ESTADÍSTICA I **Curso 2020/2021**
Grado en Matemáticas

Relación 5 de problemas

1. Un fabricante de equipo deportivo afirma que un nuevo sedal sintético tiene una resistencia media a la tensión de 8 kg. Se sabe que la resistencia a la tensión es una v.a. con distribución normal con desviación típica $\sigma = 0{,}5$. Se prueba una muestra de 50 sedales y se obtiene, para la muestra, una resistencia media de 7.8 kg.

(a) Si queremos encontrar evidencia estadística de que lo que dice el fabricante es falso, ¿cuál es la hipótesis nula y la alternativa del contraste que hay que llevar a cabo?
(b) Al nivel $\alpha = 0{,}05$, ¿podemos afirmar que la especificación del fabricante no es cierta?
(c) Calcula el p-valor del contraste. ¿Qué decisión se tomaría para $\alpha = 0{,}001$?
(d) Para el nivel de significación $\alpha = 0{,}05$, ¿cuál es la probabilidad de error de tipo 2 que se comete si realmente la resistencia media fuese de 7.8 kg?

2. La aterosclerosis coronaria (AC) es una de las principales causas de mortalidad en los países occidentales. Se cree que la oxidación del colesterol de baja densidad [ver, por ejemplo, Steinberg et al. (1989), *New Engl. J. Med.*, 320, 915-924] es un importante mecanismo en el desarrollo de la AC. Hay una interesante polémica científica sobre los supuestos efectos antioxidantes (y, por tanto, beneficiosos contra la AC) de las bebidas alcohólicas consumidas en cantidades moderadas. Algunos autores han puesto en duda estos efectos cardioprotectores, otros los han atribuido al alcohol en sí mismo y, por último, otros [ver, por ejemplo, Gorinstein et al., 2002, *Nutrition Research* 22, 659-666, para una visión más amplia de este tema] atribuyen la mayor parte de los efectos beneficiosos a las sustancias fenólicas que están contenidas en el vino tinto en mucha mayor medida que en otras bebidas alcohólicas. Según los defensores de esta última teoría, el vino tinto (consumido siempre en cantidades muy moderadas) sería mucho más cardiosaludable que las demás bebidas alcohólicas.

Se han determinado los valores de epicatequina (una sustancia fenólica) en 10 muestras de vino tinto, encontrando que la media muestral era 195.1 mg/l y el error típico 10.051. Los correspondientes valores para 10 muestras de cerveza fueron 65.5 mg/L y 3.4184. En vista de estos resultados ¿puede decirse que son significativamente diferentes los contenidos medios de esta sustancia en el vino tinto y en la cerveza?

Responde a la misma pregunta para el caso del ácido ferúlico en el que los resultados obtenidos fueron, para el vino tinto, 7.2 mg/l (media) y 0.4541 mg/l (error típico) (tamaño muestral = 10) y para la cerveza 6.8 mg/l (media) y 0.3571 mg/l (error típico) (tamaño muestral = 10).

3. Se realiza un experimento para comparar los incrementos en los niveles plasmáticos de insulina producidos por la ingesta de carne y de pescado. Para ello se midieron los incrementos (medidos en picomoles por litro) producidos en la concentración de insulina en la sangre de 6 voluntarios, 90 minutos después de comer un bistec de 250 gr. Dos días más tarde se realizó de nuevo el experimento con las mismas 6 personas, después de consumir un filete de pescado. En la siguiente tabla se indican los respectivos incrementos en la concentración de insulina producidos por la carne y por el pescado:

Persona:	1	2	3	4	5	6
Resultados con la carne:	109	106	111	105	110	108
Resultados con el pescado:	100	95	105	106	80	88

(a) ¿Proporcionan estos datos suficiente evidencia estadística, al nivel de significación 0.05, para afirmar que el incremento medio en la concentración de insulina producido por el consumo de carne es mayor que el producido por el consumo de pescado? Responde a la misma pregunta anterior utilizando un nivel de significación de 0.01. Indica las suposiciones previas necesarias para garantizar la validez del procedimiento empleado.

(b) Calcula el tamaño muestral que sería necesario para estimar el incremento medio producido por el consumo de carne, de manera que se tenga una probabilidad 0.95 de cometer, como máximo, un error de 0.2 unidades.

4. En un estudio sobre el consumo de productos lácteos en la población universitaria española, realizado a partir de una muestra de 380 mujeres y 120 hombres, se ha determinado que 260 mujeres y 87 hombres de la muestra consumen leche entera.

(a) Calcula, a partir de los datos anteriores, un intervalo de confianza de nivel 95 % para la proporción poblacional de hombres que consumen leche entera.

(b) ¿Permiten los datos afirmar a nivel $\alpha = 0{,}05$ que, en la población, la proporción de hombres que consumen leche entera es diferente a la de mujeres?

5. En una piscifactoría se desea contrastar la hipótesis nula de que el porcentaje de peces adultos que miden menos de 20 cm es como máximo del 10 %. Para ello, se toma una muestra de 6 peces y se rechaza H_0 si se encuentra más de uno con longitud inferior a 20 cm.

(a) ¿Cuál es el nivel de significación de este contraste?

(b) Calcula la potencia del contraste si en realidad hay un 20 % de peces que miden menos de 20 cm.

6. Se dispone de una muestra de vaiid de una población normal con desviación típica conocida $\sigma = 2$. Se desea contrastar la hipótesis nula $H_0 : \mu = 5$ frente a $H_0 : \mu = 6$ con una probabilidad de error de tipo I igual a 0,05 y una probabilidad de error de tipo II igual a 0,363. Si se utiliza el contraste uniformemente más potente que da el lema de Neyman-Pearson, ¿cuál es el tamaño muestral necesario?

7. Consideramos una muestra $X_1, \dots, X_n$ de vaiid de una población con función de densidad $f(x, \theta) = \theta(1-x)^{\theta-1}$, si $0 \le x \le 1$ (y 0 en el resto), donde $\theta > 0$. Se desea contrastar $H_0 : \theta = 2$ frente a $H_1 : \theta \neq 2$. Si $n = 60$ y $\prod_{i=1}^{60}(1 - x_i) = 0{,}0003$, ¿cuál es la decisión que hay que adoptar si se utiliza el comportamiento asintótico del contraste de razón de verosimilitudes?

8. Se desea contrastar la hipótesis nula de que una muestra aleatoria simple de tamaño n procede de una distribución uniforme en el intervalo [0,1] frente a la hipótesis alternativa de que procede de una distribución con función de densidad $f(x) = 2x$, si $0 \leq x \leq 1$.

(a) Si $n = 1$, es decir, se dispone de una única observación, calcula la región crítica del contraste uniformemente más potente de nivel 0,05.
(b) ¿Cuál es la probabilidad de error de tipo II de este contraste?
(c) Si $n = 12$ y $\sum_{i=1}^{12} \log x_i = -4{,}5$, ¿qué decisión hay que tomar de acuerdo con el contraste uniformemente más potente de nivel $\alpha = 0{,}05$?

9. Deseamos comparar el rendimiento de dos minas de carbón. Para esto, extraemos unas muestras de mineral en cada una de las minas: En la primera mina, extraemos mineral en 12 puntos diferentes (una tonelada en cada uno de ellos); el contenido medio de carbón por tonelada es de 150 Kg., con una cuasivarianza muestral de 900. En la segunda mina, extraemos mineral también en 12 puntos diferentes (una tonelada en cada uno de ellos); el contenido medio de carbón por tonelada es de 180 Kg., con una cuasivarianza muestral de 950. Asumiendo normalidad, se pide:

(a) ¿Se puede aceptar la igualdad de varianzas, al nivel de significación 0.10?
(b) ¿Se puede considerar estadísticamente probado, al nivel de significación 0.10, que el contenido medio de carbón por tonelada de mineral de la segunda mina es superior al de la primera?

10. Sean $X_1, \ldots, X_n$ vaiid de una población con función de densidad $f(x, \theta) = e^{-(x-\theta)}$, si $x \geq \theta$ (y 0 en caso contrario). Escribe la región crítica del contraste de razón de verosimilitudes de nivel α para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$.

11. Considera la v.a. X = cantidad de contaminación por mercurio (en p.p.m.) en los peces capturados en los ríos norteamericanos Lumber y Wacamaw (véase el fichero *contaminación por mercurio en peces)*. ¿Para qué niveles de significación puede afirmarse que el contenido medio de mercurio en los tejidos de los peces no es el mismo en los dos ríos?

12. Sea $X_1, \ldots, X_n$ una m.a.s. de una distribución F_θ tal que $F_\theta(x) = F(x - \theta)$, donde F es continua, estrictamente creciente y $F(0) = 1/2$ (es decir, F tiene mediana 0 y θ es la mediana de F_θ). Queremos contrastar $H_0 : \theta \leq 0$ frente a $H_1 : \theta > 0$. Para ello utilizamos el contraste definido por la región crítica $R = \{T_n > c\}$, donde $T_n = \#\{i : X_i > 0\}$ es el número de observaciones positivas en la muestra.

(a) ¿Cuál es la distribución de T_n? ¿Cuánto valen, en función de θ, $\mathrm{E}(T_n)$ y $\mathrm{Var}(T_n)$?
(b) Determina cuánto debe valer el valor crítico c para que el contraste tenga nivel de significación α aproximadamente.
(c) Supongamos que la muestra es de tamaño $n = 36$ y procede de una distribución normal de media θ y varianza 1. Calcula la función de potencia del contraste anterior si $\alpha = 0{,}05$.