

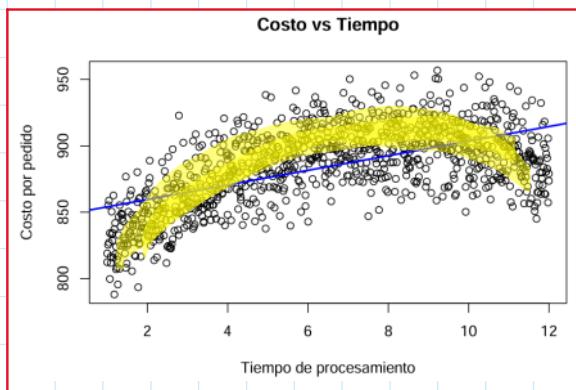
## Reaso evaluación 1

miércoles, 17 de septiembre de 2025 7:48 p. m.

### Reaso estadística II

Una empresa de logística nacional busca optimizar sus procesos internos de procesamiento de pedidos. Para cada pedido, se mide el **tiempo de procesamiento** (en horas), que incluye desde la recepción del pedido hasta que está listo para despacho. La gerencia está interesada en comprender cómo se relaciona este tiempo con el **costo total** por pedido (en miles de pesos), que refleja gastos de personal, energía y recursos operativos.

El equipo de analítica recolectó información de **1.000 pedidos** en diferentes momentos y condiciones. Los tiempos de procesamiento se distribuyen en el rango de **1 a 12 horas**, y para cada nivel de tiempo se registraron dos observaciones (réplicas), con el fin de poder aplicar pruebas de ajuste.



Regression	SSR	1	$SSR/1 = MSR$	$F_{1, n-2}$
Residuals	SSE	$n-2$	$SSE/n-2 = MSE$	$MSE/MSE$
Lack of fit	SSLDF	$m-2$	$SSLDF/m-2$	$F_{m-2, n-m}$
Pure error	SSPE	$n-m$	$SSPE/n-m$	$MSPE/MSPE$
Total	SST	$n-1$	$SST/n-1$	$MSR/MSPE$

$$SST = SSR + SSE$$

$$n-1 = 1 + n-2$$

$$SSE = 364868 + 21134 = 576002$$

$$SST = 306523 + 576002 = 882525$$

$$n-m = 500; n-2 = 998$$

$$n-2 = n-m + m-2$$

$$m-2 = 998 - 500$$

$$m-2 = 498$$

$$m = 500$$

$$MSR = SSR/1 = 306523$$

$$MSE = SSE/n-2 = 576002/998 = 577$$

$$MSLDF = SSLDF/m-2 = 364868/498 = 733$$

$$MSPE = SSPE/n-m = 21134/500 = 422$$

$$\text{F}_{\text{sig}}: MSR/MSE = 306523/577 = 531$$

$$\text{Flut}: MSLDF/MSPE = 733/422$$

$$\alpha = 0,05$$

$\text{Pr}_{\text{al}} < \alpha \rightarrow \text{Rechazar } H_0$

Lack of fit test - Anova Table				
	Sum Sq	Df	Mean Sq	F value
Regression	306523	1	306523	531.0910 < 2.2e-16 ***
Residuals	576002	998	577	
Lack of fit	364868	498	733	1.7351 4.839e-10 ***
Pure error	21134	500	422	
Total	882525	999		

A un nivel de significancia del 5%, se pue establecer que existe evidencia suficiente para rechazar  $H_0$ , de modo que existe carencia de ajuste.

```

## Regression 306523 1 306523 531.0910 < 2.2e-16 ***
## Residuals 576002 998 577
## Lack of fit 364868 498 733 1.7351 4.839e-10 ***
## Pure error 211134 500 422
## Total 882525 999
## ---

```

Se pide establecer que existe evidencia suficiente para rechazar  $H_0$ , de modo que existe carencia de ajuste (el modelo no se ajusta a una linea recta)

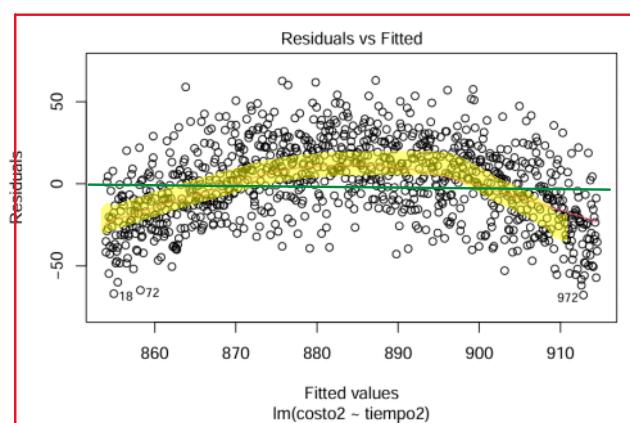
- Se busca NO rechazar  $H_0$  (LOF)

## (2). Validación supuestos del modelo

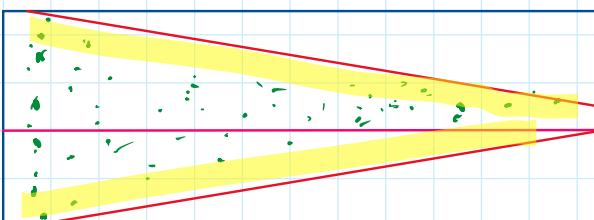
$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i ; \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) \quad i=1, \dots, n$$

### (1). Varianza constante

$Y_i vs \epsilon_i$



Varianza no constante



(1). Independientes (asume)

(2). identicamente dist. (Normal)

(3) Media cero (siempre se cumple)

(4). Varianza constante (homocedasticidad)

:  $Y_i vs \epsilon_i$  (incorrecto) : probar independencia

$$Y_i = \beta_0 + \beta_1 X_{i1}$$

• Tendencia en forma U;  $\alpha$

• Hay problemas de falta de ajuste

• No digo nada sobre la varianza

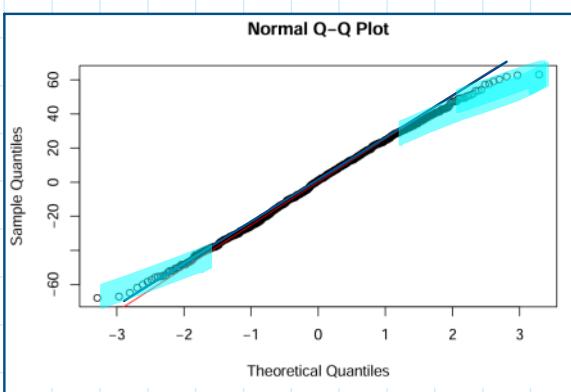
Varianza constante:

linealidad

: Heterocedasticidad

- Realizar la comprobación del supuesto de normalidad

{(1). gráfica  
(2). Shapiro-Wilk.



$H_0: \epsilon_i \sim N(0, \sigma^2) \quad i=1, \dots, n$

$H_1: \epsilon_i \neq N(0, \sigma^2)$

```

## 
## Shapiro-Wilk normality test
## 
## data: residuals(model2)
## W = 0.99684, p-value = 0.04413

```

Pval <  $\alpha = 0.05$

• A un nivel de significancia del 5%, se puede establecer que existe evidencia para rechazar  $H_0$ : Los errores NO tienen una distribución normal.

## (3). Comprobar significancia

1) D-n 2) Mcb 3)

Fsig.

(Significancia modelo)

```

## Analysis of Variance Table
## 
## Model: lm(cost2 - tiempo2)
## 
##   Term          Sum Sq  Df  F value    Pr(>F)
##   (Intercept) 531.0910  1  177.031 0.0000000
##   tiempo2     4.839e-10 1  0.0161 0.9968400
##   Residuals  882525.00 998
## 
## 
## 
## 
## 
```

4) Fstat

## 1.5.1. Comprobar significancia:

$$H_0: \beta_1 = 0$$

$$F = \frac{MSR}{MSE} \sim F_{f_1, n-2}$$

$$H_1: \beta_1 \neq 0$$

Si:

(significancia menor)

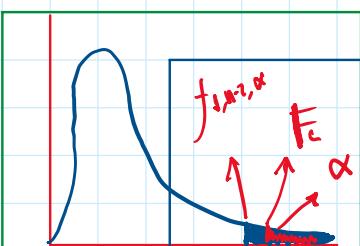
```
# Analysis of Variance Table
#
# Response: costo2
# Df Sum Sq Mean Sq F value    Pr(>F)
# tiempo2     1 306523 306523 531.09 < 2.2e-16 ***
# Residuals 998 576002   577
```

$$F = 531,09$$

$$Pr < 0,05$$

A un nivel de significancia del 5%: El modelo es significativo

Región de rechazo:  $R_i: F_c > F_{f_1, n-2, \alpha}$  : Distribución F: Cola derecha



### 4. Significación individual

$$H_0: \beta_j = 0 ; j=0,1$$

$$H_1: \beta_j \neq 0$$

$$T = \frac{\hat{\beta}_j}{\sqrt{V(\hat{\beta}_j)}} \sim t_{n-2}$$

$$T^2 = F \checkmark$$

```
## Call:
## lm(formula = costo2 ~ tiempo2)
##
## Residuals:
##   Min     1Q Median     3Q    Max 
## -67.800 -16.846   1.007  17.268  63.029
## 
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 848.4956  1.7280 491.04 <2e-16 ***
## tiempo2      5.5025  0.2388 23.05 <2e-16 ***
## ---    
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
```

```
# Residual standard error: 24.02 on 998 degrees of freedom
```

```
# Multiple R-squared:  0.3473, Adjusted R-squared:  0.3467
```

```
# F-statistic: 531.1 on 1 and 998 DF, p-value: < 2.2e-16
```

$$T = \frac{\hat{\beta}_0}{\sqrt{V(\hat{\beta}_0)}} = \frac{848,4956}{1,7280} = 491,04$$

$$F_{f_1, n-2} : n-2 = 998$$

$$\sqrt{MSE} = 24,02; MSE = 24,02^2$$

$\beta_1$ : A un nivel de significancia del 5%: Por un aumento unitario en el tiempo procesamiento (horas), el cambio promedio en el costo total es de 5,5025 (miles CUP)

### Interpretación $R^2$ :

$$R^2: \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$= 0,3473$$

(Variancia explicada por el modelo: da la relación lineal entre X y Y)

### Intervalos:

#### 1. Intervalo confianza:

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

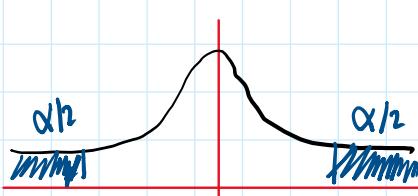
$$\beta_1:$$

```
## Call:
## lm(formula = costo2 ~ tiempo2)
##
## Residuals:
##   Min     1Q Median     3Q    Max 
## -67.800 -16.846   1.007  17.268  63.029
## 
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 848.4956  1.7280 491.04 <2e-16 ***
## tiempo2      5.5025  0.2388 23.05 <2e-16 ***
## ---    
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
```

$$Y_i = 848,4956 + 5,5025 X_{i1}$$

$$5,5025 \pm 1,64 \cdot 0,2388$$



```

## Multiple R-squared:  0.3473, Adjusted R-squared:  0.3467
## F-statistic: 531.1 on 1 and 98 DF, p-value: < 2.2e-16
##          2.5 %    97.5 %
## (Intercept) 845.104800 851.886491
## tiempo2      5.033961  5.971051

```

• Con una confianza del 95% por el momento anterior en el tiempo, en promedio los horas varían entre 5,03 y 5,97 horas

¿Será que  $\beta_1$  es significativo?

• Si, no contiene al cero

$$(5,03 : 5,97) : (-3,3)$$

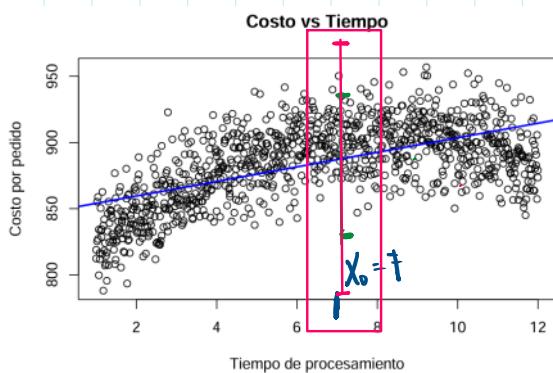
- (1). Concluir con probabilidad de hipótesis: significativa.
- (2). Concluir IC: Confianza

$$\text{Sigi: } 1 - \text{Conf}$$

## (6). Intervalo confianza respuesta media e intervalo predicción

Rep. Media:

$$Y_0 \pm t_{\alpha/2, n-2} \cdot \sqrt{MSE \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}} \right]}$$



$$Y_0 \pm t_{\alpha/2, n-2} \cdot \sqrt{MSE \left[ \frac{1}{n} + \frac{1}{S_{xx}} + \frac{(X_0 - \bar{X})^2}{S_{xx}} \right]}$$

Intervalo es más ancho

$$Y_0 \pm t_{\alpha/2, n-2} \cdot \sqrt{MSE + S.E^2(Y_0)}$$

```

##          fit      lwr      upr
## 1 887.0132 885.5041 888.5223

```

```

##          fit      lwr      upr
## 1 887.0132 839.8455 934.1808

```

IC 95%

IP

$$Y_0: 887,0132 :$$

$$Y_i = 818,4956 + 5,5025(t)$$