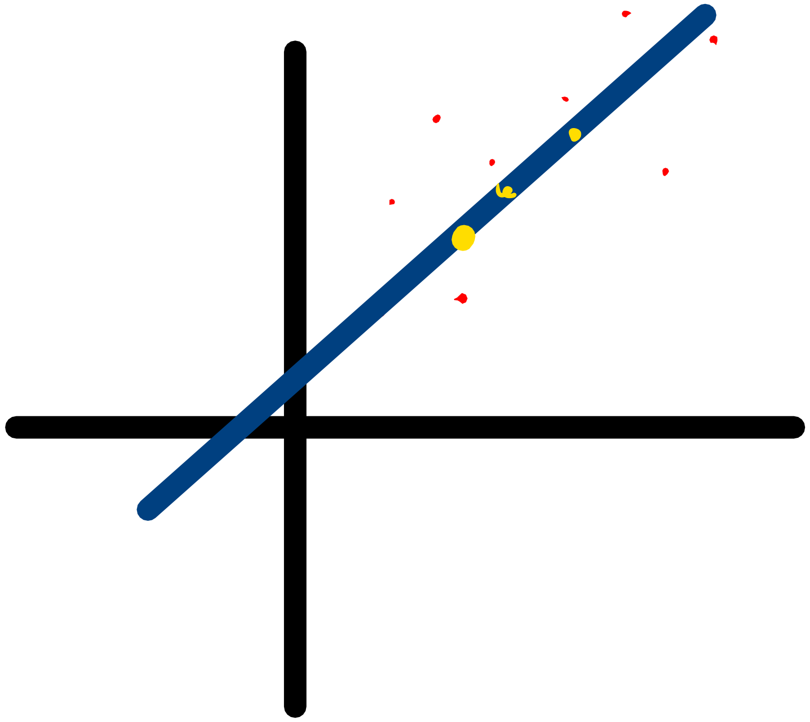


$$R^2 = 1 - \left( \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \right)$$

Proporción de variabilidad Debida a los Residuos Respecto de la variabilidad Total

$R^2 \approx 1$ cuando tengo un buen ajuste

$\approx$ chico

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

$$\hat{\beta}_1 | X = x \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

$\sigma^2 \rightarrow S_r^2$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}.$$

$$\hat{\beta}_0 | X = x \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)\right)$$

$$T = \frac{\hat{\beta}_0 - \beta_0}{S_r\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} = \frac{\hat{\beta}_0 - \beta_0}{se(\hat{\beta}_0)} \sim t_{n-2}$$

29.6376

$$S_r^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}, = \hat{\sigma}^2$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x}^2).$$

$$T = \frac{\hat{\beta}_1 - \beta_1}{\frac{S_r}{\sqrt{S_{xx}}}} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim t_{n-2}$$

0.7243

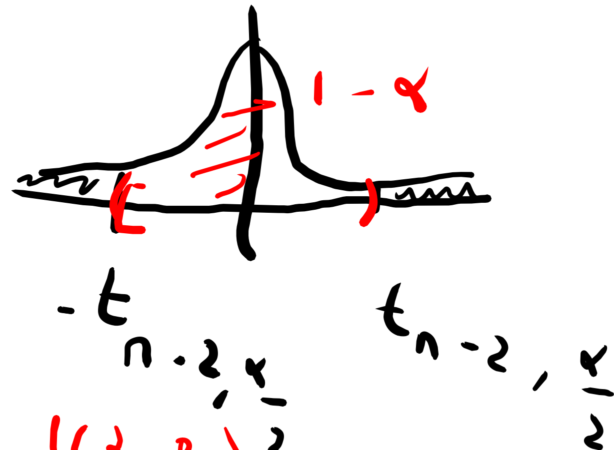
$$\left[\hat{\beta}_0 + t_{n-2,\alpha/2} S_r \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}; \hat{\beta}_0 - t_{n-2,\alpha/2} S_r \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}\right]$$

$$\left[\hat{\beta}_1 - t_{n-2,\alpha/2} \frac{S_r}{S_{xx}}; \hat{\beta}_1 + t_{n-2,\alpha/2} \frac{S_r}{S_{xx}}\right]$$

```r
confint(modelo)
```

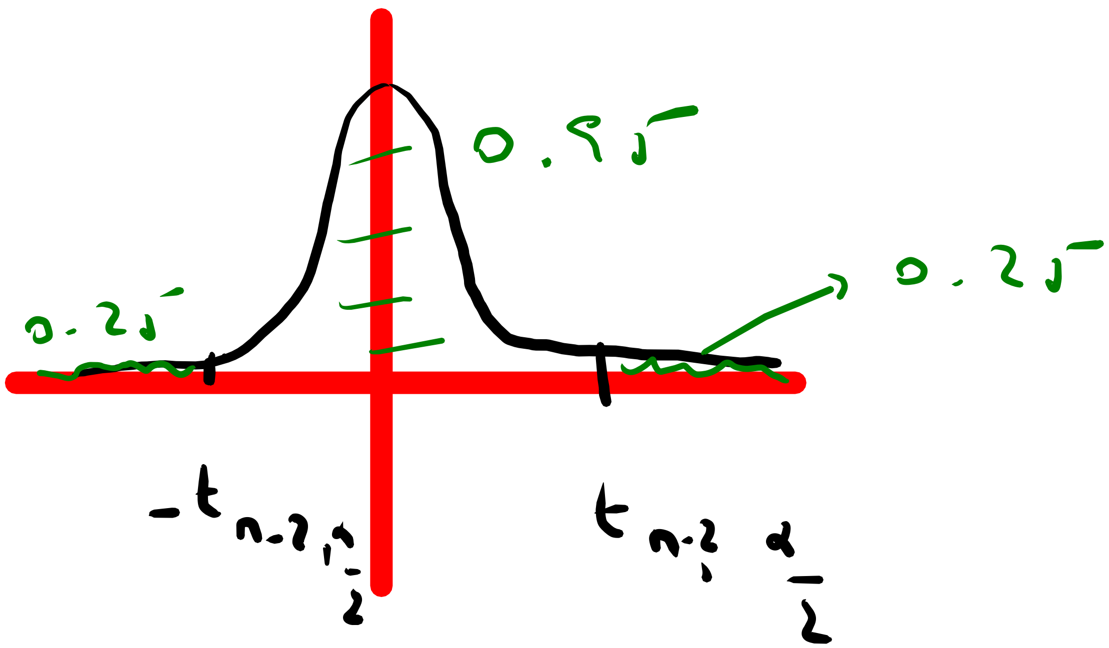
Nivel de confianza por default es 95%

```
##                  2.5 %     97.5 %
## (Intercept) 41.265155 163.885130
## edad         3.822367   6.818986
```




IC Para $\beta_0$ de nivel $1-\alpha$ es $[41.26 ; 163.88]$

$\beta_1$ $[3.82 ; 6.81]$


0.95
0.25
0.25
-t n-1, α/2
t n-1, α/2

Intervalo de confiança para $\sigma^2$

$\hat{\sigma}^2$

$$\left[\frac{(n-2)S_r^2}{\chi^2_{n-2,\alpha/2}}, \frac{(n-2)S_r^2}{\chi^2_{n-2,1-\alpha/2}}\right]$$

Test de hipótesis para $\beta_0$ y $\beta_1$

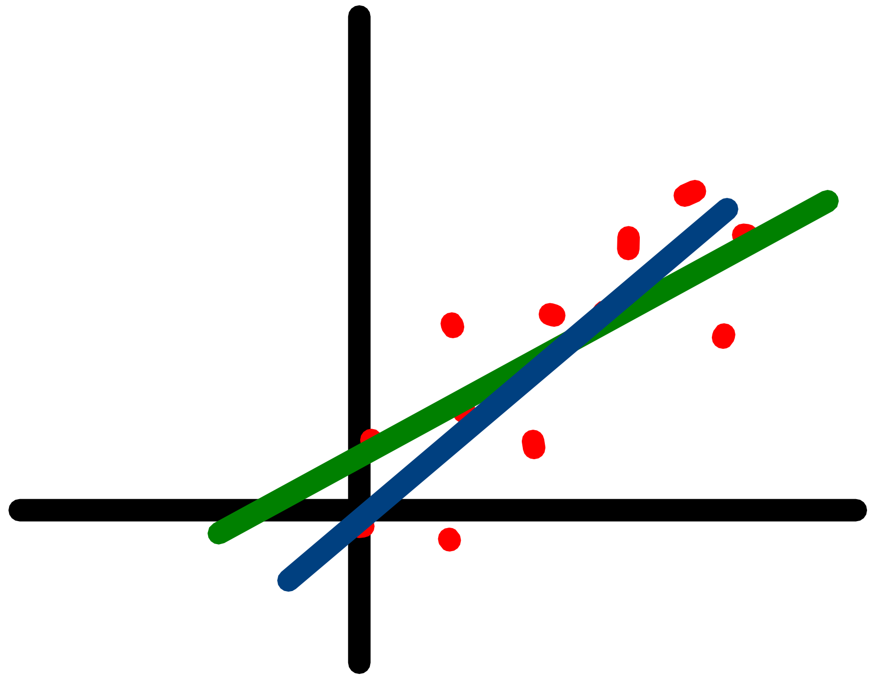
Recta de mejor ajuste

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

Test

$$H_0 \quad \beta_0 = 0 \quad \sim \quad H_1 \quad \beta_0 \neq 0$$

$$\beta_1 = 0 \quad \sim \quad \beta_1 \neq 0$$



$$\underbrace{Y = \beta_0 + \beta_1 x}$$

$$Y = \beta x$$

$\beta_0 = 0$ vs $\beta_0 \neq 0$

No Rechazar Ho

<u>Paso 2</u> estudio el modelo $Y = \beta x$

$$y = \beta_0 +$$

$\beta_1 = 0$ vs $\beta_1 \neq 0$

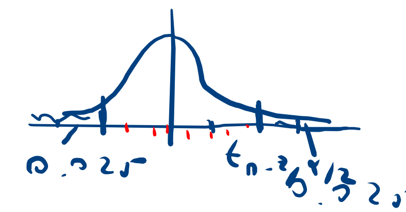
$$T = \frac{\hat{\beta}_1 - \beta_1}{\frac{S_r}{\sqrt{S_{xx}}}} = \frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \sim t_{n-2} \rightarrow \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} \sim t_{n-2} \text{ Bajo Ho}$$

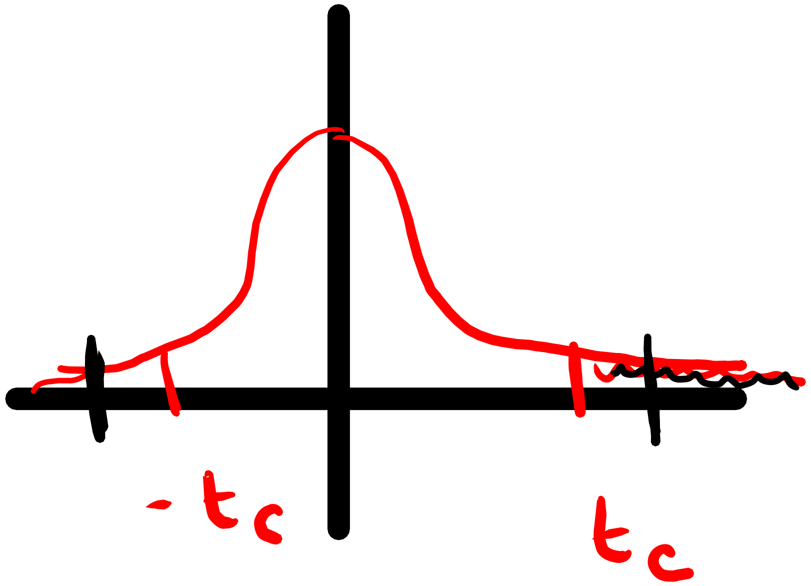
con los datos de las observaciones

$$\frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} \Rightarrow t_c = \frac{5.3207}{0.2243} = \boxed{17.3458}$$

$\alpha = 0.05$

$t_{n-2, \frac{\alpha}{2}}$

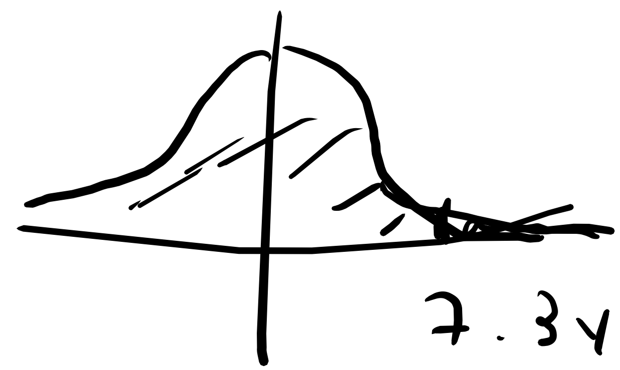





$$P(|T| > |t_c|) = p \text{ valor}$$

$$= \underbrace{P(T > 7.346)} + \underbrace{P(T < -7.346)}$$

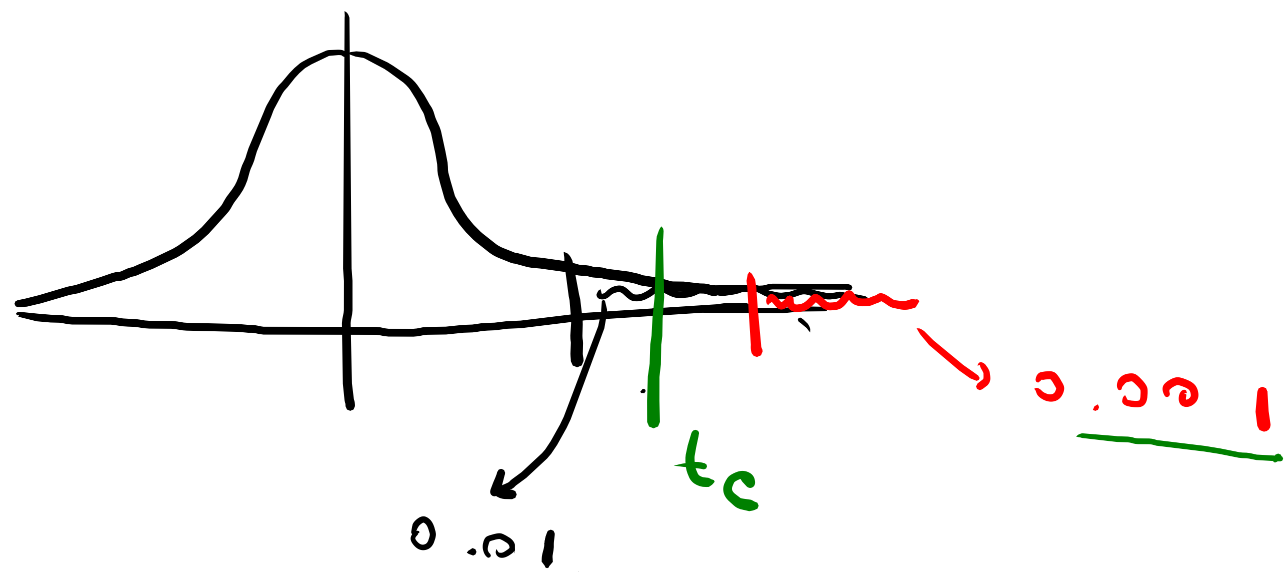
pt

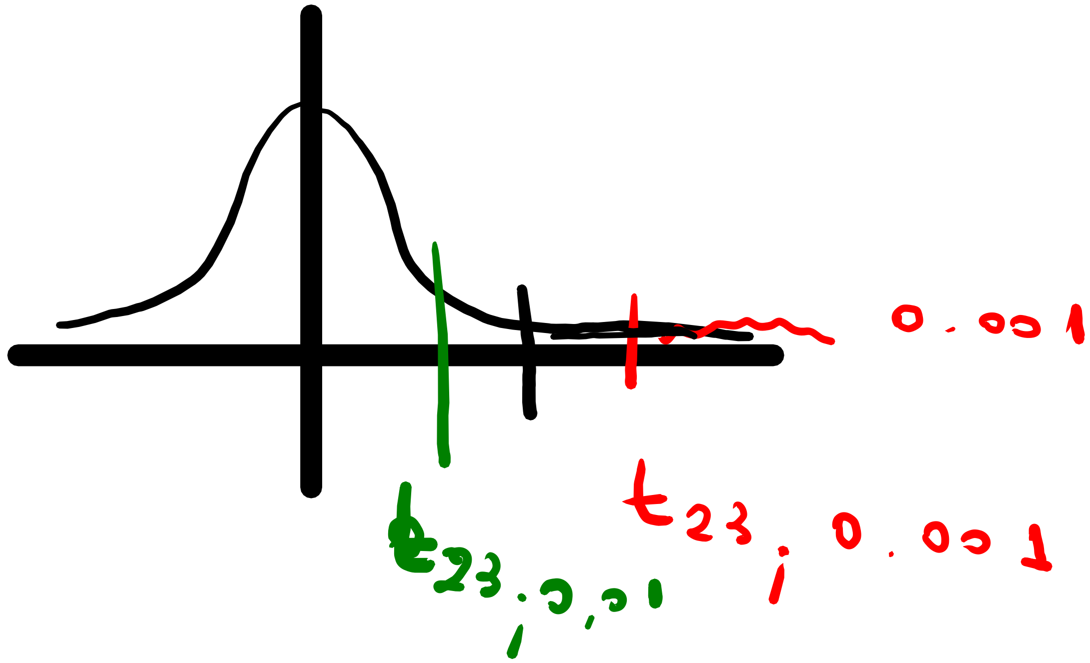


```
Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept) 102.5751    29.6376    3.461    0.00212 **
edad          5.3207     0.7243    7.346   1.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

T valor $\hat{\beta}_1$

$0.001 < 0.00212 < 0.01$






P valor $\simeq$ 0.002

Rechazo $H_0$ a un nivel del 1%.

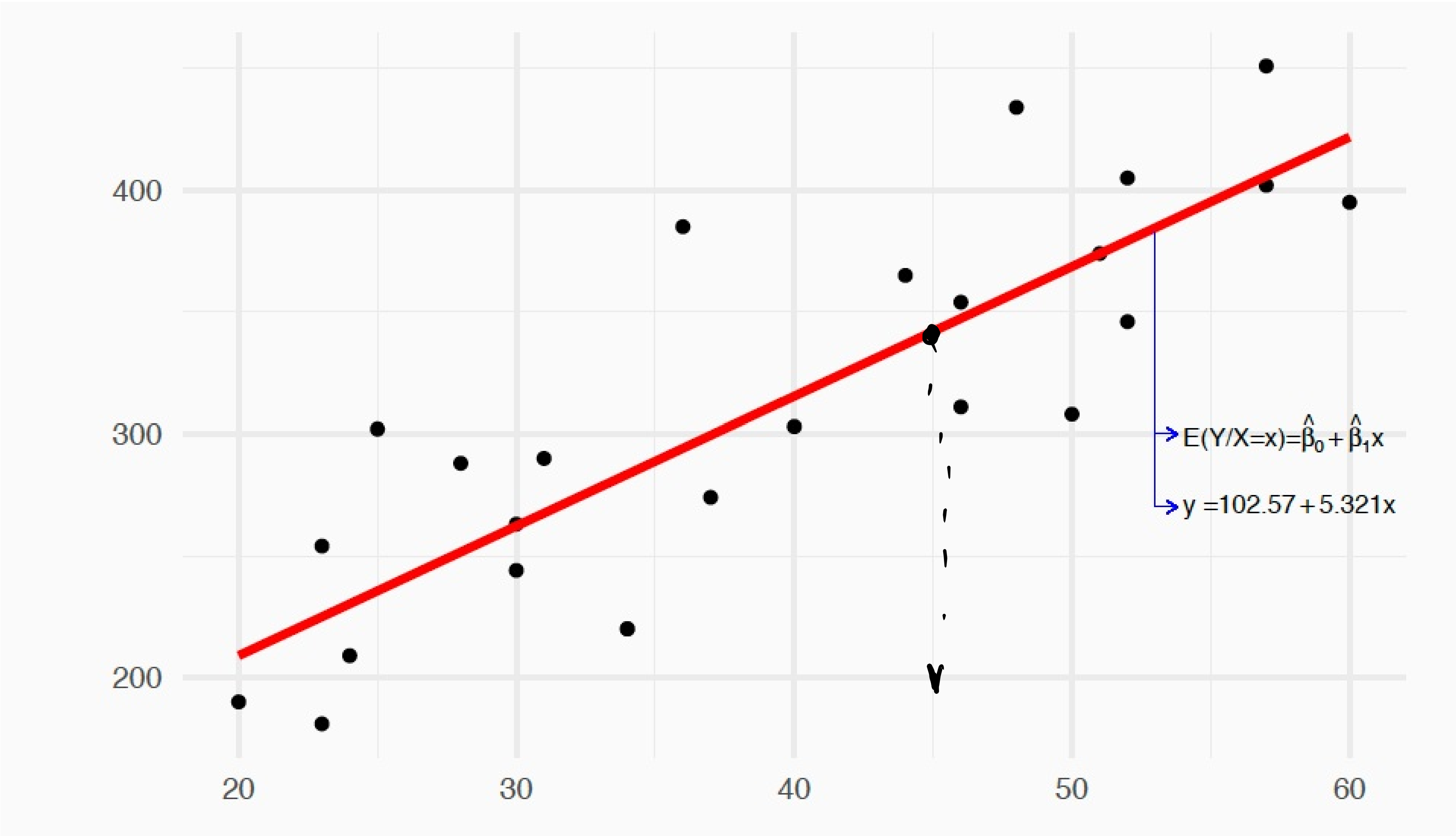
Pero no rechazo $H_0$ a un nivel del 1 por mil

El p valor correspondiente a $\beta_1$ es muy chico. Entonces rechazado $H_0$, para este test, a cualquier nivel

## Conclusión

Para este problema, ambos parámetros son significativamente $\neq 0$

Es decir, hay suficiente evidencia para afirmar que ambos parámetros son $\neq 0$

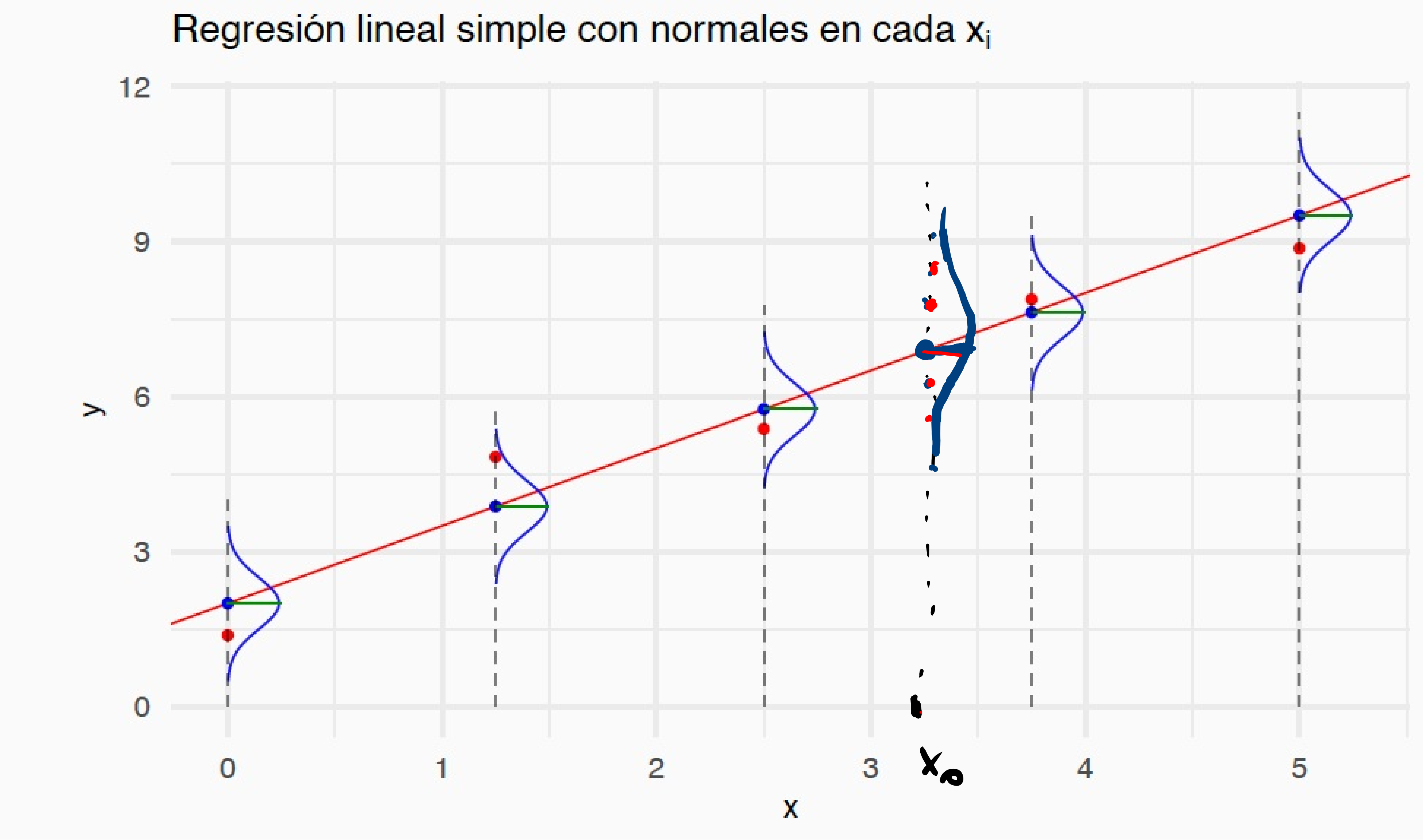


400
300
200
20
30
40
50
60
$E(Y/X=x)=\hat{\beta}_0+\hat{\beta}_1 x$
$y=102.57+5.321x$

```r
nuevas.edades <- data.frame(edad = c(30,45))
nuevas.edades
```

```
##   edad
## 1   30
## 2   45
```

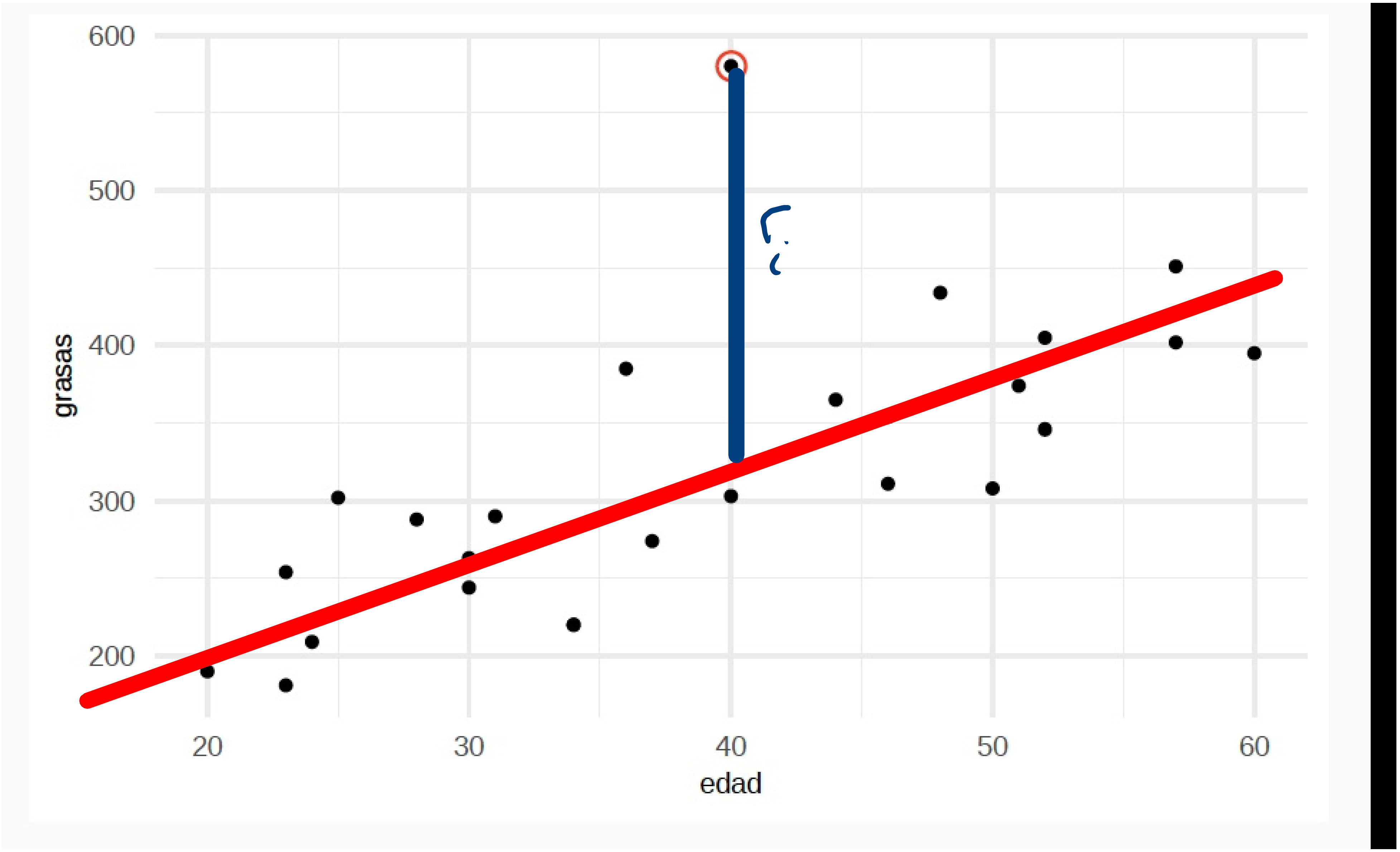
```r
predict(modelo, nuevas.edades)
```



Regresión lineal simple con normales en cada $x_i$
12
9
6
3
0
y
0
1
2
3
4
5
x
$x_0$

```r
predict(modelo, newdata=nuevas.edades, interval="confidence")
```

```r
predict(modelo, newdata=nuevas.edades, interval="prediction")
```


600
500
400
300
200
grasas
20
30
40
50
60
edad

```
> summary(modelo)

Call:
lm(formula = grasas ~ edad, data = grasas)

Residuals:
    Min      1Q  Median      3Q     Max
-63.478 -26.816  -3.854  28.315  90.881

Coefficients:
            Estimate  Std. Error  t value  Pr(>|t|)
(Intercept) 102.5751     29.6376    3.461   0.00212 **
edad          5.3207      0.7243    7.346  1.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.46 on 23 degrees of freedom
Multiple R-squared:  0.7012,    Adjusted R-squared:  0.6882
F-statistic: 53.96 on 1 and 23 DF,  p-value: 1.794e-07
```

```
> summary(regresion)

Call:
lm(formula = grasas ~ edad, data = datos)

Residuals:
   Min     1Q Median     3Q    Max
-73.34 -37.60 -13.00  19.36 254.59

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  110.324     46.304   2.383   0.0255 *
edad           5.383      1.133   4.753 7.78e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 67.97 on 24 degrees of freedom
Multiple R-squared:  0.4849,    Adjusted R-squared:  0.4634
F-statistic: 22.59 on 1 and 24 DF,  p-value: 7.784e-05
```

0.002 es chico

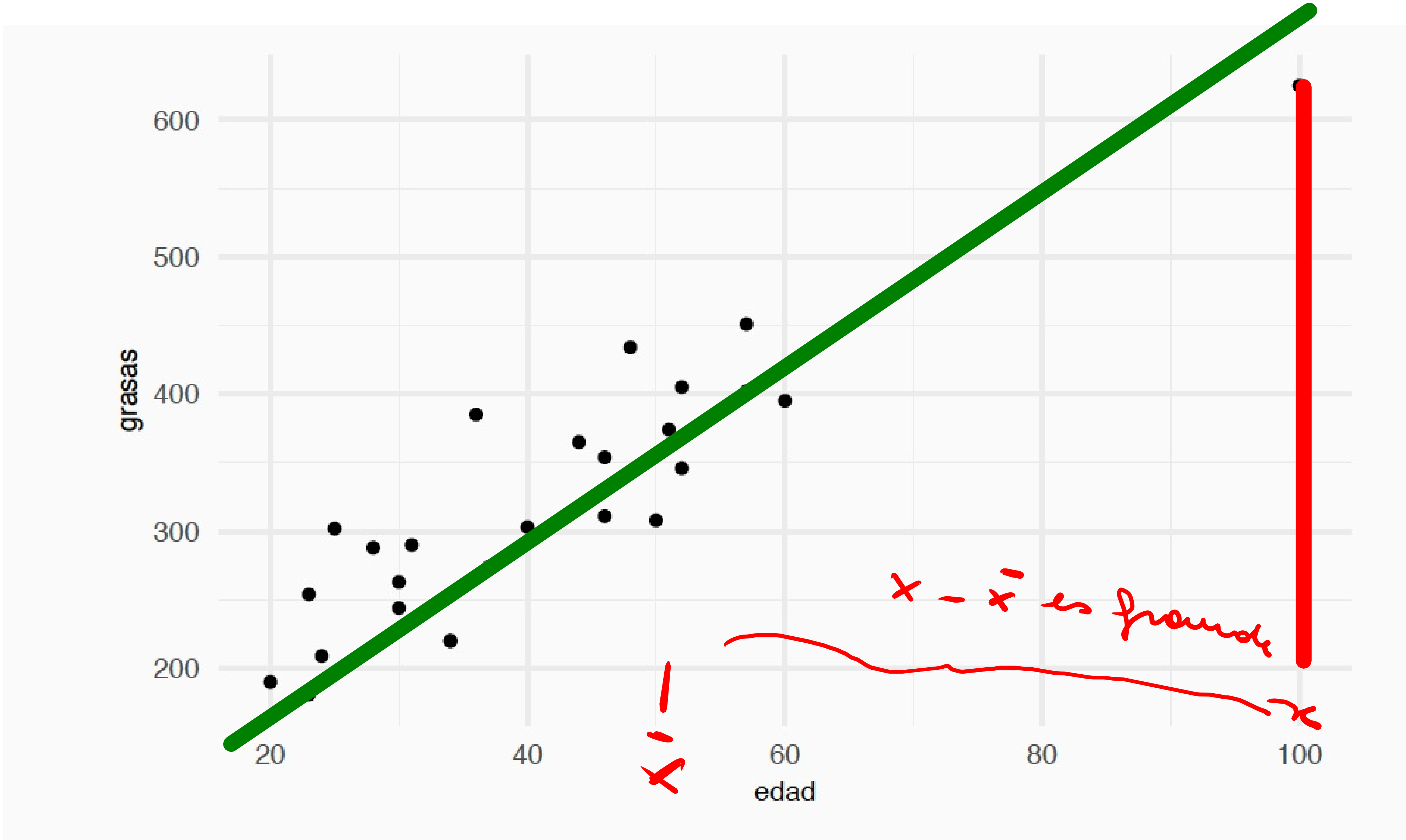
$0.01 < 0.02 < 0.05$

20 ; 25 ; 30 ; 31 ; 42 ; 50 ; 510

$$\frac{\sum x_i}{n}$$

$$\min_{\beta_0, \beta_1} \sum r_i^2$$

nombre [ nombre > ]


600
500
400
300
200
grasas
20
40
60
80
100
edad
x - x̄ es grande
x̄
x

$$\mathrm{Var}(r_i) = \sigma^2 (1 - h_{ii})$$

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

```
> summary(regresion.original) #Sin punto de alto leverage

Call:
lm(formula = grasas ~ edad, data = datos.originales)

Residuals:
    Min      1Q  Median      3Q     Max 
-63.478 -26.816  -3.854  28.315  90.881 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 102.5751    29.6376   3.461  0.00212 ** 
edad          5.3207     0.7243   7.346 1.79e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 43.46 on 23 degrees of freedom
Multiple R-squared:  0.7012,	Adjusted R-squared:  0.6882 
F-statistic: 53.96 on 1 and 23 DF,  p-value: 1.794e-07
```

```
> summary(regresion.ConLevBueno) #Con punto de alto leverage

Call:
lm(formula = grasas ~ edad, data = datos.ConLevBueno)

Residuals:
    Min      1Q  Median      3Q     Max 
-63.695 -25.312  -3.459  27.712  90.821 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 105.4709    22.4607   4.696 9.00e-05 ***
edad          5.2419     0.5029  10.423 2.18e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 42.57 on 24 degrees of freedom
Multiple R-squared:  0.8191,	Adjusted R-squared:  0.8115 
F-statistic: 108.6 on 1 and 24 DF,  p-value: 2.175e-10
```

$$\bar{h} = \frac{\sum_{i=1}^{n} h_{ii}}{n} = \frac{2}{n}$$

criterio que vamos a considerar es

$$h_{ii} > 2\bar{h} = 2 * \frac{2}{n} = \frac{4}{n}$$

en R → hatvalues

```
> summary(regresion.original) #Sin punto de alto leverage

Call:
lm(formula = grasas ~ edad, data = datos.originales)

Residuals:
    Min      1Q  Median      3Q     Max 
-63.478 -26.816  -3.854  28.315  90.881 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 102.5751    29.6376   3.461  0.00212 ** 
edad          5.3207     0.7243   7.346 1.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.46 on 23 degrees of freedom
Multiple R-squared:  0.7012,	Adjusted R-squared:  0.6882 
F-statistic: 53.96 on 1 and 23 DF,  p-value: 1.794e-07
```

```
> summary(regresion) # Con punto de alto leverage

Call:
lm(formula = grasas ~ edad, data = datos)

Residuals:
    Min      1Q  Median      3Q     Max 
-87.163 -40.453  -4.734  29.165 118.566 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  28.8926    29.0073   0.996    0.329    
edad          7.3254     0.6495  11.279 4.46e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 54.98 on 24 degrees of freedom
Multiple R-squared:  0.8413,	Adjusted R-squared:  0.8347 
F-statistic: 127.2 on 1 and 24 DF,  p-value: 4.461e-11
```

Distancia de Cook

$$D_i = \frac{r_{si}^2}{2} \frac{h_{ii}}{1 - h_{ii}}$$

```
cooks.o <- cooks.distance(modelo.o)
p <- length(coef(modelo.o))
n<-dim(grasas.o)[1]
cooks_umbral.o <- qf(0.5,p,n-p)
```

gL

F distribución de Fisher $F_{p;\ n-p}$

$F_{2;\ n-2}$