PUBLICACIONES DE 4er CURSO

Curso: 4º

Grado: Economía

Asignatura: ECONOMETRÍA III

TRANSPARENCIAS PARTE 2: Regresores estocásticos
Tema 3: Modelo lineal simple

Profesores: Antonio Aznar

Departamento de ANÁLISIS ECONÓMICO

Curso Académico 2015/16

1542

Facultad de
Economía y Empresa
Universidad Zaragoza

TEMA 3: MODELOS LINEAL SIMPLE

1. Introducción
2. Estimación Mínimo Cuadrático Ordinaria (MCO)
3. Medidas de ajuste
4. Supuestos y propiedades de los MCO
5. Contrastes e intervalos de confianza

1. Introducción

***La línea de regresión poblacional*:**

$$NotaExamen = \beta_0 + \beta_1 REM + u_i$$

REM: ratio estudiantes maestros.

β_1 = pendiente de la línea de regresión poblacional

$$= \frac{\Delta NotaExamen}{\Delta REM}$$

= cambio en la nota debido a un cambio unitario en REM

- Nos gustaría conocer el valor poblacional de β_1.

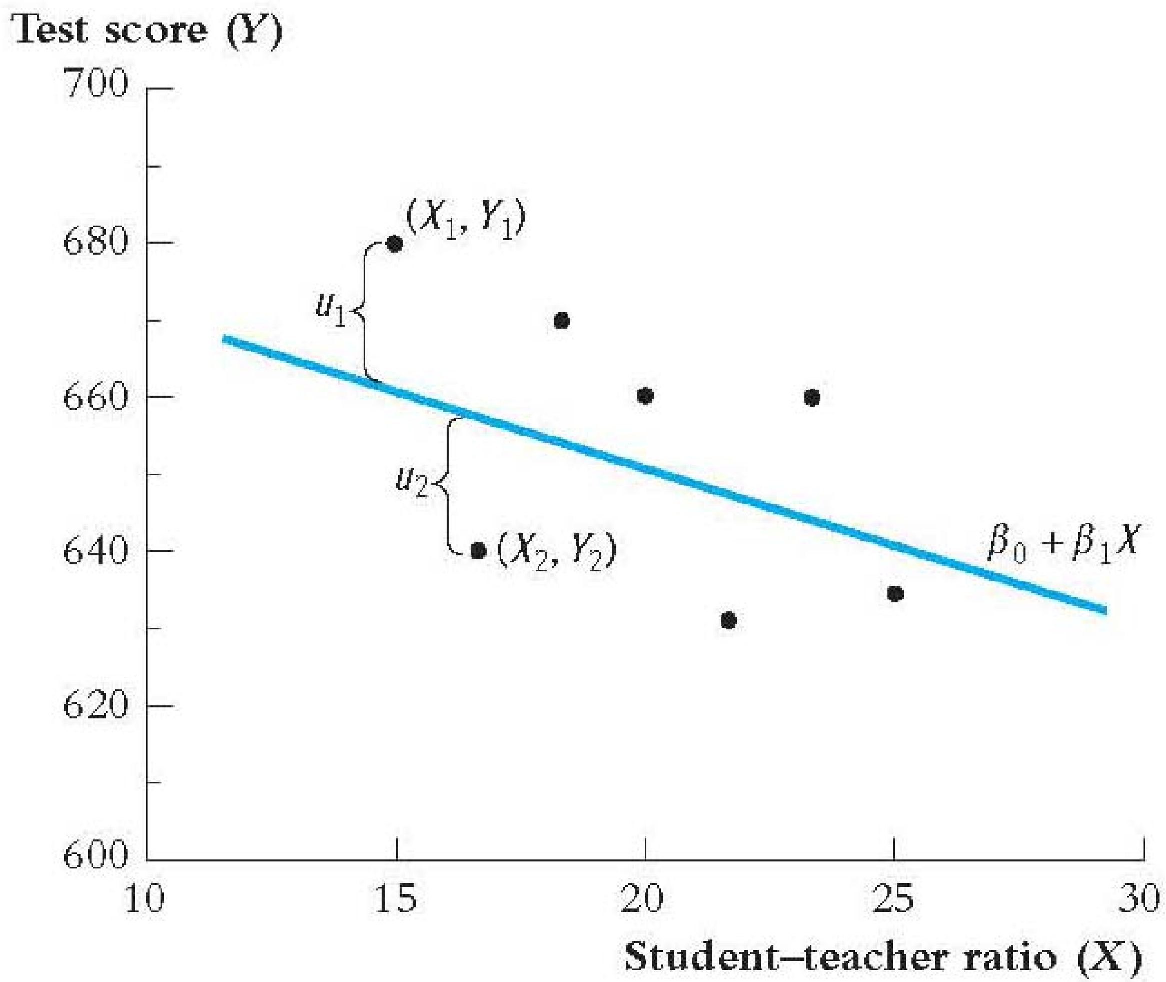
C.C. 4.1 El Modelo de Regresión Lineal Poblacional

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1,\ldots, n$$

- Tenemos n observaciones, (X_i, Y_i), $i = 1,.., n$.
- X es la ***variable independiente*** o ***regresor***
- Y es la ***variable dependiente***
- β_0 = constante
- β_1 = pendiente
- u_i = el error de regresión
- El error recoge los factores omitidos. En general, estos factores omitidos son otros factores que influyen en Y, diferentes a X. El error de regresión también incluye el posible error de observación de Y.

El modelo de regresión poblacional en un gráfico:

Observaciones de Y y X (n = 7); la línea de regresión poblacional; y el error de regresión (el “termino de error”):



2. Estimación de Mínimos cuadrados ordinarios (MCO)

¿Cómo podemos estimar β_0 y β_1 a partir de los datos?

Recordar que $\bar{Y}$ era el estimador mínimo-cuadrático de μ_Y: $\bar{Y}$ se obtiene resolviendo,

$$\min_m \sum_{i=1}^{n} (Y_i - m)^2$$

Los estimadores MCO de $\boldsymbol{\beta_0}$ **y** $\boldsymbol{\beta_1}$ resuelven

$$\min_{\beta_0, \beta_1} \sum_{1}^{n} \left[Y_i - \left(\beta_0 + \beta_1 X_i \right) \right]^2$$

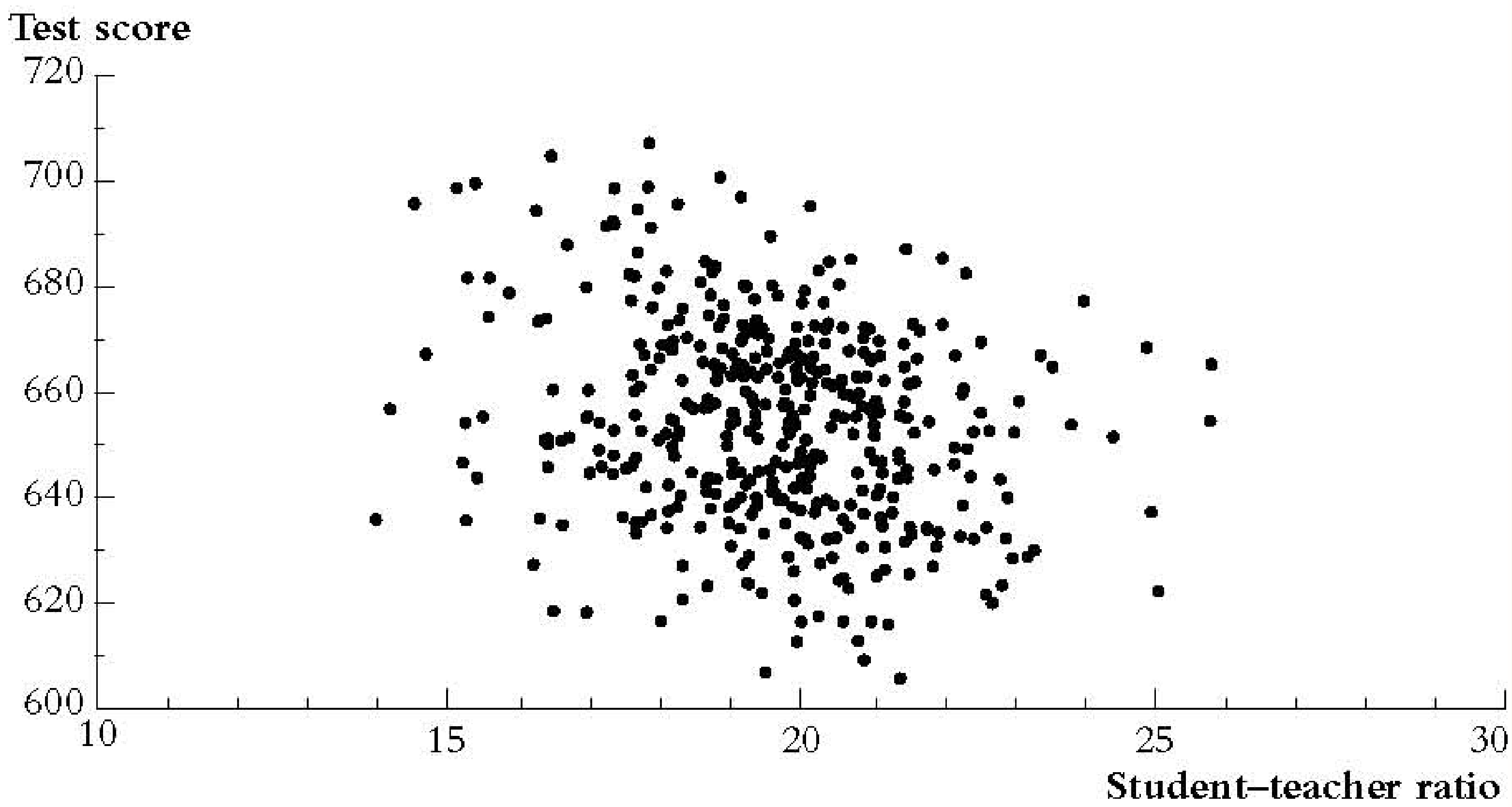
Ecuaciones Normales

Media: $2\sum_{1}^{n}\left(Y_i - \hat{m}\right) = 0 \Rightarrow \hat{m} = \frac{\sum Y_i}{n} = \overline{Y}$

Regresión: Derivando con respecto a β_0 y β_1 y reagrupando se obtiene

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{1}^{n} X_i = \sum_{1}^{n} Y_i$$

$$\hat{\beta}_0 \sum_{1}^{n} X_i + \hat{\beta}_1 \sum_{1}^{n} X_i^2 = \sum_{1}^{n} X_i Y_i$$


Test score
720
700
680
660
640
620
600
10
15
20
25
30
Student–teacher ratio

C.C.4.2 Los Estimadores MCO, Valores Predichos, y Residuos

Los estimadores MCO de los coeficientes son:

$$\hat{\beta}_1 = \frac{\sum_1^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_1^n (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

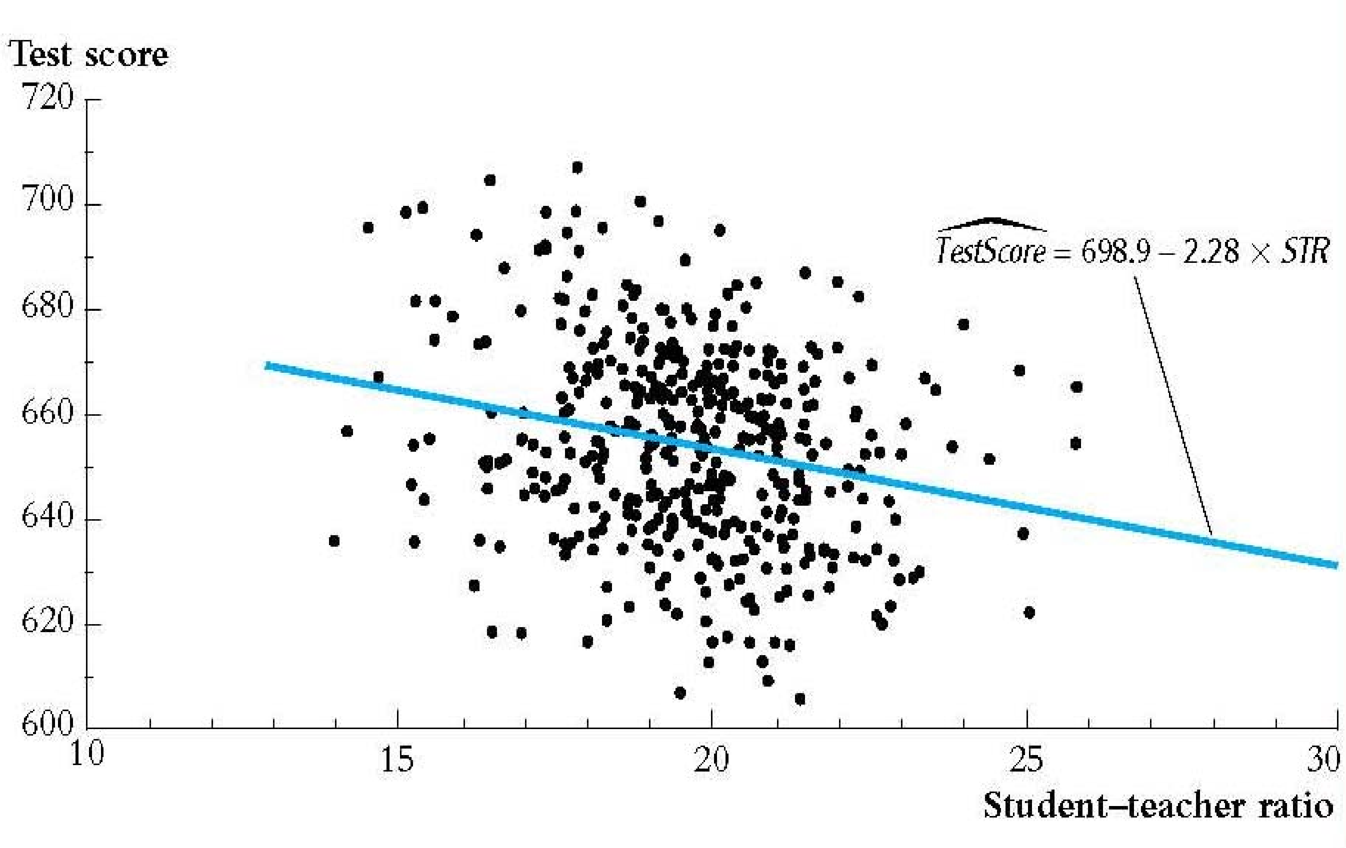
Los valores MCO predichos, $\hat{Y}_i$, y los residuos, $\hat{u}_i$ son

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad \text{i=1,2,...n}$$

$$\hat{u}_i = Y_i - \hat{Y}_i \quad \text{i=1,2,...n}$$

Las estimaciones se calculan a partir de n observaciones de las dos variables.

Aplicación al ejemplo de California: Calificación-Tamaño



Pendiente estimada $= \hat{\beta}_1 = -\ 2.28$

Constante estimada $= \hat{\beta}_0 = 698.9$

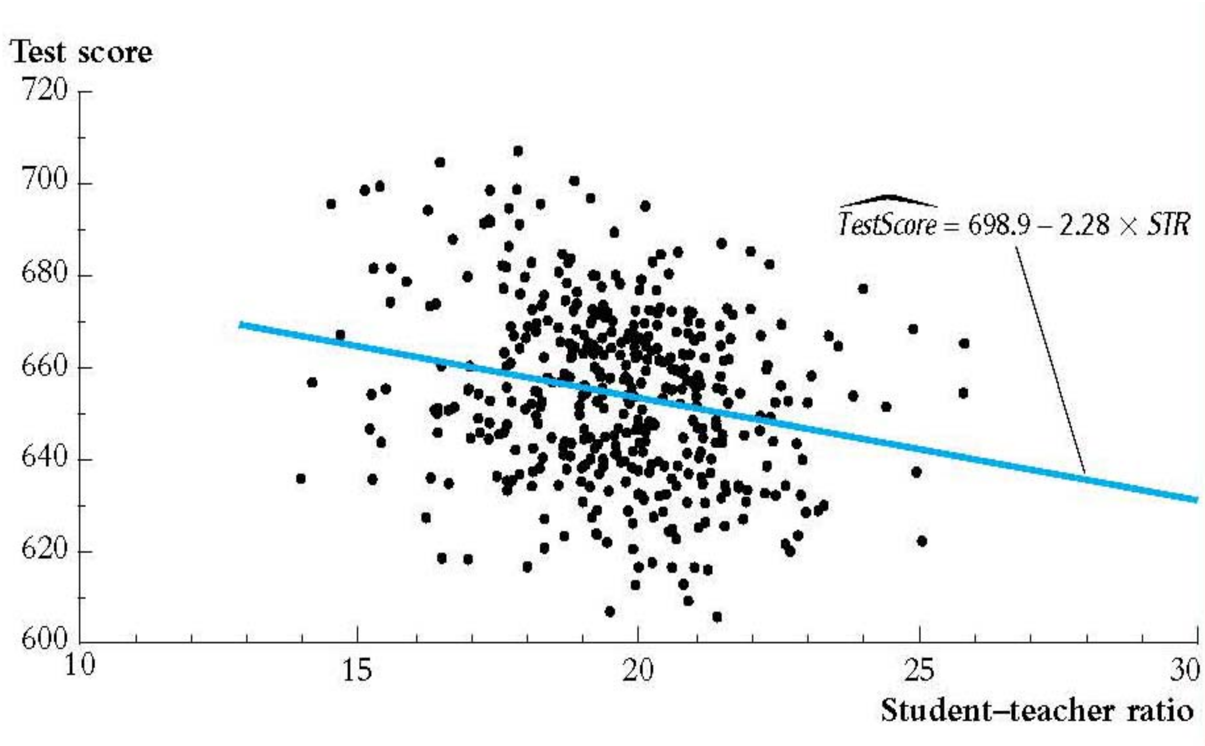
Línea de regresión esti.: $\widehat{NotaExamen} = 698.9 - 2.28 \times \mathrm{REM}$

Interpretación de las estimaciones

$$\widehat{NotaExamen} = 698.9 - 2.28 \times \text{REM}$$

- Los distritos con un estudiante más por profesor en promedio tienen una nota inferior en 2.28 .
- Esto es, $\dfrac{\Delta NotaExamen}{\Delta REM} = -2.28$
- Aquí, la constante no tiene una interpretación económica significativa.

Valores predichos y residuos:



Uno de los distritos en los datos Antelope, CA, para el que REM= 19.33 y *NotaExamen* = 657.8

Valor predicho: $\hat{Y}_{Antelope} = 698.9 - 2.28 \times 19.33 = 654.8$

residuo: $\hat{u}_{Antelope} = 657.8 - 654.8 = 3.0$

Regresión MCO (GRETL)

Modelo 1: MCO, usando las observaciones 1-420
Variable dependiente: NotaExamen

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	698.933	9.46749	73.82	6.57e-242
REM	-2.27981	0.479826	-4.751	2.78e-06

Media de la vble. dep.	654.1565	D.T. de la vble. dep.	19.05335
Suma de cuad. residuos	144315.5	D.T. de la regresión	18.58097
R-cuadrado	0.051240	R-cuadrado corregido	0.048970
F(1, 418)	22.57511	Valor p (de F)	2.78e-06
Log-verosimilitud	-1822.250	Criterio de Akaike	3648.499
Criterio de Schwarz	3656.580	Crit. de Hannan-	3651.6

$$\widehat{NotaExamen} = 698,93 - 2,28 REM$$ (Línea Estimada)

3. Medidas de Ajuste

Dos estadísticos para medir como la línea de regresión se ajusta a los datos:

- El $\boldsymbol{R^2}$ mide la fracción de la varianza de Y que se explica por X; no tiene unidades y su rango de variación va de cero (el peor ajuste) hasta uno (ajuste perfecto)

- El **error estándar de la regresión (ESR) o desviación típica** mide la magnitud del residuo tipo de la regresión en las unidades de Y.

El $\boldsymbol{R^2}$ es la fracción de la varianza muestral deY_i "explicada" por la regresión.

$Y_i = \hat{Y}_i + \hat{u}_i$ = predicción MCO +residuo MCO

$\Rightarrow$ Varianza muestral (Y) = varianza muestral($\hat{Y}_i$) +varianza muestral($\hat{u}_i$) (¿por qué?)

$\Rightarrow$ suma total de cuadrados = "explicada" SE + "residual" SR

Definición de R^2:

$$R^2 = \frac{SE}{ST} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{\hat{Y}})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

- $R^2 = 0$ significa $SE = 0$
- $R^2 = 1$ significa $SE = ST$
- $0 \le R^2 \le 1$
- Para la regresión con un solo regresor X, R^2 = el cuadrado del coeficiente de correlación entre X e Y

El error estándar de regresión (*ESR*) o desviación típica

El *ESR* mide la dispersión de la distribución de u. El E*SR* es (casi) la desviación estándar de los residuos MCO:

$$ESR = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(\hat{u}_i - \bar{\hat{u}})^2}$$

$$= \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}$$

Debido a que $\bar{\hat{u}} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$.

El *ESR*:

- Tiene las unidades de u, que son las unidades de Y.
- Mide el tamaño medio del residuo MCO (la "equivocación" media hecha por la línea de regresión MCO).
- El error cuadrático medio residual (*RMSE*) está estrechamente relacionado con el ESR:

$$\mathrm{RMS}E = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2}$$

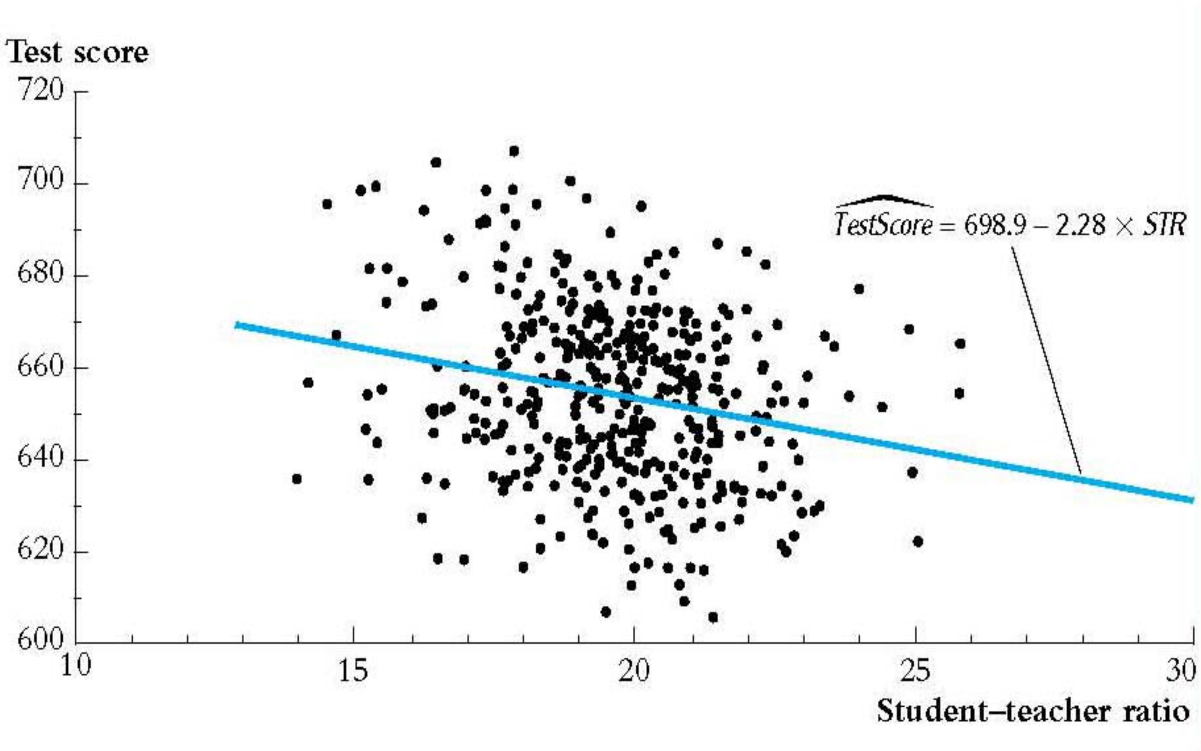
Mide lo mismo que el ESR – la diferencia es que divide por n en lugar de por (n–2).

Nota técnica: ¿Por qué divide por n–2 en lugar de por n–1?

$$\text{ES}R = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}$$

- La división por n–2 es una corrección por los grados de libertad – como la división por n–1 en s_Y^2, excepto que para ESR estimamos dos parámetros (β_0 y β_1, con $\hat{\beta}_0$ y $\hat{\beta}_1$), mientras que en s_Y^2 solamente se estima uno (μ_Y, con $\bar{Y}$).
- Cuando n es grande no importa si se divide por n, n–1, o n–2 .
- Para los detalles ver la Sección 17.4

Ejemplo de R^2 y de ESR



$$\widehat{NotaExamen} = 698.9 - 2.28 \times \text{REM}, \quad \mathbf{R^2 = .05,\ ESR = 18.6}$$

REM explica solamente una pequeña fracción de la variación en las notas. ¿Tiene esto sentido? ¿Significa esto que la REM no es importante para tomar decisiones de política?

4. Supuestos y propiedades de los MCO

¿Cuáles son las propiedades de la distribución muestral de los estimadores MCO? ¿Cuándo será $\hat{\beta}_1$ insesgado? ¿Cuál es su varianza?

Para contestar estas preguntas, necesitamos hacer algunas hipótesis acerca de como Y y X están relacionadas entre si, y acerca de como sus datos han sido recogidos (técnica de muestreo utilizada)

C.C. 4.3 Las hipótesis de los MCO (Regresor no estocástico)

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1,\ldots, n$$

1. $E(u_i) = 0$, i=1,2…n
2. $Y_i\ i = 1,2,...n$ son i.i.d. (Esfericidad)
   - $EY_iY_j = 0 \Rightarrow Eu_iu_j = 0$ (No autocorrelación)
   - $E(Y_i - EY_i)^2 = \sigma^2 \Rightarrow Eu_i^2 = \sigma^2$ (Homocedasticidad)
3. $u_i\ se\ distribuye: N\left(0,\sigma^2\right)$ (Normalidad)
4. No hay grandes atípicos ni en la variable dependiente ni en el regresor.
5. Para muestras grandes se cumple (Comentar)

$$\frac{1}{n}\sum_{1}^{n}\left(X_i - \bar{X}\right)^2 \to \sigma_X^2$$

- **La distribución muestral de $\hat{\beta}_1$**

Como $\bar{Y}$, $\hat{\beta}_1$ tiene una distribución muestral.

- ¿Qué es $E(\hat{\beta}_1)$?
    - Si $E(\hat{\beta}_1) = \beta_1$, entonces el MCO es insesgado –una Buena propiedad!
- ¿Qué es var($\hat{\beta}_1$)? (medida de incertidumbre muestral)
    - Necesitamos su formula para calcular el error estándar de $\hat{\beta}_1$.
- ¿Cuál es la distribución de $\hat{\beta}_1$ en muestras pequeñas y grandes?
    - $\hat{\beta}_1$ sigue una distribución normal

. Propiedades de los Estmadores MCO (Resuelto en cursos anteriores)

1. **Todo Tamaño Muestral**
    - Insesgadez: $E\left(\hat{\beta}_i\right) = \beta_i \quad i = 1, 2, ..k$
    - $\operatorname{var}\left(\hat{\beta}_1\right) = \dfrac{\sigma^2}{nS_1^2}, \; S_i^2 = \dfrac{1}{n}\sum_1^n\left(X_1 - \bar{X}_1\right)^2$
    - Los MCO son ELIO
2. **Muestras Grandes**
    - Consistencia: $\hat{\beta}_i \xrightarrow{p} \beta_i$
    - Asintóticamente Eficiente.

La media y varianza de la distribución muestral de $\hat{\beta}_1$

Podemos escribir:

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

$$\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u}$$

Por lo que $$Y_i - \bar{Y} = \beta_1(X_i - \bar{X}) + (u_i - \bar{u})$$

Así,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \bar{X})[\beta_1(X_i - \bar{X}) + (u_i - \bar{u})]}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\beta}_1 = \beta_1 \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Por lo que $\hat{\beta}_1 - \beta_1 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$.

Ahora $\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^{n}(X_i - \bar{X})u_i - \left[\sum_{i=1}^{n}(X_i - \bar{X})\right]\bar{u}$

$$= \sum_{i=1}^{n}(X_i - \bar{X})u_i - \left[\left(\sum_{i=1}^{n} X_i\right) - n\bar{X}\right]\bar{u}$$

$$= \sum_{i=1}^{n}(X_i - \bar{X})u_i$$

Sustituyendo $\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^{n}(X_i - \bar{X})u_i$ en la expresión de $\hat{\beta}_1 - \beta_1$:

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Por lo que

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Ahora calculamos *$E(\hat{\beta}_1)$ y var($\hat{\beta}_1$):*

$$E(\hat{\beta}_1) - \beta_1 = E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

$$= \left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})Eu_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right]$$

$= 0$ porque $Eu_i = 0$ **por la hipótesis #1**

- Así, la hipótesis #1 implica que $E(\hat{\beta}_1) = \beta_1$
- Esto es, $\hat{\beta}_1$ **es un estimador insesgado de $\boldsymbol{\beta}_1$.**

$$\operatorname{var}\left(\hat{\beta}_1\right) = \sigma^2_{\hat{\beta}_1} = E\left(\hat{\beta}_1 - E\left(\hat{\beta}_1\right)\right)^2 = E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]^2 =$$

$$\frac{1}{\left(\sum_1^n \left(X_i - \bar{X}\right)^2\right)^2}\left[\sum_1^n \left(X_i - \bar{X}\right)^2 Eu_i^2 + 2\sum_i\sum_j \left(X_i - \bar{X}\right)^2 \left(X_j - \bar{X}\right)^2 Eu_i u_j\right] =$$

$$= \frac{\sigma^2 \sum_1^n \left(X_i - \bar{X}\right)^2}{\left(\sum_1^n \left(X_i - \bar{X}\right)^2\right)^2} = \frac{\sigma^2}{\sum_1^n \left(X_i - \bar{X}\right)^2}$$

En lo que respecta a la distribución de probabilidad de $\hat{\beta}_1$ teniendo en cuenta la hipótesis #3 sobre la normalidad de las perturbaciones aleatorias, se tiene que

$$\hat{\beta}_1 - \beta_1 \sim N\left(0, \sigma^2_{\hat{\beta}_1}\right)$$

Propiedades Asintóticas

1) **Consistencia.** El estimador $\hat{\beta}_1$ es **consistente** porque se cumplen las dos condiciones suficientes:

$$\lim\left[E\left(\hat{\beta}_1\right) - \beta_1\right] = 0$$

$$\lim\left[\operatorname{var}\left(\hat{\beta}_1\right)\right]=0$$

Por lo que podemos escribir

$$p\lim\hat{\beta}_1=\beta_1 \text{ ó } \hat{\beta}_1\xrightarrow{p}\beta_1$$

2) **Eficiencia Asintótica:** Hemos visto que la distribución es

$$\hat{\beta}_1-\beta_1\sim N\left(0,\sigma^2_{\hat{\beta}_1}\right)$$

(En muestras grandes no hace falta la hipótesis #3).

Teniendo en cuenta que

$$\sigma^2_{\hat{\beta}_1}=\frac{\sigma^2}{\sum_1^n\left(X_i-\bar{X}\right)^2}=\frac{\sigma^2}{n}\frac{1}{\frac{1}{n}\sum_1^n\left(X_i-\bar{X}\right)^2}$$

Y teniendo en cuenta la hipótesis #5 podemos escribir

$$\hat{\beta}_1 - \beta_1 \sim N\left(0, \frac{\sigma^2}{n}\frac{1}{\sigma_X^2}\right)$$

O,equivalentemente,

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \sim N\left(0, \frac{\sigma^2}{\sigma_X^2}\right)$$

Sea β^* otro estimador consistente de β_1. Entonces, en muestras grandes, la distribución de probabilidad de $\sqrt{n}\left(\beta^* - \beta_1\right)$ siempre tiene una varianza mayor que $\frac{\sigma^2}{\sigma_X^2}$, que es la de $\sqrt{n}(\hat{\beta}_1 - \beta_1)$. Esto es lo que significa decir que $\hat{\beta}_1$ es asintóticamente eficiente.

C.C. 4.3 Las hipótesis de los MCO (Regresor Estocástico) (Heterocedasticidad)

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1,\ldots, n$$

1. La distribución condicional de u dado X tiene media cero, esto es, $E(u|X = x) = 0$.

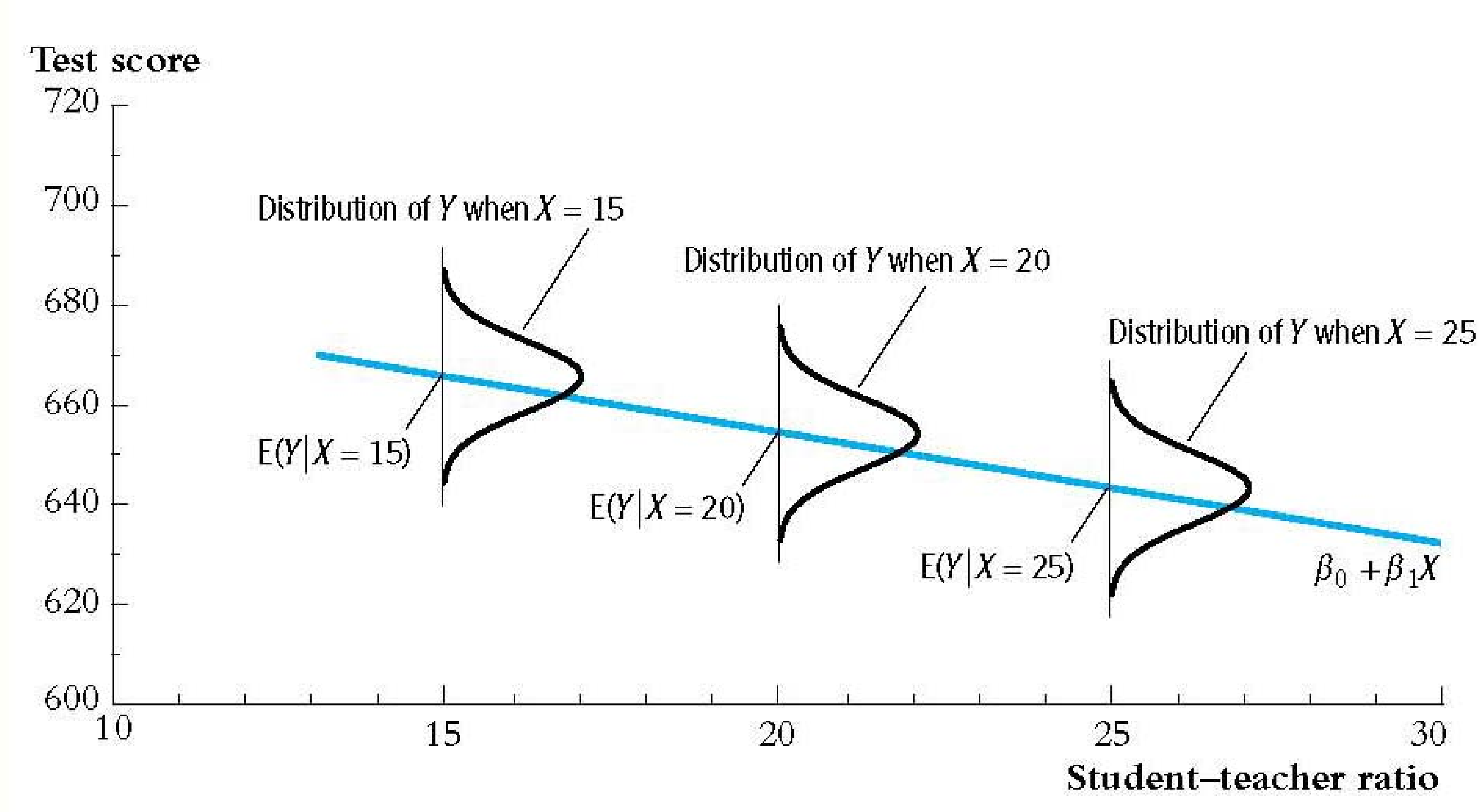
- *Esto implica que* $\hat{\beta}_1$ *es insesgado*

2. (X_i, Y_i), $i = 1,\ldots,n$, son i.i.d.
   - *Esto es verdad si* (X, Y) *se recogen siguiendo el muestreo aleatorio simple*
3. Se excluyen grandes atípicos de X *y*/o Y.
   - *Técnicamente,* X *e* Y *tienen momentos de cuarto orden finitos*
   - *Los atípicos pueden llevar a estimaciones de* $\hat{\beta}_1$

No significativas.

Hipótesis #1: $E(u|X = x) = 0$.

Para cualquier valor dado de X, la media de u es cero:



Ejemplo: $NotaExamen_i = \beta_0 + \beta_{1i}$ REM+ u_i, u_i = otros factores

- ¿Cuáles son estos otros factores?
- ¿Es $E(u|X=x) = 0$ plausible para estos otros factores?

Hipótesis #2: (X_i, Y_i), $i = 1,\ldots,n$ son i.i.d.

Se cumple automáticamente si la unidad de muestreo (individuo, distrito) se elige por muestreo aleatorio simple:

- Las unidades de muestreo son tomadas de la misma población de forma que (X_i, Y_i) están idénticamente distribuidas para todos $i = 1,\ldots, n$.
- Las unidades de muestreo son seleccionadas aleatoriamente de forma que los valores de (X, Y) para las diferentes unidades están independientemente distribuidas.

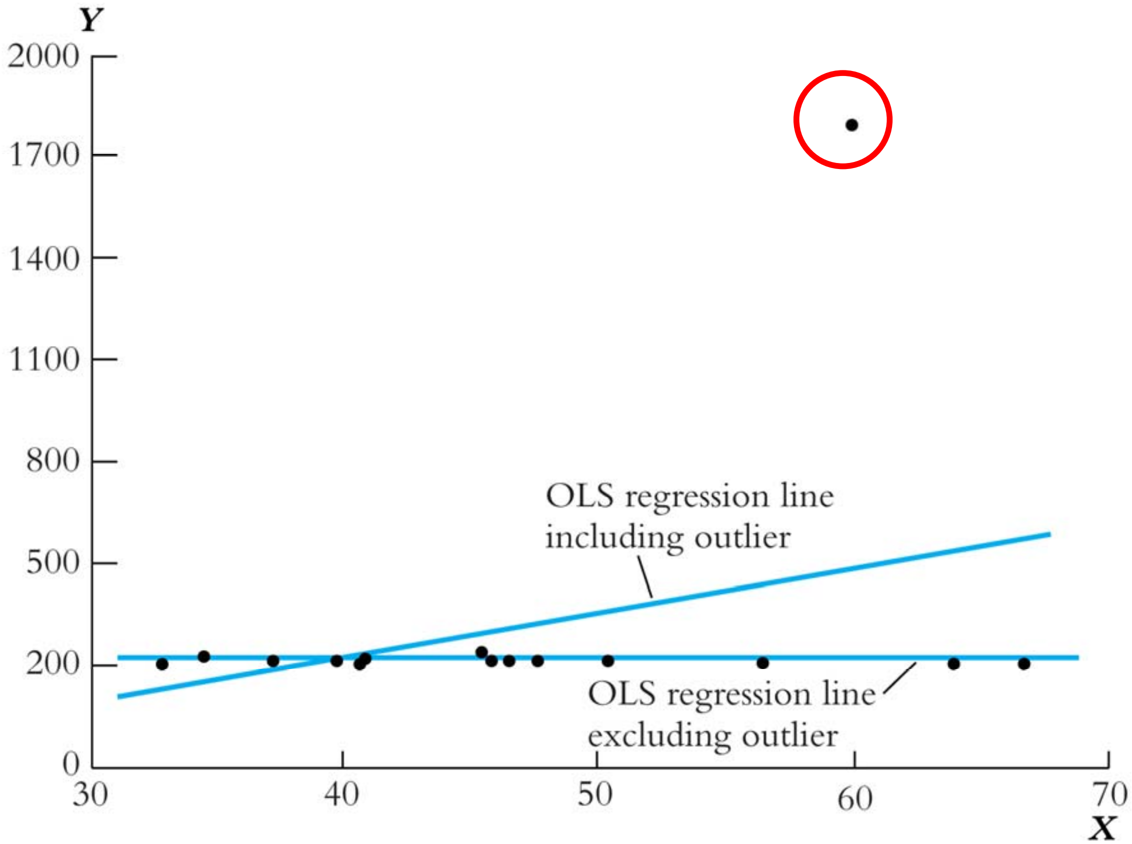
El caso principal de muestreo no-i.i.d. es cuando los datos son tomados a lo largo del tiempo para la misma unidad (datos panel y datos de series temporales– volveremos sobre esta cuestión.

Hipótesis #3: No hay atípicos grandes

Formulación técnica: $E(X^4) < \infty$ **y** $E(Y^4) < \infty$

- Un atípico grande es un valor extremo de X o Y
- A nivel técnico, si X e Y son limitados, entonces tienen momentos finitos de orden 4. (Tanto las notas como el REM, la renta familiar satisfacen este requisito)
- La relevancia de esta hipótesis es que los atípicos grandes pueden influir fuertemente los resultados – por lo que se necesita una regla para eliminarlos.
- Examina tus datos! Si tienes un atípico grande, ¿es un error tipográfico? ¿Pertenece a tus datos? ¿Por qué es atípico?

Los MCO pueden ser sensibles a un atípico:



- *¿Es el punto aislado un atípico de X o de Y?*
- En la práctica, los atípicos son pequeños errores de manipulación. ¡ Haz el gráfico de los datos!
  -

La distribución muestral de $\hat{\beta}_1$

Como $\bar{Y}$, $\hat{\beta}_1$ tiene una distribución muestral.

- ¿Qué es $E(\hat{\beta}_1)$?
  - Si $E(\hat{\beta}_1) = \beta_1$, entonces el MCO es insesgado – ¡Buena propiedad!
- ¿Qué es var($\hat{\beta}_1$)? (medida de incertidumbre muestral)
  - Necesitamos su formula para calcular el error estándar de $\hat{\beta}_1$.
- ¿Cuál es la distribución de $\hat{\beta}_1$ en muestras pequeñas?
  - En general es muy complicada
- ¿Cuál es la distribución de $\hat{\beta}_1$ en muestras grandes?
  - En muestras grandes, $\hat{\beta}_1$ sigue una distribución normal

Partiremos de la expresión a la que hemos llegado anteriormente

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Ahora calculamos $E(\hat{\beta}_1)$ y $var(\hat{\beta}_1)$:

$$E(\hat{\beta}_1) - \beta_1 = E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

$$= E\left\{E\left[\left.\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right| X_1,\ldots,X_n\right]\right\}$$

$= 0$ porque **$E(u_i|X_i=x) = 0$ por la hipótesis #1**

- Asi, la hipótesis #1 implica que $E(\hat{\beta}_1) = \beta_1$
- Esto es, $\hat{\beta}_1$ **es un estimador insesgado de β_1.**
- Para detalles ver App. 4.3

A continuación, calculamos *var*($\hat{\beta}_1$):

Escribir

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\left(\frac{n-1}{n}\right)s_X^2}$$

Donde $v_i = (X_i - \bar{X})u_i$. Si n es grande, $s_X^2 \approx \sigma_X^2$ y $\frac{n-1}{n} \approx 1$,

por lo que

$$\hat{\beta}_1 - \beta_1 \approx \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\sigma_X^2},$$

o $\quad \text{var}(\hat{\beta}_1 - \beta_1) = \text{var}(\hat{\beta}_1)$

$$= \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} v_i\right)\Big/(\sigma_X^2)^2 = \frac{\text{var}(v_i)/n}{(\sigma_X^2)^2}$$

Para la igualdad final se utiliza la hipótesis 2. Así,

$$\boxed{\text{var}(\hat{\beta}_1) = \frac{1}{n}\times\frac{\text{var}[(X_i - \mu_x)u_i]}{(\sigma_X^2)^2}} \quad .$$

Resumen

1. $\hat{\beta}_1$ es insesgado: $E(\hat{\beta}_1) = \beta_1$ –¡justo como $\bar{Y}$!
2. $\text{var}(\hat{\beta}_1)$ is inversamente proporcional a n- ¡igual que con $\bar{Y}$!

¿Cuál es la distribución muestral de $\hat{\beta}_1$?

La distribución muestral es complicada – depende de la distribución poblacional de (Y, X) – pero cuando n es grande se pueden obtener simples y buenas aproximaciones:

(1) Como $\text{var}(\hat{\beta}_1) \propto 1/n$ y $E(\hat{\beta}_1) = \beta_1$, $\hat{\beta}_1 \xrightarrow{p} \beta_1$

(2) Cuando n es grande, la distribución muestral de $\hat{\beta}_1$ es una Normal según el TCL

Recordar el TCL: suponer $\{v_i\}$, $i = 1,\ldots, n$ es i.i.d. con $E(v) = 0$ y $\text{var}(v) = \sigma^2$. Entonces, cuando n es grande, $\frac{1}{n}\sum_{i=1}^{n} v_i$ se distribuye aproximadamente como $N(0,\sigma_v^2/n)$.

Aproximación a la distribución de $\hat{\beta}_1$ en muestras grandes:

$$\hat{\beta}_1 - \beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\left(\frac{n-1}{n}\right)s_X^2} \approx \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\sigma_X^2}, \text{ where } v_i = (X_i - \bar{X})u_i$$

- Cuando n es grande, $v_i = (X_i - \bar{X})u_i \approx (X_i - \mu_X)u_i$, que es i.i.d. (*why*?) y $\text{var}(v_i) < \infty$ (*why*?). Por lo tanto, utilizando el TCL , $\frac{1}{n}\sum_{i=1}^{n} v_i$ se distribuye aproximadamente como una $N(0, \sigma_v^2 / n)$.
- Así, para n grande, $\hat{\beta}_1$ se distribuye aproximadamente como

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_v^2}{n(\sigma_X^2)^2}\right), \text{ where } v_i = (X_i - \mu_X)u_i$$

Cuanto mayor la varianza de *X*, menor es la varianza de $\hat{\beta}_1$

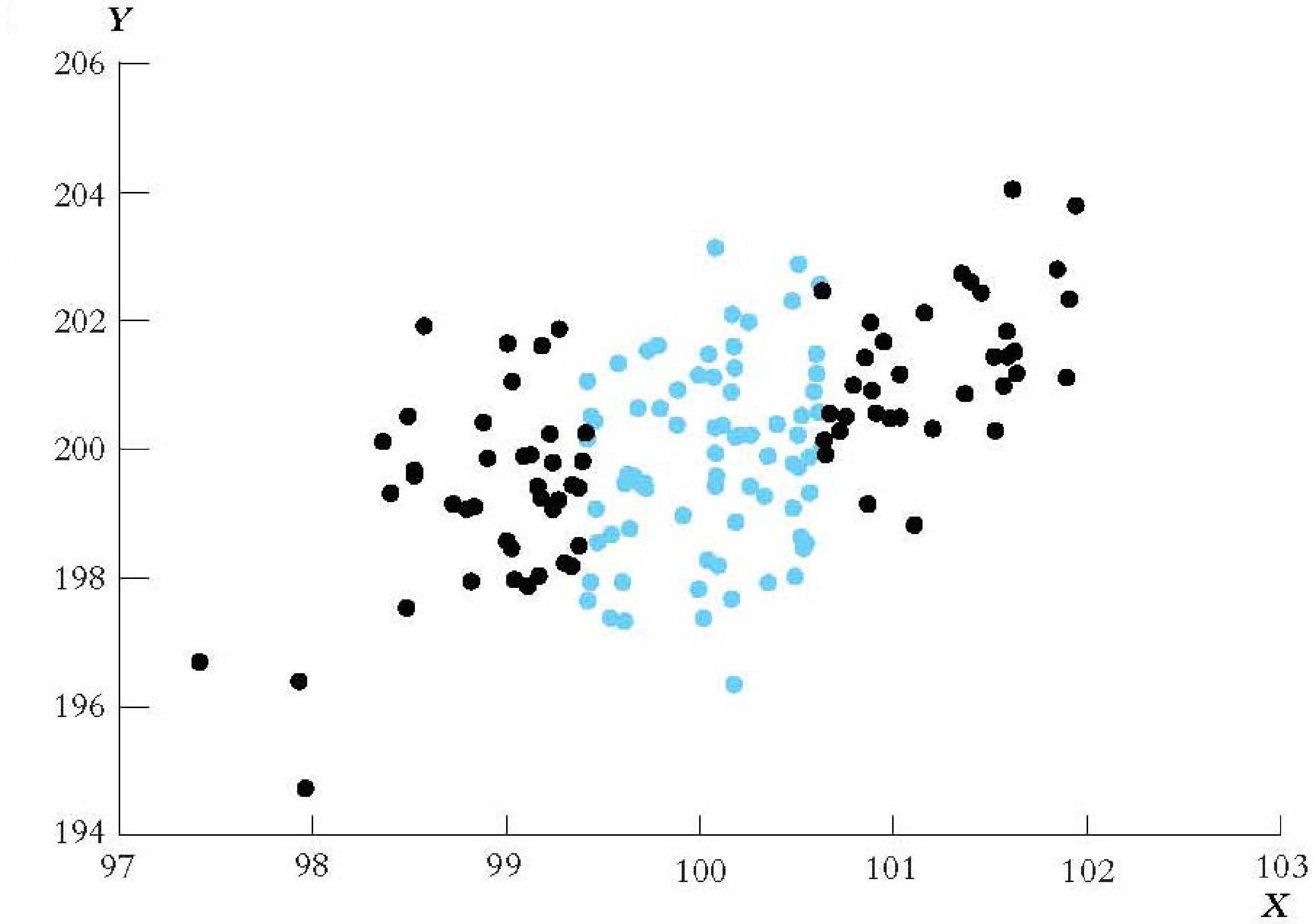
La varianza es

$$\text{var}(\hat{\beta}_1 - \beta_1) = \frac{1}{n} \times \frac{\text{var}[(X_i - \mu_x)u_i]}{(\sigma_X^2)^2}$$

donde $\sigma_X^2 = \text{var}(X_i)$. La varianza de *X* aparece (al cuadrado) en el denominador – por lo que al crecer la dispersión de *X* decrece la varianza de $\hat{\beta}_1$

Graficamente

Cuanto mayor es la varianza de *X*, *menor es la varianza de* $\hat{\beta}_1$



El número de puntos negros y azules es el mismo. ¿Cuál de ellos llevará a una menor varianza del estimador?

C.C. 4.4 Distribuciones para grandes muestras de $\hat{\beta}_0$ y $\hat{\beta}_1$:

Si las tres hipótesis de los MCO se cumplen, entonces $\hat{\beta}_0$ y $\hat{\beta}_1$ poseen distribuciones muestrales conjuntas normales para muestras grandes. La distribución muestral para n grande de $\hat{\beta}_1$ es N(β_1 ,$\sigma^2_{\hat{\beta}_1}$) donde la varianza es

$$\mathrm{var}(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\mathrm{var}[(X_i - \mu_x)u_i]}{\sigma_X^4} \propto \frac{1}{n}.$$

La distribución muestral para n grande de $\hat{\beta}_0$ es N($\beta_0, \sigma^2_{\hat{\beta}_0}$) donde la varianza es

$$\sigma^2_{\hat{\beta}_0} = \frac{1}{n} \frac{\mathrm{var}\left(H_i u_i\right)}{\left[E\left(H_i^2\right)\right]^2} \;\; \textit{donde} \quad H_i = 1 - \left[\frac{\mu_x}{E\left(X_i^2\right)}\right] X_i$$

C.C. 4.3 Las hipótesis de los MCO (Regresor Estocástico) (Homoscedasticidad)

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1,\dots, n$$

1. $E\left(u_i \setminus X_i\right) = 0$.
2. $\left(X_i, Y_i\right)$ son i.i.d.
3. $\operatorname{var}\left(u_i \setminus X_i\right) = \sigma^2$
4. Se excluyen los valores atípicos grandes

Destacar que las hipótesis #1,2 y 3 implican la esfericidad condicional

- $\operatorname{var}\left(u_i \setminus X_1, \dots, X_n\right) = \sigma^2$
- $E\left(u_i, u_j \setminus X_1, \dots X_n\right) = 0 \quad i \neq j$

Propiedades de los Esimadores MCO

$\hat{\beta}_1$ es insesgado por las mismas razones dadas en el caso heterocedástico. Pero su varianza es diferente. Hay que tener en cuenta que (siempre en muestras grandes):

$$\text{var}\left[\left(X_i-\mu_X\right)u_i\right]=E\left[\left(X_i-\mu_X\right)u_i\right]^2=E\left[\left(X_i-\mu_X\right)^2u_i^2\right]$$

$$=E\left[\left(X_i-\mu_X\right)^2\text{var}(u_i\setminus X_i)\right]=\sigma_X^2\sigma^2$$

Por lo tanto, la varianza será

$$\sigma_{\hat{\beta}_1}^2=\frac{1}{n}\times\frac{\text{var}[(X_i-\mu_x)u_i]}{\sigma_X^4}=\frac{1}{n}\frac{\sigma_X^2\sigma^2}{\sigma_X^4}=\frac{\sigma^2}{n\sigma_X^2}$$

Esta es la misma varianza del estimador MCO para el caso de regresor no estocástico. Todo lo dicho anteriormente se sigue aplicando en muestras grandes. Es decir, el estimador es consistente y asintóticamente eficiente.

5. Contrastes e intervalos de confianza

1. Contraste de hipótesis sobre β_1
2. Intervalos de confianza para β_1

   - Heterocedasticidad y homocedasticidad.
   - Distribución t-Student

1. Contraste de hipótesis sobre β_1

$$H_0 : \beta_1 = \beta_{1,0}$$

Seguir el siguiente proceso en tres etapas

1. **Definir el estadístico de contraste**

2. **Derivar la distribución de probabilidad de este estadístico bajo la hipótesis nula.**

3. **Determinar la región crítica del contraste**

2. Definir el estadístico de contraste

Utilizaremos el t-ratio definido como

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{ES\left(\hat{\beta}_1\right)}$$

En donde ES$\left(\hat{\beta}_1\right)$ es el error estándar del estimador definido como la raíz cuadrada de la varianza. Este error estándar dependerá de si estamos en el caso homocedástico o heterocedástico.

Fórmula para el ES($\hat{\beta}_1$):Heterocedasticidad

Recordar la expresión para la varianza de $\hat{\beta}_1$ (n grande):

$$\text{var}(\hat{\beta}_1) = \frac{\text{var}[(X_i - \mu_x)u_i]}{n(\sigma_X^2)^2} = \frac{\sigma_v^2}{n(\sigma_X^2)^2}$$, donde $v_i = (X_i - \mu_X)u_i$.

El estimador de la varianza de $\hat{\beta}_1$ sustituye los valores poblacionales desconocidos de σ_v^2 y σ_X^2 por estimadores definidos a partir de los datos:

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{1}{n} \times \frac{\text{estimador of } \sigma_v^2}{(\text{estimador of } \sigma_X^2)^2} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{v}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}$$

donde $\hat{v}_i = (X_i - \bar{X})\hat{u}_i$.

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{v}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}, \text{ donde } \hat{v}_i = (X_i - \bar{X})\hat{u}_i \,.$$

$$ES(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} = \text{el error está ndar de } \hat{\beta}_1$$

Esto es un poco feo, pero:

- Es menos complicado de lo que parece. El numerador estima la var(v), el denominador estima [var(X)]^2.
- ¿Por qué son n-2 los grados de libertad? Porque se han estimado dos coeficientes (β_0 y β_1).
- *El ES*($\hat{\beta}_1$) está incorporado en el software

Fórmula para el ES($\hat{\beta}_1$): Homocedasticidad

$$\text{ES}(\hat{\beta}_1) = \sqrt{\frac{1}{n} \times \frac{\dfrac{1}{n-2}\displaystyle\sum_{i=1}^{n}\hat{u}_i^2}{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2}}.$$

Por lo tanto, tenemos dos formulas para los errores estándar de $\hat{\beta}_1$.

- ***Errores estándar para homocedásticidad*** –son más simples pero solo son válidos si los errores son homocedásticos.
- **Errores Estándar robustos a la heterocedasticidad-**son válidos sean o no los errores heterocedásticos.
- Normalmente, los paquetes informáticos proporcionan los primeros por defecto y si se desea los segundos hay que indicarlo. ¡Cuidado! Porque lo que el software nos da por defecto solo es valido si los errores son homocedásticos.
- En GRETL se indica en : “Desviaciones típicas robustas”

3. Derivar la distribución del estadístico de contraste bajo la hipótesis nula.

Al hablar del estimador MCO hemos obtenido que

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\hat{\beta}_1}\right)$$

Por lo tanto, bajo la hipótesis nula tenemos

$$\hat{\beta}_1 \sim N\left(\beta_{1,0}, \sigma^2_{\hat{\beta}_1}\right)$$

Por lo que $t = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$

En muestras grandes el error estándar del estimador será un estimador consistente de la desviación típica del estimador por lo que $t \sim N(0,1)$.

En muestras pequeñas $t \sim t(student)$

4. Determinar la región crítica

Hay dos procedimientos equivalentes:

O bien se calcula el p-valor y se rechaza la hipótesis nula al 5% si dicho p-valor es inferior a 0.05

O bien se rechaza si $\left|t^{act}\right| > 1.96$ para una hipótesis alternativa bilateral $\left(\beta_1 \neq \beta_{1,0}\right)$.

Ejemplo**: Notas Examen y REM, datos de California**

Línea de regresión estimada:

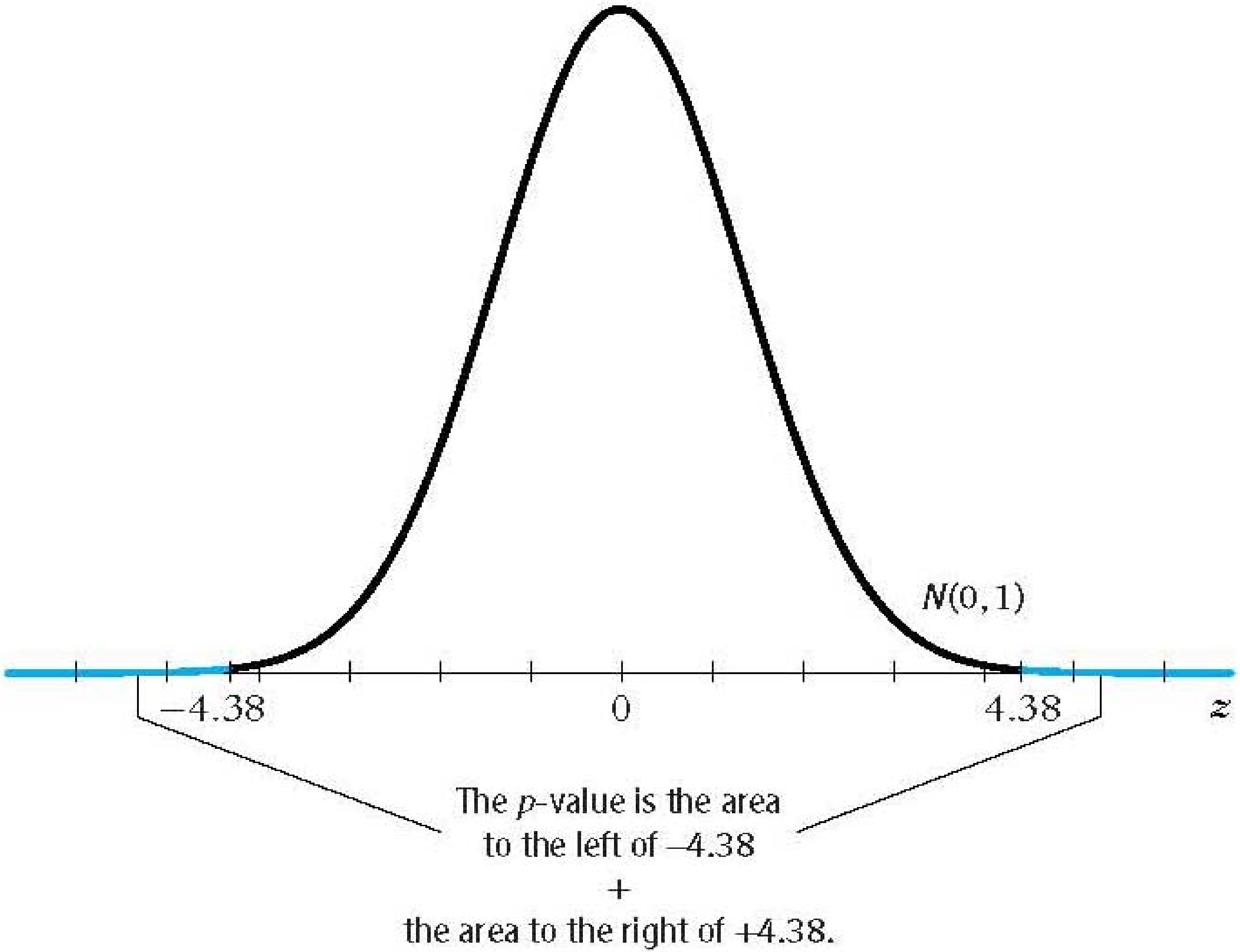
$$\widehat{NotasExamen} = 698.9 - 2.28 \times \text{REM}$$

El programa nos da los errores estándar:

$$ES(\hat{\beta}_0) = 10.4 \qquad\qquad ES(\hat{\beta}_1) = 0.52$$

$$\textit{Estadístico-t}\ \beta_{1,0} = 0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{ES(\hat{\beta}_1)} = \frac{-2.28 - 0}{0.52} = -4.38$$

- El punto crítico para un nivel de 1% para una hipótesis bilateral es 2.58, por lo que rechazamos la hipóteis nula para el nivel del 1%.
- Alternativamente, podemos calcular el p-valor…



El p-valor basado en una aproximación normal estándar es 0.00001 (10^{-5})

2. Intervalos de confianza para β_1

Recordar que un intervalo de confianza del 95% es, al mismo tiempo:

- El conjunto de valores que no puede ser rechazados con un nivel del 5% ;
- Un intervalo que es una función de los datos muestrales que contienen el verdadero valor del parámetro el 95% de las muestras repetidas.

Como el estadístico-t para β_1 es N(0,1) en muestras grandes, la construcción de un intervalo del 95% para β_1 es como el caso de la media muestral:

Intervalo de confianza del 95% para $\beta_1 = \{\hat{\beta}_1 \pm 1.96 \times \text{ES}(\hat{\beta}_1)\}$

Ejemplo de intervalo de confianza: Calificaciones y REM

Línea de regresión estimada: $\widehat{NotasExamen} = 698.9 - 2.28 \times REM$

$$ES(\hat{\beta}_0) = 10.4 \qquad\qquad ES(\hat{\beta}_1) = 0.52$$

Intervalo de confianza del 95% para $\hat{\beta}_1$:

$$\{\hat{\beta}_1 \pm 1.96 \times ES(\hat{\beta}_1)\} = \{-2.28 \pm 1.96 \times 0.52\}$$

$$= (-3.30, -1.26)$$

Las siguientes dos afirmaciones son equivalentes *(¿por qué?)*

- El intervalo del 95% no incluye el cero;
- La hipótesis $\beta_1 = 0$ se rechaza al nivel del 5%

Una forma concisa (y convencional) para presentar los resultados de la regresión:

Poner los errores estándar debajo de los correspondientes coeficientes estimados.

$$\widehat{NotaExamen} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28}\times\text{REM},\ R^2 = .05,\ \text{E}SR = 18.6$$

<u>Esta expresión proporciona mucha información</u>

- La línea de regresión estimada es

$$\widehat{NotaExamen} = 698.9 - 2.28\times\text{REM}$$

- El error estándar de $\hat{\beta}_0$ es 10.4
- The standard error of $\hat{\beta}_1$ is 0.52
- El R^2 es .05; el error estándar de la regresión es 18.6

Regresión MCO: GRETL

```
Modelo 1: MCO, usando las observaciones 1-420
Variable dependiente: NotasExamen
Desviaciones típicas robustas ante heterocedasticidad, variante HC1

             Coeficiente   Desv. Típica   Estadístico t    Valor p
  ------------------------------------------------------------------
  const        698.933       10.3644          67.44        9.49e-227
  REM           -2.27981      0.519489        -4.389       1.45e-05

Media de la vble. dep.  654.1565   D.T. de la vble. dep.   19.05335
Suma de cuad. residuos  144315.5   D.T. de la regresión    18.58097
R-cuadrado              0.051240   R-cuadrado corregido    0.048970
F(1, 418)               19.25943   Valor p (de F)          0.000014
Log-verosimilitud      -1822.250   Criterio de Akaike      3648.499
Criterio de Schwarz     3656.580   Crit. de Hannan-Quinn

3651.693
```

SO:

$$\widehat{NotaExamen} = 698.9 - 2.28\times\text{REM}, \, , R^2 = \mathbf{.05}, \, ESR = \mathbf{18.6}$$

$$(10.4) \; (0.52)$$

t ($\beta_1 = 0$) = **–4.38**, p-valor = **0.000** (bilateral)

Intervalo de confianza al 95% bilateral para β_1 es (–3.30, –1.26)

INTERPRETACIÓN DE COEFICIENTES

Los logaritmos pueden utilizarse para transformar tanto la variable dependiente (Y) como la independiente (X) (la variable que se transforma debe de ser positiva).

Caso	**Especificación**	**Interpretación de $\hat{\beta}_1$**
Lin-log	$Y_i = \beta_0 + \beta_1 ln(X_i) + u_i$	Un 1% de cambio en X está asociado con un cambio en Y de 0.01 β_1
Log-lin	$ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$	Un cambio en X de una unidad está asociado con un cambio en Y de 100 β_1%
Log-log	$ln(Y_i) = \beta_0 + \beta_1\ ln(X_i) + u_i$	Un 1% de cambio en X está asociado con un cambio en Y de β_1%, por lo que β_1 es la elasticidad de y respecto a X.