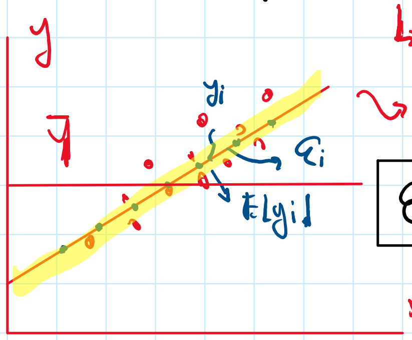
Primer taller

jueves, 7 de noviembre de 2024 3:43 p. m.

$$\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i} \; : \; \boxed{k=1}$$

$$\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i} \; ; \; \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2):$$

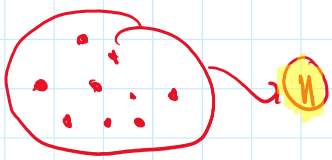
↳ No son cuantificables



$E[Y_i] = E[\beta_0 + \beta_1 X_{i1} + \varepsilon_i] = \boxed{\beta_0 + \beta_1 X_{i1}}$

$\boxed{\varepsilon_i = y_i - E[y_i]}$; $\boxed{\hat{\varepsilon}_i = y_i - \hat{y}_i}$

(1). MCO : $\hat{\beta}_0$; $\hat{\beta}_1$ (V.a)



N

Inferencia: Generalización

$\boxed{\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}}$: $\boxed{Y = bx + c}$

(Recta de regresión aj).

(1). Realizar un RLS:

Y:	Overall Score
X:	GI

↳ Government Integrity. (unidades)

Score (puntaje) : 0 – 100.

Operador Pipe: ▷ : %>%

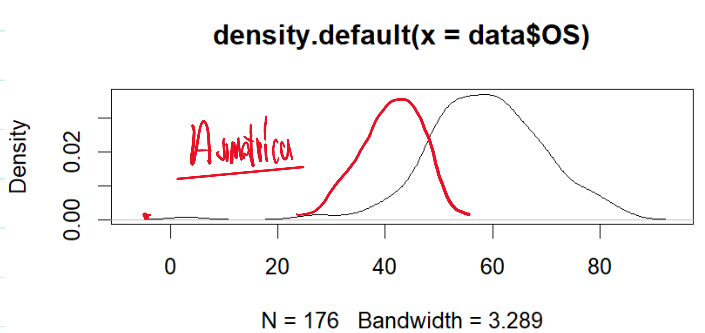
Ctrl + Shift + M

Cos(sen(x)) ; x ▷ sen() ▷ cos()

n = 176

$$\boxed{Y_i \sim N(\beta_0 + \beta_1 X_{i1}, \sigma^2)}$$

El modelo sí podría ser adecuado.



(2). Punto 2: $\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i}$; $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2):$

Hallar una recta ajustada: $\boxed{\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}}$

- Esta recta siempre se interpretará en términos de Promedios. (Esta recta es un promedio)

• $\hat{\beta}_0 \sim N(\beta_0, \sum^n x_i^2 \sigma^2)$; $\hat{\beta}_1 \sim N(\beta_1, \sigma^2/S_{xx})$

Esta recta siempre se interpretará en términos de Promedios. (Esta recta es un promedio)

- $\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sum_{i=1}^{n} x_i^2 \sigma^2}{n S_{xx}}\right)$; $\hat{\beta}_1 \sim N(\beta_1, \sigma^2 / S_{xx})$ → Nos permiten estimar los parámetros

- $\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$; $\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})} = \frac{S_{xy}}{S_{xx}}}$; $\boxed{Y \sim X}$ Regresión

```
(Intercept)
   40.40907
> beta_1
         GI
0.4104104
>
```

$\hat{\beta}_0$, $\hat{\beta}_1$; $\bar{y}$: $\bar{x}$

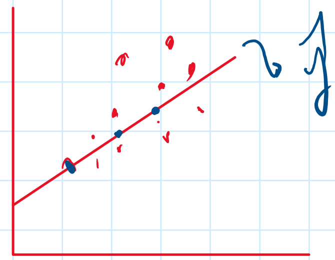
```
[1] 58.64318
> mean_x
[1] 44.42898
```

```
        1         2         3         4         5
6.8516950 5.1204639 7.9627059 9.4668097 3.9470903
```

$\hat{\varepsilon}_1$ $\hat{\varepsilon}_2$ ...

```
       1        2        3        4        5        6
76.64830 77.87954 74.63729 70.53319 75.25291 80.38304
```

$\hat{y}_1$ $\hat{y}_2$ ... :



(3). Punto 3: Parámetros — Prueba de hipótesis — $\boxed{\text{(Probar para } \beta_0)}$

$\begin{cases} H_0: \text{No hay diferencias significativas respecto a algo} \\ H_1: \text{Sí hay} \end{cases}$: $\begin{cases} H_0: \beta_0 = c \\ H_1: \beta_1 \neq c \end{cases}$; $\boxed{c = 0}$

$\boxed{\text{significancia: } c = 0}$ $\boxed{\text{(Prueba T-student)}}$

1, $\boxed{\text{Prueba de dos colas}}$

$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sum_{i=1}^{n} x_i^2 \sigma^2}{n S_{xx}}\right)$ → La suma de los errores ε_i

→ $\boxed{MSE}$: Distribuida T. :

$T = \frac{\hat{\beta}_0 - c}{\sqrt{Var[\hat{\beta}_0]}}$

$Var[\hat{\beta}_0] : \frac{\sum_{i=1}^{n} x_i^2 \sigma^2}{n S_{xx}}$; MSE

$T = \frac{\hat{\beta}_0 - 0}{\sqrt{Var[\hat{\beta}_0]}} \sim t_{n-2}$ Bajo H_0 cierta → $\boxed{\text{En el caso de una regresión lineal simple.}}$

$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$ → Dos estimadores

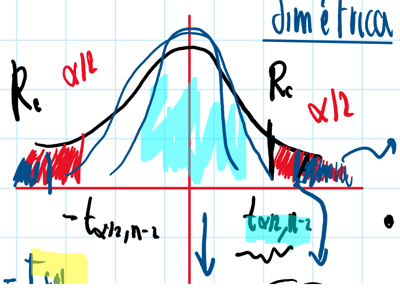
- Rechazar o no rechazar H_0

(1). $R_c : \{ |T_{cal}| > t_{\alpha/2, n-2} \}$ → $\boxed{\text{Bajo } H_0}$

(2). Valor P : $P(|t_{n-2}| > |T_{calc}|)$

T_{calc} : $\boxed{T \text{ pero cuando ya se evalúa}}$

Simétrica



$\boxed{\text{Valor p}}$

- $\boxed{\alpha = P(\text{Error tipo I}) = P(\text{Rechazar } H_0 \mid H_0 \text{ cierta})}$

$-t_{\alpha/2,n-2}$ $\quad t_{\alpha/2,n-2}$

$-T_{cal}$

R_a T_{cal} ~ percentil

- $\boxed{\alpha = P(\text{Error tipo I}) = P(\text{Rechazar } H_0 \mid H_0 \text{ cierta})}$

↳ se asemeja a un umbral:

$\boxed{\alpha = 0{,}05 \quad \alpha = 0{,}1 \quad \alpha = 0{,}01}$

$R_c: \{ T_{calc} > t_{\alpha/2,n-2} \ \vee \ -T_{calc} < -t_{\alpha/2,n-2} \}$

(*). En resumen:

$\boxed{\begin{array}{l} (1).\ VP = P(|t_{n-2}| > |t_{calc}|) < \alpha \\ (2).\ \text{Región } |T_{cal}| > |t_{\alpha/2,n-2}| \end{array}}$

Probar la significancia de un parámetro, para saber si tiene un efecto en el modelo de regresión. Si es susceptible de interpretación.

$n = 176$

En la práctica: Para β_0: $T_{cal} = \dfrac{40{,}40907 - 0}{S.E(\hat{\beta}_0)} = \dfrac{40{,}40907}{0{,}41041} = 34{,}27$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 40.40907    1.17916   34.27   <2e-16 ***
GI           0.41041    0.02389   17.18   <2e-16 ***
```

$\hat{\beta}_0 \quad \hat{\beta}_1$

$\boxed{P(|t_{176-2}| > 34{,}27) = 2e\text{-}16 = 0} < \alpha$

$\boxed{\alpha = 0{,}05}$

- Se puede concluir, que existe suficiente evidencia para rechazar H0. El parámetro B0 es significativamente diferente de 0.

↳ Tiene un efecto en el modelo.

$\boxed{\text{La covariable } X, \text{ debe contener al cero}}$

$\boxed{X: \text{Área de casa}}$ $\quad X = 0$.

$\boxed{(1).\ \text{Significancia}; (2).\ R_c.\ (\beta_0)}$

$\boxed{X: \text{no tiene al cero: No es interpretable}}$

$\boxed{(\text{En el caso de } \beta_1):}$ $\quad T_{cal}: \dfrac{0{,}4104 - 0}{S.\hat{e}(\hat{\beta}_1)} = 17{,}18 : \boxed{VP = 0} < \alpha.$

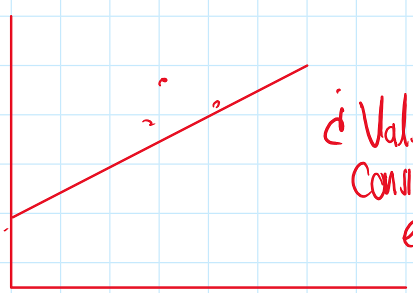
$\boxed{\text{Cuantifica el efecto de } X_i}$

El parámetro B1 es significativo. Es significativamente diferente de cero ----> La integridad del gobierno tiene un efecto sobre el promedio del overall score.

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i$

- En el caso de regresión LS.

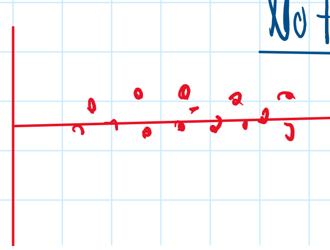
Validar la significancia de B1, equivale a probar la significancia del modelo en general, es decir, la significancia de la regresión (el efecto de la covarible).

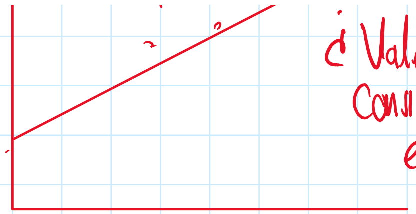


¿Vale la pena considerar este efecto?

No tengo pendiente

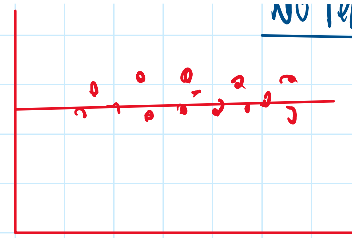
β_0





¿Vale la pena considerar este efecto?

β_0



No tengo pendiente

(4). Punto 4: I.C para los parámetros ¿Son significativos?

Confianza: Grado de certeza (seguridad)

β_0

(+)

: 95% Confianza

α: Conf: $1-\alpha$.

- El ancho y precisión del int. varían en función del nivel de con.

95% ( )
90% ( )
99% ( )

```
        2.5 %   97.5 %
(Intercept) 38.08178 42.73636
> confint_beta1
        2.5 %    97.5 %
GI 0.3632559 0.4575649
```

Para β_0: $\boxed{\hat{\beta}_0 \pm t_{\alpha/2,\, n-2} \cdot \hat{s.e}(\hat{\beta}_0)}$ $\boxed{\hat{\beta}_1 \pm t_{\alpha/2,\, n-2} \cdot \hat{s.e}(\hat{\beta}_1)}$

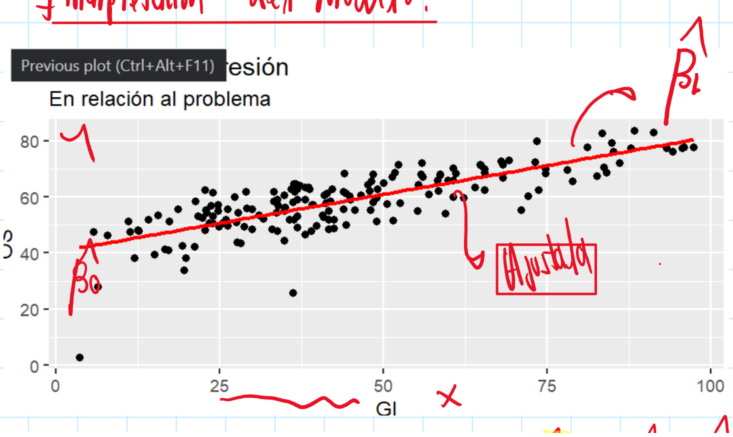
$\Downarrow$

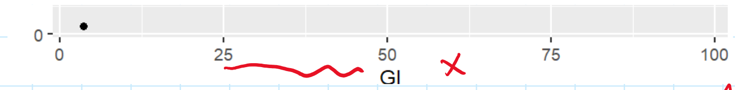
(38,08178 ; 42,73636) (0,363 ; 0,457)

- No deben contener al cero

- Con un nivel de confianza del 95%, se puede determinar que el valor para B1 está comprendido entre 0,363 y 0,457. Con un 95% de confianza, es posible determinar que el puntaje de integridad de gobierno tiene un efecto comprendido entre 0,363 y 0,457 unidades. → Sobre el OS

Interpretación del Modelo:





Interpretación problema:

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

$\widehat{OS}_i = 40,40907 + 0,4104\, GI_i$

¿Qué significa?

OS: Overal Score

GI: Goverment Integrity

Interpretación $\hat{\beta}_1$: Por un aumento unitario del puntaje de integridad del gobierno (GI) el cambio promedio del puntaje general del Índice de libertad económica es de 0,4104 $\hat{\beta}_1$

$E[Y_i]$

Interpretar $\hat{\beta}_0$:

el valor promedio de Y (Overal score) cuando X=0 → puntaje de integridad del gobierno es cero