Equipo 6
Álvaro Acedo 174052
Diana Santiago 175325
Jimena Reyes 173361
Paulina Mazariegos 171929
Pamela Goya 171789

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 $i = 1, \dots, n$. Determine:

(i) ¿Cuáles son lineales en los parámetros β_0 y β_1?

Lineal en los parámetros: las derivadas de la ecuación del modelo, respecto a cada uno de sus parámetros no dependen de ningún parámetro.

(ii) ¿Cuáles son lineales en las variables Y y X?

Lineal en las variables: comprobamos al expresar el modelo como combinación lineal de las variables con "algún tipo de parámetros".

(iii) ¿Cuáles podrían considerarse modelos de regresión lineal simple? (posiblemente con v. transformadas)

Ser modelo de regresión lineal simple: es necesario que el modelo sea lineal en los parámetros y que tenga una variable explicativa.

a) $Y_i = \beta_0 + \beta_1 / x_i + \varepsilon_i$

i) linealidad en los parámetros

$\frac{\partial Y_i}{\partial \beta_0} = 1$ la derivada de la ecuación que describe al modelo, respecto a cada uno de sus parámetros, no depende de ningún parámetro.

$\frac{\partial Y_i}{\partial \beta_1} = \frac{1}{x_i}$ $\therefore$ Es lineal en los parámetros β_0 y β_1

ii) linealidad en las variables

$(1, 1/x_i) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \beta_0 + \beta_1 / x_i$

Los incisos (i) y (iii) están bien.

$\therefore$ Es lineal en las variables

El inciso (ii) está mal porque el modelo no es linea

iii) ¿Es modelo de regresión lineal simple?

☑ Lineal en los parámetros

☑ Tiene una variable explicativa

$\therefore$ Es un modelo de regresión lineal simple

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

i) linealidad en los parámetros

$\frac{\partial Y_i}{\partial \beta_0} = 1$ la derivada de la ecuación que describe al modelo, respecto a cada uno de sus parámetros, no depende de ningún parámetro.

$\frac{\partial Y_i}{\partial \beta_1} = \log(x_i)$ $\therefore$ Es lineal en los parámetros β_0 y β_1

ii) linealidad en las variables

Los incisos (i) y (iii) están bien.

$(1, \log(x_i)) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \beta_0 + \beta_1 \log(x_i)$

El inciso (ii) está mal porque el modelo no es lin

$\therefore$ Es lineal en las variables

iii) ¿Es modelo de regresión lineal simple?

☑ Lineal en los parámetros

☑ Tiene una variable explicativa

$\therefore$ Es un modelo de regresión lineal simple

c) $y_i = \beta_0 x_i^{\beta_1}$

i) linealidad en los parámetros

$$\frac{\partial y_i}{\partial \beta_0} = x_i^{\beta_1}$$

$$\frac{\partial y_i}{\partial \beta_1} = \beta_0 x_i^{\beta_1} \log(x_i)$$

Al menos una derivada de la ecuación que describe al modelo, respecto a cada uno de sus parámetros, depende de otro parámetro.

∴ No es lineal en los parámetros β_0 y β_1

ii) linealidad en las variables

El modelo no se puede expresar como combinación lineal de las variables con los parámetros.

∴ No es lineal en las variables

Están bien los tres incisos.

iii) ¿Es modelo de regresión lineal simple?

☐ Lineal en los parámetros

☑ Tiene una variable explicativa

∴ No es un modelo de regresión lineal simple

d) $\log(y_i) = \log(\beta_0) + \beta_1 \log(x_i) + \varepsilon_i$

$$\Rightarrow \quad e^{\log(y_i)} = e^{\log(\beta_0)} e^{\beta_1 \log(x_i)} e^{\varepsilon_i}$$

$$\Rightarrow \quad y_i = \beta_0 x_i^{\beta_1} e^{\varepsilon_i} \qquad e^{\varepsilon_i} = \varepsilon_i^*$$

i) linealidad en los parámetros

$$\frac{\partial y_i}{\partial \beta_0} = x_i^{\beta_1} e^{\varepsilon_i}$$

$$\frac{\partial y_i}{\partial \beta_1} = \beta_0 e^{\varepsilon_i} x_i^{\beta_1} \log(x_i)$$

Al menos una derivada de la ecuación que describe al modelo, respecto a cada uno de sus parámetros, depende de otro parámetro.

∴ No es lineal en los parámetros β_0 y β_1

ii) linealidad en las variables

El modelo no se puede expresar como combinación lineal de las variables con los parámetros.

∴ No es lineal en las variables

Los tres incisos están bien.

iii) ¿Es modelo de regresión lineal simple?

☐ Lineal en los parámetros

☑ Tiene una variable explicativa

∴ No es un modelo de regresión lineal simple

e) $y_i = \beta_0 + \beta_1 \varepsilon_i$

i) linealidad en los parámetros

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

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

La derivada de la ecuación que describe al modelo, respecto a cada uno de sus parámetros, no depende de ningún parámetro.

∴ Es lineal en los parámetros β_0 y β_1

ii) linealidad en las variables

$$(1, 1)\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \beta_0 + \beta_1 \quad \longrightarrow \text{"constante"}$$

El modelo no se puede expresar como combina

∴ Es lineal en las variables

iii) ¿Es modelo de regresión lineal simple?

☑ Lineal en los parámetros

☐ Tiene una variable explicativa

Notamos que tiene ninguna variable explícita (x_i).

∴ 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. Sea b_1 el estimador mínimo cuadrático de β_1, tq la Suma de Cuadrados Residual (SCR) es mínima.

Notar que: $SCR = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

Para obtener los puntos críticos:

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

$$\iff \sum_{i=1}^{n} y_i - nb_0 - b_1 \sum_{i=1}^{n} x_i = 0 \qquad \text{(I)}$$

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

$$\iff \sum_{i=1}^{n} x_i y_i - b_0 \sum_{i=1}^{n} x_i - b_1 \sum_{i=1}^{n} x_i^2 = 0 \qquad \text{(II)}$$

(I) y (II) forman un sistema de ecuaciones de la forma:

$$\underbrace{\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} b_0 \\ b_1 \end{pmatrix}}_{\bar{x}} = \underbrace{\begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}}_{b}$$

(Los índices de las sumas recorren $i = 1$ hasta $i = n$

Sea $A\bar{x} = b \Longrightarrow \bar{x} = A^{-1} b$

Notar que

$$A^{-1} = \frac{1}{n\sum x_i^2 - \sum x_i \sum x_i} \begin{pmatrix} \sum x_i & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

$$\Longrightarrow A^{-1} b = \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i \sum y_i - \sum x_i \sum x_i y_i \\ -\sum x_i \sum y_i + n \sum x_i y_i \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \bar{x}$$

$$\Longrightarrow b_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

P.d. $\dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

Notar que:

(*) $\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum \left[x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y} \right]$

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

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

$$= \sum x_i y_i + (\sum x_i)(\sum y_i)\left[\cancel{\frac{1}{n}} - \cancel{\frac{1}{n}} - \frac{1}{n}\right] = \underline{\underline{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}}$$

(**) $\sum (x_i - \bar{x})^2 = \sum \left[x_i^2 - 2x_i\bar{x} + \bar{x}^2 \right] = \sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2$

$$= \sum x_i^2 - 2\left(\frac{\sum x_i}{n}\right)\sum x_i + n\left(\frac{\sum x_i}{n}\right)^2 = \sum x_i^2 - \frac{2}{n}(\sum x_i)^2 + \frac{1}{n}(\sum x_i)^2$$

$$= \underline{\underline{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}}$$

Sustituyendo:

$$(*)\quad \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2} = \frac{\sum X_iY_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{1}{n}(\sum X_i)^2} \cdot \left(\frac{n}{n}\right) = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} \quad (**)$$

Multiplicamos por 1

Bien

$$\Rightarrow b_1 = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2} \quad \blacksquare$$

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

Como $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = b_0 + b_1 X_i$

$$\Rightarrow \sum(Y_i-\hat{Y}_i)\hat{Y}_i = \sum(Y_i\hat{Y}_i - \hat{Y}_i^2) = \sum Y_i(b_0 + b_1X_i) - (b_0+b_1X_i)^2$$

$$= b_0\sum Y_i + b_1\sum X_iY_i - \sum[b_0^2 + 2b_0b_1X_i + (b_1X_i)^2]$$

$$= b_0\sum Y_i + b_1\sum X_iY_i - nb_0^2 - b_0b_1\sum X_i - b_0b_1\sum X_i - b_1\sum X_i^2$$

$$= b_0\underbrace{\left[\sum Y_i - nb_0 - b_1\sum X_i\right]}_{A} + b_1\underbrace{\left[\sum X_iY_i - b_0\sum X_i - \sum X_i^2\right]}_{B}$$

Recordar las ecuaciones normales:

$$\sum_{i=1}^{n} Y_i - nb_0 - b_1\sum_{i=1}^{n} X_i = 0 \quad \Rightarrow A = 0$$

$$\sum_{i=1}^{n} X_iY_i - b_0\sum_{i=1}^{n} X_i - b_1\sum_{i=1}^{n} X_i^2 = 0 \quad \Rightarrow B = 0$$

Bien

$$\Rightarrow \sum(Y_i-\hat{Y}_i)\hat{Y}_i = b_0(A) + b_1(B) = 0 \quad \blacksquare$$

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

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

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.

$\sum x_i = 11.32$

entonces $\bar{X} = \dfrac{11.32}{200} = 0.0567$

$\sum y_i = 20.72$

entonces $\bar{y} = \dfrac{20.72}{200} = 0.1036$

además, $S_{xy} = \dfrac{\sum x_i y_i - \sum x_i \overbrace{\sum y_i / n}^{\bar{y}}}{n-1}$

$= \dfrac{22.13 - 11.34(0.1036)}{200-1} \approx 0.1053$

$S_x^2 = \dfrac{\sum x_i^2 - (\sum x_i)^2/n}{n-1}$

$= \dfrac{12.16 - (11.34)^2/200}{200-1} \approx 0.0579$

luego, $b_1 = \dfrac{S_{xy}}{S_x^2} \approx 1.8195$

entonces $b_0 = \bar{y} - b_1 \bar{x}$

≈ 0.0004

Bien

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

$$S_y^2 = \frac{\sum y_i^2 - (\sum y_i)^2 / n}{n-1}$$

$$= \frac{84.96 - (20.72)^2/200}{200-1} \approx 0.4161$$

luego,

$$a_1 = \frac{S_{xy}}{S_y^2} \approx 0.2530$$

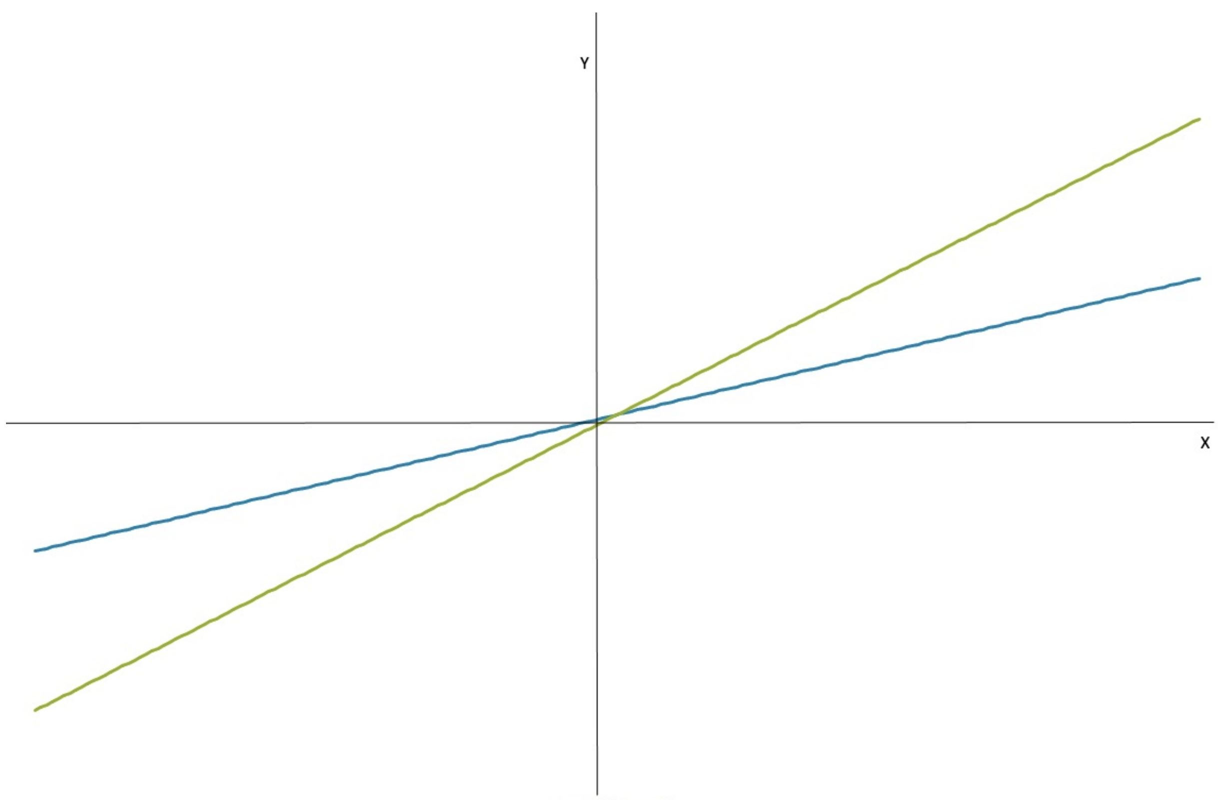
$$a_0 = \bar{x} - b_1 \bar{y}$$

$$\approx 0.0305$$

$$\therefore \hat{X}_i = 0.0305 + 0.2530\, Y_i$$

Bien

entonces $$Y_i = \frac{\hat{X}_i - 0.0305}{0.2530}$$



Notemos que 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.

Por otro lado, para obtener Y_i a través de $\hat{X}_i$ minimizamos las diferencias al cuadrado entre los valores observados de X y sus valores estimados, i.e. minimizamos distancias horizontales. Bien

Por lo tanto, es de esperarse que estas rectas estimadas de Y tengan diferentes pendientes. Como se puede ver en la gráfica, la pendiente de $\hat{Y}_i$ es menor que la de Y_i.