ESTADÍSTICA APLICADA II

Tarea No. 1 Dr. Víctor M. Guerrero

Ago-Dic, 2021

1. En los siguientes modelos, el término ε_i representa el error aleatorio y se supone que se cuenta con una muestra de n datos, de manera que i = 1,..., n. **Determine:** (i) ¿Cuáles son lineales en los parámetros β_0 y β_1? (ii) ¿Cuáles son lineales en las variables Y y X? (iii) ¿Cuáles podrían considerarse modelos de regresión lineal simple? (posiblemente con variables transformadas).

(a) $Y_i = \beta_0 + \beta_1/X_i + \varepsilon_i$

(b) $Y_i = \beta_0 + \beta_1 \log(X_i) + \varepsilon_i$

(c) $Y_i = \beta_0 X_i^{\beta_1}$

(d) $\log(Y_i) = \log(\beta_0) + \beta_1 \log(X_i) + \varepsilon_i$

(e) $Y_i = \beta_0 + \beta_1 \varepsilon_i$

2. **Demuestre** que el estimador mínimo-cuadrático de β_1 en un modelo de regresión lineal simple, se puede escribir como

$$b_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

y que

$$\sum (Y_i - \hat{Y}_i)\hat{Y}_i = 0 .$$

3. A partir de una muestra de n = 200 parejas de observaciones, se calcularon las siguientes cantidades:

$$\sum X_i = 11.34, \quad \sum Y_i = 20.72, \quad \sum X_i^2 = 12.16, \quad \sum Y_i^2 = 84.96 \quad \text{y} \quad \sum X_i Y_i = 22.13$$

Con base en estas cantidades, **estime** las dos regresiones

$$\hat{Y}_i = b_0 + b_1 X_i \qquad \text{y} \qquad \hat{X}_i = a_0 + a_1 Y_i .$$

Deduzca una recta estimada para Y a partir de la segunda ecuación.
Grafique las dos rectas estimadas de Y en la misma gráfica y **comente** acerca de ellas, en particular acerca de cómo se podrían interpretar las mismas y por qué difieren.

1. En los siguientes modelos, el término ε_i representa el error aleatorio y se supone que se cuenta con una muestra de n datos, de manera que i = 1,..., n. **Determine:** (i) ¿Cuáles son lineales en los parámetros β_0 y β_1? (ii) ¿Cuáles son lineales en las variables Y y X? (iii) ¿Cuáles podrían considerarse modelos de regresión lineal simple? (posiblemente con variables transformadas).

(a) $$Y_i = \beta_0 + \beta_1/X_i + \varepsilon_i$$

i) linealidad en los parámetros:

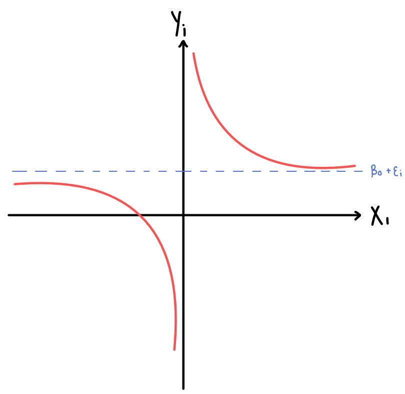
$$\frac{\partial Y_i}{\partial \beta_0} = 1$$

$$\frac{\partial Y_i}{\partial \beta_1} = \frac{1}{X_i}$$

no hay interacción entre los parámetros ∴ el modelo es lineal en los parámetros

ii) linealidad en las variables:

$$Y_i = \beta_0 + \frac{\beta_1}{X_i} + \varepsilon_i$$



∴ no es lineal en las variables

iii) ¿Es modelo de regresión lineal simple?

- es lineal en los parámetros ✓
- solo tiene una variable explicativa (X_i) ✓

∴ sí es un modelo de regresión lineal simple

(b) $$Y_i = \beta_0 + \beta_1 \log(X_i) + \varepsilon_i$$

i) linealidad en los parámetros:

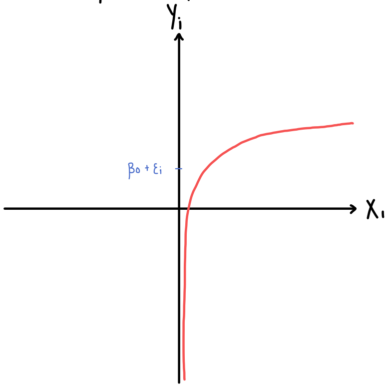
$$\frac{\partial Y_i}{\partial \beta_0} = 1$$

$$\frac{\partial Y_i}{\partial \beta_1} = \ln(X_i)$$

no hay interacción entre los parámetros ∴ el modelo es lineal en los parámetros

ii) linealidad en las variables:

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + \varepsilon_i$$



$\therefore$ no es lineal en las variables

iii) ¿Es modelo de regresión lineal simple?

- es lineal en los parámetros ✓
- solo tiene una variable explicativa (X_i) ✓

$\therefore$ sí es un modelo de regresión lineal simple

(c) $Y_i = \beta_0 X_i^{\beta_1}$

i) linealidad en los parámetros:

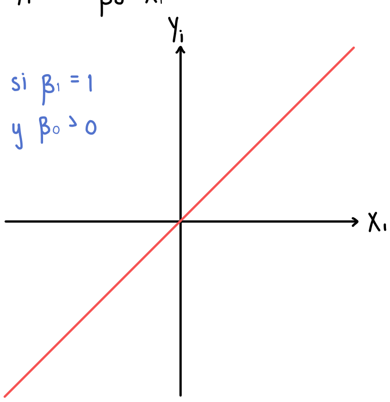
$$\frac{\partial Y_i}{\partial \beta_0} = X_i^{\beta_1}$$

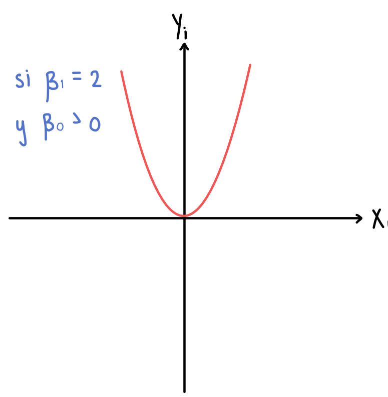
$$\frac{\partial Y_i}{\partial \beta_1} = \beta_0 \cdot \beta_1 X_i^{(\beta_1 - 1)}$$

hay interacción entre los parámetros $\therefore$ el modelo no es lineal en los parámetros

ii) linealidad en las variables:

$$Y_i = \beta_0 \cdot X_i^{\beta_1}$$





$\therefore$ si $\beta_1 = 1 \implies$ sí es lineal en las variables

si $\beta_1 > 1 \implies$ no es lineal en las variables

iii) ¿Es modelo de regresión lineal simple?

- es lineal en los parámetros ✗
- solo tiene una variable explicativa (X_i) ✓

$\therefore$ no es un modelo de regresión lineal simple

(d) $$\log(Y_i) = \log(\beta_0) + \beta_1 \log(X_i) + \varepsilon_i$$

$$\ln(Y_i) = \ln(\beta_0) + \ln(X_i^{\beta_1}) + \varepsilon_i$$
$$= \ln(\beta_0 \cdot X_i^{\beta_1}) + \varepsilon_i$$
$$\Longrightarrow Y_i = \beta_0 \cdot X_i^{\beta_1} \cdot e^{\varepsilon_i}$$

i) linealidad en los parámetros:

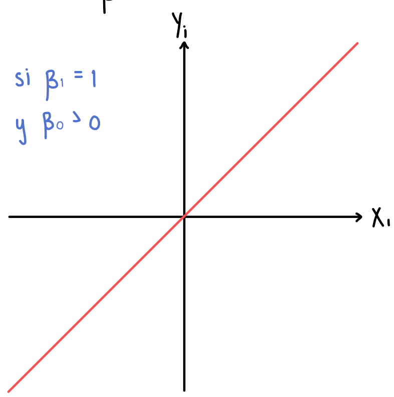
$$\frac{\partial Y_i}{\partial \beta_0} = X_i^{\beta_1} \cdot e^{\varepsilon_i}$$

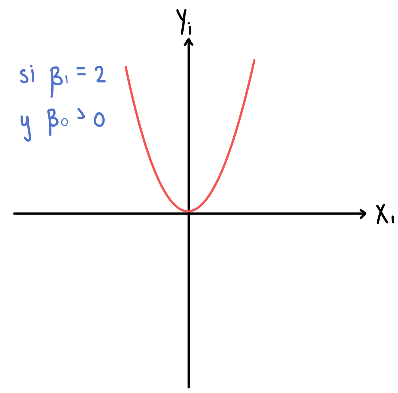
$$\frac{\partial Y_i}{\partial \beta_1} = \beta_0 \cdot \beta_1 \cdot e^{\varepsilon_i} \cdot X_i^{(\beta_1 - 1)}$$

hay interacción entre los parámetros $\therefore$ el modelo no es lineal en los parámetros

ii) linealidad en las variables:

$$Y_i = \beta_0 \cdot X_i^{\beta_1} \cdot e^{\varepsilon_i}$$





$\therefore$ si $\beta_1 = 1 \Longrightarrow$ sí es lineal en las variables
si $\beta_1 > 1 \Longrightarrow$ no es lineal en las variables

iii) ¿Es modelo de regresión lineal simple?

- es lineal en los parámetros X
- solo tiene una variable explicativa (X_i) ✓

$\therefore$ no es un modelo de regresión lineal simple

(e) $$Y_i = \beta_0 + \beta_1 \varepsilon_i$$

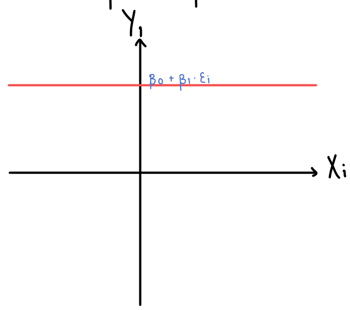
i) linealidad en los parámetros:

$$\frac{\partial Y_i}{\partial \beta_0} = 1$$

$$\frac{\partial Y_i}{\partial \beta_1} = \varepsilon_i$$

ii) linealidad en las variables:

$$Y_i = \beta_0 + \beta_1 \cdot \varepsilon_i$$



El modelo no se puede expresar como combinación lineal ni de los parámetros con las variables, ni viceversa. $\therefore$ No es lineal en los parámetros, ni en las variables.

iii) ¿Es modelo de regresión lineal simple?

- es lineal en los parámetros x
- solo tiene una variable explicativa (X_i) x

$\therefore$ no es un modelo de regresión lineal simple

2. **Demuestre** que el estimador mínimo-cuadrático de β_1 en un modelo de regresión lineal simple, se puede escribir como

$$b_1 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2}$$

y que

$$\sum(Y_i - \hat{Y}_i)\hat{Y}_i = 0.$$

i) P.d:

$$b_1 = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2} \quad \begin{matrix} \} * \\ \} * \end{matrix}$$

Dem:

Sea b_1 el estimador mínimo cuadrático de β_1 $\therefore$ la suma de cuadrados residual (SCR) es mínima, i.e: $SCR = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$ es mínima

$\Longrightarrow$ derivamos e igualamos a cero para encontrar los puntos críticos.

$$\text{I)}\ \left.\frac{\partial SCR}{\partial \hat{\beta}_0}\right|_{\substack{\hat{\beta}_0 = b_0 \\ \hat{\beta}_1 = b_1}} = 0 \iff \sum_{i=1}^{n} 2(y_i - b_0 - b_1 x_i)(-1) = 0$$

$$\iff \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0$$

$$\iff \sum_{i=1}^{n} y_i - nb_0 - b_1 \sum_{i=1}^{n} x_i = 0$$

$$\text{II)}\ \left.\frac{\partial SCR}{\partial \hat{\beta}_1}\right|_{\substack{\hat{\beta}_0 = b_0 \\ \hat{\beta}_1 = b_1}} = 0 \iff \sum_{i=1}^{n} 2(y_i - b_0 - b_1 x_i)(-x_i) = 0$$

$$\iff \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)x_i = 0$$

$$\iff \sum_{i=1}^{n} x_i y_i - b_0 \sum_{i=1}^{n} x_i - b_1 \sum_{i=1}^{n} x_i^2 = 0$$

I) y II) forman el siguiente sistema de ecuaciones.

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

$$\Longrightarrow \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \underbrace{\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}^{-1}}_{*} \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

$$* = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{pmatrix}$$

$$= \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

$$= \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i \\ -\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i + n\sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

$$\Longrightarrow b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Además,

$$* = \sum(x_i - \overline{x})(y_i - \overline{y})$$

$$= \sum(x_i y_i - x_i \overline{y} - \overline{x} y_i + \overline{x}\,\overline{y})$$

$$= \sum x_i y_i - \overline{y}\sum x_i - \overline{x}\sum y_i + n\overline{x}\,\overline{y}$$

$$= \sum x_i y_i - \frac{\sum y_i}{n}\sum x_i - \frac{\sum x_i}{n}\sum y_i + \cancel{n}\left(\frac{\sum x_i}{\cancel{n}}\frac{\sum y_i}{n}\right)$$

$$= \sum x_i y_i - \cancel{\frac{\sum y_i}{n}\sum x_i} - \frac{\sum x_i}{n}\sum y_i + \cancel{\sum x_i \frac{\sum y_i}{n}}$$

$$= \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$= \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n}$$

$$
\begin{aligned}
* \quad &= \Sigma (x_i - \bar{x})^2 \\
&= \Sigma (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \Sigma x_i^2 - 2\bar{x}\Sigma x_i + n\bar{x}^2 \\
&= \Sigma x_i^2 - 2\bar{x}\Sigma x_i + \cancel{n}\left(\bar{x}\,\frac{\Sigma x_i}{\cancel{n}}\right) \\
&= \Sigma x_i^2 - \cancel{2}\bar{x}\Sigma x_i + \cancel{\bar{x}\Sigma x_i} \\
&= \Sigma x_i^2 - \bar{x}\Sigma x_i \\
&= \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n} \\
&= \frac{n\Sigma x_i^2 - (\Sigma x_i)^2}{n}
\end{aligned}
$$

$$
\begin{aligned}
\Longrightarrow b_1 &= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \\
&= \frac{\dfrac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n}}{\dfrac{n\Sigma x_i^2 - (\Sigma x_i)^2}{n}} \\
&= \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} \quad //.
\end{aligned}
$$

ii) P.d:

$$\Sigma (y_i - \hat{y}_i)\hat{y}_i = 0$$

Dem:

$$
\begin{aligned}
\Sigma (y_i - \hat{y}_i)\hat{y}_i &= \Sigma (y_i\hat{y}_i - \hat{y}_i^2) \\
&= \Sigma y_i\hat{y}_i - \Sigma \hat{y}_i^2 \\
&= \Sigma y_i(b_0 + b_1 x_i) - \Sigma (b_0 + b_1 x_i)^2 \\
&= b_0\Sigma y_i + b_1\Sigma x_i y_i - \Sigma (b_0^2 + 2b_0 b_1 x_i + b_1^2 x_i^2) \\
&= b_0\Sigma y_i + b_1\Sigma x_i y_i - n b_0^2 + b_2 b_1\Sigma x_i + b_0 b_1 \Sigma x_i + b_1^2\Sigma x_i^2 \\
&= b_0\big(\underbrace{\Sigma y_i - n b_0 + b_1\Sigma x_i}_{A \rightarrow 0}\big) + b_1\big(\underbrace{\Sigma x_i y_i + b_0\Sigma x_i + b_1\Sigma x_i^2}_{B \rightarrow 0}\big)
\end{aligned}
$$

Recordemos las ecuaciones normales

$A = \Sigma y_i - n b_0 + b_1\Sigma x_i = 0$

$B = \Sigma x_i y_i + b_0\Sigma x_i + b_1\Sigma x_i^2 = 0$

$$= 0 \quad //.$$

3. A partir de una muestra de n = 200 parejas de observaciones, se calcularon las siguientes cantidades:

$$\sum X_i = 11.34, \quad \sum Y_i = 20.72, \quad \sum X_i^2 = 12.16, \quad \sum Y_i^2 = 84.96 \quad \text{y} \quad \sum X_i Y_i = 22.13$$

Con base en estas cantidades, **estime** las dos regresiones

$$\hat{Y}_i = b_0 + b_1 X_i \qquad \text{y} \qquad \hat{X}_i = a_0 + a_1 Y_i .$$

Deduzca una recta estimada para Y a partir de la segunda ecuación.
Grafique las dos rectas estimadas de Y en la misma gráfica y **comente** acerca de ellas, en particular acerca de cómo se podrían interpretar las mismas y por qué difieren.

$\hat{Y}_i = b_0 + b_1 X_i$

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \bar{X} \sum X_i} = 1.8195$$

$$b_0 = \bar{Y} - b_1 \bar{X} = 0.0004$$

$$\therefore \hat{Y}_i = 0.0004 + 1.8195 X_i$$ //

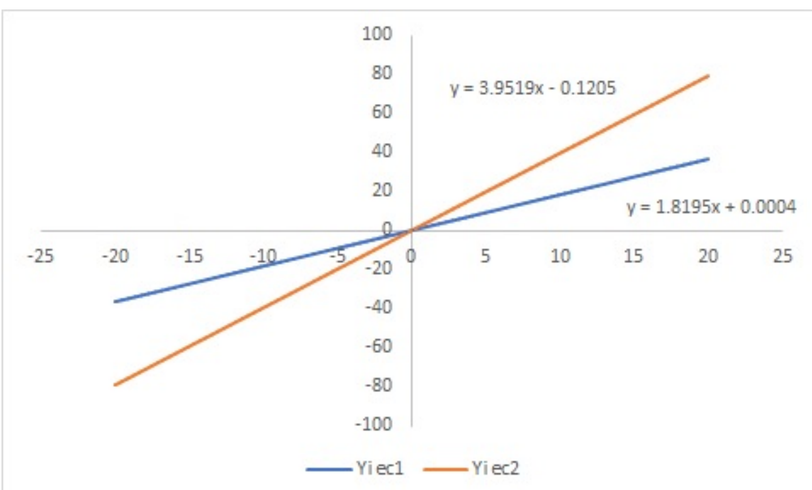
$\hat{X}_i = a_0 + a_1 Y_i$

$$a_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2} = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum Y_i^2 - \bar{Y} \sum Y_i} = 0.25304$$

$$a_0 = \bar{X} - a_1 \bar{Y} = 0.0304$$

$$\therefore \hat{X}_i = 0.0304 + 0.25304 Y_i$$

$$\Longrightarrow Y_i = -0.1205 + 3.9519 \hat{X}_i$$ //



Para obtener $\hat{Y}_i$, minimizamos las diferencias al cuadrado entre los valores observados de Y y los valores estimados de Y; i.e: estamos minimizando distancias verticales.
Para obtener $\hat{X}_i$, minimizamos las diferencias al cuadrado entre los valores observados de X y los valores estimados de X; i.e: estamos minimizando distancias horizontales.
$\therefore$ Es de esperarse que estas rectas estimadas de Y tengan diferentes pendientes.