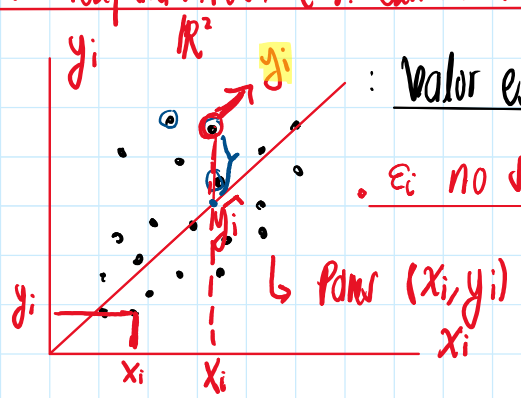
Primer taller

martes, 2 de septiembre de 2025 7:56 p. m.

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

(1). Independientes; (2). Idénticamente distribuidos; (3). Media cero; (4). Varianza constante σ^2



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

- ε_i no son observables: son estimables: residual $\boxed{e_i = y_i - \hat{y}_i}$

(1). y_i son las observaciones crudas (puntos)
(2). $\hat{y}_i$ son las observaciones ajustadas sobre la línea recta.

- Minimizar los residuales: Minimizando la varianza.

$\boxed{\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}}$: Determinístico: $\boxed{\hat{\beta}_0, \hat{\beta}_1 : \text{MLE, OLS}}$

MLE ↓ $\boxed{\text{Método estadístico}}$; OLS → $\boxed{\text{Método matemático}}$

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

$\boxed{\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sum_{i=1}^{n} X_i^2 \sigma^2}{n S_{xx}}\right)}$; $\boxed{n = \#\text{ datos a considerar}}$

$\boxed{\hat{\beta}_1 \sim N(\beta_1, \sigma^2 / S_{xx})}$

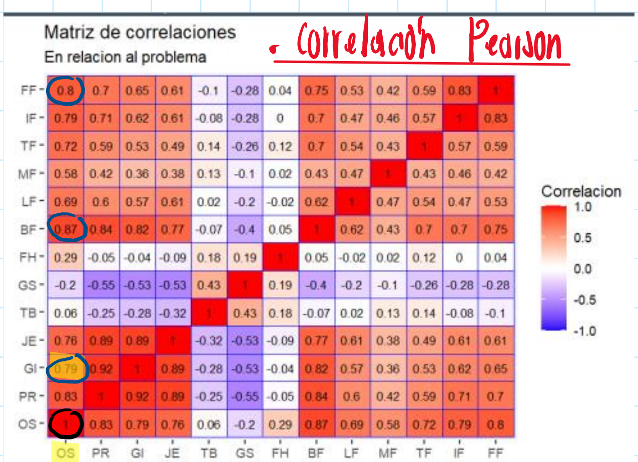
- $\boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}}$

$\bar{Y}$: Promedio sobre la respuesta. → Promedio sobre las observaciones (Variable Ind). : $\bar{X}$: Promedio sobre X. (covariables, variable independiente).

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) y_i}{\sum_{i=1}^{n} (X_i - \bar{X})} = \frac{S_{xy}}{S_{xx}}}$$

1. Realice un **breve análisis descriptivo**. ¿Un modelo de regresión lineal simple podría ser adecuado en este caso? ¿Por qué?

(1). Variables independientes altamente correlacionadas con la variable independiente podría ser útil.



	OS	PR	GI	JE	TB	GS	FH	BF	LF	MF	TF	IF	FF
FF	0.8	0.7	0.65	0.61	-0.1	-0.28	0.04	0.75	0.53	0.42	0.59	0.83	1
IF	0.79	0.71	0.62	0.61	-0.08	-0.28	0	0.7	0.47	0.46	0.57	1	0.83
TF	0.72	0.59	0.53	0.49	0.14	-0.26	0.12	0.7	0.54	0.43	1	0.57	0.59
MF	0.58	0.42	0.36	0.38	0.13	-0.1	0.02	0.43	0.47	1	0.43	0.46	0.42
LF	0.69	0.6	0.57	0.61	0.02	-0.2	-0.02	0.62	1	0.47	0.54	0.47	0.53
BF	0.87	0.84	0.82	0.77	-0.07	-0.4	0.05	1	0.62	0.43	0.7	0.7	0.75
FH	0.29	-0.05	-0.04	-0.09	0.18	0.19	1	0.05	-0.02	0.02	0.12	0	0.04
GS	-0.2	-0.55	-0.53	-0.53	0.43	1	0.19	-0.4	-0.2	-0.1	-0.26	-0.28	-0.28
TB	0.06	-0.25	-0.28	-0.32	1	0.43	0.18	-0.07	0.02	0.13	0.14	-0.08	-0.1
JE	0.76	0.89	0.89	1	-0.32	-0.53	-0.09	0.77	0.61	0.38	0.49	0.61	0.61
GI	0.79	0.92	1	0.89	-0.28	-0.53	-0.04	0.82	0.57	0.36	0.53	0.62	0.65
PR	0.83	1	0.92	0.89	-0.25	-0.55	-0.05	0.84	0.6	0.42	0.59	0.71	0.7
OS	1	0.83	0.79	0.76	0.06	-0.2	0.29	0.87	0.69	0.58	0.72	0.79	0.8

- Correlación Pearson

- $\boxed{\text{La variable independiente seleccionada va a ser GI.}}$

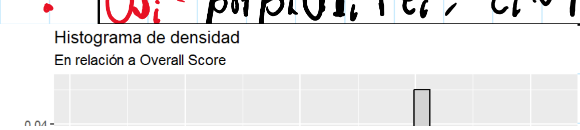
- Se va ajustar con OS como respuesta

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

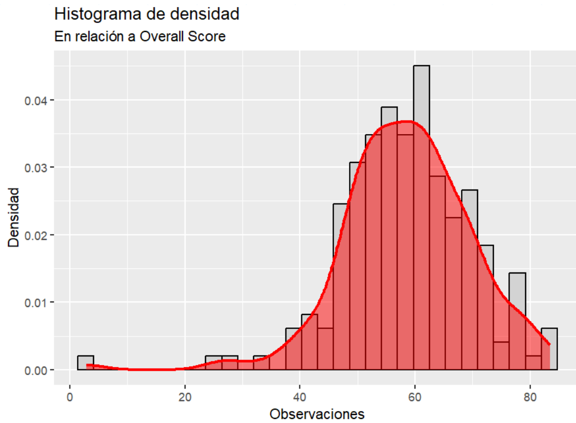
$\boxed{\bullet \text{ La correlación } -1 \leq \rho_{xy} \leq 1}$

(1). Relación lineal perfecta

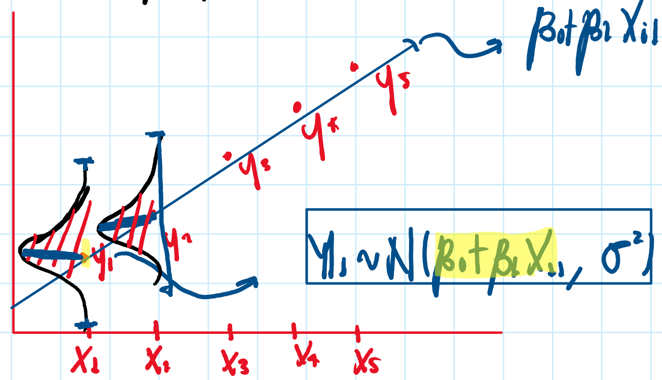
- $\boxed{OS_i = \beta_0 + \beta_1 GI_i + \varepsilon_i ;\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)}$: $\boxed{\hat{OS}_i = \hat{\beta}_0 + \hat{\beta}_1 GI_i}$: $\boxed{\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}}$



- $\boxed{Y_i \sim N(\beta_0 + \beta_1 X_{i1}, \sigma^2)}$; $i = 1, \dots, 5$



$\boxed{Y_i \sim N(\beta_0 + \beta_1 X_{i1}, \sigma^2)}$; $i = 1, \ldots, 5$



$\boxed{Y_1 \sim N(\beta_0 + \beta_1 X_{11}, \sigma^2)}$

2. Escriba la ecuación del modelo de regresión lineal, considerando los supuestos asociados. Obtenga los valores calculados de los parámetros, $\overline{y}, \hat{y}, \hat{\varepsilon}_i$ y analícelos.

$\boxed{OS_i = \beta_0 + \beta_1 GI_i + \varepsilon_i ;\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)}$: $\boxed{\hat{OS}_i = \hat{\beta_0} + \hat{\beta_1}\, GI_i}$: $\boxed{\hat{Y}_i = \hat{\beta_0} + \hat{\beta_1} X_{i1}}$

$\boxed{lm:\ Y \sim X}$: $\boxed{e_i = Y_i - \hat{Y}_i}$

$\boxed{\text{. Diferencia entre las observaciones puntuales y la línea ajustada}}$

```
Call:
lm(formula = OS ~ GI, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-38.987  -3.688   0.955   4.466  12.634

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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.812 on 174 degrees of freedom
Multiple R-squared:  0.6291,    Adjusted R-squared:  0.6269
F-statistic: 295.1 on 1 and 174 DF,  p-value: < 2.2e-16
```

$\hat{\beta_0}$ $\hat{\beta_1}$

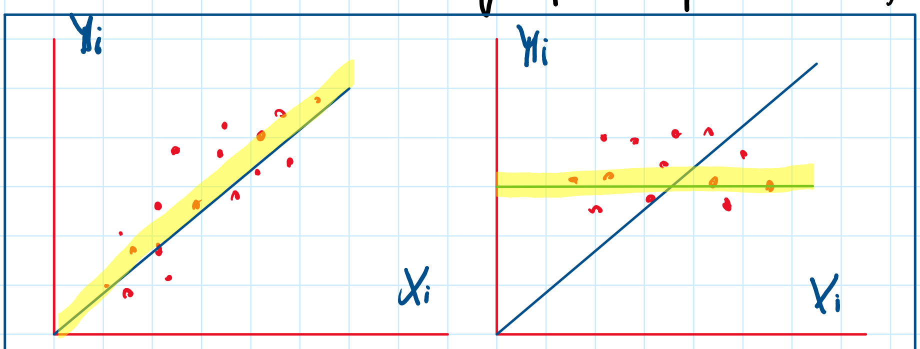
$\boxed{\alpha = 0{,}05}$

. Recta de regresión ajustada: $\boxed{\hat{Y}_i = 40{,}409 + 0{,}41041\, X_{i1}}$

3. Determine si los parámetros del modelo β_0, β_1 son significativos, considerando $\alpha = 0.05$. Realice una interpretación en relación al problema. **¿Estos parámetros tienen sentido?**

. Problema significancia: (1). Principio de parsimonia:

. Visión econometrista: lo mejor posible pero de la forma más sencilla posible.



$\boxed{\text{. Revisar si el efecto de una variable es significativo sobre el modelo}}$

. Repaso de EI:

(1). Valor P: $\boxed{P(|t_{n-2}| > |T_0|)}$ En el caso de $\hat{\beta_0}$: $\boxed{T_0 : \dfrac{\hat{\beta_0} - 0}{\sqrt{Var(\hat{\beta})}} \sim t_{n-2}}$

. Prueba de hipótesis

$\boxed{\begin{cases} H_0: \beta_0 = 0 \\ H_1: \beta_0 \neq 0 \end{cases}}$: Hipótesis de doble cola

. Se rechaza la hipótesis nula a un nivel de significancia α ($\alpha = 0{,}05$) del 5%, cuando el

. Se rechaza la hipótesis nula a un nivel de significancia α ($\alpha = 0{,}05$) del 5%, cuando el $Vp < \alpha$

- Como se rechaza la hipótesis nula: Los parámetros son significativos.
- El intercepto es significativo y la covariable GI es significativa

$$\begin{cases} H_0: \beta_j = 0 & j = 0, \dots, k : \; j = 0, 1 \\ H_1: \beta_j \neq 0 \end{cases}$$

j: Notación utilizada para los parámetros

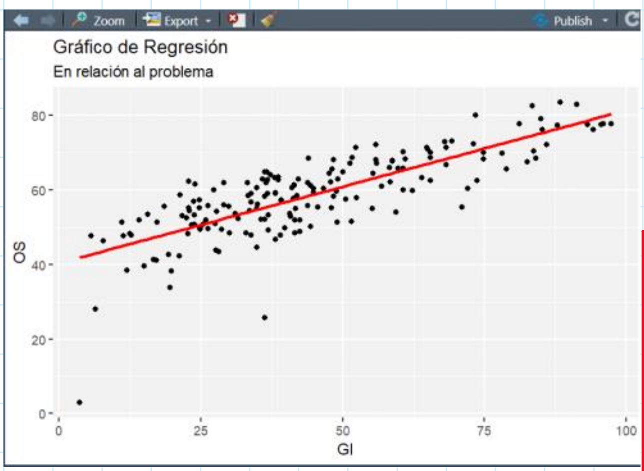
- (2). Región de rechazo: $R_c = \{ |T_0| > t_{\alpha/2, n-2} \}$ $\alpha = 0{,}1$ $\alpha = 0{,}01$

- Una variable solo puede ser interpretable si es significativa: Por aumento unitario en la covariable GI, el cambio promedio en la respuesta OS es de 0,41041 unidades

4. Calcule un intervalo de confianza -considerando $\alpha = 0.05$- para ambos parámetros. ¿Puede concluir a partir de este resultado si los parámetros son significativos?

Intervalo de confianza:

$$\hat{\beta}_j \pm t_{\alpha/2, n-2} \cdot \sqrt{\widehat{var}(\hat{\beta}_j)}$$




```
> # Confint (para intervalos de confianza)
> confint(model)
                 2.5 %     97.5 %
(Intercept) 38.0817785 42.7363582
GI           0.3632559  0.4575649
>
```

- Con una confianza del 95% se puede determinar que, por un aumento unitario en GI, el cambio promedio en OS, está comprendido entre 0,363 y 0,457 unidades

- Si se cambia la confianza, cambia el intervalo (amplitud)

. $\alpha = 0{,}1$: confianza 90%

LI LS

95% LI LS