Tarea #1

EQUIPO #3

Sofía Alejandra Díaz Miranda 172360
David Isaac López Romero 173993
Sofía Oliva Ruiz 164595
Adriana Alavez Lujano 163480
Diego Carlos Krafft de Silva 173246

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) Es lineal en los parámetros

$$\frac{dY_i}{d\beta_0} = 1 \quad , \quad \frac{dY_i}{d\beta_1} = \frac{1}{X_i}$$

ii) No es lineal en X_i

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

$$\iff Y_i - \beta_0 - \varepsilon_i = \frac{\beta_1}{X_i}$$

$$\iff X_i = \frac{\beta_1}{Y_i - \beta_0 - \varepsilon_i}$$

Bien los tres incisos.

No es lineal en Y_i

iii) Si consideramos $Z_i = \frac{1}{X_i} \quad \forall\, i = 1, \ldots, n$ entonces:

$Y_i = \beta_0 + \beta_1 Z_i + \varepsilon_i$ la cual es regresión lineal simple porque
- lineal en sus parámetros
- solo hay una variable explicativa

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

i) Es lineal en los parámetros

$$\frac{dY_i}{d\beta_0} = 1 \quad , \quad \frac{dY_i}{d\beta_1} = \log(X_i)$$

ii) No hay relación lineal en las variables

$$\frac{dY_i}{dX_i} = \frac{\beta_1}{X_i}$$

Bien los tres incisos.

iii) Si consideramos $z_i = \log(x_i)$ entonces

$Y_i = \beta_0 + \beta_1 z_i + \varepsilon_i$ es regresión lineal simple

- relación lineal en sus parámetros
- solo hay una variable explicativa

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

i) No es lineal en sus parámetros

$$\frac{dY_i}{d\beta_0} = X_i^{\beta_1} \quad , \quad \frac{dY_i}{d\beta_1} = \beta_0 \ln(\beta_1) X_i^{\beta_1}$$

Bien los incis

ii) En general, no hay relación lineal entre X e Y.
Solo se puede cuando $\beta_1 = 1$.

iii) $Y_i = \beta_0 X_i^{\beta_1} \underset{Y_i > 0}{\iff} \log(Y_i) = \log(\beta_0) + \beta_1 \log(X_i)$

Si consideramos $\tilde{Y}_i = \log(Y_i)$, $\tilde{X}_i = \log(X_i)$ y $\tilde{\beta}_0 = \log(\beta_0)$ tenemos

$\tilde{Y}_i = \tilde{\beta}_0 + \beta_1 \tilde{X}_i$ que es regresión lineal simple

No es modelo de regresión porque no tiene error alea

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

i) $$\frac{dY_i}{d\beta_0} = \frac{1}{\beta_0} \quad , \quad \frac{dY_i}{d\beta_1} = \log(X_i)$$

No es lineal en los parámetros porque se tiene $\log(\beta_0)$ y por esto no es combinación lineal de parámetros.

ii) No es lineal entre X e Y

iii) Si consideramos $\tilde{Y}_i = \log(Y_i)$, $\tilde{X}_i = \log(X_i)$, $\tilde{\beta}_0 = \log(\beta_0)$ tenemos

$$\tilde{Y} = \tilde{\beta}_0 + \beta_1 \tilde{X}_i + \varepsilon_i$$

que es modelo de regresión lineal

Bien los tres incisos.

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

i) $\frac{dY_i}{d\beta_0} = 1$, $\frac{dY_i}{d\beta_1} = \varepsilon_i$

Es lineal en los parámetros No es lineal en los parámetros po

ii) No hay relación de X con

iii) Como no es modelo de regresión lineal simple, si $x_i = \varepsilon_i$, entonces

Si se hace esto entonces no habría error aleatorio y no podría ser u

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

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

Dem: Consideremos a $Y = \hat{\beta}_0 + \hat{\beta}_1 X_i$ y sea

$$SCR = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

Objetivo: $\min\limits_{(\beta_0, \beta_1) \in \mathbb{R}^2} SCR(\beta_0, \beta_1)$

$$\frac{\partial(SC)}{\partial \beta_0}(\beta_0, \beta_1) = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-1)$$

$$= -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

$$= -2 \left\{ \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 X_i \right\}$$

$$= -2 \left\{ n\overline{Y} - n\hat{\beta}_0 - n\hat{\beta}_1 \overline{X} \right\}$$

$$\frac{\partial(SC)}{\partial \beta_1}(\beta_0, \beta_1) = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-X_i)$$

$$= -2 \left\{ \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} \hat{\beta}_0 X_i - \sum_{i=1}^{n} \hat{\beta}_1 X_i^2 \right\}$$

$$= -2 \left\{ \sum_{i=1}^{n} X_i Y_i - n\hat{\beta}_0 \overline{X} - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \right\}$$

Así, igualamos ambas ecuaciones a cero y resulta que

$$\begin{cases} \frac{\partial(SC)}{\partial\beta_1} = 0 \\ \frac{\partial(SC)}{\partial\beta_0} = 0 \end{cases} \iff \begin{cases} n\bar{y} - n\hat{\beta}_0 - n\hat{\beta}_1\bar{x} = 0 \\ \sum_{i=1}^{n} x_i y_i - n\hat{\beta}_0\bar{x} - \hat{\beta}_1\sum_{i=1}^{n} x_i^2 = 0 \end{cases}$$

$$\iff \begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \\ \sum_{i=1}^{n} x_i y_i = n\hat{\beta}_0\bar{x} + \hat{\beta}_1\sum_{i=1}^{n} x_i^2 \end{cases}$$

Despejamos la segunda ecuación, tras sustituir la primera

$$\sum_{i=1}^{n} x_i y_i = n(\bar{y} - \hat{\beta}_1\bar{x})\bar{x} + \hat{\beta}_1\sum_{i=1}^{n} x_i^2$$

$$= n\bar{x}\bar{y} - n\hat{\beta}_1\bar{x}^2 + \hat{\beta}_1\sum_{i=1}^{n} x_i^2$$

$$= n\bar{x}\bar{y} + \hat{\beta}_1\left\{-n\bar{x}^2 + \sum_{i=1}^{n} x_i^2\right\}$$

Así, $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$

y así, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$

Ahora, veamos que son mínimos, con la Hessiana

$$\frac{\partial^2(SC)}{\partial\beta_0^2} = -2(-n) = 2n$$

$$\frac{\partial^2(SC)}{\partial\beta_1\partial\beta_0} = -2(-n\bar{x}) = 2n\bar{x} = \frac{\partial^2(SC)}{\partial\beta_0\partial\beta_1}$$

$$\frac{\partial^2(SC)}{\partial\beta_1^2} = -2\sum_{i=1}^{n}(-x_i^2) = 2\sum_{i=1}^{n} x_i^2$$

Construimos la Hessiana

$$H(SC)(\beta_0,\beta_1) = \begin{pmatrix} 2n & 2n\bar{x} \\ 2n\bar{x} & 2\sum x_i^2 \end{pmatrix}$$

Utilizamos el criterio de Sylvester con menores principales para ver si es definida positiva.

$$|H_1| = 2n > 0$$

$$|H| = 4n \sum x_i^2 - 4n\bar{x}^2 = 4n\left(\sum x_i^2 - n\bar{x}^2\right)$$

Veamos si $|H| > 0$.

$$\sum x_i^2 - n\bar{x}^2 = \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

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

$$= \sum\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right) = \sum (x_i - \bar{x})^2 > 0$$

Así pues, $|H| = 4n\left(\sum x_i^2 - n\bar{x}^2\right) > 0$, por lo que

$$b_0 = \bar{y} - b_1\bar{x}, \quad b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \quad \text{minimizan } SC$$

Ahora bien, P.D. $b_1 = \dfrac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

Ya vimos que $\sum x_i^2 - n\bar{x}^2 = \sum (x_i - \bar{x})^2$, así basta estudiar el numerador,

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

$$= \sum x_i y_i - n\bar{y}\frac{1}{n}\sum x_i - n\bar{x}\frac{1}{n}\sum y_i + \sum \bar{x}\bar{y}$$

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

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

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

De esto modo, vemos que

$$\begin{cases} \sum x_i y_i - n\bar{x}\bar{y} = \sum (x_i - \bar{x})(y_i - \bar{y}) \\ \sum x_i^2 - n\bar{x}^2 = \sum (x_i - \bar{x})^2 \end{cases}$$

Así, $b_1 = \dfrac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

Asimismo,

$$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{y}^2} = \frac{n\left\{\sum x_i y_i - n\frac{1}{n}\sum x_i \frac{1}{n}\sum y_i\right\}}{n\left\{\sum x_i^2 - n\frac{1}{n}\sum x_i \frac{1}{n}\sum x_i\right\}}$$

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

Bien.

Otra forma

Con, eta vo $b_1 = \frac{S_{xy}}{S_{x^2}}$

$$b_1 = \frac{\left(\dfrac{\sum x_i y_i - \sum x_i \left\{\frac{1}{n}\sum y_i\right\}}{\cancel{n-1}}\right)}{\left(\dfrac{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}{\cancel{n-1}}\right)} = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}$$

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

Ahora 2.2

$$\left(\begin{array}{l} b_0 = \bar{y} - b_1 \bar{x} \\ b_1 = \frac{S_{xy}}{S_{xx}} \end{array} \right)$$

P.D. $\sum (y_i - \hat{y}_i)\hat{y}_i = 0$

$$\sum (y_i - \hat{y}_i)\hat{y}_i = \sum e_i \hat{y}_i = \sum e_i (b_1 + b_0 x_i)$$

$$= b_1 \sum e_i + b_0 \sum e_i x_i \overset{(*)}{\underset{(**)}{=}} 0$$

Vemos que

$$\sum e_i = \sum (y_i - \hat{y}_i) = \sum (y_i - b_0 - b_1 x_i) = \sum y_i - n b_0 - b_1 \sum x_i = 0 \ldots (*)$$

Primera ecuación normal

Asimismo,

$$\sum e_i x_i = \sum x_i (y_i - \hat{y}_i) = \sum x_i (y_i - b_0 - b_1 x_i)$$

$$= \sum x_i y_i - b_0 \sum x_i - b_1 \sum x_i^2 = \sum x_i y_i - b_0 n \bar{x} - b_1 \sum x_i^2$$

$$= \sum x_i y_i - (\bar{y} - b_1 \bar{x}) n \bar{x} - b_1 \sum x_i^2$$

$$= \sum x_i y_i - n \bar{x}\bar{y} + b_1 n \bar{x}^2 - b_1 \sum x_i^2$$

$$= \sum x_i y_i - n \bar{x}\bar{y} - b_1 \left(\sum x_i^2 - n\bar{x}^2\right)$$

$$= \cancel{\sum x_i y_i - n \bar{x}\bar{y}} - \left\{ \frac{\cancel{\sum x_i y_i - n\bar{x}\bar{y}}}{\cancel{\sum x_i^2 - n\bar{x}^2}} \right\} \left(\cancel{\sum x_i^2 - n\bar{x}^2} \right)$$

Bien

$$= 0 \ldots (**)$$

$$= \frac{S_{xy}}{S_{xx}} \sum (x_i - \bar{x}) y_i - \frac{S_{xy}}{S_{xx}} \bar{y} \sum (x_i - \bar{x}) + S_{xy}$$

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.

Sabemos que para obtener b_1, debemos calcular

$$b_1 = \frac{S_{xy}}{S^2_x}$$

Donde $S_{xy} = \dfrac{\sum X_i Y_i - \sum X_i \sum Y_i / n}{n-1}$ Cov. muestral de X y Y

y $S^2_x = \dfrac{\sum X^2_i - (\sum X_i)^2/n}{n-1}$ Varianza muestral de X

$\Rightarrow$ con los datos que nos proporcionan

$$S_{xy} = \frac{(22.13) - (11.34)\frac{(20.72)}{200}}{200-1} = 0.105302392$$

Esta parte está mal. -3

$$S^2_x = \frac{(12.16) - \frac{(12.16)^2}{200}}{200-1} = 0.0573903115б$$

$$b_1 = \frac{0.105302392}{0.05739031156} = 1.834846146$$

y para calcular $b_0 = \bar{Y} - b_1 \bar{X}$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{20.72}{200} = 0.1036$$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{11.34}{200} = 0.0567$$

$$b_0 = 0.1036 - 1.834846146\,(0.0567) = 0.000435776478 2$$

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

Falta la X. -1

$$\therefore \quad \underline{\hat{Y} = 0.0004357764782 + 1.834846146} \quad (1)$$

Ahora para calcular a_0 y a_1 necesitamos

$$a_1 = \frac{S_{xy}}{S^2 y}$$

Donde,

$$S^2 y = \frac{\sum Y^2_i - (\sum Y_i)^2/n}{n-1} \qquad \text{Varianza muestral de } Y$$

$$= \frac{84.96 - \frac{(84.96)^2}{200}}{200-1} = 0.2455728241$$

$$a_1 = \frac{0.105302392}{0.2455728241} = 0.4287933748$$

Y para calcular $a_0 = \overline{X} - a_1 \overline{Y}$

$$\Rightarrow a_0 = 0.0567 - 0.4287933748\,(0.1036)$$

$$= 0.0567 - 0.04442299363 = 0.01227700637$$

$$\Rightarrow \hat{X}_i = a_0 + a_1 \hat{Y}_i$$

$$\therefore \quad \underline{\hat{X} = 0.01227700637 + 0.4287933748\,\hat{Y}}$$

Ahora deducimos la recta para $\hat{Y}$, despejando

$$0.4287933748\,\hat{Y} = \hat{X} - 0.0122700637$$

$$\hat{Y} = \frac{\hat{X} - 0.0122700637}{0.4287933748}$$

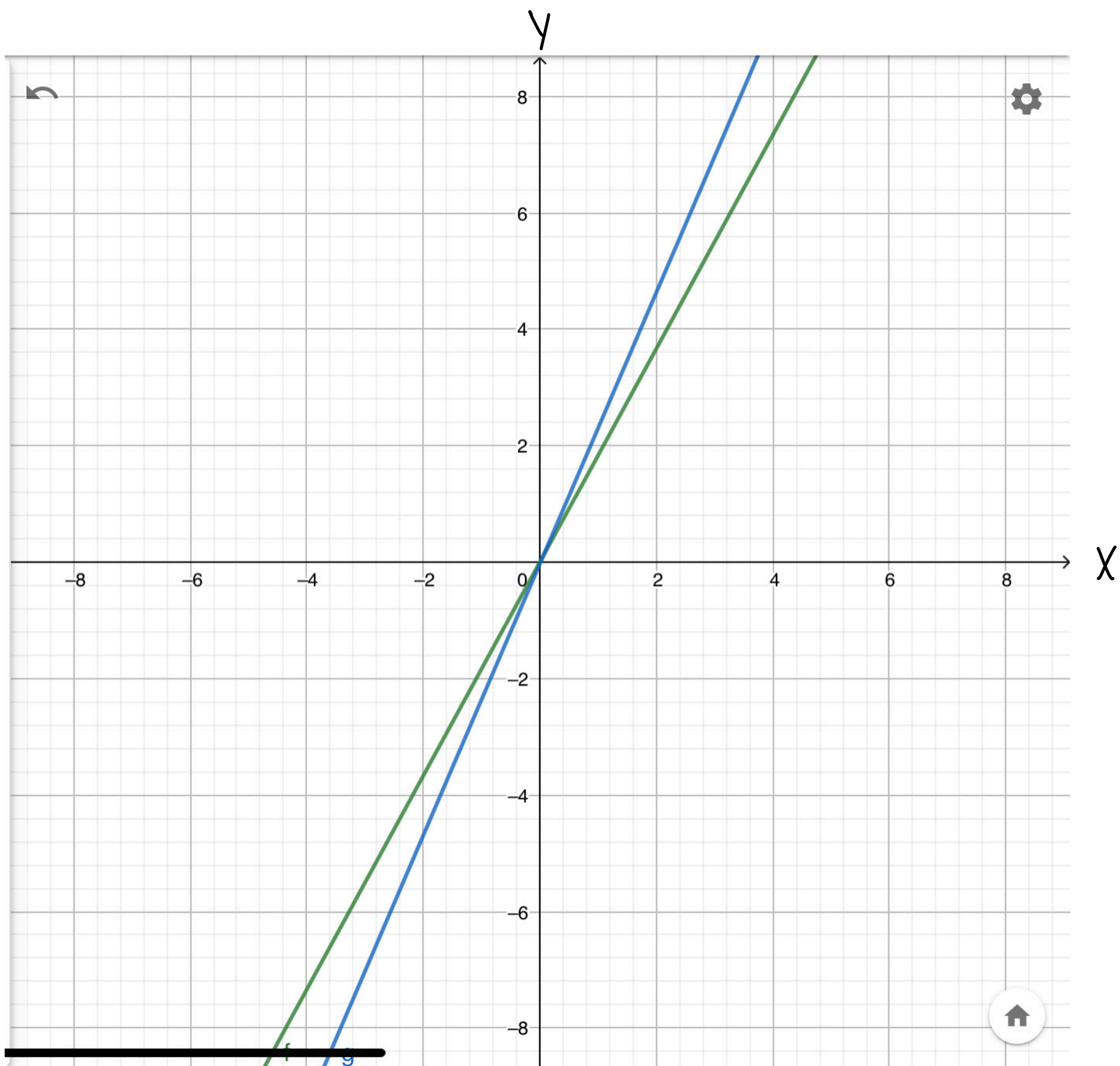
Esto está mal porque se acarreó el erro

(2')

Ahora graficamos las dos rectas de $\hat{Y}$ (1) y (2) en la misma gráfica

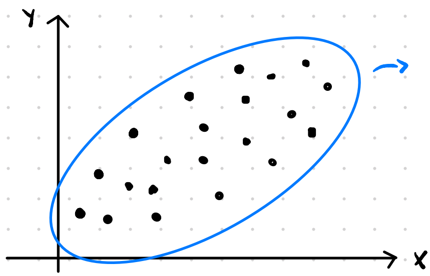
$$\hat{Y} = 0.0004357764782 + 1.834846146\,\bar{X} \quad (1)$$

$$\hat{Y} = \frac{\hat{X} - 0.0122700637}{0.4287933748} \quad (2')$$



Al estimar a Y a partir de X y a X a partir de Y resulta en rectas distintas.

Para verlo mejor, supongamos un diagrama de dispersión.

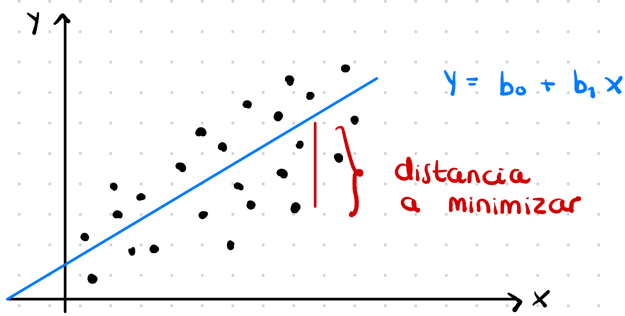


En la nube se encuentra toda la muestra.

Cuando estimamos a Y a través de X por el modelo $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ utilizamos mínimos cuadrados óptimos

$$\min_{(\beta_0, \beta_1) \in \mathbb{R}^2} SCR = \min_{(\beta_0, \beta_1) \subset \mathbb{R}^2} \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

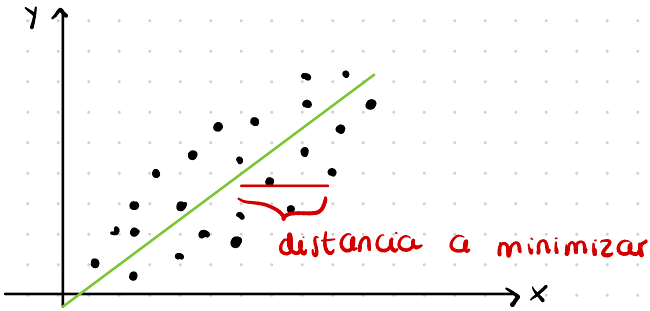
que minimiza la distancia entre una recta de estimación con los valores Y_i como se puede observar



Por otro lado, si estimamos X a partir de Y con $\hat{X}_i = \alpha_0 + \alpha_1 Y_i$ hacemos la misma estimación por minimización de cuadrados, pero se tiene

$$\min_{(a_0, a_1) \in \mathbb{R}^2} \widetilde{SCR} = \min_{(a_0, a_1) \in \mathbb{R}^2} \sum_{i=1}^{n} (X_i - (\alpha_0 + \alpha_1 Y_i))^2$$

que minimiza la distancia entre una recta de estimación con los valores X_i, como se puede observar a continuación



* Como se minimizan distancias diferentes se obtienen rectas diferentes.

Bien

$$\overbrace{\hat{\beta}_1 = \frac{\operatorname{Cov}(x,y)}{\operatorname{Var}(x)}}^{y \text{ on } x} \qquad \overbrace{\hat{\beta}_1 = \frac{\operatorname{Cov}(y,x)}{\operatorname{Var}(y)}}^{x \text{ on } y}$$

↳ Como son dos maneras de estimar tenemos valores diferentes de pendientes. Así, aunque tengan la misma relación lineal, tienen distinta recta de estimación.