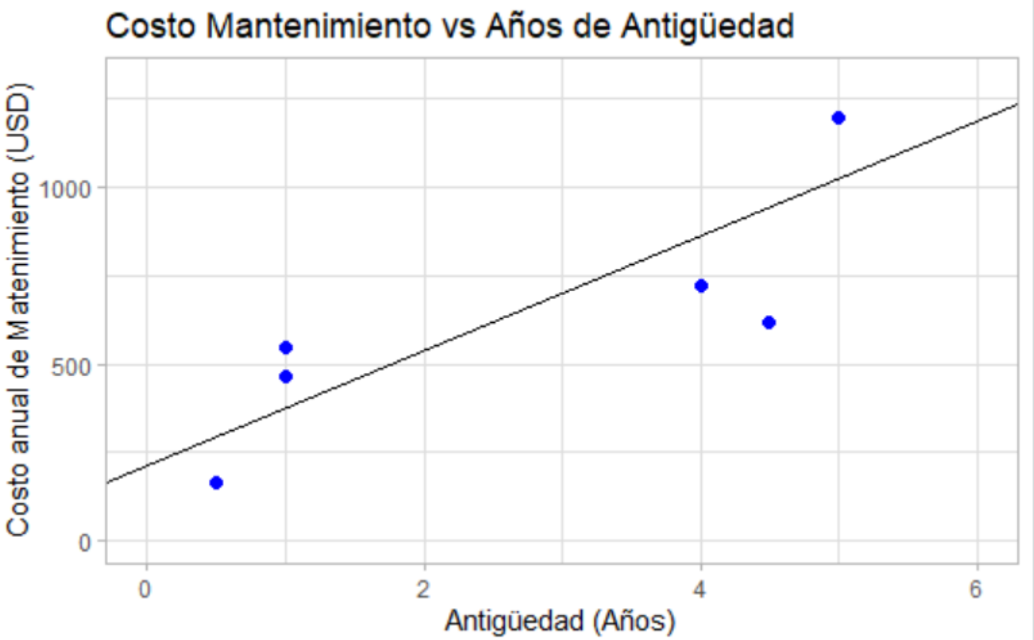
Regresión Lineal Simple

Mariano Bonoli Escobar
Javier Gil

CASO DE DISCUSION I

El Gerente de Logística desea construir un modelo para presupuestar el costo anual de mantenimiento de los auto-elevadores. En base a su experiencia cree que la variable de mayor relevancia es la antigüedad del equipo. Un analista de la gerencia recopiló la siguiente información para 7 equipos:

Antigüedad [años]	4.5	1	1	5	0.5	4	6
Costo anual U$S	619	549	466	1194	163	723	1345



Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Objetivo

Objetivo

Encontrar un modelo que se ajuste a los datos, estimar sus parámetros y poder efectuar predicciones de la variable explicada $\tilde{y}$ en función de los valores de la variable explicativa x.

$$\tilde{y} = f\left(x \middle| \theta_1, \ldots, \theta_p\right) + \tilde{\varepsilon}$$

$\tilde{y}$: Variable explicada o de respuesta. Se considera aleatoria.

x: Variable explicativa (o independiente). Se considera como NO aleatoria.

$\theta_1, \ldots, \theta_p$: Parámetros del modelo (p parámetros)

$\tilde{\varepsilon}$: Perturbación o error. Es una variable aleatoria

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Transformaciones

$$\tilde{y} = f\left(x \middle| \theta_1, \ldots, \theta_p\right) + \tilde{\varepsilon}$$

$$f\left(x \middle| \beta_0 ; \beta_1\right) = \beta_0 + \beta_1 \; x$$

$$f\left(x \middle| \beta_0 ; \beta_1\right) = \beta_0 + \beta_1 \; x^2$$

$$f\left(x \middle| \beta_0 ; \beta_1\right) = \beta_0 + \beta_1 \operatorname{sen}\left(x\right)$$

$$f\left(x \middle| \beta\right) = \beta \sqrt[3]{x}$$

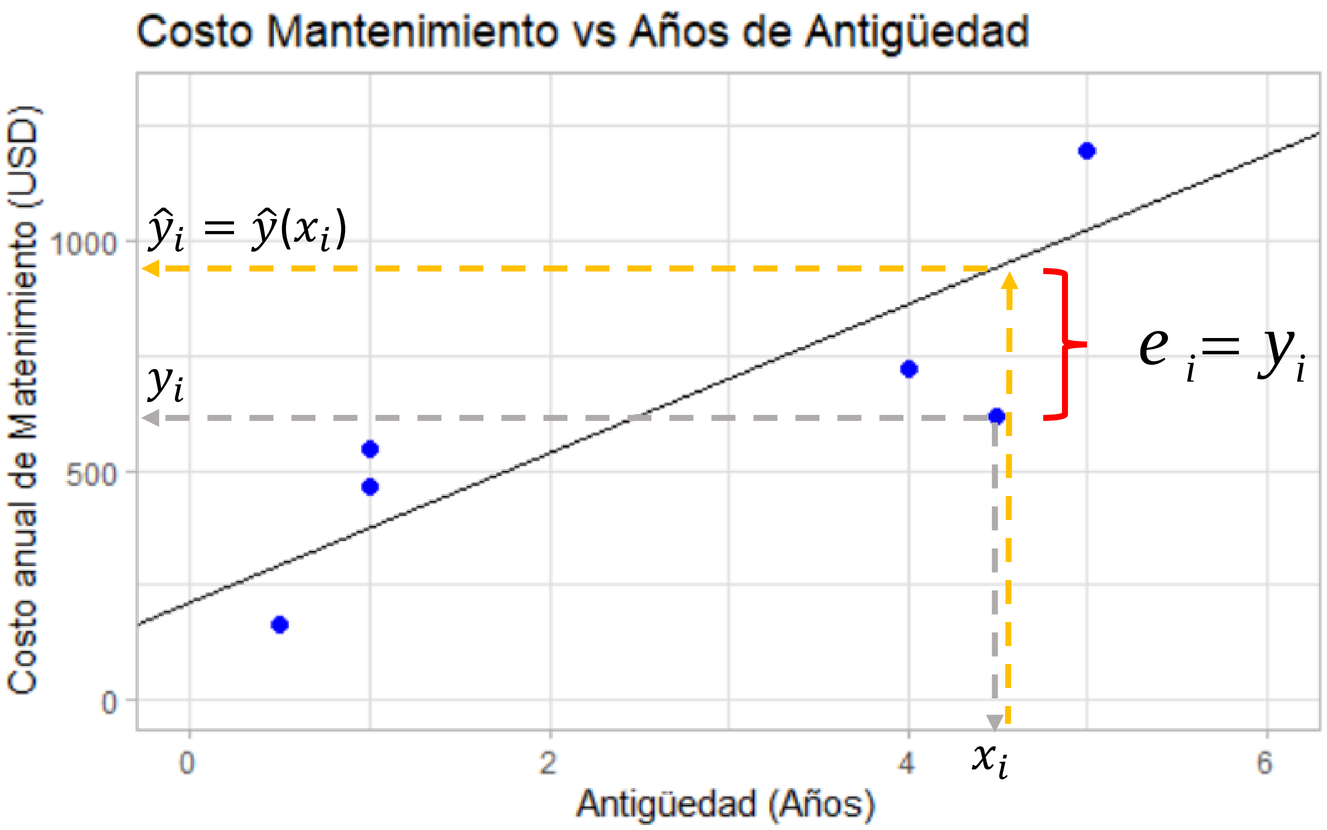
$$y = \alpha \, x^{\beta} \; \varepsilon$$

$$\underbrace{\ln y}_{y'} = \ln \alpha \; + \beta \; \underbrace{\ln x}_{x'} + \ln \varepsilon$$

La variable x puede estar bajo formas no lineales

Regresión Lineal Simple

Definiciones



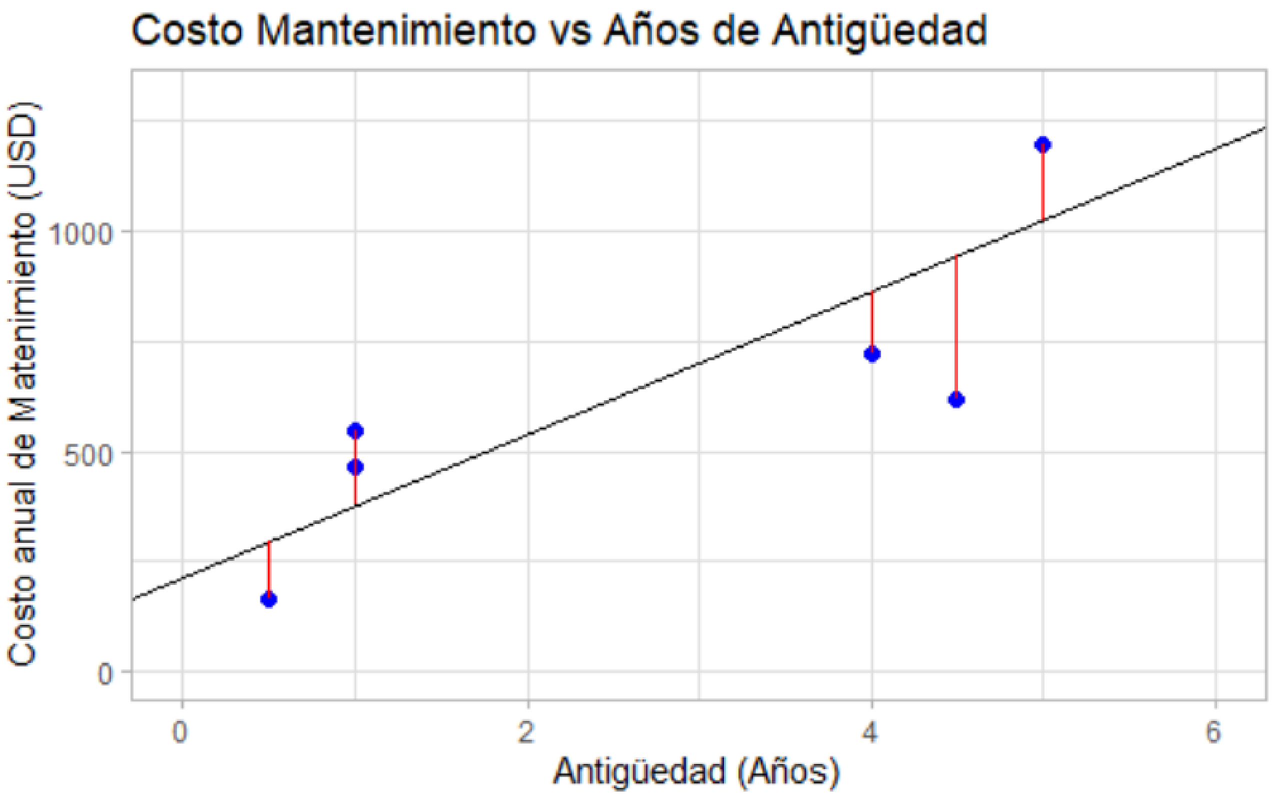
$$\hat{y} = b_o + b_1 x$$

$$e_i = y_i - \hat{y}_i = y_i - (b_o + b_1 x_i)$$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Método de Mínimos Cuadrados



Criterio de Gauss:

$$Q = \sum e_i^2 \quad \rightarrow \quad \min$$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Método de Mínimos Cuadrados

Criterio de Gauss:

$$Q = \sum e_i^2 \quad \rightarrow \quad \min$$

$$Q = \sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} \left(y_i - b_0 - b_1\, x_i\right)^2$$

$$\left.\begin{aligned} \frac{\partial Q}{\partial b_o} &= 0 \\ \frac{\partial Q}{\partial b_1} &= 0 \end{aligned}\right\} \Rightarrow \begin{aligned} \sum_{i=1}^{n} y_i &= n\, b_o + b_1 \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i &= b_o \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 \end{aligned} \Rightarrow \begin{aligned} b_1 &= \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)\left(x_i - \bar{x}_i\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2} = \frac{S_{xy}}{S_{xx}} \\ b_0 &= \bar{y} - b_1 \bar{x} \end{aligned}$$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

- Modelo

$$\hat{y}_i = b_0 + b_1 x_i$$

- Estimación

$$b_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2} = \frac{SC_{xy}}{SC_{xx}}$$

$$b_0 = \overline{y} - b_1 \overline{x}$$

$$SC_{xx} = \sum (x_i - \overline{x})^2 = \sum x_i^2 - n\overline{x}^2$$

$$SC_{yy} = \sum (y_i - \overline{y})^2 = \sum y_i^2 - n\overline{y}^2$$

$$SC_{xy} = \sum (x_i - \overline{x})(y_i - \overline{y}) = \sum x_i y_i - n\overline{xy}$$

Mariano Bonoli Escobar
Javier Gil

CASO DE DISCUSION I

El Gerente de Logística desea construir un modelo para presupuestar el costo anual de mantenimiento de los auto-elevadores. En base a su experiencia cree que la variable de mayor relevancia es la antigüedad del equipo. Un analista de la gerencia recopiló la siguiente información para 7 equipos:

Antigüedad [años]	4.5	1	1	5	0.5	4	6
Costo anual U$S	619	549	466	1194	163	723	1345

$$S_{xy} = \sum_{i=1}^{n} x_i \, y_i - n\,\bar{x}\,\bar{y} = 4914{,}2857$$

$$S_{xx} = \sum_{i=1}^{n} x_i^{\,2} - n\,\bar{x}^2 = 30{,}3571$$

$$S_{yy} = \sum_{i=1}^{n} y_i^{\,2} - n\,\bar{y}^2 = 1029465{,}4286$$

$$\hat{y} = 213{,}9 + 161{,}9x$$

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,87907124
Coeficiente de determinación R^2	0,77276625
R^2 ajustado	0,7273195
Error típico	216,30039
Observaciones	7

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	1	795536,1345	795536,134	17,0037732	0,00914211
Residuos	5	233929,2941	46785,8588		
Total	6	1029465,429			

	Coeficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%
Intercepción	213,941176	148,0094266	1,44545642	0,20795505	-166,529167	594,41152
Antigüedad	161,882353	39,25788114	4,12356317	0,00914211	60,9667568	262,797949

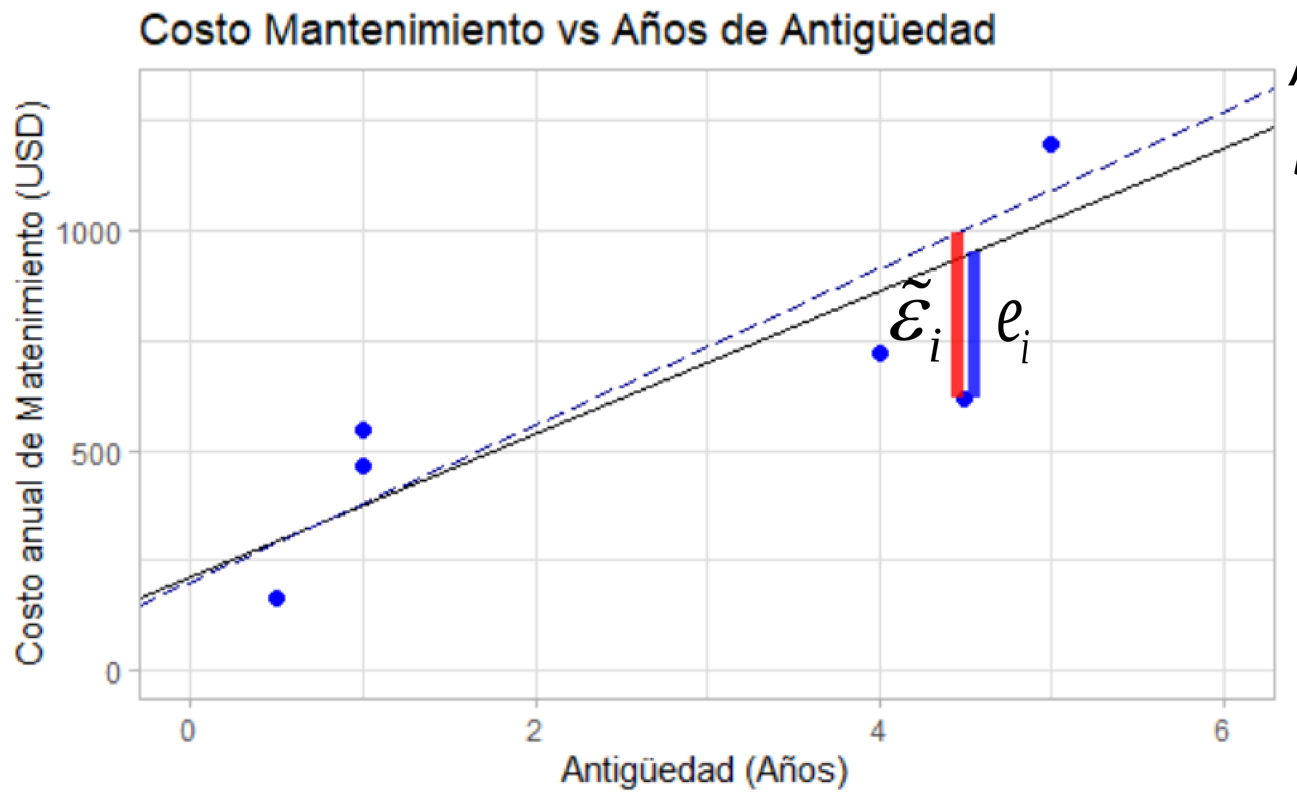
Mariano Bonoli Escobar
Javier Gil

Especificación del modelo

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Especificación del modelo



$\beta_o + \beta_1 x$ Recta de regresión poblacional

$b_o + b_1 x$ Recta de regresión estimada

$$\hat{y}_i = b_o + b_1 x_i$$

$$\tilde{y}_i = \beta_o + \beta_1 x_i + \tilde{\varepsilon}_i$$

Mariano Bonoli Escobar
Javier Gil

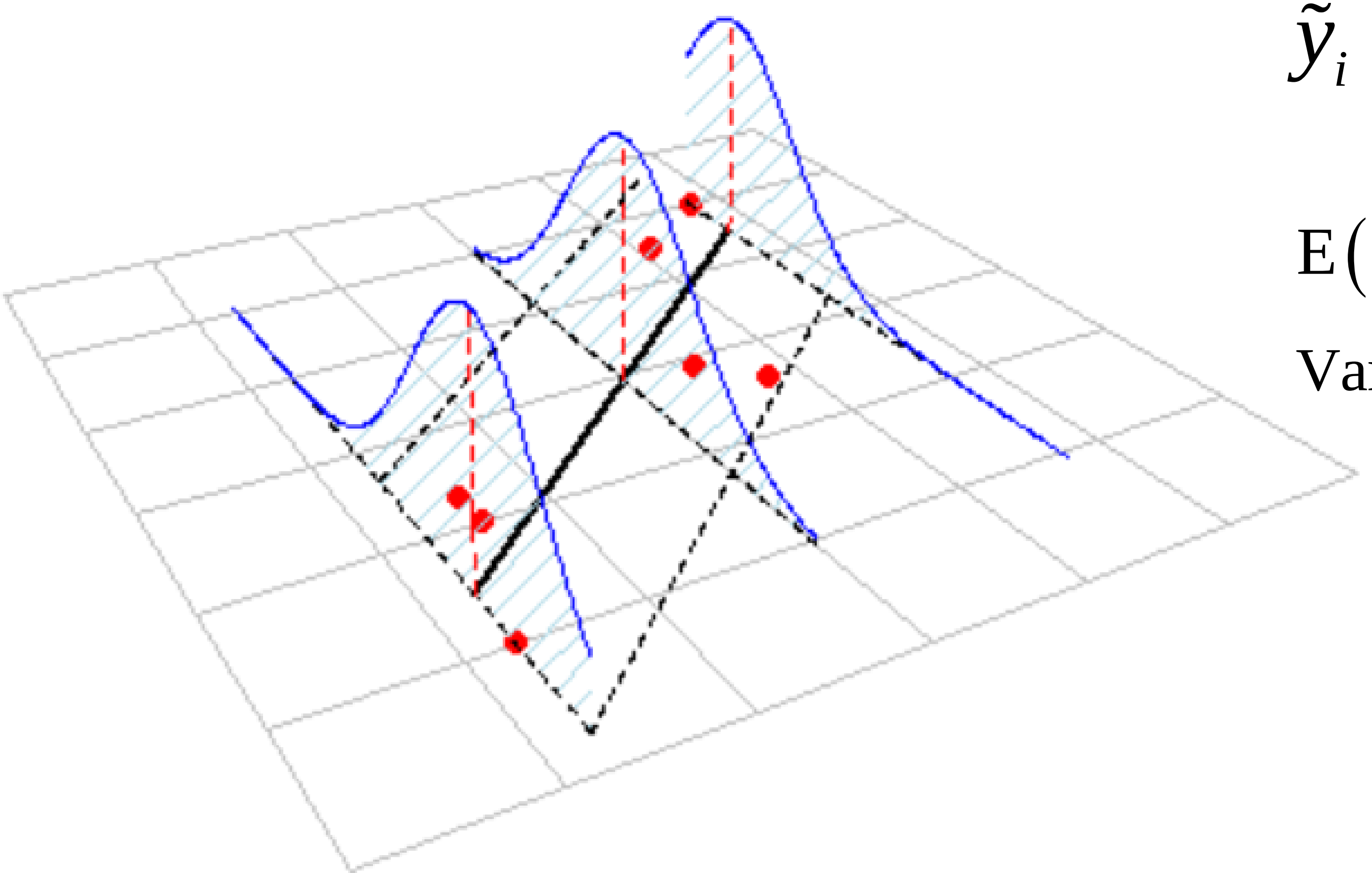
Regresión Lineal Simple

Supuestos del modelo de Regresión

- $\mathrm{E}\left(\tilde{\varepsilon}_i\right) = 0$ — Ausencia de vicio
- $\mathrm{D}^2\left(\tilde{\varepsilon}_i\right) = \sigma_\varepsilon^2 = \sigma^2$ — Homocedasticidad
- $\tilde{\varepsilon}_i \sim Normal\left(0;\ \sigma_\varepsilon^2\right)$ — Normalidad de los residuos
- $\mathrm{Cov}(\tilde{\varepsilon}_i ; \tilde{\varepsilon}_j) = 0$ — Ausencia de autocorrelación
- x_i no son aleatorios

Regresión Lineal Simple

Especificación del modelo



$$\tilde{y}_i = \beta_0 + \beta_1 \, x_i + \tilde{\varepsilon}_i$$

$$\mathrm{E}\left(\tilde{y}_i\right) = \mathrm{E}\left(\tilde{y} \middle| x_i\right) = \beta_o + \beta_1 \, x_i = \mu_i$$

$$\mathrm{Var}\left(\tilde{y}_i\right) = \sigma^2 \quad \forall i : 1..n$$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Propiedades de los estimadores

$$b_1 = \frac{\sum_{i=1}^{n} x_i\,\tilde{y}_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2} = \frac{\sum_{i=1}^{n} x_i\,\tilde{y}_i - \bar{x}\sum_{i=1}^{n}\tilde{y}_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2} = \frac{\sum_{i=1}^{n} x_i\,\tilde{y}_i - \sum_{i=1}^{n}\bar{x}\,\tilde{y}_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\tilde{y}_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$\boxed{\begin{gathered} b_1 \sim Normal\left(\mathrm{E}(b_1);\, \mathrm{D}(b_1)\right) \\ b_0 \sim Normal\left(\mathrm{E}(b_0);\, \mathrm{D}(b_0)\right) \end{gathered}}$$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Propiedades de los estimadores

$$E(b_1) = E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})\tilde{y}_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2}\right] = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2}\right] = \beta_1$$

$$E(b_0) = E(\bar{y} - b_1\bar{x}) = \beta_0$$

$$\boxed{b_0 \text{ y } b_1 \text{ son estimadores insesgados de } \beta_0 \text{ y } \beta_1 \text{ respectivamente}}$$

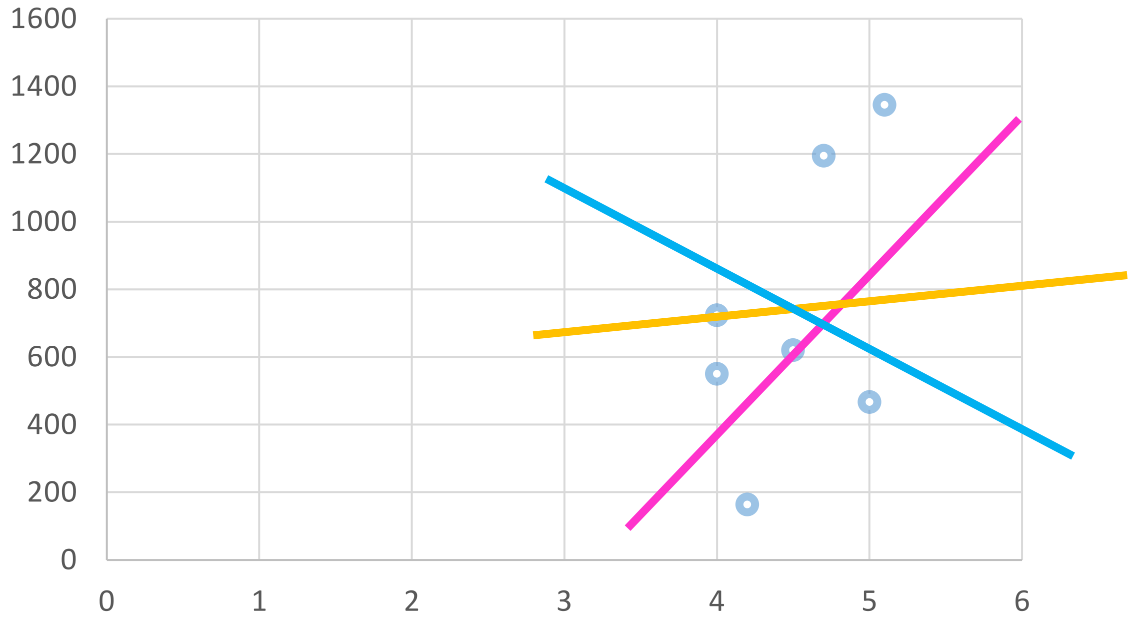
Regresión Lineal Simple

Propiedades de los estimadores

Se puede demostrar que:

$$D^2(b_1) = \frac{\sigma^2}{S_{xx}} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$D^2(b_0) = \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)\sigma^2$$



Si los valores de la variable explicativa se encuentran muy próximos la estimación de la pendiente es “inestable”

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Varianza residual

$$\operatorname{Var}(\tilde{\varepsilon}_i) = \sigma^2 \quad \forall \quad i:1..n$$

Se estima por:

$$S^2 = \frac{Q}{n-2}$$

$$Q = S_{yy} - b_1 S_{xy}$$

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,87907124
Coeficiente de determinación R^2	0,77276625
R^2 ajustado	0,7273195
Error típico	216,30039
Observaciones	7

ANÁLISIS DE VARIANZA

	Grados de libertac	Suma de cuadrados	ıedio de los cuadr	F	ılor crítico de ı
Regresión	1	795536,1345	795536,134	17,0037732	0,00914211
Residuos	5	233929,2941	46785,858		
Total	6	1029465,429			

	Coeficientes	Error típico	Estadístico t	Probabilidad	Inferior 95% S
Intercepción	213,941176	148,0094266	1,44545642	0,20795505	-166,529167
Antigüedad	161,882353	39,25788114	4,12356317	0,00914211	60,9667568

$$Q = 233929$$

$$S^2 = 46785$$

$$S = 216,3$$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Distribuciones de los estimadores

$$b_1 \sim Normal\left(\beta_1;\ \frac{\sigma}{\sqrt{S_{xx}}}\right) \quad \Rightarrow \quad \frac{b_1-\beta_1}{S_{b_1}} \sim t_{n-2} \qquad \text{con} \quad S_{b_1} = \frac{S}{\sqrt{S_{xx}}}$$

$$b_0 \sim Normal\left(\beta_0;\ \sqrt{\frac{1}{n}+\frac{\overline{x}^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}\,\sigma\right) \quad \Rightarrow \quad \frac{b_0-\beta_0}{S_{b_0}} \sim t_{n-2} \qquad \text{con} \quad S_{b_0} = \sqrt{\frac{1}{n}+\frac{\overline{x}^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}\,S$$

$$\frac{\nu\ S^2}{\sigma^2} \sim \chi^2_{n-2}$$

Mariano Bonoli Escobar
Javier Gil

Validación de modelos

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple
Mecanismos de validación de modelos

1. Procedimiento I (condición necesaria)
   - Coeficiente de correlación
   - Coeficiente de determinación

2. Procedimiento II (condición suficiente)
   - Test de significación de los coeficientes de regresión

Mariano Bonoli Escobar
Javier Gil

Validación de Modelos

Test de Significación

$$H_o)\ \beta_1 = 0$$

$$H_1)\ \beta_1 > 0$$

$$H_1)\ \beta_1 < 0$$

$$H_1)\ \beta_1 \neq 0$$

¿Tenemos conocimientos extra estadísticos?

$$\frac{b_1 - \beta_1}{S_{b_1}} \sim t_{(n-2)}$$

$$S_{b_1} = \frac{S}{\sqrt{S_{xx}}}$$

Si se rechaza Ho concluimos que x e y tienen un cierto grado de asociación LINEAL

Mariano Bonoli Escobar
Javier Gil

CASO DE DISCUSION I

El Gerente de Logística desea construir un modelo para presupuestar el costo anual de mantenimiento de los auto-elevadores. En base a su experiencia cree que la variable de mayor relevancia es la antigüedad del equipo. Un analista de la gerencia recopiló la siguiente información para 7 equipos:

Antigüedad [años]	4.5	1	1	5	0.5	4	6
Costo anual U$S	619	549	466	1194	163	723	1345

$$H_0)\ \beta_1 \leq 0$$

$$H_1)\ \beta_1 > 0$$

$$\frac{b_1}{S_{b_1}} = \frac{161{,}88}{39{,}26} = 4{,}12$$

$$CR: t_{obs} = \frac{b_1 - 0}{S_{b_1}} \geq t_{crit} = t_{(n-2;1-\alpha)} = 1{,}48$$

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,87907124
Coeficiente de determinación R^2	0,77276625
R^2 ajustado	0,7273195
Error típico	216,30039
Observaciones	7

Niveles de significación a posteriori para test bilateral

ANÁLISIS DE VARIANZA

	Grados de libertac	Suma de cuadrados	iedio de los cuadr	F	alor crítico de F
Regresión	1	795536,1345	795536,134	17,0037732	0,00914211
Residuos	5	233929,2941	46785,8588		
Total	6	1029465,429			

	Coeficientes	Error típico	Estadístico t	Probabilidad	ıferior 95%	Superior 95%
Intercepción	213,941176	148,0094266	1,44545642	0,20795505	166,529167	594,41152
Antigüedad	161,882353	39,25788114	4,12356317	0,00914211	60,9667568	262,797949

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple
Coeficiente de Correlación

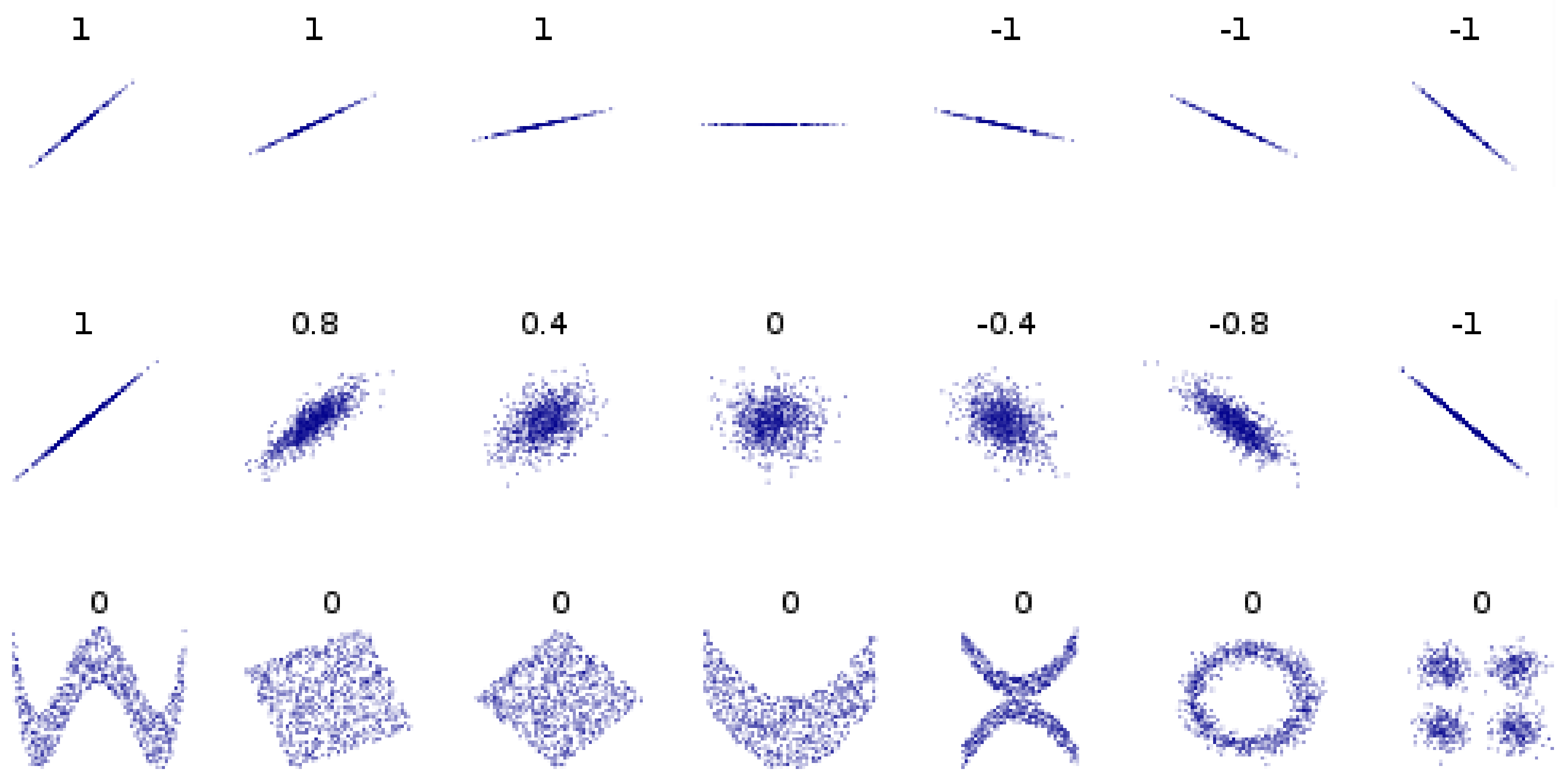
Mide del **grado** de **dependencia lineal** entre dos variables cuantitativas

$$r = \frac{SCxy}{\sqrt{SCxx \;\; SCyy}} = \frac{\sum x_i y_i - n\,\bar{x}\,\bar{y}}{\sqrt{\left(\sum x_i^{\,2} - n\,\bar{x}^2\right)\left(\sum y_i^{\,2} - n\,\bar{y}^2\right)}}$$

Mariano Bonoli Escobar
Javier Gil

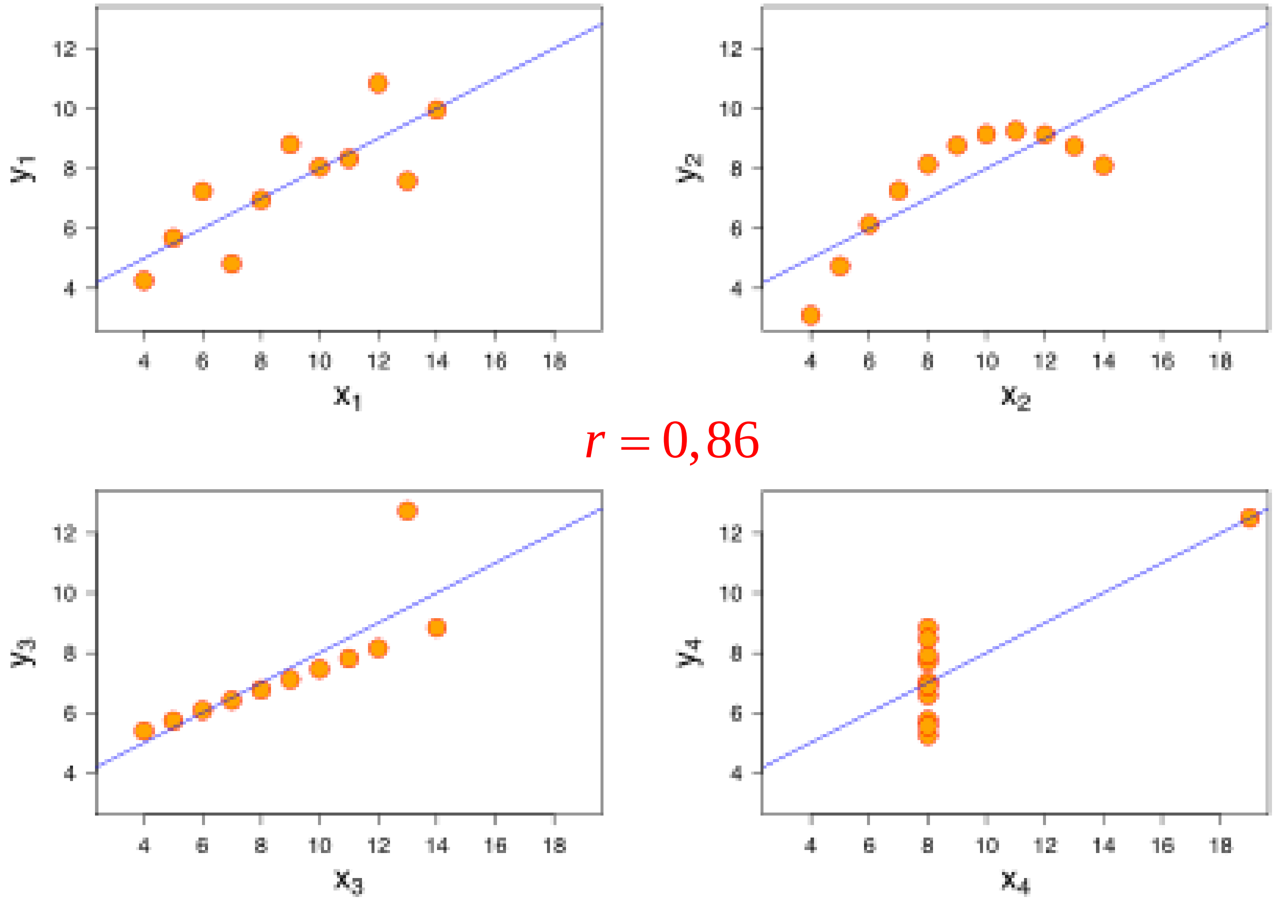
Regresión Lineal Simple

Coeficiente de Correlación



Mariano Bonoli Escobar
Javier Gil

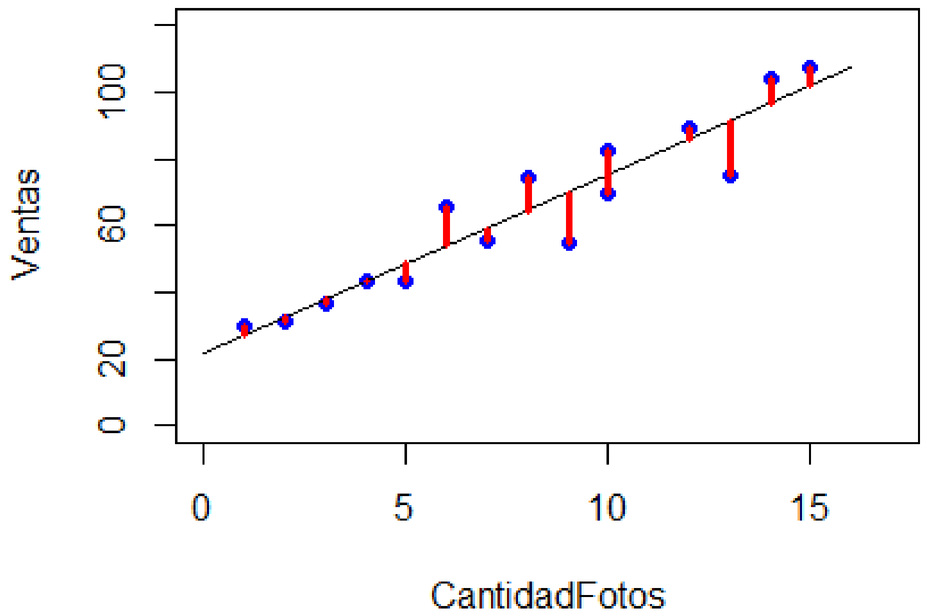
Regresión Lineal Simple
Coeficiente de Correlación

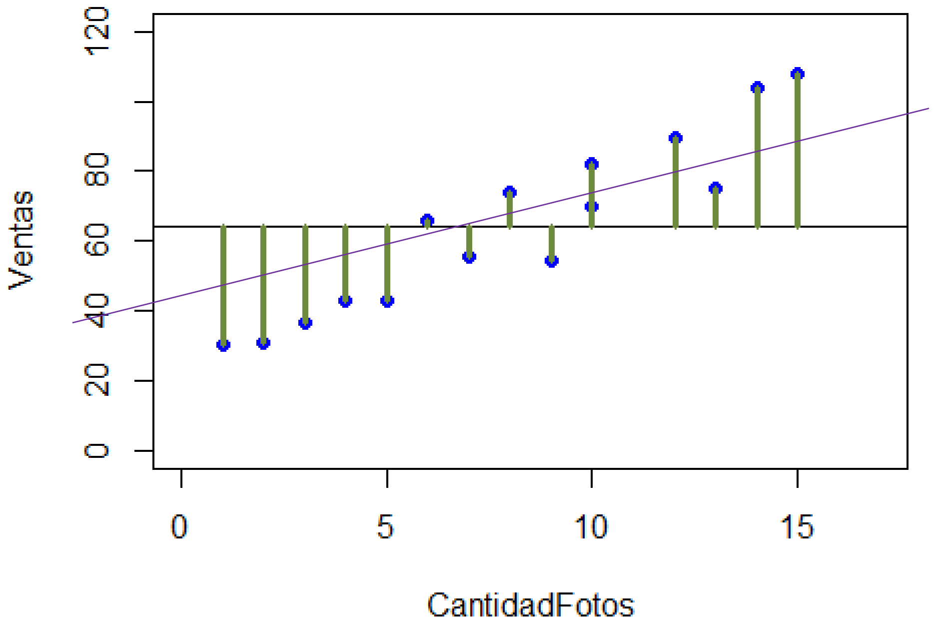


$r = 0,86$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple
Coeficiente de determinación





$$R^2 = 1 - \frac{Q}{T} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple
Fuentes de variabilidad

Variabilidad Total $$SC_{Total} = T = \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2 = SCy$$

Variabilidad No explicada $$SC_{Residual} = Q = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 = \sum_{i=1}^{n} e_i^{\ 2} = \sum_{i=1}^{n} y_i^2 - b_0 \sum_{i=1}^{n} y_i - b_1 . \sum_{i=1}^{n} x_i y_i = Q$$

Variabilidad Explicada $$SC_{Regresion} = H = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y} \right)^2$$

	Suma Cuadrados	Grados libertad	Cuadrados medios	CM Esperado
Regresión	$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	1	$CM_{\mathrm{Regr}} = b_1 S_{xy} / 1$	$\sigma^2 + \beta_1^{\ 2} S_{xx}$
Residual	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$n-2$	$CM_{\mathrm{Res}} = (S_{yy} - b_1 S_{xy})/(n-2)$	σ^2
Total	$\sum_{i=1}^{n}(y_i - \bar{y})^2$	$n-1$		

$$T = H + Q$$

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

Variación total = Variación explicada + Variación Residual

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Coeficiente de determinación

$$R^2 \geq 0,90$$ **Si queremos hacer pronósticos confiables**

$$R^2 \geq 0,80$$ **Procesos físicos e industriales**

$$R^2 \geq 0,70$$ **Economía**

$$R^2 \geq 0,50$$ **Sociología**

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple
Bondad de Ajuste

Un buen modelo debe contener

- Coeficiente de correlación con valor absoluto alto
- Coeficiente de determinación alto
- Su coeficiente de regresión $\beta 1$ deben ser significativamente distinto de cero

Mariano Bonoli Escobar
Javier Gil

Inferencia sobre $\tilde{y}$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Inferencia sobre $\mathrm{E}(\tilde{y}|x) = \beta_0 + \beta_1 x$

Estimación puntual $\hat{\mathrm{E}}\left(\tilde{y}|x_o\right) = b_0 + b_1\, x_0$

$$\mathrm{E}\left(\tilde{y}|x_o\right) \sim Normal\left(\beta_o + \beta_1 x_o\ ;\ \sigma\sqrt{\frac{1}{n} + \frac{\left(x_o - \overline{x}\right)^2}{S_{xx}}}\right)$$

$$\frac{\left(b_0 + b_1\, x_0\right) - \left(\beta_o + \beta_1 x_o\right)}{S\sqrt{\frac{1}{n} + \frac{\left(x_o - \overline{x}\right)^2}{S_{xx}}}} \sim t_{(n-2)}$$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Inferencia sobre $\tilde{y}|x_0 = \beta_0 + \beta_1 x + \tilde{\varepsilon}$

Estimación puntual $\quad \hat{y}\big|x_o = b_0 + b_1\, x_0$

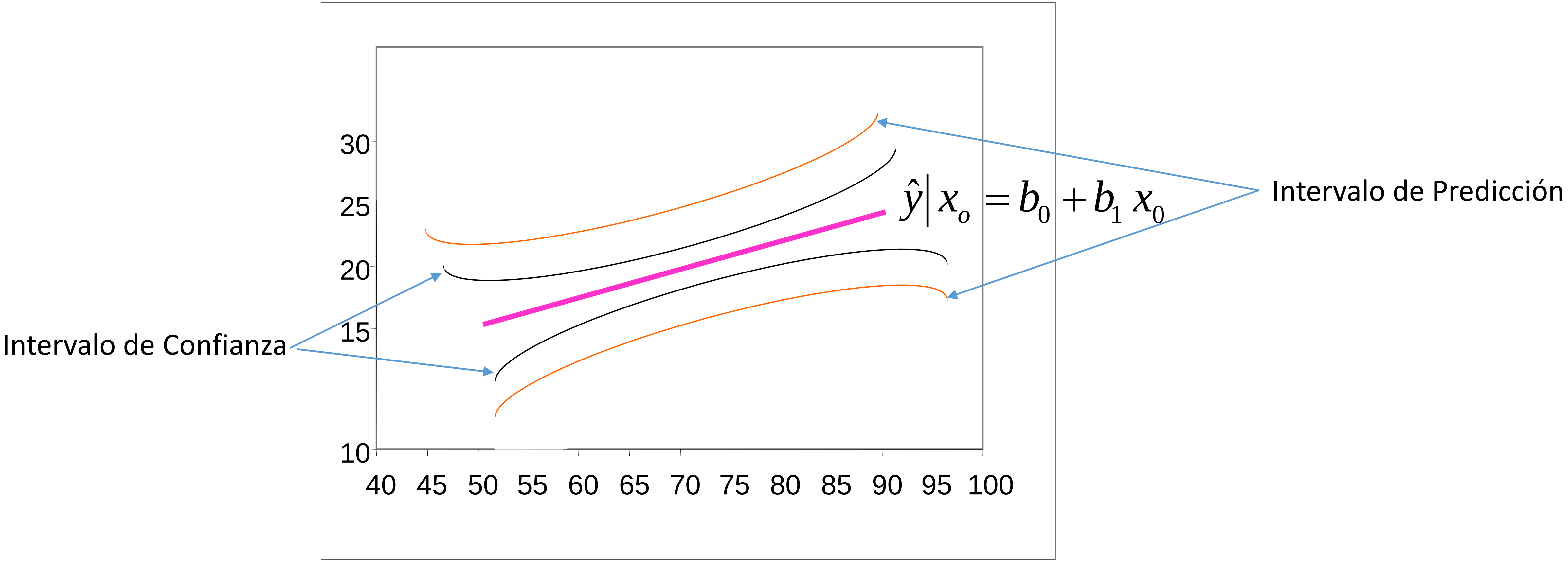
$$\tilde{y}\big|x_o \sim Normal\left(\beta_o + \beta_1 x_o \,;\, \sigma\sqrt{1+\frac{1}{n}+\frac{\left(x_o-\overline{x}\right)^2}{S_{xx}}}\right)$$

$$\frac{\left(b_0 + b_1\, x_0\right) - \left(\beta_o + \beta_1 x_o\right)}{S\sqrt{1+\frac{1}{n}+\frac{\left(x_o-\overline{x}\right)^2}{S_{xx}}}} \sim t_{(n-2)}$$

Mariano Bonoli Escobar
Javier Gil

Regresión Lineal Simple

Inferencia sobre $\tilde{y}|x_0 = \beta_0 + \beta_1 x + \tilde{\varepsilon}$



Mariano Bonoli Escobar
Javier Gil

CASO DE DISCUSION I

El Gerente de Logística desea construir un modelo para presupuestar el costo anual de mantenimiento de los auto-elevadores. En base a su experiencia cree que la variable de mayor relevancia es la antigüedad del equipo. Un analista de la gerencia recopiló la siguiente información para 7 equipos.

- Estime con 90% la el costo anual de mantenimiento esperado para un equipo de 5,5 años de antigüedad
- ¿Cuál será el mayor costo de mantenimiento a pagar para un equipo de 5,5 años de antigüedad con 1% de probabilidad de ser superado?

Antigüedad [años]	4.5	1	1	5	0.5	4	6
Costo anual U$S	619	549	466	1194	163	723	1345

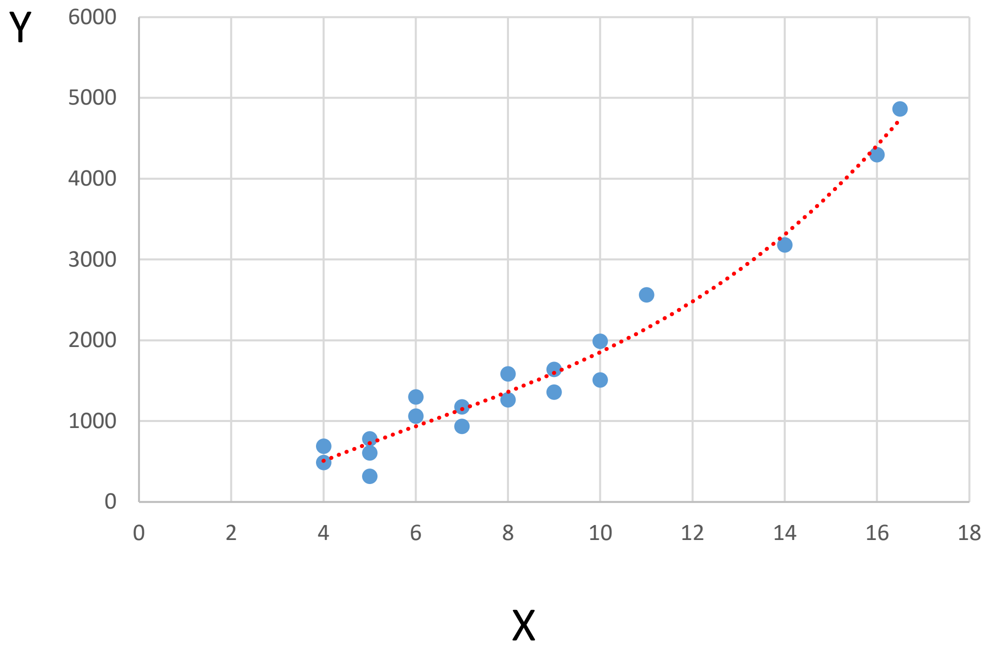
Resolución con Infostat

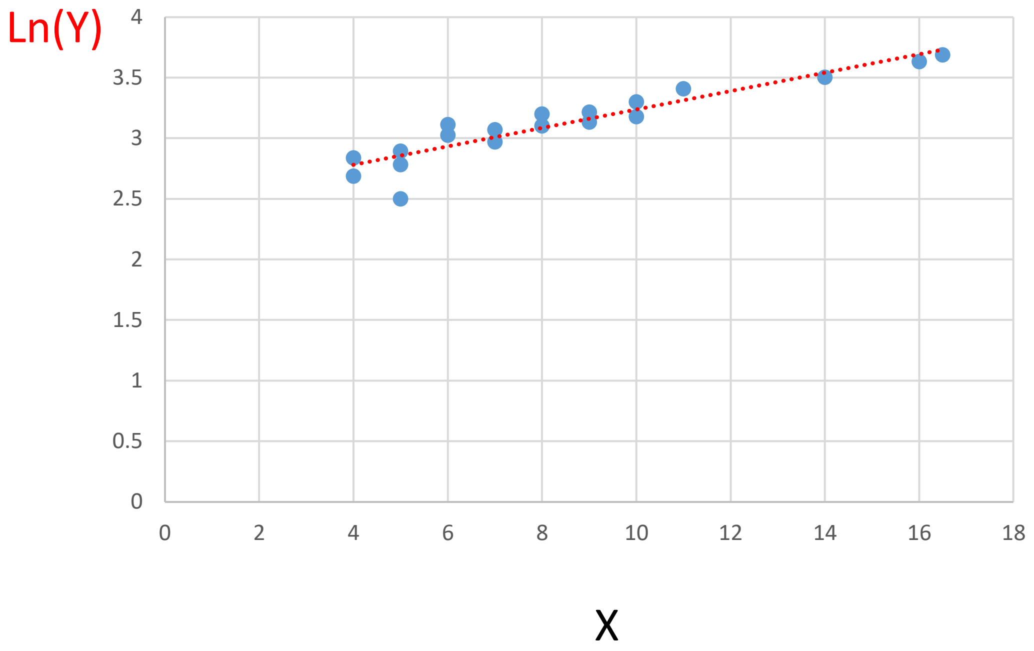
Mariano Bonoli Escobar
Javier Gil

Trasformaciones Linealizantes

Mariano Bonoli Escobar
Javier Gil

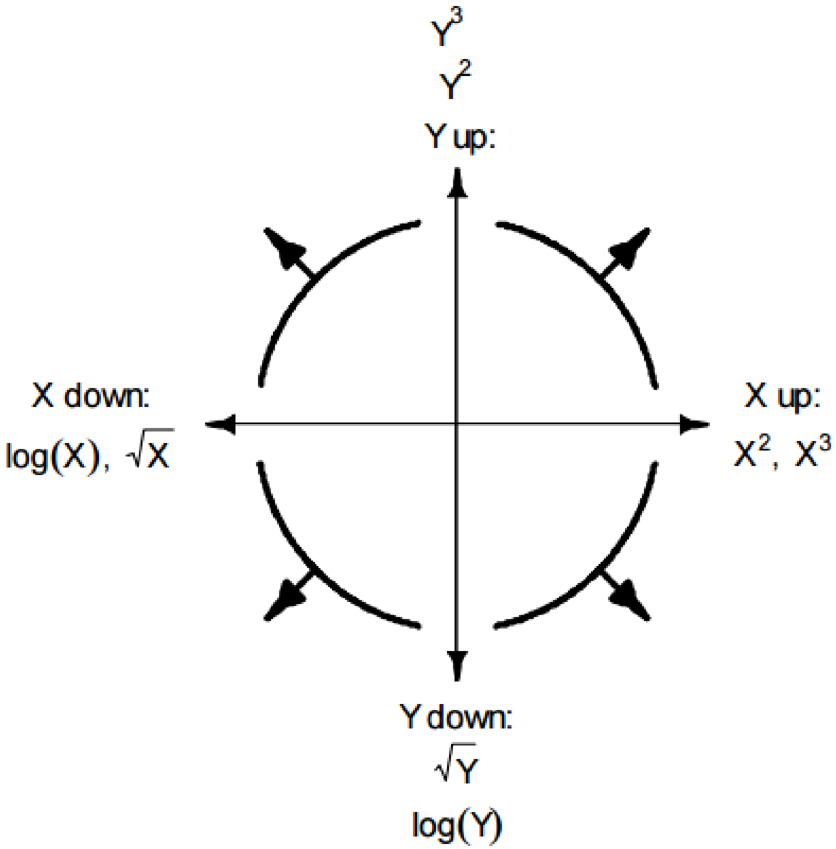
Regresión Lineal Simple
Transformaciones Linealizantes





Mariano Bonoli Escobar
Javier Gil

Transformaciones Linealizantes
Regla de Mosteller y Tukey



Mariano Bonoli Escobar
Javier Gil

Transformaciones Linealizantes
Interpretación de los coeficientes

$$y_i = \alpha\, x_i^{\beta} \varepsilon_i \quad \rightarrow \quad \underbrace{ln(y_i)}_{y_i'} = \underbrace{ln(\alpha)}_{\beta_0} + \underset{\beta_1}{\beta}\, \underset{x_i'}{ln(x_i)} + \underset{\varepsilon_i'}{ln(\varepsilon_i)}$$

$$\beta_1 = \frac{\partial ln(y)}{\partial ln(x)} = \frac{dy/y}{dx/x} = \text{Elasticidad de } y \text{ sobre } x$$

Variación porcentual promedio de y por unidad porcentual de variacion en x

$$y_i = \alpha\ e^{\beta x i} \varepsilon_i \quad \rightarrow \quad \underbrace{ln(y_i)}_{y_i'} = \underbrace{ln(\alpha)}_{\beta_0} + \underbrace{\beta}_{\beta_1}\ x_i + \underbrace{ln(\varepsilon_i)}_{\varepsilon_i'}$$

$$\beta_1 = \frac{\partial ln(y)}{\partial x} = \frac{dy/y}{dx} = \text{Variación porcentual promedio de } y \text{ por unidad de } x$$

Mariano Bonoli Escobar
Javier Gil

Regresión con intercepto conocido

Mariano Bonoli Escobar
Javier Gil

Regresión con intercepto conocido
Concepto

- En ocasiones contamos con conocimiento a priori sobre el valor de la ordenada al origen β_0
- En estos casos lo mas apropiado es adoptar el valor de β_0 conocido y estimar el resto de los parámetros del modelo, ganando un grado de libertad y un modelo mas parsimonioso

Mariano Bonoli Escobar
Javier Gil

Regresión con intercepto conocido
Expresiones

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \rightarrow \quad \underbrace{y_i - \beta_0}_{Z_i} = \beta_1 x_i + \varepsilon_i \quad \rightarrow \quad \hat{Z}_i = b_1\, x_i$$

$$Q = \sum_{i=1}^{n} z_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i z_i\right)^2}{\sum_{i=1}^{n} x_i^2} \qquad S^2 = \frac{Q}{v} \qquad v = n-1$$

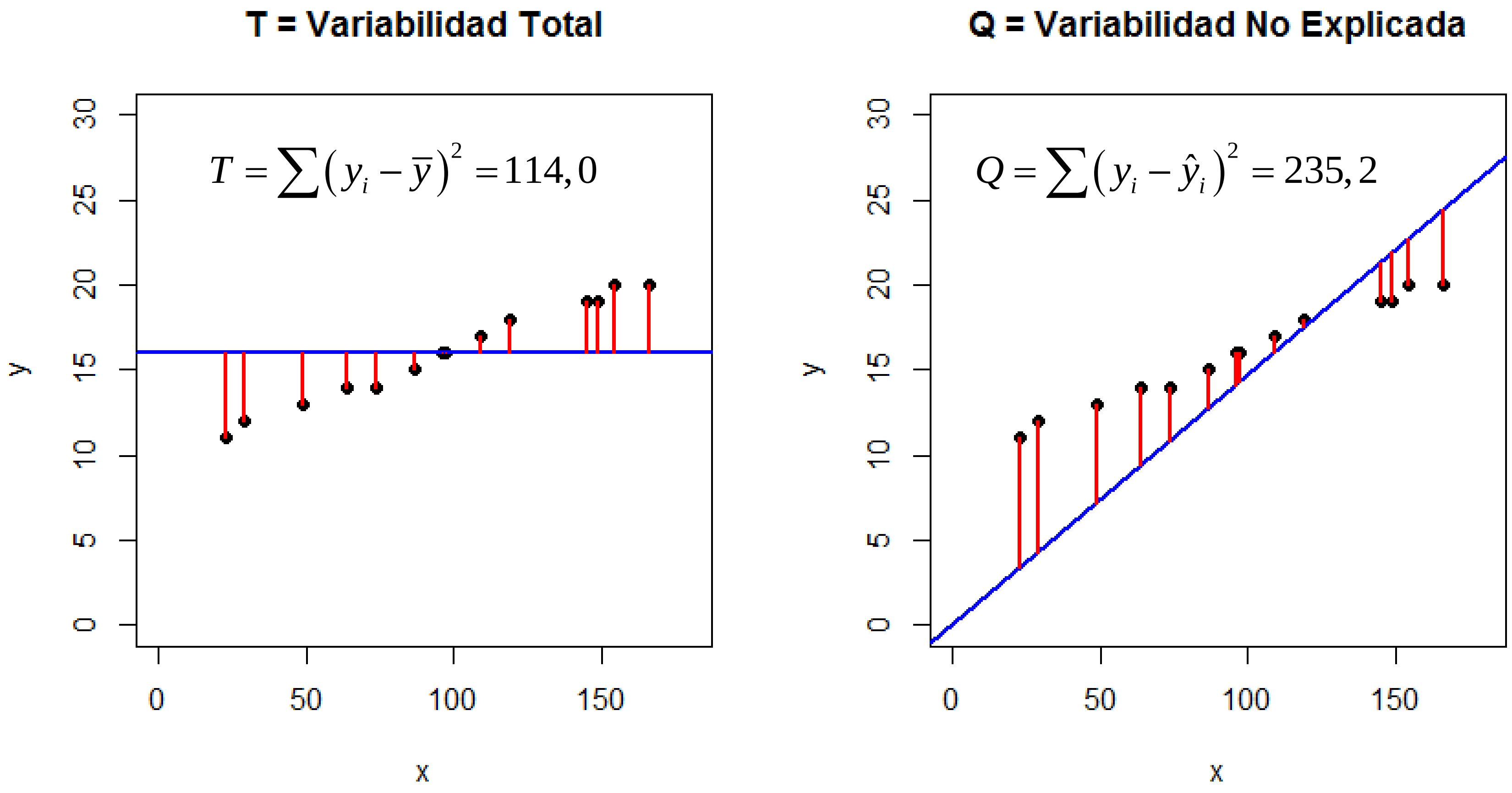
$$b_1 = \frac{\sum_{i=1}^{n} x_i z_i}{\sum_{i=1}^{n} x_i^2}$$

$$\hat{D}^2(b_1) = S^2 \frac{x_0^2}{\sum_{i=1}^{n} x_i^2}$$

$$\hat{D}^2\left(E(z|x_0)\right) = S^2 \frac{x_0^2}{\sum_{i=1}^{n} x_i^2}$$

Regresión con intercepto conocido

Uso incorrecto

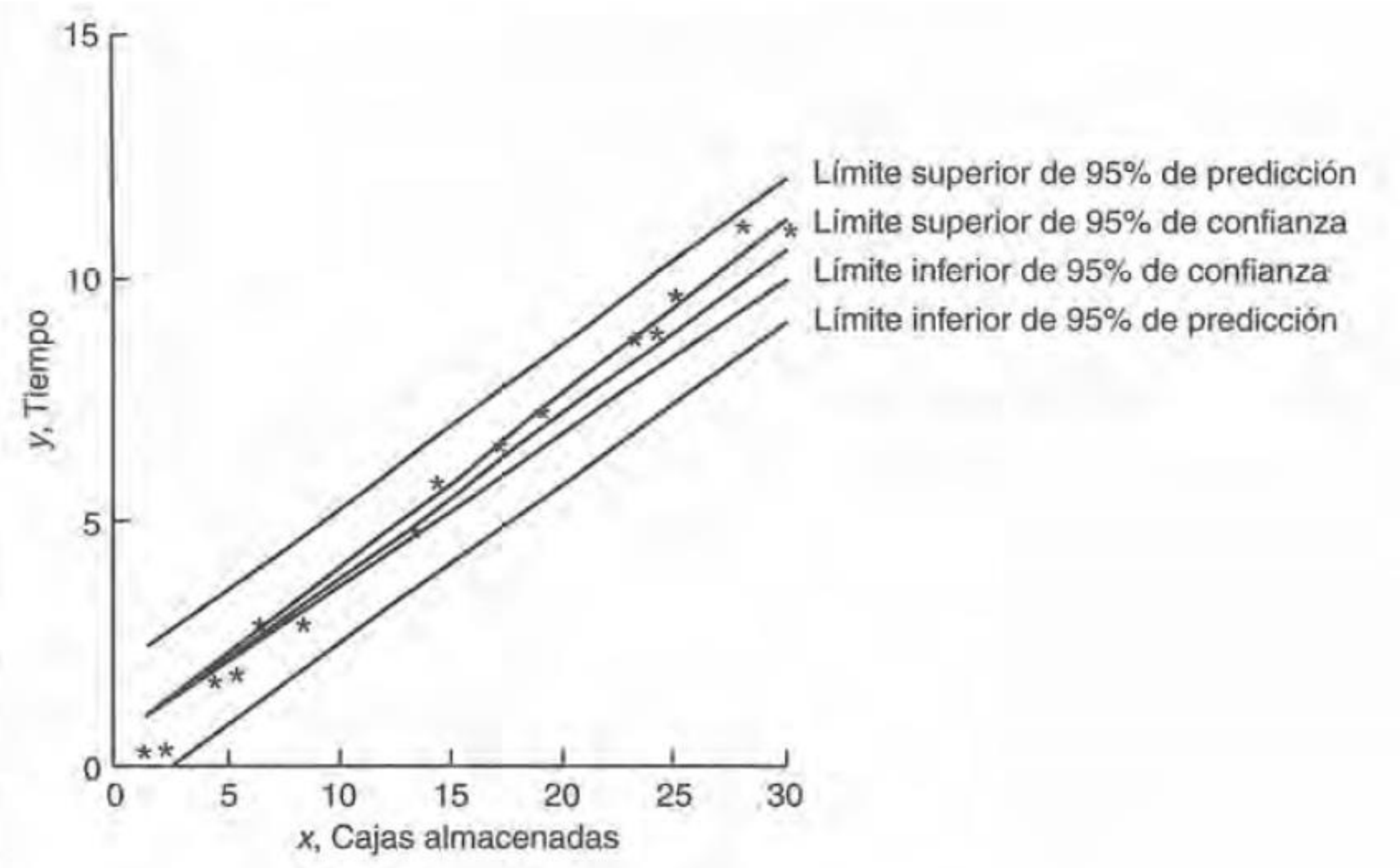


$$R^2 = 1 - \frac{Q}{T} = 1 - \frac{235,2}{114,0} = -1,06$$

Mariano Bonoli Escobar
Javier Gil

Regresión con intercepto conocido

Intervalos de Confianza y Predicción



Mariano Bonoli Escobar
Javier Gil