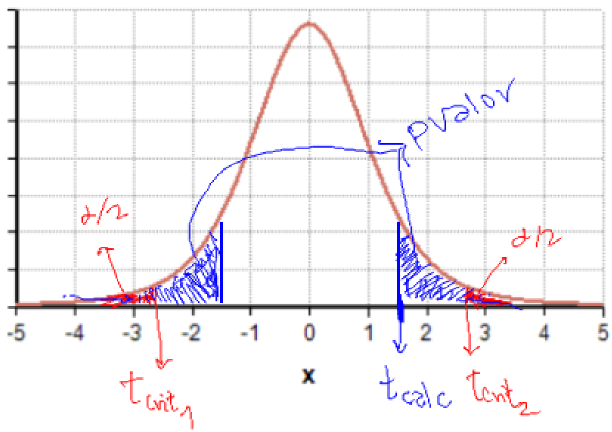
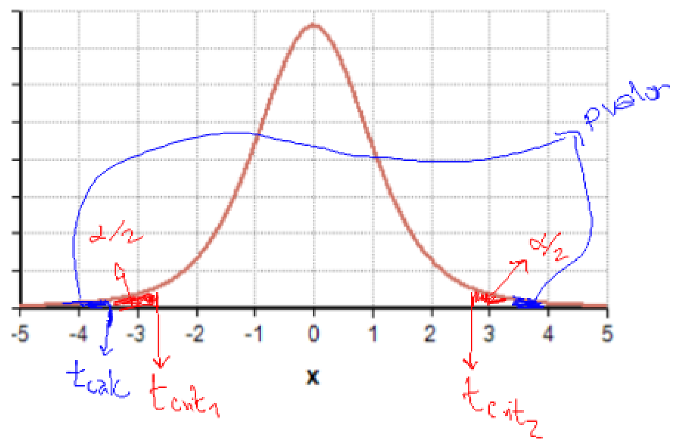
Prueba de hipótesis sobre correlación

$H_0: \rho = 0$

$H_1: \rho \neq 0$





t_{calc} cae en la zona de no rechazo de H_0

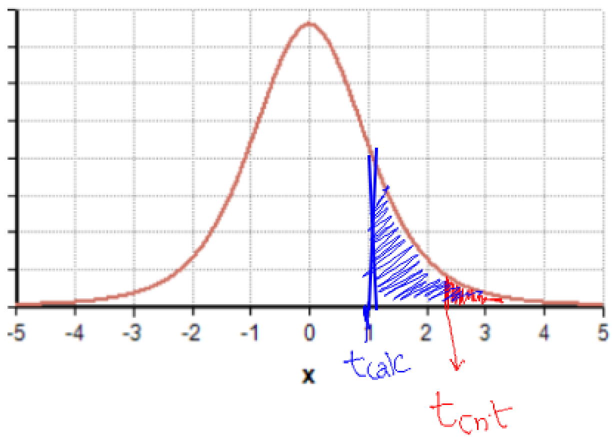
pvalor $> \alpha \Rightarrow$ No se rechaza H_0

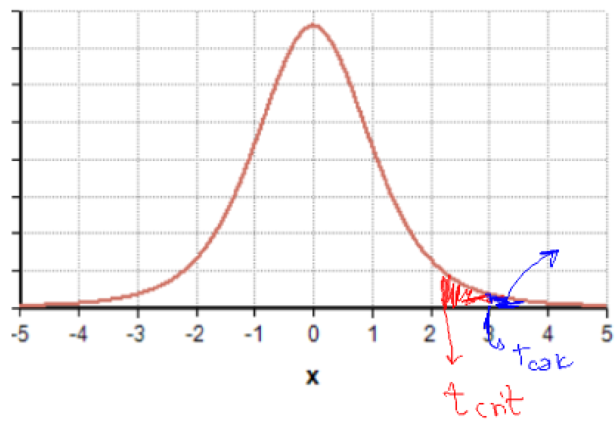
t_{calc} cae en la zona de rechazo de H_0

pvalor $< \alpha \Rightarrow$ se rechaza H_0

$H_0: \rho \leq 0$

$H_1: \rho > 0$





t_{calc} cae en la zona de No rechazo de H_0

$pV > \alpha \Rightarrow$ No se rechaza H_0

t_{calc} cae en la zona de rechazo de H_0

$pV < \alpha \Rightarrow$ Se rechaza H_0

$H_0 : \rho \leq 0.7$ ×

$H_1 : \rho > 0.7$ ✓

$\alpha = 0.05$

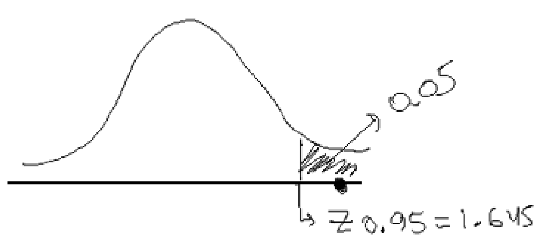
$$Z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$

sigue una distribución normal:

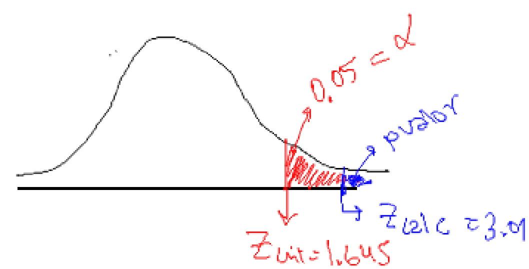
$$Z_r \sim N\left(\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n-3}\right)$$

En nuestro ejercicio : $r = 0.877$, $n = 40$

$$Z_{calc} = \frac{\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\sqrt{\frac{1}{n-3}}} = \frac{\frac{1}{2}\ln\left(\frac{1.877}{0.123}\right) - \frac{1}{2}\ln\left(\frac{1.7}{0.3}\right)}{\sqrt{\frac{1}{37}}} = 3.01$$



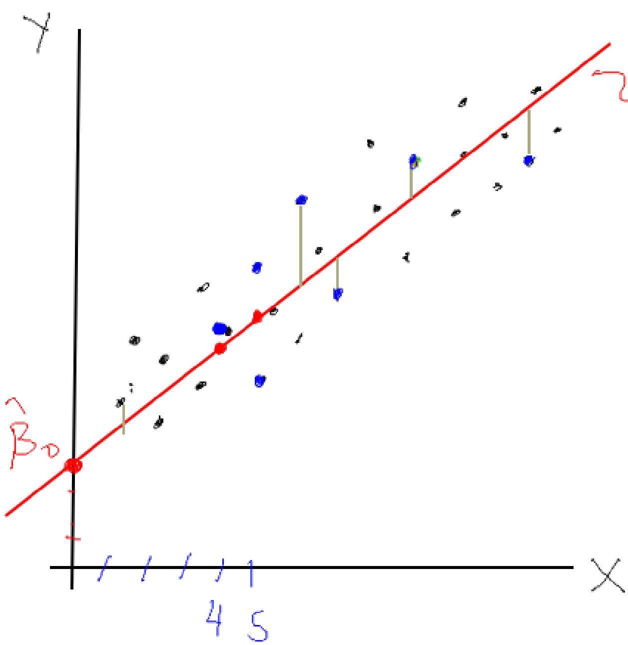
se rechaza H_0



REGRESIÓN

$Y \longleftarrow X$

endógena   exógena



$y = mx + b$

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = 3 + 4X_i$

$\varepsilon_i = Y_i - \hat{Y}_i$

$(Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \varepsilon_i)$

Y vs $\hat{Y}$

Y: Valor observado o real

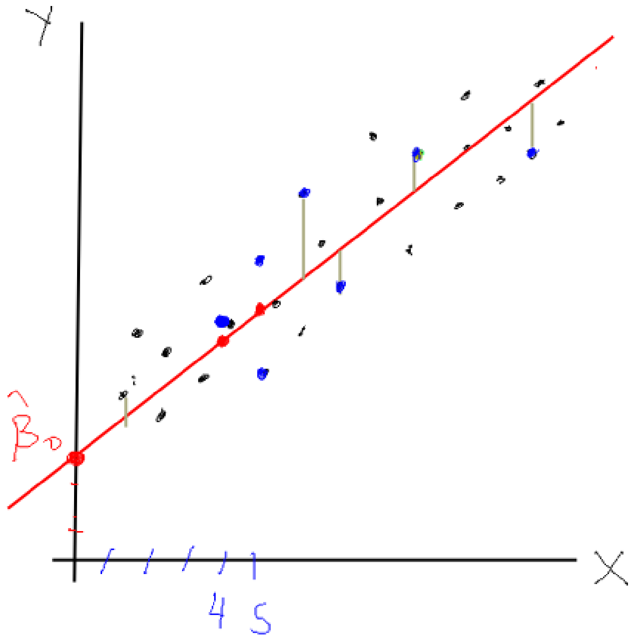
$\hat{Y}$: Valor estimado (MEDIA CONDICIONAL)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n$$

Mínimos cuadrados ordinarios

El método de mínimos cuadrados busca estimar los parámetros del modelo minimizando la suma de los cuadrados de los residuos.

errores



$$\varepsilon_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

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

$$\frac{\partial S}{\partial \hat{\beta}_0} = 0 \quad \Rightarrow \hat{\beta}_0 = \ldots$$

$$\frac{\partial S}{\partial \hat{\beta}_1} = 0 \quad \Rightarrow \hat{\beta}_1 = \ldots$$

Máxima verosimilitud

verosímil = creíble, compatible, aceptable, admisible, plausible

Asumiendo que los errores siguen una distribución normal,

$$\varepsilon_i \sim N(0, \sigma^2)$$

3, 3.6, 2.9, 4, 2.5

$N(\mu = 3, \sigma^2 = 1)$ $\quad$ $N(\mu = 3.8, \sigma^2 = 1)$

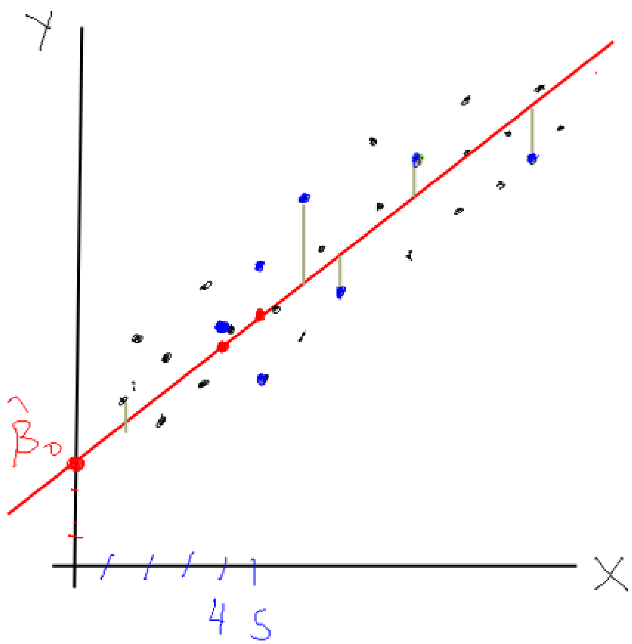
se tiene que:

$$Y_i | \beta_0, \beta_1, \sigma^2 \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$$

$$N(3 + 4X_i, 1)$$

$$N(3.5 + 4.2X_i, 1.2)$$

$$\vdots$$



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

$$X \longrightarrow X+1$$

$$\hat{Y}^* = \hat{\beta}_0 + \hat{\beta}_1 (X+1)$$

$$\hat{Y}^* - \hat{Y} = \hat{\beta}_1$$

cuando $X \uparrow 1$ unidad, $\hat{y} \uparrow \hat{\beta}_1$ unidades $(\hat{\beta}_1 > 0)$

X = Horas de estudio

Y = Puntaje en el examen

$\hat{\beta}_0 = 8.99$ es el puntaje promedio en el examen cuando el alumno no estudia (cero horas de estudio)

$\hat{\beta}_1 = 0.64$: Por cada hora adicional de estudio, el puntaje promedio obtenido en el examen se incrementa en 0.64 puntos.

X = Área de un inmueble comercial (m^2)

Y = Precio de alquiler mensual (\$)

$$\overset{\$}{\hat{Y}} = \overset{\$}{800} + \overset{\$/m^2}{10}\,\overset{m^2}{X}$$

$\hat{\beta}_1 = 10$ dólares$/m^2 \Rightarrow$ por cada metro cuadrado adicional, el precio promedio aumenta en 10 dólares

$\hat{\beta}_0 = 800$ dólares $\Rightarrow$ el inmueble no puede tener un área de 0 m^2 por lo tanto beta0 no tiene interpretación

Recordando

$$IC(\mu) = \hat{\mu} \pm \underbrace{t_{1-\frac{\alpha}{2},n-1} \times \frac{s}{\sqrt{n}}}_{\text{Margen de error}} \rightarrow \text{error estándar}$$

$$IC(\beta_j) = \hat{\beta}_j \pm \underbrace{t_{1-\frac{\alpha}{2},n-2} \times S_{\hat{\beta}_j}}_{\text{margen de error}} \rightarrow \text{error estándar}$$

resumen

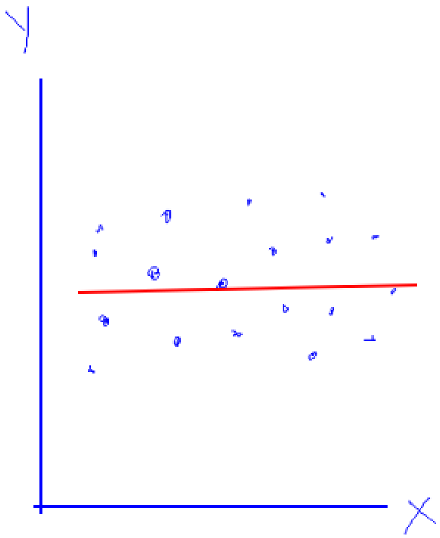
```r
modelo |> tidy()
```

$\hat{\beta}_0$ $S_{\hat{\beta}_0}$

```
# A tibble: 2 x 5
  term          estimate std.error statistic  p.value
  <chr>            <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)      8.99     0.588       15.3 8.13e-18
2 horas_estudio    0.641    0.0570      11.3 1.16e-13
```

$\hat{\beta}_1$ $S_{\hat{\beta}_1}$

```r
resumen = modelo |> tidy()
b0  = resumen$estimate[1]
b1  = resumen$estimate[2]
sb0 = resumen$std.error[1]
sb1 = resumen$std.error[2]
```


Y
X