a la suma de la esperanza de los errores al cuadrado, cada uno de los cuales es igual a la varianza del error, por lo cual esa esperanza de la suma de errores al cuadrado es n veces la varianza del error..

Así pues

$$n\sigma^2 = E\left(\sum e_i^2\right) + \sigma^2 \sum\left[\frac{\sum X_i^2/n + X_i^2 - 2X_i \sum X_i/n}{\sum(X_i - \overline{X})^2}\right]$$

$$= E\left(\sum e_i^2\right) + 2\sigma^2 \tag{2.43}$$

de donde se sigue que

$$E\left(\sum e_i^2\right) = (n - 2)\sigma^2. \tag{2.44}$$

Por consiguiente, lo que se obtiene es

$$E\left(\frac{SCR}{g.l.}\right) = \sigma^2, \tag{2.45}$$

es decir,

$$S^2 = \hat{\sigma}^2 = \frac{SCR}{g.l.} = \frac{SCR}{n-2} \tag{2.46}$$

es insesgado. Dicho estimador sirve entonces para obtener

$$\widehat{Var}(b_1) = \frac{S^2}{\Sigma(X_i - \overline{X})^2} \tag{2.47}$$

e igualmente

$$\widehat{Var}(b_0) = \frac{S^2 \Sigma X_i^2}{n\Sigma(X_i - \overline{X})^2} \quad . \tag{2.48}$$

Mientras que

$$S = \sqrt{\frac{SCR}{n - 2}} \tag{2.49}$$

es el **error estándar de la regresión**.

2.5 Pronóstico

La fórmula del pronóstico puntual es

$$\hat{Y}_i = \overline{Y} + b_1(X_i - \overline{X}) . \tag{2.50}$$

Lo que se pronostica con $\hat{Y}_i$ es el valor medio de Y, para un valor dado X_i . De hecho, para un valor dado X_0 se tiene que

$$\hat{Y}_0 = b_0 + b_1 X_0 = \overline{Y} + b_1\left(X_0 - \overline{X}\right) \tag{2.51}$$

así que $\hat{Y}_0$ es una variable aleatoria.

Esto en realidad lo que indica es que para cada X_0 se tiene toda una distribución de valores de Y_0 , con media $E(Y_0) = \beta_0 + \beta_1 X_0$, de forma que $\hat{Y}_0 = \hat{E}\left(Y_0\right)$, como se muestra en la Figura 2.9.

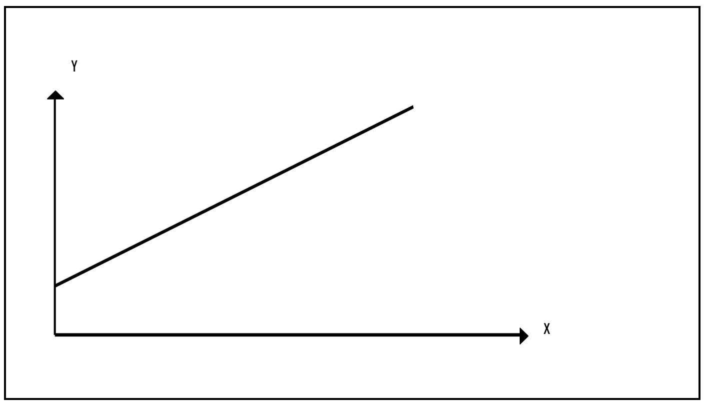


Figura 2.9. Pronósticos con el modelo de regresión lineal simple, con β_0 y β_1 conocidos.

En esta gráfica, las distribuciones tienen la media que sigue y la varianza constante,

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{y} \quad Var(Y_i) = \sigma^2 , \tag{2.52}$$

debido a que β_0 y β_1 están fijos, por lo cual **las distribuciones tienen la misma forma**. Si en lugar de β_0 y β_1 se usaran b_0 y b_1, la varianza cambiaría, como se verá más adelante.

Como $\hat{Y}_0$ es una variable aleatoria, tiene asociada una varianza, que en este caso resulta ser

$$Var\left(\hat{Y}_0\right) = Var\left(\overline{Y}\right) + \left(X_0 - \overline{X}\right)^2 Var\left(b_1\right) + 2Cov\left[\overline{Y}, b_1\left(X_0 - \overline{X}\right)\right]$$

$$= \frac{\sigma^2}{n} + \left(X_0 - \overline{X}\right)^2 \left[\frac{\sigma^2}{\sum \left(X_i - \overline{X}\right)^2} \right] \tag{2.53}$$

la cual será mínima cuando $X_0 = \overline{X}$. El efecto de varianzas distintas para pronósticos con parámetros estimados se aprecia en la Figura 2.10.

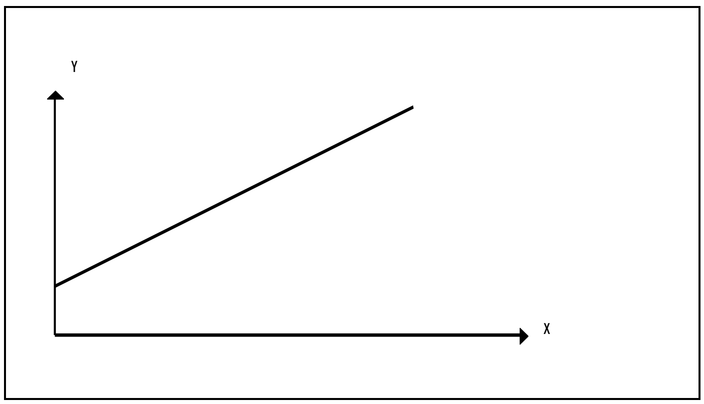


Figura 2.10. Ilustración del efecto de varianza cambiante

Si se desea pronosticar un "valor individual" de Y para $X = X_0$, también se usa

$$\hat{Y}_0 = \overline{Y} + b_1(X_0 - \overline{X}) \tag{2.54}$$

pero ahora la varianza del error de este pronóstico, o sea de $e_0 = Y_0 - \hat{Y}_0$, se obtiene como

$$\mathrm{Var}(e_0) = \mathrm{Var}(Y_0) + \mathrm{Var}(\hat{Y}_0) - 2\mathrm{Cov}(Y_0, \hat{Y}_0)$$

$$= \sigma^2 + \mathrm{Var}\left(\hat{Y}_0\right)$$

$$= \sigma^2 \left[1 + \frac{1}{n} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum \left(X_i - \overline{X}\right)^2} \right] . \tag{2.55}$$

Nótese que $\mathrm{Cov}(Y_0, \hat{Y}_0) = 0$ se sigue del hecho de que $\hat{Y}_0$ es función lineal de $Y_1,...,Y_n$ y de que Y_0 no está correlacionado con esas Y's

Ejemplo: Con los datos numéricos, se sabe que

$$b_0 = 1,\ \ b_1 = 1.75,\ \ \sum x_i^2 = 40,\ \ \sum e_i^2 = 1.5 \ \text{ y } \ S^2 = 0.5$$

Entonces

$$\hat{\mathrm{V}}\mathrm{ar}(b_1) = S^2/\sum x_i^2 = 0.5/40 = 0.0125 \quad \Rightarrow \quad \hat{\mathrm{e}}\mathrm{e}(b_1) = \sqrt{0.0125} = 0.1118$$

$$\hat{\mathrm{V}}\mathrm{ar}(b_0) = S^2\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum x_i^2}\right) = 0.5\left(\frac{1}{5} + \frac{16}{40}\right) = 0.3 \quad \Rightarrow \quad \hat{\mathrm{e}}\mathrm{e}(b_0) = \sqrt{0.3} = 0.5477$$

donde los errores estándar (ee) de los parámetros estimados se obtienen como la raíz cuadrada de las respectivas varianzas.

Además se tiene que

$$\hat{\mathrm{V}}\mathrm{ar}(Y_0) = 0.5\left[\frac{1}{5} + \frac{(X_0 - 4)^2}{40}\right]$$

de aquí, si se desea un pronóstico de Y para $X_0 = 10$, se obtiene

$$\hat{Y}_0 = 1 + 1.75(10) = 18.5$$

con $\hat{\mathrm{V}}\mathrm{ar}\left(\hat{Y}_0\right) = = 0.5\left(\frac{1}{5} + \frac{36}{40}\right) = 0.55$ y $\hat{\mathrm{V}}\mathrm{ar}\left(e_0\right) = S^2 + \hat{\mathrm{V}}\mathrm{ar}\left(\hat{Y}_0\right) = 0.5 + 0.55 = 1.05$

2.6 Estimación por Máxima Verosimilitud

Si se supone que los errores del modelo de regresión lineal simple son normales, entonces se tiene que

$$Y_i \sim N(\beta_0 + \beta_1 X_i\,,\,\sigma^2) \quad \text{con } Y_i \text{ y } Y_j \text{ independientes si } i \neq j. \tag{2.56}$$

Se tiene pues que la funcion de densidad para cada Y_i es

$$f\,(Y_i) = \frac{1}{\sqrt{2\Pi}\sigma}\, e^{-\frac{1}{2\sigma^2}\left(Y_i - \beta_0 - \beta_1 X_i\right)^2} \quad \text{para } i = 1,\dots, n \tag{2.57}$$

de tal forma que la función de densidad conjunta resulta ser

$$L = \prod_{i=1}^{n} f\,(Y_i) = \left(\sqrt{2\pi}\sigma\right)^{-n} e^{-\frac{1}{2\sigma^2}\sum\left(Y_i - \beta_0 - \beta_1 X_i\right)^2} \tag{2.58}$$

cuando β_0, β_1 y σ^2 son conocidos.

En cambio, cuando lo que se conoce son los valores de las Y’s, L es una función de verosimilitud, que debe ser maximizada en función de los parámetros desconocidos β_0 , β_1 y σ^2.

Nótese que al maximizar L respecto a β_0 y β_1, esencialmente se minimiza $\Sigma\left(Y_i - \beta_0 - \beta_1 X_i\right)^2$, en consecuencia los estimadores de MCO resultan ser también de MV, cuando la normalidad se cumple. Esto implica en particular, que los estimadores b_0 y b_1 tendrán las propiedades de los estimadores de MV, o sea, serán consistentes, asintóticamente eficientes, invariantes bajo transformaciones uno a uno y estarán en función de estadísticos suficientes.

Por otro lado, el estimador de MCO de σ^2 se obtiene al igualar a cero la siguiente derivada parcial, evaluada en $\beta_0 = b_0$, $\beta_1 = b_1$ y $\sigma^2 = \tilde{\sigma}^2$

$$\frac{\partial \log(L)}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum\left(Y_i - \beta_0 - \beta_1 X_i\right)^2 = 0 \qquad (2.59)$$

lo cual da por resultado

$$\tilde{\sigma}^2 = \sum\left(Y_i - b_0 - b_1 X_i\right)^2 \big/ n = (n\text{-}1/n)\ S^2 \qquad (2.60)$$

el cual tiene sesgo y por ello se descarta, así que se prefiere trabajar con el estimador insesgado S2.

2.7 Inferencia estadística

Debido a que b_0 y b_1 son combinaciones lineales de las Y's, se sigue que, al distribuirse los errores del modelo como una normal, entonces

$$b_i \sim N(\ \beta_i\ ,\ Var(b_i)\ ) \quad \text{para } i = 0, 1. \qquad (2.61)$$

Esto permite hacer inferencia sobre β_0 y β_1. Para ello se sabe que

$$\frac{b_i - \beta_i}{\sqrt{Var(b_i)}} \sim N(\ 0, 1\ ), \qquad i = 0, 1. \qquad (2.62)$$

con la varianza del estimador conocida.

Además, se usará el hecho de que

$$\frac{\sum e_i^2}{\sigma^2} = \frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{(n-2)} \qquad (2.63)$$

con $\sum e_i^2$ una variable aleatoria independiente de b_0 y b_1.

De esta forma se obtiene, para b_1

$$t = \underbrace{\left(\frac{b_1 - \beta_1}{\sigma / \sqrt{\sum x_i^2}} \right)}_{N(0,1)} \Big/ \underbrace{\sqrt{\frac{(n-2)S^2}{\sigma^2} / (n-2)}}_{\sqrt{\frac{\chi^2_{(g.l.)}}{g.l.}}}$$

$$= \frac{b_1 - \beta_1}{S / \sqrt{\sum x_i^2}} \sim t_{(n-2)} \qquad (2.64)$$

Es decir, se tiene un cociente de una variable Normal estandarizada entre la raíz cuadrada de una variable χ^2 con n-2 grados de libertad, dividida entre sus grados de libertad, lo cual proporciona una distribución t de Student con n-2 grados de libertad. De manera similar se obtiene la conclusión para b_0, que conduce al resultado general

$$\frac{b_i - \beta_i}{\hat{e}e(b_i)} \sim t_{(n-2)} \quad \text{para } i = 0, 1 \qquad (2.65)$$

donde

$$\hat{e}e(b_0) = S\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}} \quad \text{y} \quad \hat{e}e(b_1) = S\sqrt{\frac{1}{\sum x_i^2}}\,. \qquad (2.66)$$

Por lo tanto, a manera de ejemplo, un intervalo de 100(1-α)% de confianza para β_1 estará dado por

$$\beta_1 \in \left\{ b_1 \pm t_{(n-2),\alpha/2}\, \hat{e}e(b_1) \right\} \qquad (2.67)$$

Conviene recordar que la forma de la distribución t de Student es como se muestra en la Figura 2.11.

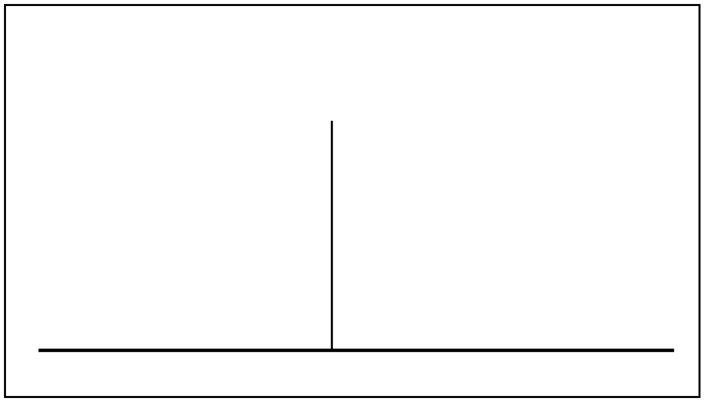


Figura 2.11. Forma de la ditribución $t_{(n-2)}$

También se pueden construir intervalos de confianza para los pronósticos de valores medios, con la expresión siguiente

$$E(Y_0) \in \left\{ \hat{Y}_0 \pm t_{(n-2),\alpha/2}\, \hat{e}e\left(\hat{Y}_0\right) \right\} \qquad (2.68)$$

que proporciona una confianza del 100(1-α)%.

De igual forma pueden calcularse intervalos de predicción para valores individuales de Y, mediante

$$Y_0 \in \left\{ \hat{Y}_0 \pm t_{(n-2),\alpha/2}\, \hat{e}e\left(e_0\right) \right\} \;. \qquad (2.69)$$

De esta forma se pueden calcular las llamadas bandas de pronóstico de la Figura 2.12, que deben ser interpretadas como intervalos de predicción para un pronóstco a la vez.

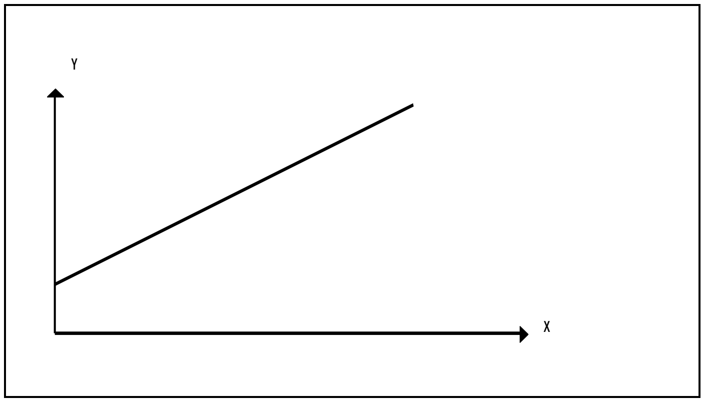


Figura 2.12. Bandas de predicción para los pronósticos de valores individuales

Por otro lado, también se pueden realizar pruebas de hipótesis del tipo

$$H_0 : \beta_1 = \beta_1^* \quad \text{vs.} \quad H_A : \beta_1 \neq \beta_1^* \tag{2.70}$$

en donde H_0 representa la hipótesis de trabajo o nula (lo que se desea rechazar) y H_A la hipótesis alternativa, mientras que β_1^* es un valor dado que resulta de interés. Para ello se usa el estadístico de prueba

$$\frac{b_1 - \beta_1^*}{\hat{e}e(b_1)} \sim t_{(n-2)} \tag{2.71}$$

y, a un nivel α de significancia, se declara un rechazo de H_0 cuando

$$\left| \frac{b_1 - \beta_1^*}{\hat{e}e(b_1)} \right| > t_{(n-2),\alpha/2}. \tag{2.72}$$

2.8 Análisis de varianza

La fórmula del análisis de varianza (ANOVA por sus siglas en inglés) surge de la descomposición de la discrepancia residual mostrada en la Figura 2.13

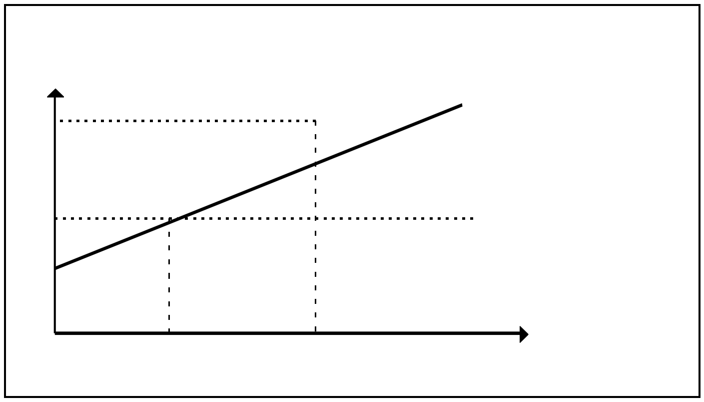


Figura 2.13. Descomposición del residuo

que se expresa algebráicamente como sigue

$$(Y_i - \hat{Y}_i) = (Y_i - \overline{Y}) - (\hat{Y}_i - \overline{Y}) \tag{2.73}$$

La ecuación anterior conduce a

$$
\begin{aligned}
\sum\left(Y_i - \hat{Y}_i\right)^2 &= \sum\left(Y_i - \overline{Y}\right)^2 - 2\sum\left(Y_i - \overline{Y}\right)\left(Y_i - \hat{Y}_i\right) + \sum\left(Y_i - \hat{Y}_i\right)^2 \\
&= \sum\left(Y_i - \overline{Y}\right)^2 - 2\sum\left(Y_i - \overline{Y}\right)b_1\left(X_i - \overline{X}\right) + \sum\left(Y_i - \hat{Y}_i\right)^2 \\
&= \sum\left(Y_i - \overline{Y}\right)^2 - 2b_1^2\sum\left(X_i - \overline{X}\right)^2 + \sum\left(Y_i - \hat{Y}_i\right)^2 \\
&= \sum\left(Y_i - \overline{Y}\right)^2 - 2\sum\left(\hat{Y}_i - \overline{Y}\right)^2 + \sum\left(Y_i - \hat{Y}_i\right)^2 \\
&= \sum\left(Y_i - \overline{Y}\right)^2 - \sum\left(\hat{Y}_i - \overline{Y}\right)^2 \qquad (2.74)
\end{aligned}
$$

así que

$$\Sigma\left(Y_i - \overline{Y}\right)^2 = \Sigma\left(Y_i - \hat{Y}_i\right)^2 + \Sigma\left(\hat{Y}_i - \overline{Y}\right)^2 \qquad (2.75)$$

$$SCT = SCR + SCE$$

Suma de cuadrados total	=	Suma de cuadrados alrededor de la regresión	+	Suma de cuadrados debida a la regresión
(alrededor de $\overline{Y}$)		(sin explicar)		(explicada)

Al aplicar el análisis de regresión, lo ideal sería que SCE fuera mucho mayor que SCR, pues de esta manera se estaría diciendo que la parte explicada por el modelo es mayor que la parte no explicada por éste, es decir,

$$\frac{SCE}{SCT} \approx 1 \quad \text{o bien} \quad \frac{SCR}{SCT} \approx 0. \qquad (2.76)$$

Otra medida de bondad de ajuste es el **Coeficiente de Determinación**, el cual se define de la siguiente forma

$$R^2 = 1 - \frac{SCR}{SCT}$$

$$= \frac{SCE}{SCT} = \frac{\Sigma\left(\hat{Y}_i - \overline{Y}\right)^2}{\Sigma\left(Y_i - \overline{Y}\right)^2} = \frac{\Sigma \hat{y}_i^2}{\Sigma y_i^2} = \frac{\sum\left(b_1 x_i\right)^2}{\sum y_i^2} = b_1^2 \frac{\sum x_i^2}{\sum y_i^2}. \qquad (2.77)$$

Cabe hacer mención que en regresión lineal simple, R corresponde al coeficiente de correlación previamente mencionado, es decir $R = r_{XY}$.

A su vez, se puede obtener el **Coeficiente de Determinación Ajustado por Grados de Libertad**, que sirve para cancelar la influencia que ejerce el número de variables independientes sobre R^2 y que está dado por

$$\overline{R}^2 = 1 - \frac{SCR/(n-2)}{SCT/(n-1)} \tag{2.78}$$

Este último coeficiente es útil principalmente en regresión lineal múltiple y su interpretación surge de la siguiente expresión

$$\overline{R}^2 = 1 - \frac{\widehat{Var}(\varepsilon)}{\widehat{Var}(Y)} \tag{2.79}$$

En el Cuadro 2.3 se muestra la Tabla de ANOVA, la cual es útil para resumir la información acerca de la descomposición de la varianza en sus diferentes fuentes.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F
Regresión	SCE	1	CME = SCE	
Residual	SCR	n-2	$CMR = \frac{SCR}{n-2}$	$F = \frac{CME}{CMR}$
TOTAL	SCT	n-1		

Cuadro 2.3. Tabla de ANOVA.

Debido al supuesto de normalidad, se sabe que

$$\frac{SCT}{\sigma^2} \sim \chi^2_{n-1} \tag{2.80}$$

y

$$\frac{SCR}{\sigma^2} \sim \chi^2_{n-2}\,. \tag{2.81}$$

Además, **si la hipótesis H_0: $\beta_1 = 0$ es cierta**, se tiene que

$$\frac{SCE}{\sigma^2} \sim \chi^2_{1} \tag{2.82}$$

con SCE y SCR independientes.

Por lo mencionado anteriormente, si $\beta_1 = 0$, se concluye que

$$F = \frac{SCE/\sigma^2 \times 1}{SCR/\sigma^2 \times (n-2)} = \frac{CME}{CMR} \sim F_{(1,n-2)} \tag{2.83}$$

De esta manera, el estadístico F de la tabla ANOVA sirve para probar la hipótesis nula

$$H_0: \beta_1 = 0 \qquad \text{vs.} \qquad H_A: \beta_1 \neq 0. \tag{2.84}$$

Para ello se compara contra valores de tablas de puntos porcentuales $F_{(1,n-2),\alpha}$, en donde α es el nivel de significancia (la forma de la distribución F se muestra en la Figura 2.14). El procedimiento señalado equivale a probar

$$H_0: Y_i = \beta_0 + \varepsilon_i \quad \text{vs.} \quad H_A: Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{2.85}$$

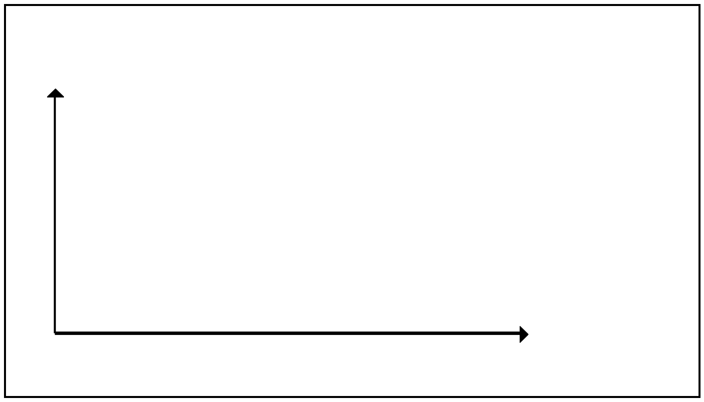


Figura 2.14. Forma de la distribución $F_{(1,n-2)}$.

Ahora se verá un ejemplo numérico con los datos numéricos utilizados previamente.

Ejemplo: Con los datos numéricos, se sabe que

$$b_0 = 1,\ \ b_1 = 1.75,\ \ S^2 = 0.5\,,\ \ \hat{e}e(b_1) = 0.1118 \text{ y } \ \hat{e}e(b_0) = 0.5477$$

de manera que un intervalo del 95% de confianza para β_1 está dado por

$$1.75 \pm t_{(3),0.025}(0.1118) = 1.75 \pm 3.182(0.1118) = (1.39,\ 2.11)$$

y la hipótesis $H_0: \beta_0 = 0$ vs. $H_A: \ \beta_0 \neq 0$ se prueba con el estadístico

$$t = \frac{b_0 - 0}{\hat{e}e(b_0)} = \frac{1}{0.5477} = 1.826 < t_{(3),0.025} = 3.182$$

el cual indica que la ordenada al origen (o intercepto) no es significativamente distinta de cero, al nivel de significancia del 5%.

Por su lado, la correspondiente tabla de Análisis de Varianza resulta ser

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F
X	SCE = 122.5	1	122.5	
Residual	SCR = 1.5	3	0.5	F = 245
Total	SCT = 124	4	--	

Al comparar el estadístico calculado $F = 122.5/0.5 = 245.0$ con $F_{(1,3),0.05} = 10.1$ se concluye el rechazo de $H_0\colon \beta_1 = 0$ al nivel de significancia del 5%. En otras palabras, β_1 se puede considerar diferente de 0 y por ello se dice que la regresión que usa a X como variable explicativa, resulta ser significativa al 5%.

Además, el coeficiente de determinación, que **indica la proporción de variabilidad de Y explicada por el modelo** es

$$R^2 = \frac{SCE}{SCT} = \frac{122.5}{124} = 0.988$$

Asimismo, el coeficiente de determinación ajustado por grados de libertad toma el valor

$$\overline{R}^2 = 1 - \frac{SCR/(n-2)}{SCT/(n-1)} = 1 - \frac{0.5}{31} = 0.984$$

que resulta ser menor que R^2, como era de esperar, por la penalización involucrada.

2.9 Análisis de residuos

El análisis de residuos sirve para verificar los supuestos del modelo. Como se supuso que

$$\varepsilon_i \sim N(0,\sigma^2) \tag{2.86}$$

es de esperar entonces que, aproximadamente,

$$e_i \sim N(0,S^2) \quad \text{o} \quad \frac{e_i}{S} \sim N(0,1)\,. \qquad (2.87)$$

Se sugiere entonces realizar gráficas de los residuos estandarizados, para tratar de identificar patrones como los mostrados en las figuras 2.15 a 2.18.

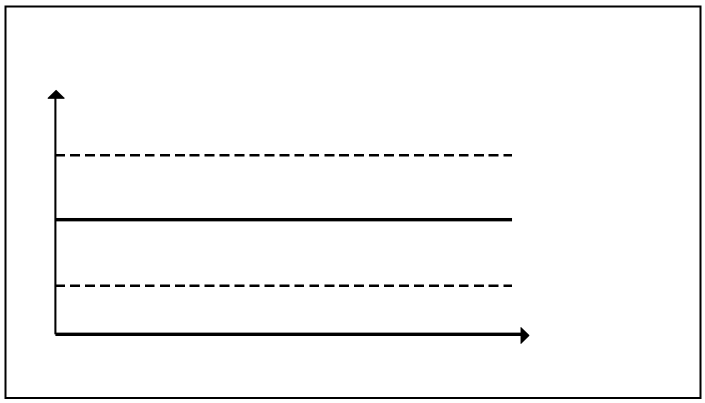


Figura 2.15. Gráfica de residuos con patrón adecuado

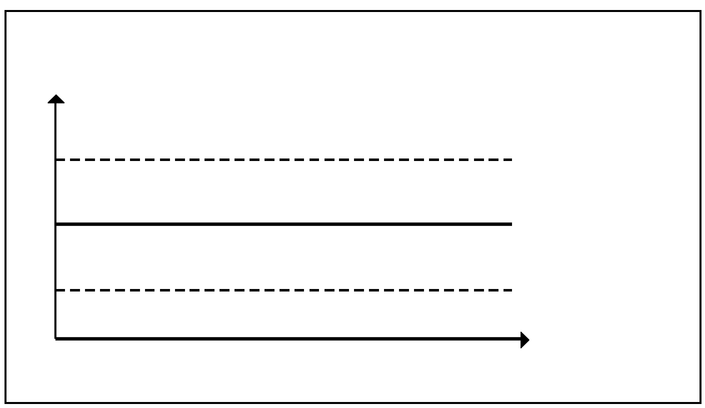


Figura 2.16. Gráfica de residuos con efecto lineal remanente

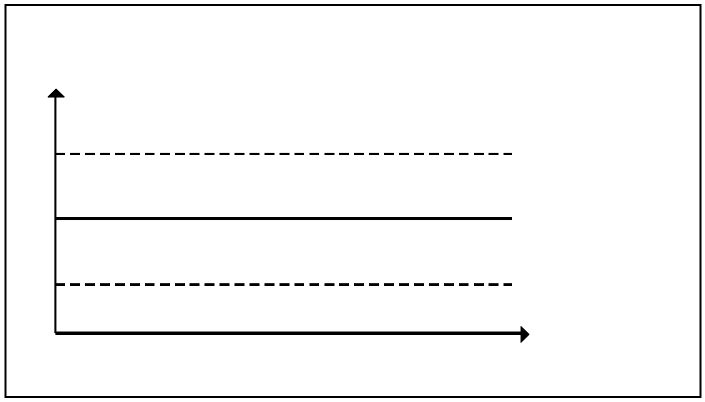


Figura 2.17. Gráfica de residuos con efecto cuadrático

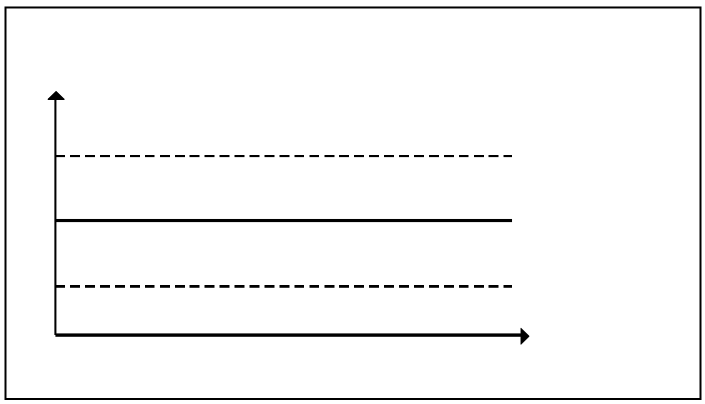


Figura 2.18. Gráfica de residuos con varianza no-constante

En particular, no conviene graficar los residuos contra las observaciones de Y, porque es previsible que exista correlación positiva entre estas variables, según se observa en lo que sigue (si el modelo contiene ordenada al origen)

$$r_{e\,Y} = \frac{\sum (e_i - \bar{e})(Y_i - \bar{Y})}{\sqrt{SCR}\sqrt{SCT}}$$

$$= \frac{\sum e_i Y_i}{\sqrt{SCR\,SCT}}$$

$$= \frac{\sum\left(Y_i - \hat{Y}_i\right)\left(Y_i - \hat{Y}_i\right)}{\sqrt{SCR\ SCT}} \quad \text{porque} \sum\left(Y_i - \hat{Y}_i\right)\hat{Y}_i = 0$$

$$= \frac{SCR}{\sqrt{SCR\ SCT}}$$

$$= \sqrt{SCR/SCT}$$

$$= \sqrt{1 - R^2} \;>\; 0 \qquad \text{(excepto cuando } R^2 = 1\text{)} . \tag{2.88}$$

Estadísticos que presenta la salida del paquete E-Views

Logaritmo de la Verosimilitud (log-likelihood)

Se calcula en el supuesto de Normalidad para los errores y se evalúa en los valores estimados de los parámetros. Como

$$L = (2\pi)^{-n/2}\hat{\sigma}^{-n}\exp[-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2] \quad \text{con} \quad \hat{\sigma}^2 = SCR/n$$

$$= (2\pi)^{-n/2}(SCR/n)^{-n/2}\exp[-\frac{1}{2SCR/n}SCR]$$

entonces se obtiene

$$\log(L) = (-n/2)\log(2\pi) - (n/2)\log(SCR/n) - n/2]$$

$$= (-n/2)[\log(2\pi) + \log(SCR/n) + 1]$$

Este estadístico permite realizar Pruebas de Razón de Verosimilitudes al comparar la diferencia de valores de log(L) para dos modelos distintos.

Criterio de Información de Akaike (AIC)

El estadístico se calcula como

$$AIC = -2\log(L)/n + 2p/n \quad \text{con} \quad p = \text{número de parámetros en el modelo}$$

Con este estadístico se puede comparar dos modelos alternativos y se penaliza el que tenga más parámetros, pues se prefiere el modelo que proporcione el mínimo valor del AIC.

Criterio de Schwarz

En este caso el estadístico proporciona una forma que penaliza aún más a los modelos que contengan un mayor número de parámetros. Su uso es similar al del AIC y se calcula como

$$SC = -2\log(L)/n + [p\log(n)]/n \quad \text{con} \quad p = \text{número de parámetros en el modelo}$$

Criterio de Hannan y Quinn

$$HQ = -2\log(L)/n + 2p\log[\log(n)]/n$$