ESTADÍSTICA II (2022/23). Grado en Matemáticas
Parcial, 25 de noviembre de 2022. SOLUCIONES

Problema 1: a) La función de distribución de una v.a. exponencial de parámetro 1 es $F(x) = 1 - e^{-x}$, si $x > 0$.

$$\begin{array}{lll} 0.25 = 1 - e^{-Q_1} & \Rightarrow & Q_1 = -\log 0.75 = 0.288 \\ 0.5 = 1 - e^{-Q_2} & \Rightarrow & Q_2 = \log 2 = 0.693 \\ 0.75 = 1 - e^{-Q_3} & \Rightarrow & Q_3 = -\log 0.25 = 1.386 \end{array}$$

Partición del espacio muestral de la exponencial, $[0,\infty)$: $A_1 = [0, 0.288)$, $A_2 = [0.288, 0.693)$, $A_3 = [0.693, 1.386)$, $A_4 = [1.386, \infty)$, con $p_j = \mathbb{P}(A_i) = 0.25$ y $e_j = 26/4 = 6.5$, $i = 1, \ldots, 4$, y frecuencias observadas $o_1 = 7$, $o_2 = 5$, $o_3 = 6$ y $o_4 = 8$.
Estadístico del contraste con hipótesis nula: $H_0 : F = F_0$, siendo F_0 una exp(1):

$$\chi^2 = \sum_{i=1}^{4} -26 = 26.769 - 26 = 0.769.$$

Región de rechazo $R = \{\chi^2 > \chi^2_{3;0.05} = 7.81\}$. No rechazamos H_0 a nivel $\alpha = 0.05$.

b) `FX <- pexp(X)`

Función de distribución F_0 de la exp(1) evaluada en los estadísticos de orden $x_{(i)}$.

```
n <- length(X)
```

$n =$ tamaño muestral.

```
v <- 1:n
w <- 0:(n-1)
FX - w/n
```

Vector de componentes $F_0(x_{(i)} - (i-1)/n$, $i = 1, \ldots, n$.

```
v/n - FX
```

Vector de componentes $i/n - F_0(x_{(i)}$, $i = 1, \ldots, n$.

```
ks.test(X,"pexp")
```

Hace el mismo contraste que en (a) con la misma muestra, pero con el procedimiento de Kolmogorov-Smirnov, cuyo estadístico del contraste es `D` $= \text{máx}(D_n^+, D_n^-) = 0.11542$, siendo

$$D_n^+ = \max_{1\le i\le n} \big(i/n - F_0(x_{(i)}\big) = 0.11542 \quad \text{y} \quad D_n^- = \max_{1\le i\le n} \big(F_0(x_{(i)} - (i-1)/n\big) = 0.06657.$$

Como el p-valor es mayor que todos los niveles de significación habituales, no hay razón para rechazar H_0, lo cual es coherente con el resultado de (a).

Problema 2: a) Modelo con 3 regresores: para $\epsilon \sim N(0,\sigma^2)$

$$\texttt{Pct.BF} = \beta_0 + \beta_1\,\texttt{Age} + \beta_2\,\texttt{Height} + \beta_3\,\texttt{Abdomen} + \epsilon.$$

Contraste global de la regresión: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ (ningún regresor influye sobre la respuesta en el modelo anterior).
Región de rechazo $R = \{F > F_{3;246;\alpha=0.01} \simeq F_{3;200;\alpha=0.01} = 3.88\}$, siendo el estadístico del contraste

$$F = \frac{0.7151}{1-0.7151}\frac{246}{3} = 205.82.$$

Rechazamos H_0 a nivel $\alpha = 0.01$, que evidentemente es mucho mayor que el p-valor. Concluimos que al menos uno de los tres regresores es significativo en el modelo, seguramente uno o los dos que son individualmente significativos, `Height` y `Abdomen`.
b) En el modelo con 3 regresores:

$$s_R^2 = 4.454^2 \Rightarrow \text{TSS} = \frac{4.454^2\,246}{1-0.7151} = 17129.44 \Rightarrow \bar{R}_3^2 = 1 - \frac{(1-0.7151)249}{246} = 0.7116$$

En el modelo con 1 regresor:

$$F = 23.65 \Rightarrow R_1^2 = \frac{23.65}{248+23.65} = 0.0871 \Rightarrow \bar{R}_1^2 = 1 - \frac{(1-0.0871)249}{248} = 0.0834.$$

Dado que $\bar{R}_3^2 >> \bar{R}_1^2$, el modelo `reg3` es preferible al `reg1`, lo cual es razonable porque ya hemos dicho que `Height` y `Abdomen` son individualmente significativos (y `Age` no lo es).
c) El contraste de hipótesis que se obtendría con `anova(reg1,reg3)` es $H_0 : \beta_2 = \beta_3 = 0$ (modelo `reg1` correcto) frente a $H_1 : \beta_2 \neq 0$ y/o $\beta_3 \neq 0$.
Región de rechazo a nivel $\alpha = 0.01$: $R = \{F > F_{2;246;\alpha=0.01} \simeq F_{2;200;\alpha=0.01} = 4.71\}$, siendo el estadístico del contraste

$$F = \frac{(7.941^2\,248 - 4.454^2\,246)/2}{4.454^2} = 271.16.$$

Por tanto, rechazamos H_0 a nivel del 1 % y evidentemente el p-valor será mucho menor que 0.01. Concluimos que el modelo `reg3` es más explicativo que el `reg1`, lo cual es coherente con lo concluido en (b).

Problema 3: a) Observemos que $X_{11} - \bar{X}_{1\bullet} = (X_{11} - X_{12})/2 = -(X_{12} - \bar{X}_{1\bullet})$. Por tanto, la distributión de $X_{11} - \bar{X}_{1\bullet}$ condicionada por $X_{12} - \bar{X}_{1\bullet}$ es una distribución degenerada (con varianza 0 y esperanza $-(X_{12} - \bar{X}_{1\bullet})$).
b) Observemos que

$$Y = (X_{11} - X_{12})^2 + (X_{21} - X_{22})^2$$

Definimos el vector aleatorio

$$\boldsymbol{\xi} = \begin{pmatrix} X_{11} - X_{12} \\ X_{21} - X_{22} \end{pmatrix} \sim N\left(\mathbf{0}_2, 2\mathbf{I}_2\right).$$

Entonces $Y = 2(\boldsymbol{\xi} - \mathbb{E}\boldsymbol{\xi})'(\mathbb{V}(\boldsymbol{\xi}))^{-1}(\boldsymbol{\xi} - \mathbb{E}\boldsymbol{\xi}) \sim 2\chi_2^2$, que es una gamma(1/4,1), porque la χ_2^2 es una gamma(1/2,1). Pero este último detalle no era necesario decirlo.