Estimación del modelo lineal con dos variables: el método de mínimos cuadrados ordinarios (MCO)

Mariana Marchionni
marchionni.mariana@gmail.com

Econometría I - FCE - UNLP
www.econometria1unlp.com

Temario de la clase

El modelo lineal con 2 variables o modelo lineal simple

$$Y_i = \alpha + \beta X_i + \mu_i \qquad i = 1, ...N$$

- Y: variable dependiente; X: variable explicativa.
- α, β: parámetros desconocidos.
- μ_i: término aleatorio, *no observable*
  (carácter aleatorio por cuestiones de modelización).
- Muestra aleatoria (MA) (Y_i, X_i) con i=1, ...N.
- Datos: una realización de la muestra aleatoria.

Estimación

Objetivo: estimar α y β a partir de la MA (Y_i, X_i) $i = 1, ...N$.

Notación y definiciones:

- $\hat{\alpha}$ y $\hat{\beta} \longrightarrow$ estimadores de α y β
- $\hat{Y}_i \equiv \hat{\alpha} + \hat{\beta} X_i \longrightarrow$ es la estimación de Y_i que podemos obtener a partir de X_i si contamos con $\hat{\alpha}$ y $\hat{\beta}$
- $e_i \equiv Y_i - \hat{Y}_i = Y_i - (\hat{\alpha} + \hat{\beta} X_i) \longrightarrow$ error de estimación o residuo. Veremos que es una suerte de estimación de μ_i

Función de regresión y regresión estimada

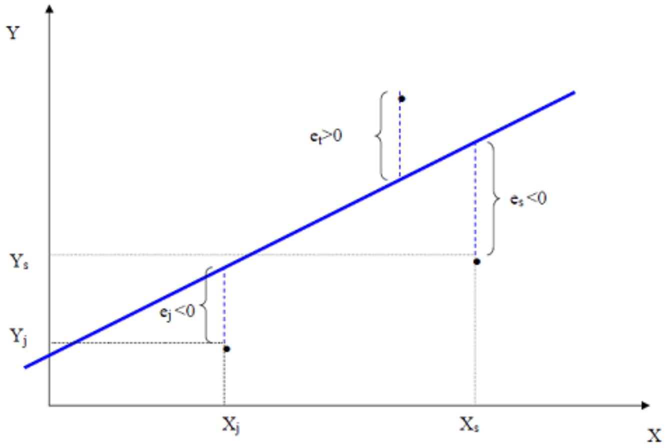
- Tomemos el modelo lineal $Y_i = \alpha + \beta X_i + \mu_i$
- La **función de regresión** es $\alpha + \beta X_i$, ¿Interpretación?
  - Si fuera cierto que $E[\mu_i] = 0$, entonces $E[Y_i] = \alpha + \beta X_i$
- Recién definimos $\hat{Y}_i \equiv \hat{\alpha} + \hat{\beta} X_i$. Notar que se puede interpretar como la **regresión estimada**
- Entonces
  - $\hat{Y}_i$ es el estimador del valor esperado de Y_i cuando $E[\mu_i] = 0$
  - y podemos interpretar al error de estimación ($e_i \equiv Y_i - \hat{Y}_i$) como una especie de estimador de μ_i

Representación gráfica

- Consideremos una realización de la muestra aleatoria $(Y_i, X_i) \qquad i = 1, \ldots N$
- Son N datos o puntos que podemos representar en el plano Y, X
- En el mismo plano podemos representar la función de regresión estimada $\hat{Y}_i \equiv \hat{\alpha} + \hat{\beta} X_i$
- Elegir valores de $\hat{\alpha}$ y $\hat{\beta}$ es equivalente a elegir una línea en el plano Y, X

Representación gráfica

La línea azul es $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$.



¿Cómo elegimos la línea?

- Parece razonable elegir $\hat{\alpha}$ y $\hat{\beta}$ de manera que la línea elegida pase lo más cerca posible de todos los puntos
- ¿Qué significa que la línea esté cerca de los puntos?

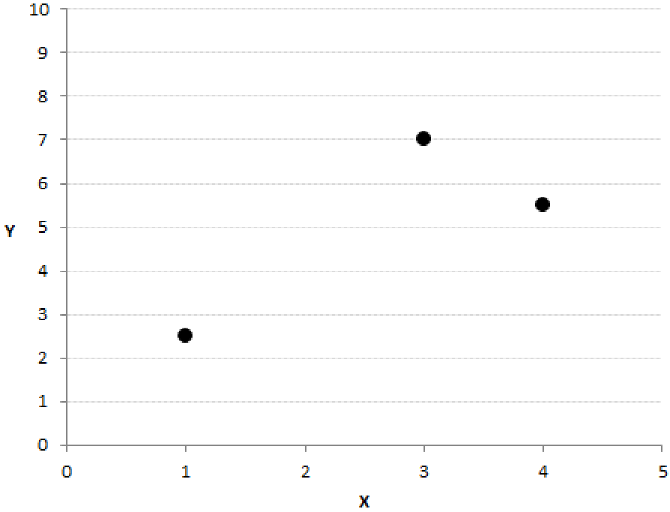
¿Cómo elegimos la línea?

Consideremos la **Suma de los Residuos Cuadráticos (SRC)**:

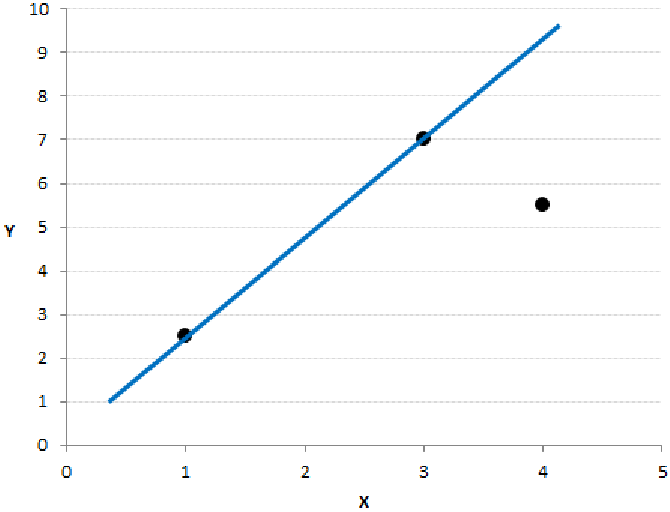
$$SRC(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} [Y_i - (\hat{\alpha} + \hat{\beta} X_i)]^2$$

- Es una medida de **error global de estimación** (función de penalización). SRC grande significa que los errores de estimación son grandes.
- SRC ≥ 0; SRC es cero si y solo si...
- SRC depende de la elección de $\hat{\alpha}$ y $\hat{\beta}$: **adoptaremos el criterio de elegir $\hat{\alpha}$ y $\hat{\beta}$ que minimicen SRC.**
- ¿Por qué errores al cuadrado?

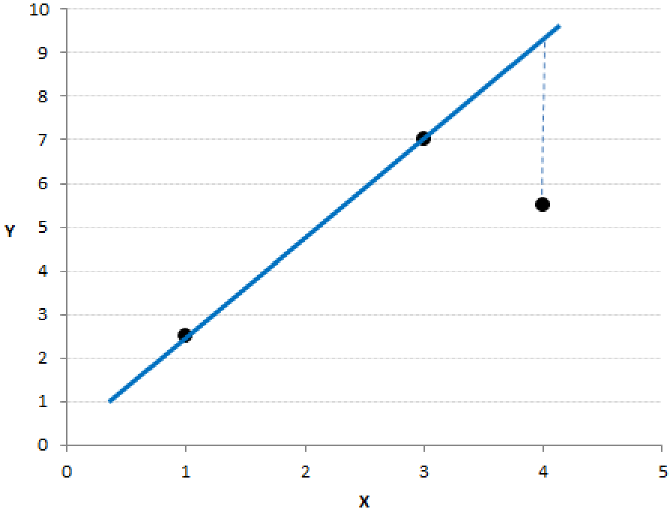
¿Cómo elegimos la línea?



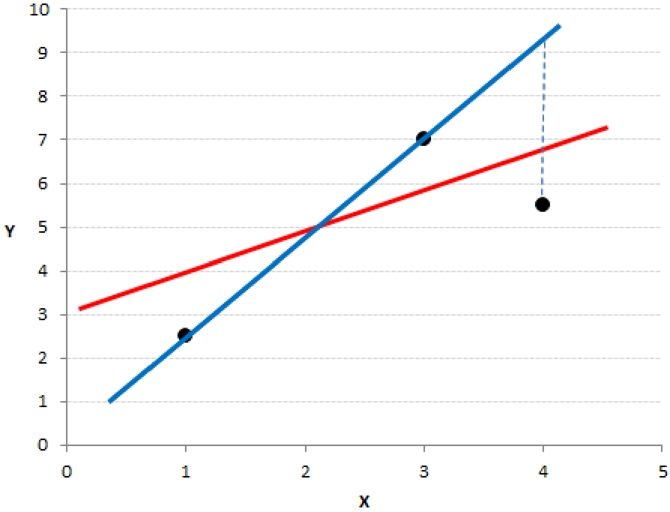
¿Cómo elegimos la línea?



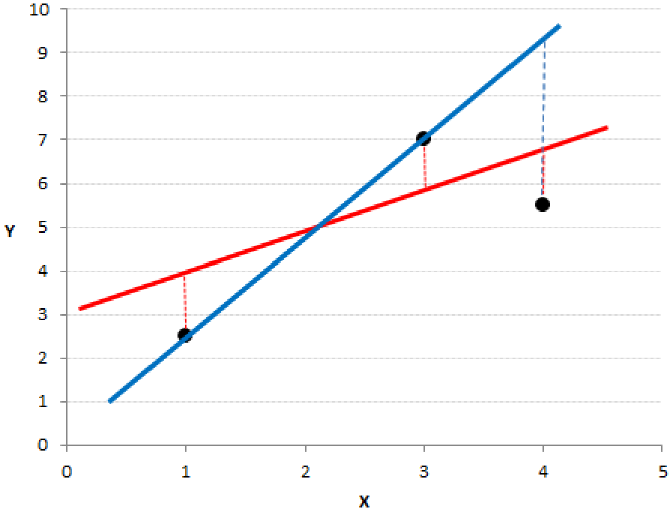
¿Cómo elegimos la línea?



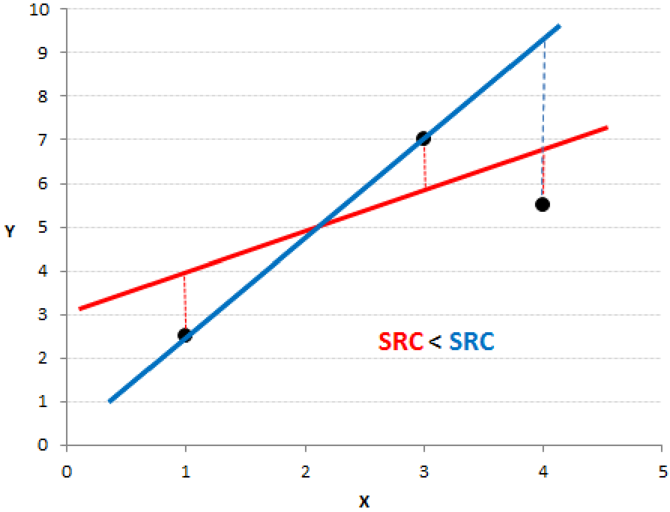
¿Cómo elegimos la línea?



¿Cómo elegimos la línea?



¿Cómo elegimos la línea?



¿Cómo elegimos la línea?

- Entonces, la función *SRC*
  - pondera positivamente *cada error de estimación individual*, es decir, el error de estimación de *cada* observación.
  - la ponderación crece más que proporcionalmente con la magnitud de cada error individual.

- Otros candidatos para función de error global
  - La suma de los errores: $\sum_{i=1}^{N} e_i$
    ¿Qué problema tiene?
  - La suma del valor absoluto de los errores: $\sum_{i=1}^{N} |e_i|$
    No es diferenciable y eso nos complica, pero se usa.

- Nos quedamos con *SRC*. El objetivo será elegir los estimadores que minimizan ese error global.

Mínimos Cuadrados Ordinarios (MCO)

Criterio de MCO: elegir $\hat{\alpha}$ y $\hat{\beta}$ de manera de minimizar SRC.

Formalmente:

$$Min_{\hat{\alpha},\hat{\beta}}\ SRC(\hat{\alpha},\hat{\beta}) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} [Y_i - (\hat{\alpha} + \hat{\beta} X_i)]^2$$

$$CPO: \begin{cases} (1) & \frac{\partial SRC}{\partial \hat{\alpha}} = 0 \\ \\ (2) & \frac{\partial SRC}{\partial \hat{\beta}} = 0 \end{cases}$$

$$Min_{\hat{\alpha},\hat{\beta}}\ SRC(\hat{\alpha},\hat{\beta}) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N}[Y_i - (\hat{\alpha} + \hat{\beta}X_i)]^2$$

$$CPO : \begin{cases} (1) \quad \frac{\partial SRC}{\partial \hat{\alpha}} = -2\sum_{i=1}^{N}[Y_i - (\hat{\alpha} + \hat{\beta}X_i)] = 0 \\ \\ (2) \quad \frac{\partial SRC}{\partial \hat{\beta}} = -2\sum_{i=1}^{N} X_i[Y_i - (\hat{\alpha} + \hat{\beta}X_i)] = 0 \end{cases}$$

- Notar que
  - (1) implica que $\sum_{i=1}^{N} e_i = 0$
  - (2) implica que $\sum_{i=1}^{N} X_i e_i = 0$.

Partimos de la CPO (1) y reacomodamos términos:

$$(1) \quad \frac{\partial SRC}{\partial \hat{\alpha}} = -2\sum_{i=1}^{N}[Y_i - (\hat{\alpha} + \hat{\beta}X_i)] = 0$$

$$\sum_{i=1}^{N} Y_i = \sum_{i=1}^{N}(\hat{\alpha} + \hat{\beta}X_i)$$

$$\sum_{i=1}^{N} Y_i = N\hat{\alpha} + \hat{\beta}\sum_{i=1}^{N} X_i$$

Dividimos por N:

$$\frac{\sum_{i=1}^{N} Y_i}{N} = \hat{\alpha} + \hat{\beta}\frac{\sum_{i=1}^{N} X_i}{N}$$

$$\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X} \ \Rightarrow \ \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \ (3)$$

Ahora reacomodamos términos en la CPO (2):

$$(2) \quad \frac{\partial SRC}{\partial \hat{\beta}} = -2\sum_{i=1}^{N} X_i[Y_i - (\hat{\alpha} + \hat{\beta}X_i)] = 0$$

$$\sum_{i=1}^{N} X_i Y_i = \hat{\alpha}\sum_{i=1}^{N} X_i + \hat{\beta}\sum_{i=1}^{N} X_i^2$$

Usamos (3) $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$ para sustituir $\hat{\alpha}$:

$$\sum_{i=1}^{N} X_i Y_i = \bar{Y}\sum_{i=1}^{N} X_i - \hat{\beta}\bar{X}\sum_{i=1}^{N} X_i + \hat{\beta}\sum_{i=1}^{N} X_i^2$$

$$\sum_{i=1}^{N} X_i Y_i - \bar{Y}\sum_{i=1}^{N} X_i = \hat{\beta}\left(\sum_{i=1}^{N} X_i^2 - \bar{X}\sum_{i=1}^{N} X_i\right)$$

$$\Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{N} X_i Y_i - \bar{Y}\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} X_i^2 - \bar{X}\sum_{i=1}^{N} X_i} = \frac{\sum_{i=1}^{N} X_i Y_i - N\bar{Y}\bar{X}}{\sum_{i=1}^{N} X_i^2 - N\left(\bar{X}\right)^2} \ (4)$$

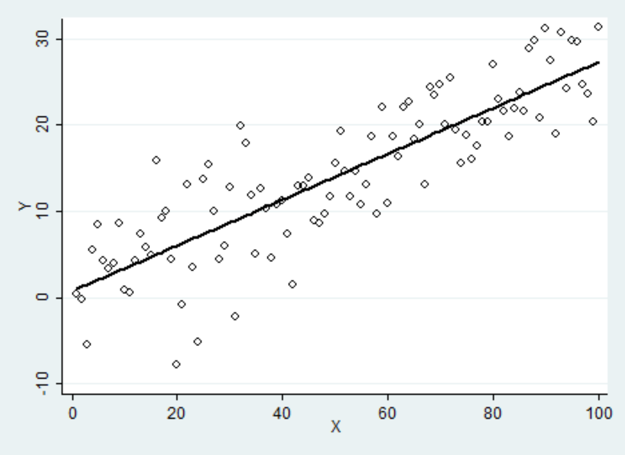
Los estimadores MCO

$$(3)\ \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$(4)\ \hat{\beta} = \frac{\sum_{i=1}^{N} X_i Y_i - N\bar{Y}\bar{X}}{\sum_{i=1}^{N} X_i^2 - N\bar{X}^2}$$

La regresión estimada por MCO

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$$



Notación: variables en forma de desvíos

- Denotemos: $x_i = X_i - \bar{X}$ y $y_i = Y_i - \bar{Y}$
- Entonces (4) puede escribirse como:

$$(4') \qquad \hat{\beta} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$$

Propiedades algebraicas

Recordemos las CPO del problema de Min SRC:

$$CPO : \begin{cases} (1) \quad \frac{\partial SRC}{\partial \hat{\alpha}} = \sum_{i=1}^{N} [Y_i - (\hat{\alpha} + \hat{\beta} X_i)] = 0 \\ \\ (2) \quad \frac{\partial SRC}{\partial \hat{\beta}} = \sum_{i=1}^{N} X_i [Y_i - (\hat{\alpha} + \hat{\beta} X_i)] = 0 \end{cases}$$

- Llamamos **propiedades algebraicas** de los estimadores a las que se derivan de las CPO.

Algunas de las propiedades

1. $\sum_{i=1}^{N} e_i = 0$
   - Se obtiene directamente de la CPO (1)
   - ¿Intuición?
   - ¿Qué pasa si el modelo no incluye una constante α?
2. $\sum_{i=1}^{N} e_i X_i = 0$
   - Se obtiene de la CPO (2)
   - ¿Intuición? Implica que $Cov(X, e) = 0$

Algunas de las propiedades

3. $\hat{Y}(\bar{X}) = \bar{Y}$
   - la regresión estimada pasa por las medias muestrales
   - dados los estimadores MCO $\hat{\alpha}$ y $\hat{\beta}$, $\hat{Y}$ es una función de X_i. Luego, $\hat{Y}(X_i) = \hat{\alpha} + \hat{\beta}X_i \Rightarrow \hat{Y}(\bar{X}) = \hat{\alpha} + \hat{\beta}\bar{X} = \bar{Y}$
     donde la última igualdad se sigue de la CPO (1)

4. $\hat{\beta} = r_{Y,X}{}^{s_Y}/s_X$ (relación entre el coeficiente de regresión y el de correlación)

5. $\bar{Y} = \bar{\hat{Y}}$ (la media de las estimaciones de Y coincide con $\bar{Y}$)

6. $\hat{\beta}$ es una función lineal de las Y_i. Es decir, se puede escribir $\hat{\beta} = \sum \omega_i Y_i$, donde ω_i son numeros reales no aleatorios, que dependen exclusivamente de X_i, y no son todos son iguales a cero.

7. $r_{\hat{Y},e} = 0$ (la correlación entre las estimaciones de Y y los residuos es nula)
   - Ver demostraciones en cap. 1.4 de notas de clase.

Bondad del ajuste del modelo

- Regresión estimada como resumen de los datos: ¿cuán buen resumen es?
- Puede mostrarse que si se estima por MCO:

$$\underbrace{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}_{STC} = \underbrace{\sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2}_{SEC} + \underbrace{\sum_{i=1}^{N} e_i^2}_{SRC}$$

$$STC = SEC + SRC$$

STC: Suma Total de Cuadrados, *SEC*: Suma Explicada de Cuadrados, *SRC*: Suma de los Residuos al Cuadrado

  - Ver demostración en cap. 1.7 de notas de clase.

Bondad del ajuste del modelo

- Partimos de la descomposición de la STC:

$$\underbrace{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}_{STC} = \underbrace{\sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2}_{SEC} + \underbrace{\sum_{i=1}^{N} e_i^2}_{SRC}$$

- Dividimos por STC:

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

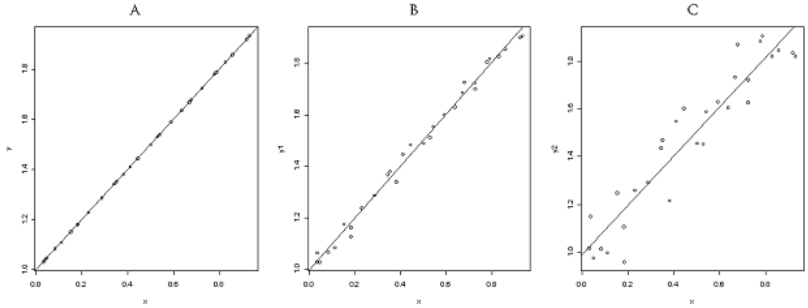
- Proponemos como medida de bondad del ajuste:

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

Propiedades del R^2

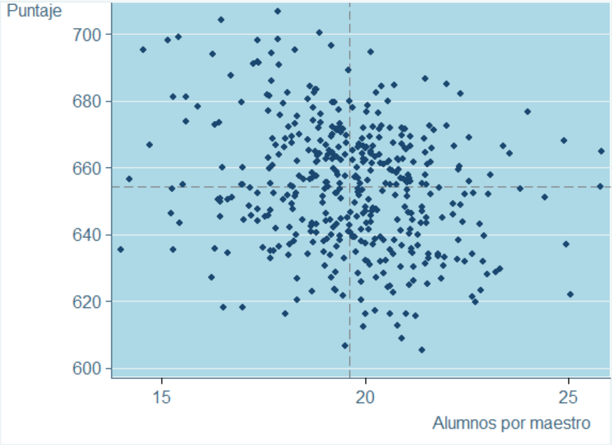
$$R^2 = \frac{SEC}{STC} = 1 - \frac{SRC}{STC}$$

- $0 \leq R^2 \leq 1$
  - $R^2 = 1$: relación lineal exacta entre Y y X.
  - $R^2 = 0$: cero correlación entre Y y X.
    la regresión estimada es una recta horizontal que coincide con $\bar{Y}$ ($\hat{\beta} = 0$)
- STC depende sólo de los datos, pero SEC y SRC dependen del modelo y del método de estimación.
  Dados los datos y el modelo, MCO minimiza SRC, es decir, maximiza R^2.



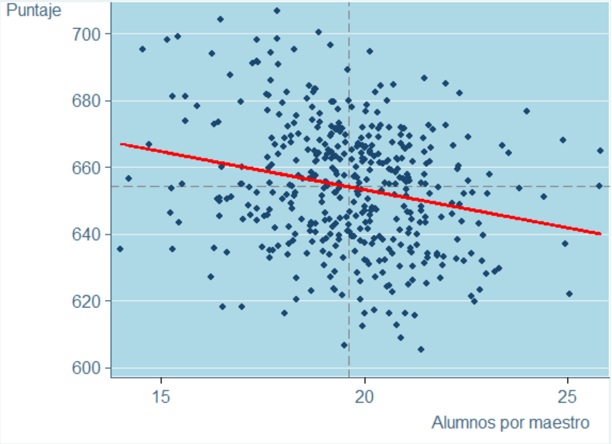
¿Cuán lejos está la recta estimada de los puntos?
En ese sentido A es bueno, B es peor, C es aún peor.
De hecho, en A el ajuste es perfecto: $R^2 = 1$.

Ejemplo: desempeño académico y tamaño de clase



Ejemplo: desempeño académico y tamaño de clase

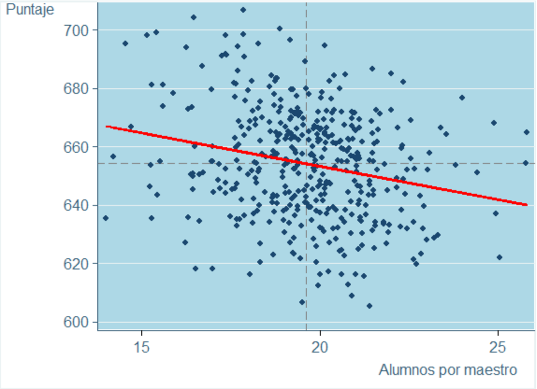
Regresión estimada: $\widehat{puntaje}_i = 698{,}9 - 2{,}28 \times ratio\_alma_i$



En el ejemplo de rendimiento académico

$$R^2 = 0.049$$

sólo 5 % de la variabilidad de los puntajes es explicada por el tamaño de la clase



Ejemplo: función de consumo keynesiana

- En su *Teoría general del interés, la ocupación y el dinero* (1936) Keynes estableció la existencia de una relación estable entre el consumo y el ingreso agregados:

$$C = f(I)$$

- La propensión marginal a consumir es positiva y menor que 1:
$$0 < \frac{dC}{dI} < 1$$
- La propensión media a consumir cae cuando aumenta el ingreso:
$$\frac{d\left({}^{C}/_{I}\right)}{dI} < 0$$
- El consumo también depende de otras situaciones y de las necesidades subjetivas, hábitos, etc. de una determinada sociedad

Modelo teórico

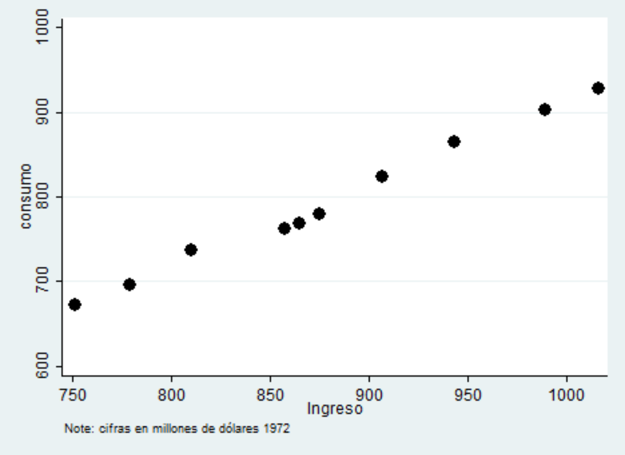
- La formulación teórica básica de la función de consumo es la lineal:

$$C = a + bI$$

- Los postulados keynesianos se satisfacen si se cumple:
  - Propensión marginal a consumir menor que 1 $\Rightarrow 0 < b < 1$
  - Propensión media decreciente en el ingreso $\Rightarrow a > 0 \qquad \left(\Rightarrow \frac{d(C/I)}{dI} = -a/I^2 < 0\right)$
- Estas proposiciones teóricas dan las bases para un estudio empírico

Una aplicación empírica

Datos de consumo e ingreso disponible de USA, 1970s
(Greene, *Econometric Analysis*, 4th edition, examples 1.1 & 6.1)



Modelo empírico

- Los datos se ajustan sólo aproximadamente a una relación lineal exacta $\Rightarrow$ la relación determinística es inadecuada
- Para representar los datos proponemos el siguiente modelo lineal:

$$C_i = a + bI_i + \mu_i \;\; con\; i = 1970, 1971, ..., 1979$$

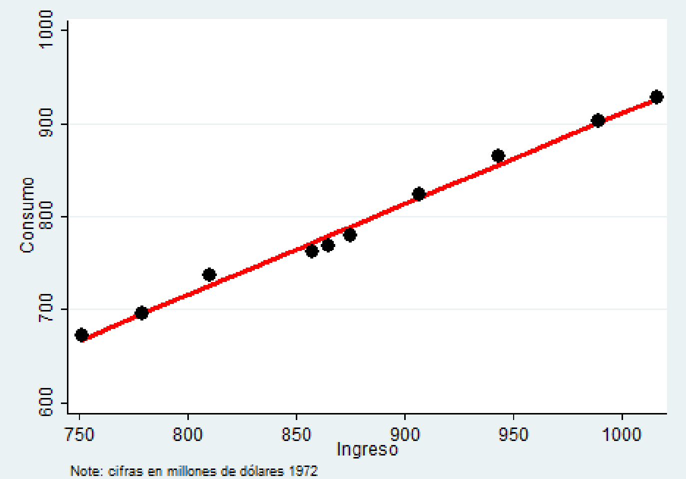
- $a + bI_i$ es la **función de regresión**: es la parte del consumo que se determina sistemáticamente a partir del ingreso
- μ_i resume el efecto de los "otros factores" que afectan al consumo agregado en cada año i más allá del ingreso. Se lo considera como aleatorio por una cuestión de modelización. No debería haber nada sistemático en el efecto de μ sobre el consumo.

La hoja con datos y cálculos

Año	Ingreso disponible	Consumo	Consumo Estimado	Errores	Errores2
1970	751.6	672.1	668.44	3.66	13.42
1971	779.2	696.8	695.46	1.34	1.78
1972	810.3	737.1	725.92	11.18	125.01
1973	864.7	767.9	779.19	-11.29	127.50
1974	857.5	762.8	772.14	-9.34	87.25
1975	874.9	779.4	789.18	-9.78	95.65
1976	906.8	823.1	820.42	2.68	7.19
1977	942.9	864.3	855.77	8.53	72.76
1978	988.8	903.2	900.72	2.48	6.16
1979	1015.7	927.6	927.06	0.54	0.29
sumatoria	8792.4	7934.3		0.00	537.01
medias	879.24	793.43			

La regresión estimada por MCO

$$\hat{C}_i = -67.6 + 0.98\, I_i$$

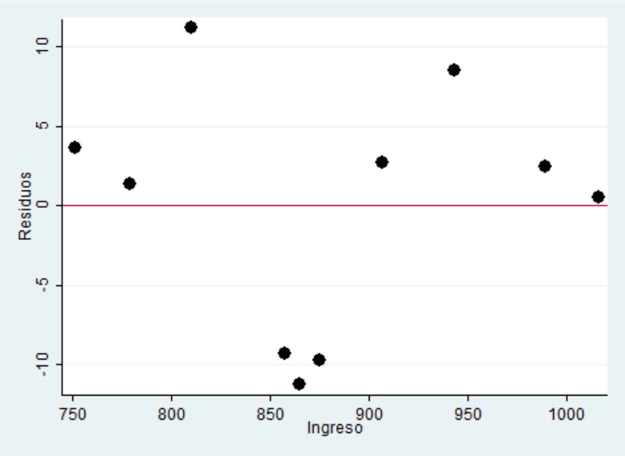


Interpretación de los resultados

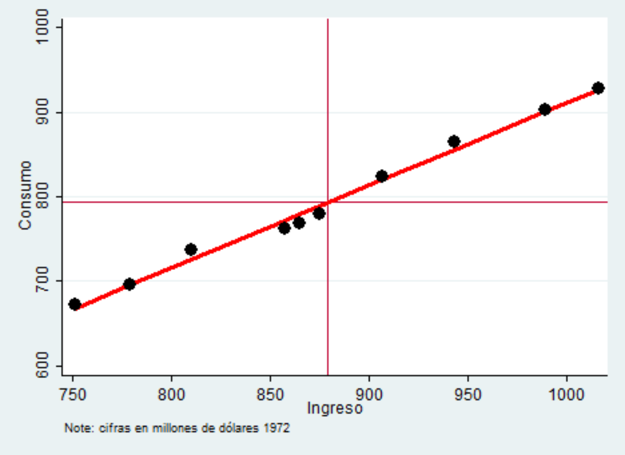
$$\hat{C}_i = -67.6 + 0.98\, I_i$$

- $\hat{b} = 0.98$: por cada aumento de un millón de dólares del ingreso disponible, el consumo aumentó en promedio en 0.98 millones de dólares en USA durante los 70s. **Propensión marginal al consumo**
- $\hat{a} = -67.6$: es el valor del consumo que deberíamos esperar si el ingreso fuera cero. **Consumo autónomo**
- Cuidado con las conclusiones que surgen de extrapolar la relación a valores de las variables fuera del rango de los datos
- Estimaciones puntuales: necesario evaluar mediante tests estadísticos

Graficamos los errores de estimación: suman cero y no están correlacionados con el ingreso

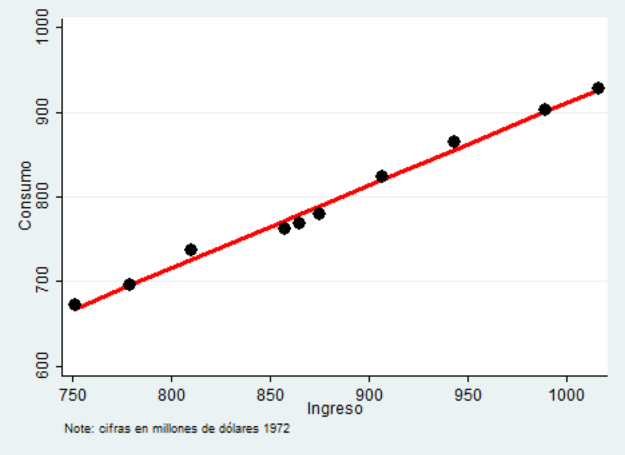


La regresión estimada pasa por las medias muestrales



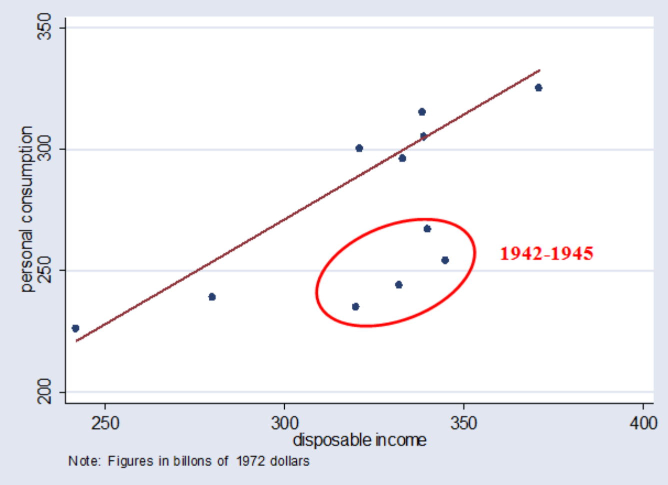
Bondad del ajuste $R^2 = 0.99$

el 99% de la variabilidad del consumo en los 70s en USA puede explicarse por la variabilidad del ingreso disponible



Mismas variables, otra década

Consumo e ingreso, USA, 1940s



Bibliografía para esta clase

- Notas de clase: capítulo 1.
- Wooldridge: capítulo 2.