

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

Nombre: _____

**En todos los contrastes de hipótesis se deberá plantear
claramente la hipótesis nula H_0 .
Todas las respuestas deberán ser razonadas.**

Problema 1: (3.5 puntos) Consideramos la muestra (ordenada de menor a mayor)

0.03 0.03 0.04 0.05 0.08 0.22 0.28 0.35 0.35 0.44 0.60 0.65 0.73
0.74 0.87 0.92 0.97 1.27 1.41 1.52 1.71 1.96 2.30 2.48 2.97 3.63

a) Queremos saber si la muestra proviene de una v.a. X con distribución exponencial de parámetro 1, es decir, con función de densidad $f(x) = e^{-x}$ si $x > 0$. Realizar un contraste χ^2 al nivel 0.05 utilizando la partición del espacio muestral de X dada por los cuatro intervalos:

$$[0, Q_1) \quad [Q_1, Q_2) \quad [Q_2, Q_3) \quad [Q_3, \infty),$$

siendo Q_1 , Q_2 y Q_3 los tres cuartiles poblacionales.

b) Explica qué hace el siguiente código de R, completa el valor sustituido con interrogantes y extrae conclusiones.

```
X <- c(0.03,0.03,0.04,0.05,0.08,0.22,0.28,0.35,0.35,0.44,0.60,0.65,0.73,
      0.74,0.87,0.92,0.97,1.27,1.41,1.52,1.71,1.96,2.30,2.48,2.97,3.63)
FX <- pexp(X)
n <- length(X)
v <- 1:n
w <- 0:(n-1)
FX - w/n
  [1] 0.02955 -0.00891 -0.03771 -0.06661 -0.07696 0.00517 0.01345 0.02608
  [9] -0.01238 0.00981 0.06657 0.05488 0.05655 0.02289 0.04259 0.02456
 [17] 0.00553 0.06532 0.06355 0.05052 0.04990 0.05145 0.05359 0.03164
 [25] 0.02562 0.01195
v/n - FX
  [1] 0.00891 0.04737 0.07617 0.10508 0.11542 0.03329 0.02501 0.01238
  [9] 0.05084 0.02866 -0.02811 -0.01642 -0.01809 0.01558 -0.00413 0.01390
 [17] 0.03293 -0.02686 -0.02509 -0.01206 -0.01144 -0.01299 -0.01513 0.00682
 [25] 0.01284 0.02652
ks.test(X,"pexp")
data: X
D = , p-value = 0.8791
alternative hypothesis: two-sided
```

Problema 2: (3.5 puntos) Se desea predecir el porcentaje Pct.BF de grasa corporal en un varón en base a su edad (Age), a su altura (Height) y el perímetro de su abdomen. A continuación detallamos los modelos de regresión lineal ajustados, así como extractos de sus resultados (no tienen por qué estar completos).

```
reg3 <- lm(Pct.BF~Age+Height+Abdomen,data=Datos)
summary(reg3)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-6.56460	8.14938	-0.806	0.421
Age	0.03059	0.02415	1.266	0.207
Height	-0.55691	0.11532	-4.829	2.41e-06 ***
Abdomen	0.68673	0.02954	23.249	< 2e-16 ***

Residual standard error: 4.454 on 246 degrees of freedom

Multiple R-squared: 0.7151

```
reg1 <- lm(Pct.BF~Age,data=Datos)
summary(reg1)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.35029	1.85449	5.581	6.25e-08 ***
Age	0.19342	0.03977	4.863	2.06e-06 ***

Residual standard error: 7.941 on 248 degrees of freedom

F-statistic: 23.65 on 1 and 248 DF, p-value: 2.055e-06

- En el modelo con los 3 regresores, realiza el contraste global de la regresión al nivel de significación del 1% y extrae conclusiones.
- Determina los coeficientes de determinación ajustados \bar{R}^2 en los modelos con 1 y con 3 regresores. En base a los valores obtenidos explica cuál de los dos modelos te parece mejor.
- Realiza el contraste de hipótesis que se obtendría con `anova(reg1,reg3)` y extrae conclusiones.

Problema 3: (3 puntos) Sea $\mathbf{X} = (X_1, X_2, X_3, X_4)' \sim N_4(\boldsymbol{\mu}, \mathbf{I}_4)$, donde

$$\mathbf{X} = \begin{pmatrix} X_{11} \\ X_{12} \\ X_{21} \\ X_{22} \end{pmatrix} \quad \text{y} \quad \boldsymbol{\mu} = \begin{pmatrix} a_1 \\ a_1 \\ a_2 \\ a_2 \end{pmatrix}.$$

Definimos $\bar{X}_{i\bullet} = (X_{i1} + X_{i2})/2$.

- a) Determinar la distribución de $X_{11} - \bar{X}_{1\bullet}$ condicionada por $X_{12} - \bar{X}_{1\bullet}$.
- b) Determinar la distribución de $Y = 2 \sum_{i=1}^2 \sum_{j=1}^2 (X_{ij} - \bar{X}_{i\bullet})^2$.