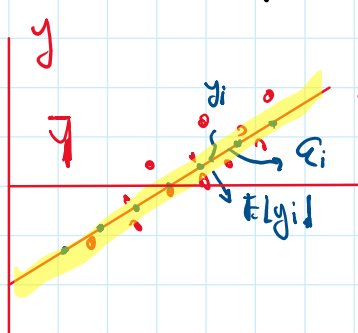


$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i \quad : \quad k=1$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i \quad ; \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

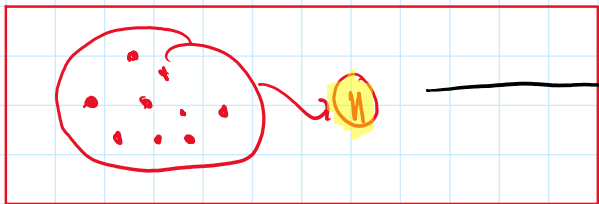


↳ No son identifiable

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

$$\epsilon_i = y_i - E[y_i] \quad ; \quad \hat{\epsilon}_i = y_i - \hat{y}_i$$

$$(1). \text{MCO} : \beta_0^1, \beta_1^1 \quad (V.M)$$



Inferencia: Generalización

$$Y_i = \beta_0^1 + \beta_1^1 X_{i1} \quad : \quad Y = bx + c$$

(Recta de regresión aj.)

(1). Realizar un RLS:

Y: Overall Score
X: G1

Operator Pipe:

▷ : (% > %)
ctrl + shift + m

N = 176

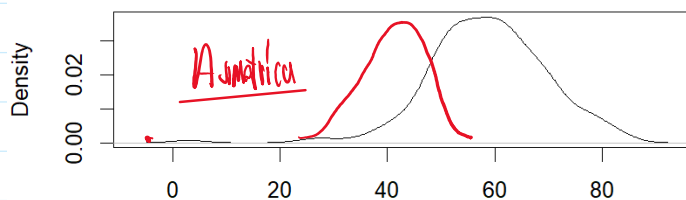
$$Y_i \sim N(\beta_0 + \beta_1 X_{i1}, \sigma^2)$$

Gov (sen(x))
↳ Government Integrity (Cinco años)
Score (puntuaje): 0 - 100.

sen() > cos()

El modelo si podría ser adecuado.

density.default(x = data\$OS)



N = 176 Bandwidth = 3.289

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

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

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

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

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

$$\beta_0 \sim N\left(\beta_0, \frac{\sum_{i=1}^n x_i^2 \sigma^2}{n S_{yy}}\right) ; \beta_1 \sim N\left(\beta_1, \sigma^2 / S_{xx}\right)$$

→ No permiten estimar los parámetros

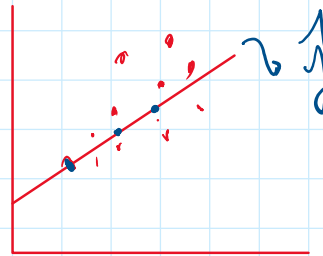
$$\beta_0 = \bar{y} - \beta_1 \bar{x} ; \beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} ; 1 \sim x \text{ Regresión}$$

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

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

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

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



(3). Ponto 3: Parámetros

Pruebas de hipótesis

(Probar por β_0)

H_0 : No hay diferencias significativas respecto a algo : $H_0: \beta_0 = c ; c=0$
 H_1 : Sí hay : $H_1: \beta_1 \neq c$

significancia: $C=0$

(Prueba t-student).

1. Prueba de dos colas

$$\beta_0 \sim N\left(\beta_0, \frac{\sum_{i=1}^n x_i^2 \sigma^2}{n S_{yy}}\right) \rightarrow \text{La suma de los errores } e_i$$

MSE : Distribución T :

$$T = \frac{\beta_0 - c}{\sqrt{\text{Var}(\beta_0)}}$$

$$\text{Var}(\beta_0) = \frac{\sum_{i=1}^n x_i^2 \sigma^2}{n S_{yy}}$$

$$T = \frac{\beta_0 - 0}{\sqrt{\text{Var}(\beta_0)}} \sim t_{n-2}$$

En el caso de una regresión lineal simple.

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Dos estimadores

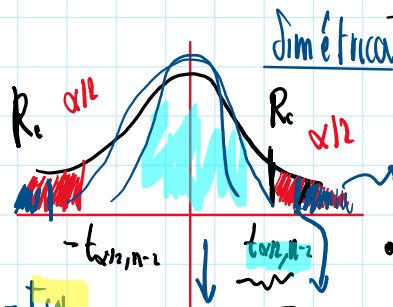
• Rechazar o no rechazar H_0 .

(1). $R_c = \{ |T_{calc}| > t_{\alpha/2, n-2} \} \rightarrow \text{Bajo } H_0$

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

T_{calc} : T pero cuando ya te evalúen

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



$\alpha = P(\text{Error tipo I}) = P(\text{Rechazar } H_0 | H_0 \text{ cierta})$
 $\alpha = 0,05$
 $\alpha = 0,1$
 $\alpha = 0,01$

$R_c: T_{calc} > t_{\alpha/2, n-2} \vee -T_{calc} < -t_{\alpha/2, n-2}$
 se arregla a un umbral

(*) En resumen:

- (1). $VP = P(H_0 > |t_{calc}|) < \alpha$
- (2). Región $|T_{calc}| > t_{\alpha/2, n-2}$

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

En la práctica: Para β_0 : $T_{calc} = \frac{40,40907 - 0}{0,41041} = \frac{40,40907}{0,41041} = 98,46$

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 ***

$P(|t_{346-2}| > 34,27) = 2e-16 = 0 < \alpha$

$\alpha = 0,05$

- Se puede concluir, que existe suficiente evidencia para rechazar H_0 . El parámetro B_0 es significativamente diferente de 0.

tiene un efecto en el modelo

La covariable X , debe contener al cero

X : Área de casa $X=0$.

(1). Significancia; (2). R^2 . (β_0)

X : no tiene al cero: No es interpretable

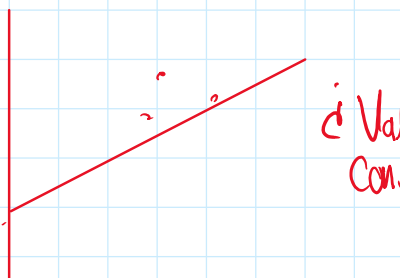
(En el caso de β_1): $T_{calc} = \frac{0,41041 - 0}{0,02389} = 17,18$: $VP = 0 < \alpha$.

El parámetro B_1 es significativo. Es significativamente diferente de cero ----> La integridad del gobierno tiene un efecto sobre el promedio del overall score.

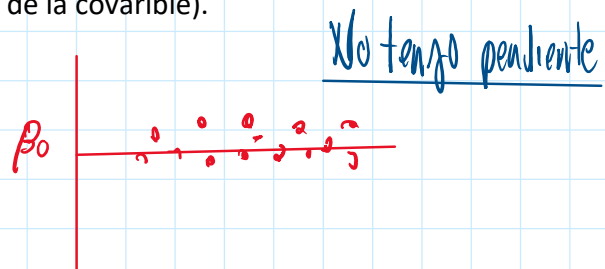
$\hat{y} = \beta_0 + \beta_1 X_i$
Cuantifica el efecto de X_i

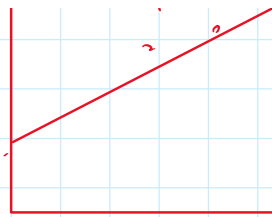
En el caso de regresión LS.

Validar la significancia de B_1 , equivale a probar la significancia del modelo en general, es decir, la significancia de la regresión (el efecto de la covariable).



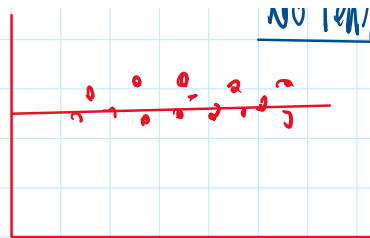
¿Vale la pena considerar este efecto?





¿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)

El ancho y precisión del Int. Varían en función del nwe/d 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 :

$$\hat{\beta}_0 \pm t_{n-2, \alpha/2} \cdot \text{S.E.}(\hat{\beta}_0)$$

$$\hat{\beta}_1 \pm t_{n-2, \alpha/2} \cdot \text{S.E.}(\hat{\beta}_1)$$

(38,08178; 42,73636)

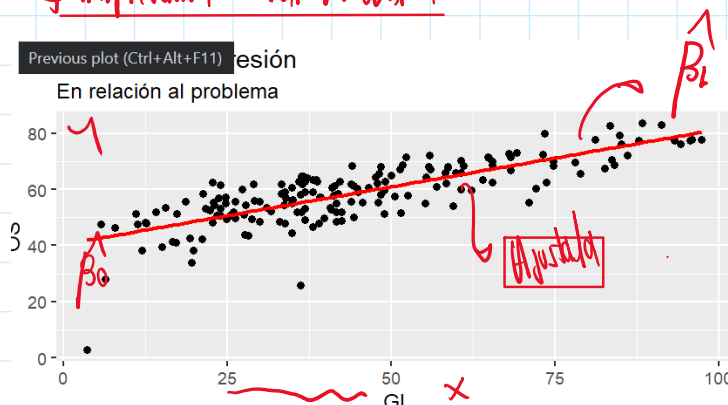
(0,363; 0,457)

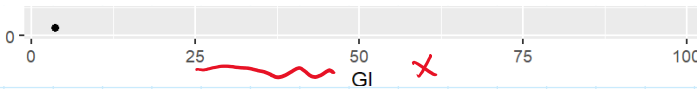
• No deben contener al cero

- Con un nivel de confianza del 95%, se puede determinar que el valor para β_1 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:

$$Y_i = \beta_0 + \beta_1 X_i$$

$$OS_i = 40,40901 + 0,4104 GI_i$$

¿Qué significa?

OS: Overall Score

GI: Government Integrity

Interpretación β_1 :

$E(Y_i)$

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.

Interpretación β_0 :

el valor promedio de (Overall score) cuando $X=0$ es el puntaje de integridad del gobierno es cero.