# FINAL ENERO 2020

**1.** $\mathbb{X} = (X_1, X_2)^T$ se distribuye como:

En $\Pi_0$: $f_0(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$

En $\Pi_1$: $f_1(x,y) = \frac{1}{2e\pi} \cdot \mathbb{1}_C$, con $C \equiv$ "rectángulo centrado en el origen de lado vertical $= 2$, y lado horizontal $= e\pi$".

$P_0 = P_1$, por lo que:

$$R_1 = \{(x,y) \in \mathbb{R}^2 : f_1(x,y) \geq f_0(x,y)\}$$

nos restringimos a $C$

$$f_1(x,y) \geq f_0(x,y) \iff \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \leq \frac{1}{2e\pi} \iff$$

$$\iff e^{-\frac{1}{2}(x^2+y^2)} \leq e^{-1} \iff -\frac{1}{2}(x^2+y^2) \leq -1 \iff$$

$$\iff \boxed{x^2+y^2 \geq 2}$$ circunferencia de radio $\sqrt{2} \approx 1'414$ centrada en el $(0,0)$.

Por tanto, en $C$:

$$\{(x,y) \in C : x^2+y^2 \geq 2\} \subset R_1$$

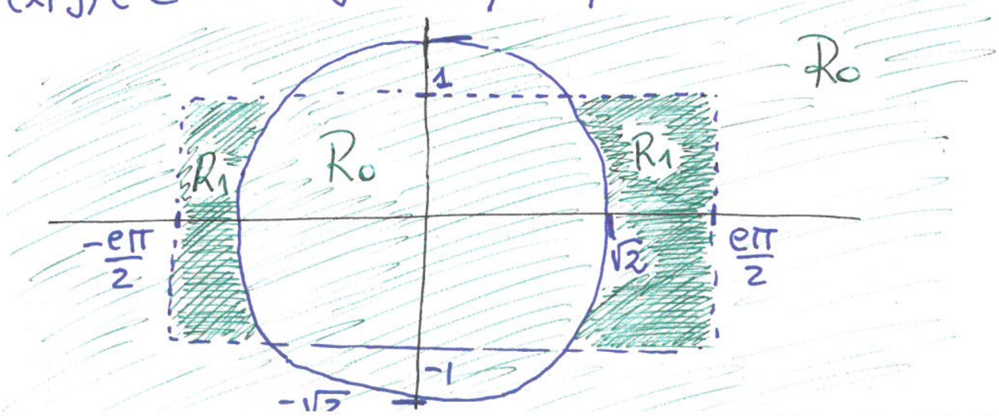
$$\{(x,y) \in C : x^2+y^2 < 2\} \subset R_0$$

Fuera de $C$: $f_1(x,y) = 0$ $\quad f_0(x,y) > 0 \quad \forall (x,y) \notin C$

Conclusión:

$$R_1 = \{(x,y) \in C : x^2+y^2 \geq 2\}$$

$$R_0 = \{(x,y) \in C : x^2+y^2 < 2\} \cup \{(x,y) \in \mathbb{R}^2 : (x,y) \notin C\}$$



2.

$n = 100$

| Observados | Probs | Esperados = n. Probs |
|---|---|---|
| 8 | $\frac{1}{4} - P_1$ | $11'\overset{\frown}{1}$ |
| 28 | $\frac{1}{4} + P_1$ | $38'\overset{\frown}{8}$ |
| 10 | $\frac{1}{4} - P_2$ | $7'8$ |
| 54 | $\frac{1}{4} + P_2$ | $42'2$ |

Antes de completar esta tabla necesitamos estimar $P_1$ y $P_2$ por máxima verosimilitud

Estimación de $P_1$ y $P_2$ por maxVERO:

$$\text{VERO}(P_1, P_2) = \left(\frac{1}{4} - P_1\right)^8 \cdot \left(\frac{1}{4} + P_1\right)^{28} \cdot \left(\frac{1}{4} - P_2\right)^{10} \cdot \left(\frac{1}{4} + P_2\right)^{54} \Longrightarrow$$

$$\Longrightarrow \log \text{VERO}(P_1, P_2) = 8 \log\left(\frac{1}{4} - P_1\right) + 28 \log\left(\frac{1}{4} + P_1\right) + 10 \log\left(\frac{1}{4} - P_2\right) + 54 \log\left(\frac{1}{4} + P_2\right)$$

$$\Longrightarrow \frac{d \log \text{VERO}(P_1, P_2)}{dP_1} = \frac{-8}{\frac{1}{4} - P_1} + \frac{28}{\frac{1}{4} + P_1} = 0 \Longleftrightarrow -2 - 8P_1 + 7 - 28P_1 = 0$$

$$\Longleftrightarrow P_1 = \frac{5}{36} = 0'13\overset{\frown}{8} \simeq \boxed{0'139}$$

$$\Longrightarrow \frac{d \log \text{VERO}(P_1, P_2)}{dP_2} = \frac{-10}{\frac{1}{4} - P_2} + \frac{54}{\frac{1}{4} + P_2} = 0 \Longleftrightarrow -2'5 - 10P_2 + 13'5 - 54P_2 = 0$$

$$\Longleftrightarrow P_2 = \frac{11}{64} \approx \boxed{0'172}$$

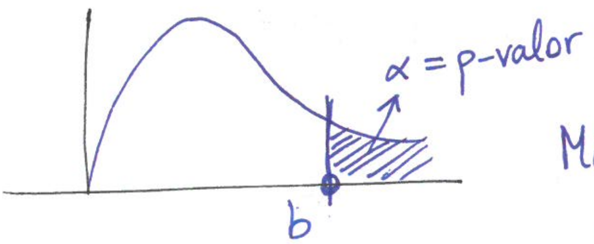
Ahora completamos la tabla

$$b = \frac{(8 - 11'1)^2}{11'1} + \frac{(28 - 38'9)^2}{38'9} + \frac{(10 - 7'8)^2}{7'8} + \frac{(54 - 42'2)^2}{42'2} =$$

$$= 0'866 + 3'054 + 0'621 + 3'3 = 7'841$$

función de distribución $\chi^2_1$

$$\alpha = 1 - \chi^2_{\underbrace{1}_{= 4 - 1 - \underbrace{2}_{\text{hemos estimado } P_1 \cdot P_2}}}(7'841)$$




Mirando en la tabla:

$\alpha \approx 0'5\%$

Existe una gran evidencia estadística en contra de Ho porque el p-valor es muy pequeño (menor que el usual $\alpha = 5\%$) $\Longrightarrow$ RECHAZO Ho

$\Longrightarrow$ el modelo teórico NO se ajusta a la muestra observada.

$\boxed{3.}$ $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1, \dots, n$

$$\varepsilon_i \sim \underbrace{f(x)}_{\text{densidad}} = \begin{cases} x+1 & \text{si } x \in [-1,0] \\ 1-x & \text{si } x \in (0,1] \end{cases}$$

$\beta_0$ es el habitual

$\hat{\beta}_0 = \bar{Y} - \frac{\bar{x}}{V_x} \operatorname{cov}_{x,Y}$

linealidad $\mathbb{E}(\cdot)$ + $X$ muestra

a) $\mathbb{E}(\hat{\beta}_0) = \mathbb{E}(\bar{Y}) - \frac{\bar{x}}{V_x} \cdot \frac{1}{n} \sum_i (x_i - \bar{x}) \mathbb{E}(Y_i - \bar{Y}) = \circledast$

$$\mathbb{E}(Y_i - \bar{Y}) = \mathbb{E}(Y_i) - \mathbb{E}(\bar{Y})$$

$$\mathbb{E}(Y_i) = \mathbb{E}(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 + \beta_1 x_i + \mathbb{E}(\varepsilon_i)$$

$$\mathbb{E}(\varepsilon_i) = \int_{-1}^{0} x(x+1)\,dx + \int_0^1 x(1-x)\,dx = \left[\frac{x^3}{3} + \frac{x^2}{2}\right]_{-1}^{0} + \left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} = 0$$

$$\Longrightarrow \mathbb{E}(Y_i) = \beta_0 + \beta_1 x_i$$

$$\mathbb{E}(\bar{Y}) = \frac{1}{n} \sum_i \mathbb{E}(Y_i) = \beta_0 + \beta_1 \bar{x}$$

$$\circledast = \beta_0 + \beta_1 \bar{x} - \frac{\bar{x}}{V_x} \cdot \frac{1}{n} \sum_i (x_i - \bar{x})(\cancel{\beta_0} + \beta_1 x_i - \cancel{\beta_0} - \beta_1 \bar{x}) =$$

$$= \beta_0 + \beta_1 \bar{x} - \frac{\bar{x}}{V_x} \cdot \frac{1}{n} \sum_i (x_i - \bar{x}) \beta_1 (x_i - \bar{x}) =$$

$$= \beta_0 + \beta_1 \bar{x} - \frac{\bar{x}}{V_x} \cdot \beta_1 \cdot \underbrace{\frac{1}{n} \sum_i (x_i - \bar{x})^2}_{= V_x} = \beta_0 + \cancel{\beta_1 \bar{x}} - \cancel{\beta_1 \bar{x}} = \beta_0$$

$$\Longrightarrow \underline{\mathbb{E}(\hat{\beta}_0) = \beta_0 \text{ insesgado}}$$

b) Se cumplen todas las hipótesis que se utilizaron en clase para hallar $V(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{n V_x}\right]$

- modelo habitual
- $\hat{\beta}_0$ estim. min. cuad. habitual
- $\varepsilon_i$ de media 0, varianza $\sigma^2$ e independientes.

Solo tenemos que sustituir $\sigma^2$ en la fórmula anterior por $V(\varepsilon_i)$:

$$V(\varepsilon_i) = \int_{-1}^{0} x^2 \cdot (x+1)\,dx + \int_0^1 x^2 (1-x)\,dx = \frac{1}{6} \Longrightarrow \boxed{V(\hat{\beta}_0) = \frac{1}{6}\left[\frac{1}{n} + \frac{\bar{x}^2}{n V_x}\right]}$$

$\boxed{4.}$ Nos ayudamos del corolario 2 del tema 1. Veamos que satisfacen todas sus hipótesis:

- $\mathbb{X} \sim N(\mu, \sigma^2 Id) \longrightarrow$ se satisface, con $\mu = \begin{pmatrix} 1 \\ -\frac{\sqrt{2}}{2} \end{pmatrix}$ y $\sigma^2 = 3$ ✓
- B simétrica e idempotente $\longrightarrow B = \begin{pmatrix} 1/3 & \sqrt{2}/3 \\ \sqrt{2}/3 & 2/3 \end{pmatrix}$ claramente simétrica e idempotentente $\iff B^2 = B$ : $\begin{pmatrix} 1/3 & \sqrt{2}/3 \\ \sqrt{2}/3 & 2/3 \end{pmatrix}\begin{pmatrix} 1/3 & \sqrt{2}/3 \\ \sqrt{2}/3 & 2/3 \end{pmatrix} = \begin{pmatrix} 1/3 & \sqrt{2}/3 \\ \frac{\sqrt{2}}{3} & 2/3 \end{pmatrix}$ ✓
- $\mu^T B \mu = 0$ : $\left(1 \quad -\frac{\sqrt{2}}{2}\right)\begin{pmatrix} 1/3 & \sqrt{2}/3 \\ \sqrt{2}/3 & 2/3 \end{pmatrix}\begin{pmatrix} 1 \\ -\frac{\sqrt{2}}{2} \end{pmatrix} = \left(1 \quad -\frac{\sqrt{2}}{2}\right)\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$ ✓

$$\overset{\text{COROLARIO 2}}{\Longrightarrow} \quad Z = \frac{1}{\sigma^2}\mathbb{X}^T B \mathbb{X} = \boxed{\frac{1}{3}\mathbb{X}^T B \mathbb{X} \sim \chi^2_1}$$ , ya que traza $(B) = 1$

$\boxed{5.}$ $Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \varepsilon_i \qquad i = 1, \dots, n$

a) $IC_{95\%}(\beta_1) = \hat{\beta}_1 \pm t_{\{n-k-1;\, \alpha/2\}} S_R \sqrt{q_{11}}$

$n = 4 \quad , \quad k = 2 \quad , \quad \alpha = 5\%$

Podemos conseguir $\hat{\beta}_1$ de $\hat{\beta} = (X^T X)^{-1} X^T \mathbb{Y}$. No necesitamos calcular todo $\beta$, lleva con la segunda coordenada:

$$\hat{\beta}_1 = \frac{1}{50}\left(2 \quad 16 \quad -17 \quad -1\right)\begin{pmatrix} 3 \\ 2 \\ 0 \\ 2 \end{pmatrix} = \frac{18}{25}$$

Podemos conseguir $q_{11}$ de la matriz $(X^T X)^{-1}$, que sería el elemento de la 2ª fila y 2ª columna: $q_{11} = \frac{11}{50} \Rightarrow \sqrt{q_{11}} = \frac{\sqrt{22}}{10}$

Para hallar $S_R$ procedemos de la siguiente forma:

$$\bar{Y} = \frac{3+2+2}{4} = \frac{7}{4} \Longrightarrow TSS = \left(3 - \frac{7}{4}\right)^2 + \left(2 - \frac{7}{4}\right)^2 + \left(2 - \frac{7}{4}\right)^2 + \left(\frac{7}{4}\right)^2 = \frac{19}{4}$$

$$R^2 = 1 - \frac{RSS}{TSS} \Longrightarrow \frac{RSS}{TSS} = 1 - R^2 \Rightarrow RSS = TSS(1 - R^2) = \frac{19}{4}(1 - 0'730526) \Rightarrow$$

$$\Rightarrow RSS = 1'28$$

$$S_R^2 = \frac{RSS}{n-k-1} = \frac{1'28}{1} = 1'28 \Longrightarrow S_R = \frac{4\sqrt{2}}{5} \approx 1'131$$

Por tanto, el intervalo de confianza sería:

$$\boxed{IC_{95\%}(\beta_1)} = \frac{18}{25} \pm \overbrace{t_{\{1;\,2'5\%\}}}^{=12'706}\, 1'131 \cdot \frac{\sqrt{22}}{10} = \boxed{\frac{18}{25} \pm 6'74}$$

b) $x_0 = (1,1)^T \longrightarrow \tilde{x}_0 = (1,1,1)^T$

$$IC_{95\%}(\text{valor de } Y|x_0) = \hat{y}_0 \pm t_{\{n-k-1;\,2'5\%\}} \cdot S_R \cdot \sqrt{1 + \tilde{x}_0^T (X^T X)^{-1} \tilde{x}_0}$$

$$IC_{95\%}(\text{media de } Y|x_0) = \hat{y}_0 \pm t_{\{n-k-1;\,2'5\%\}} \cdot S_R \cdot \sqrt{\tilde{x}_0^T (X^T X)^{-1} \tilde{x}_0}$$

$$\text{Longitud } IC_{95\%}(\text{valor } Y|x_0) = 2\, t_{\{n-k-1;\,2'5\%\}}\, S_R \sqrt{1+\tilde{x}_0^T (X^T X)^{-1} \tilde{x}_0} := L_1$$

$$\text{Longitud } IC_{95\%}(\text{media } Y|x_0) = 2\, t_{\{n-k-1;\,2'5\%\}}\, S_R \sqrt{\tilde{x}_0^T (X^T X)^{-1} \tilde{x}_0} := L_2$$

$$\Longrightarrow \frac{L_1}{L_2} = \frac{\sqrt{1+\tilde{x}_0^T (X^T X)^{-1} \tilde{x}_0}}{\sqrt{\tilde{x}_0^T (X^T X)^{-1} \tilde{x}_0}}$$

$$\tilde{x}_0^T (X^T X)^{-1} \tilde{x}_0 = (1\ 1\ 1) \cdot \frac{1}{50} \begin{pmatrix} 69 & -3 & -31 \\ -3 & 11 & -3 \\ -31 & -3 & 19 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{50} \cdot 25 = \frac{1}{2}$$

$$\Longrightarrow \boxed{\frac{L_1}{L_2} = \frac{\sqrt{1'5}}{\sqrt{0'5}} = \sqrt{3}}$$