Econometría I (EC402)-II semestre de 2013
Clase #8 - Bondad de ajuste

Andrés M. Castaño

Ingeniería Comercial
Universidad Católica del Norte
Septiembre 23 de 2013

Precisión o errores estándar de los mínimos cuadrados estimados

- Varianza de los errores:
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \mu_i^2}{n-2}$$
$$\sum \hat{\mu_i}^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2$$

- $$Var(\hat{\beta}_1) = \frac{\sum X_i^2 \sigma^2}{N \sum x_i^2}$$

- $$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

- Qué son los grados de libertad?

Características

- $Var(\hat{\beta}_2)$ es directamente proporcional a $\hat{\sigma}^2$ pero inversamente proporcional a la $\sum x_i$.

- $Var(\hat{\beta}_1)$ es directamente proporcional a $\hat{\sigma}^2$ y a $\sum X_i$ pero inversamente proporcional a la $\sum x_i$ y al tamaño de la muestra.

- Los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$, varian entre cada muestra, tienen cierta dependencia:
$$Cov(\hat{\beta}_1, \hat{\beta}_2) = \bar{X} \frac{\sigma^2}{\sum x_i^2}$$

Bondad del Ajuste MCO

- Buscamos una medida que nos indique si el ajuste MCO de la nube de puntos es o no un buen ajuste.

- Es decir, una medida que nos indique el grado de importancia de las discrepancias (o residuos) entre los valores de la variable dependiente observados Y_i y los ajustados (o estimados) $\hat{Y}_i$.

- Sea la FRP o modelo econométrico.

$$Y_i = \beta_1 + \beta_2 X_i + \mu_i; i = 1, 2, ..., n$$

- IDEA: Distintos valores observados de Y_i se deben fundamentalmente a que X_i adopta distintos valores.

Bondad del Ajuste MCO

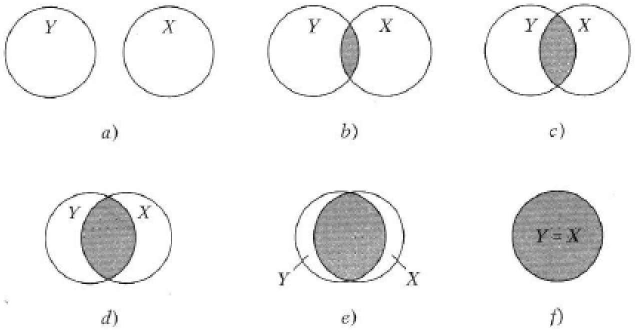
- En otras palabras: con el análisis de regresión lineal buscamos explicar la variabilidad de Y , entre distintos individuos (corte transversal) o distintos momentos en el tiempo (series de tiempo) a través de la variabilidad de X.

- Dado que el modelo incorpora μ_i, la variabilidad de Y incluye un elemento aleatorio (no explicado por X).

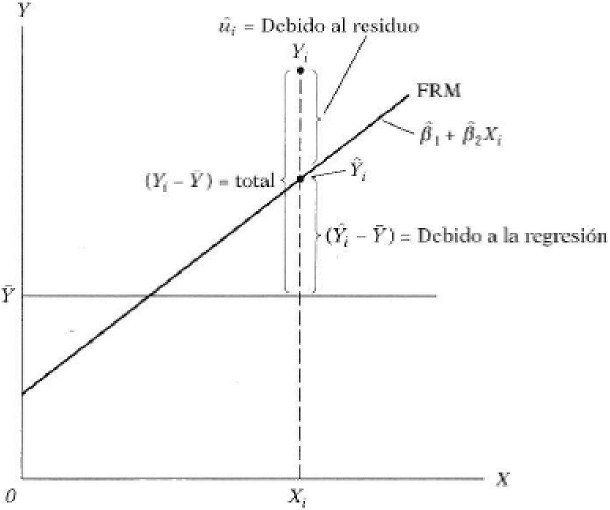
- Al estimar el modelo, ajustamos una recta MCO a una nube de puntos. El criterio tiene que ver con variabilidad: minimizamos la variabilidad de los residuos.

- Esto sugiere que la variabilidad total de las Yi observadas será solo en parte explicada por el modelo.

Bondad del Ajuste MCO



Bondad del Ajuste MCO



Bondad del Ajuste MCO

- La variabilidad total de la Y_i observadas se puede expresar a través de la suma total de cuadrados (STC) o suma cuadrática total (SCT):

$$SCT = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Descomponemos la variabilidad en:

$$STC = \sum_{i=1}^{n}(\hat{Y}_i + \mu_i - \bar{Y})^2 = \sum_{i=1}^{n}((\hat{Y}_i - \bar{Y}) + \mu_i)^2$$

$$= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y}) + \sum_{i=1}^{n}\mu_i^2 + 2\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})\mu_i$$

$$= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}\mu_i^2 + 2\sum_{i=1}^{n}\mu_i\hat{Y}_i - 2\bar{Y}\sum_{i=1}^{n}\mu_i$$

Bondad del Ajuste MCO

$$\textbf{SCT} = \underbrace{\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2}_{\substack{\text{SUMA CUADRÁTICA} \\ \text{EXPLICADA} \\ \textbf{SCE} \\ \text{(mide la variabilidad} \\ \text{de los valores} \\ \text{ajustados)}}} + \underbrace{\sum_{i=1}^{n} e_i^2}_{\substack{\text{SUMA CUADRÁTICA} \\ \text{DE LOS RESIDUOS} \\ \textbf{SCR} \\ \text{(mide la variabilidad} \\ \text{de los residuos)}}}$$

$$\textbf{SCT} = \textbf{SCE} + \textbf{SCR}$$

Puede ser escrito igual como:

$$STC = SEC + SRC$$

Bondad del Ajuste MCO

- Si se satisface el supuesto de que las X_i no son todas iguales, $STC > 0$ y por lo tanto.

$$\frac{STC}{STC} = \frac{SEC}{STC} + \frac{SRC}{STC}$$

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

- $\frac{SEC}{STC}$ es la proporción de la variabilidad total explicada por el modelo $= R^2$ o coeficiente de determinación:

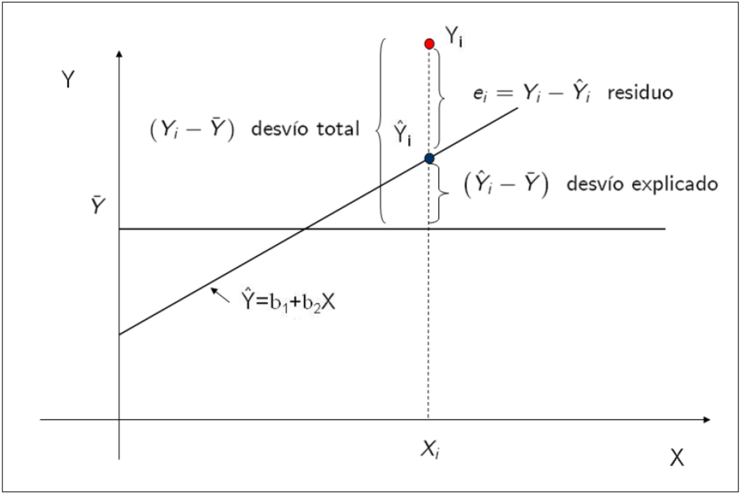
$$R^2 = \frac{SEC}{STC}$$

ó bien:

$$R^2 = 1 - \frac{SRC}{STC}$$

- $\frac{SRC}{STC}$ es la proporción de la variabilidad total NO explicada por el modelo.

Bondad del Ajuste MCO



Coeficiente de Determinación

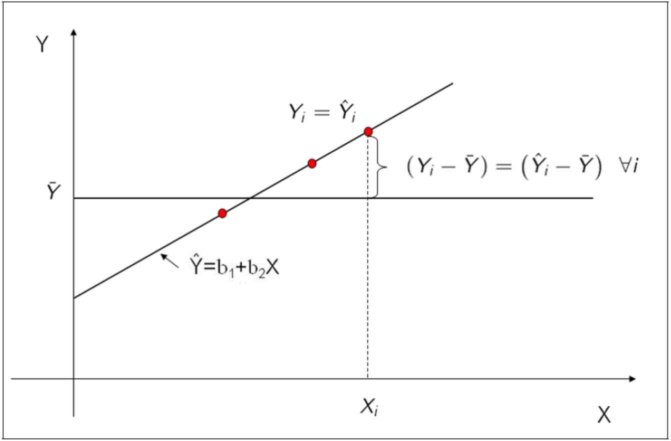
- El coeficiente de determinación R^2 es una medida de bondad del ajuste MCO.

Propiedades:

- $R2 \geq 0$, pues relaciona variabilidades (desvíos elevados al cuadrado).
- $0 \leq R2 \leq 1$ y $R^2 x100$ es el % de variabilidad de Y explicada por el modelo.

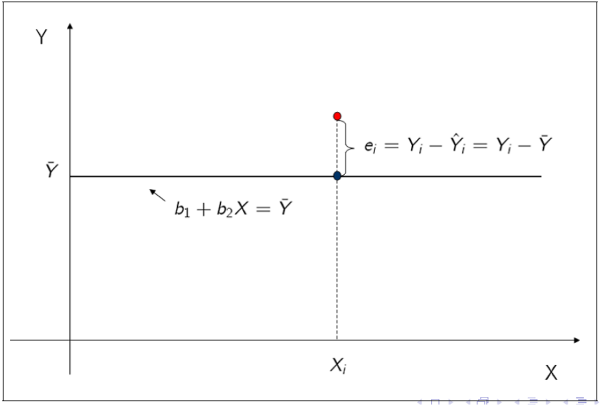
Coeficiente de Determinación

$R^2 = 1$. Todas las observaciones están sobre la recta MCO $\Longrightarrow$ $SEC = STC$ y $SRC = 0$, el ajuste es perfecto.



Coeficiente de Determinación

$R2 = 0$. La recta MCO es la recta $\bar{Y}$. $SRC = STC$ y $SEC = 0$. Conocer el valor de X no aporta ninguna información sobre la variabilidad de Y . El modelo no explica nada.



Coeficiente de Determinación

- En general, $0 < R2 < 1$ y a mayor R^2, mejor ajuste.

- El R^2 del modelo simple es igual al coeficiente de correlación muestral entre X_i y Y_i. Demostración (Tarea): El coeficiente de correlación muestral se define como:

$$r_{xy} = \frac{Cov_Muestral(X,Y)}{S_X S_Y} = \frac{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{n-1}}{\sqrt{S_X^2}\sqrt{S_Y^2}}$$

$$= \frac{S_{XY}}{S_X S_Y}$$

$$-1 \leq r_{XY} \leq 1$$

$$R^2 = 1 - \frac{SRC}{STC}$$

Patrones de correlación