Resumen primer parcial

sábado, 13 de septiembre de 2025 1:29 p. m.

Fundamentos del modelo regresión

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

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

Algunas propiedades importantes

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

MLE → Método estadístico; OLS → Método matemático

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

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

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

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

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

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

Pruebas de hipótesis:

$\boxed{T_{0,j}^2 = F_{0,1}}$: $\boxed{\beta_1}$: Significancia individual

$\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases}$; $j = 0, 1$: $\boxed{T: \dfrac{\hat{\beta}_j}{\sqrt{\mathrm{Var}[\hat{\beta}_j]}} \sim t_{n-2}}$: $P(|t_{n-2}| > |T_0|) < \alpha$

- Rechaza H_0

Significancia regresión

$\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases}$; $j = 1$: $\boxed{F: \dfrac{SSR/1}{SSE/n-2} = \dfrac{MSR}{MSE} \sim f_{1,n-2}}$

$P(f_{1,n-2} > F_0) < \alpha$: Rechaza H_1

$SST = S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$; $gl(SST) = n-2$

$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$; $gl(SSR) = 1$

$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$; $gl(SSE) = n-1$

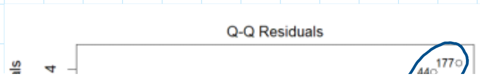
Validación de supuestos

Normalidad errores

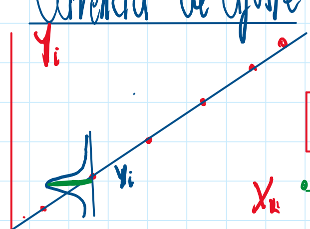
Shapiro-wilk: $\begin{cases} H_0: \varepsilon_i \sim \text{Normal} \\ H_1: \varepsilon_i \nsim \text{Normal} \end{cases}$ $i = 1, \dots, n$

Pval < α

- Existe, a un nivel de significancia del 5% evidencia suficiente para rechazar H_0, así, se concluye que los errores no tienen una distribución normal.



Carencia de ajuste



Y_i; $i = 1, \dots, n$

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

- el modelo es acertado

Problemas de ajuste

$\boxed{\begin{matrix} H_0: E[Y_i | X_i] = \beta_0 + \beta_1 X_{i1} \\ H_1: E[Y_i | X_i] \neq \beta_0 + \beta_1 X_{i1} \end{matrix}}$

$\boxed{F_{LOF} = \dfrac{SSLOF/(m-2)}{SSPE/(n-m)} = \dfrac{MSLOF}{MSPE} \sim f_{m-2,\,n-m}}$

- SSLOF: Suma carencia ajuste
- SSPE: Suma cuadrados error puro

$SSPE = \sum_{i=1}^{m} \sum_{j=1} (y_{ij} - \bar{Y}_i)^2$; $SSLOF: \sum_{i=1}^{m} n_i (\bar{y}_i - \hat{y}_i)$

- $\boxed{\begin{matrix} SSE = SSPE + SSLOF \\ n-2 = m-2 + n-m \end{matrix}}$

n = # Observaciones en mi conjunto de datos (muestra)
m = # niveles

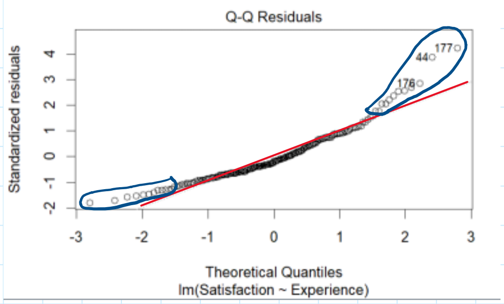
Análisis de varianza que incorpora la prueba de falta de ajuste en el modelo de RLS

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F Calculado
Regresión	SSR	1	$MSR = \frac{SSR}{1} = SSR$	$F_0 = \frac{MSR}{MSE}$
Error	SSE	$n-2$	$MSE = \frac{SSE}{n-2}$	
Falta de Ajuste	$SSLOF$	$m-2$	$MSLOF = \frac{SSLOF}{m-2}$	$F_0 = \frac{MSLOF}{MSPE}$
Error Puro	$SSPE$	$n-m$	$MSPE = \frac{SSPE}{n-m}$	
Total	SST	$n-1$		

Intervalos:

(1). Confianza (parámetros)

$\boxed{\hat{\beta}_j \pm t_{\alpha/2,\,n-2} \cdot \sqrt{\mathrm{Var}[\hat{\beta}_j]}}$ $j = 0, 1$



$$\hat{\beta}_j \pm t_{\alpha/2,\, n-2} \cdot \sqrt{\widehat{Var}[\hat{\beta}_j]} \qquad j = 0, 1$$

(2). Respuesta media

$$\hat{Y}_0 \pm t_{n-2} \cdot \sqrt{MSE\left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}}\right]}$$

(3). Predicción respuesta

$$\hat{Y}_0 \pm t_{n-2} \cdot \sqrt{MSE\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}}\right]}$$

¿Cómo identificar los elementos de cada tabla?

Tabla resumen

e_i

$\alpha = 0{,}05$

```
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) 40.40987    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
```

$\sqrt{\widehat{Var}[\hat{\beta}_0]}$

$\hat{\beta}_0$ $\hat{\beta}_1$

T_0

$\sqrt{MSE}$ $F = \frac{MSR}{MSE}$ R^2 $n-2$

Significancia regresión

```
> anova(modelo1)
Analysis of Variance Table

Response: OS
           Df  Sum Sq Mean Sq F value    Pr(>F)
GI          1 13694.2 13694.2  295.09 < 2.2e-16 ***
Residuals 174  8074.9    46.4
---
```

SSR

$n-2$ SSE MSE MSR $F = \frac{MSR}{MSE}$

Carencia ajuste

MSR MSE

SSR SSE

```
> lack_fit_test(modelo)
Lack of fit test - Anova Table
              Sum Sq  Df Mean Sq   F value     Pr(>F)
Regression   20790.7   1 20790.7 1300.7898 < 2.2e-16 ***
Residuals     3004.8 188    16.0
Lack of fit   2130.9  98    21.7    2.2392 6.69e-05 ***
Pure error     873.9  90     9.7
Total        23795.5 189
---
```

$F = \frac{MSR}{MSE}$

$P_{val} < \alpha$: Rechazo H_0

- A un nivel de significancia del 5%, el modelo presenta carencia ajuste

$F_{LoF} = \frac{MSLOF}{MSPE}$

$SSLOF$ $SSPE$ SSE

$\boxed{n-m}$ $\boxed{m-2}$ $\boxed{n-1}$ $\boxed{n-2}$

$MSLOF$
$MSPE$