ESTADÍSTICA II **Curso 2016/2017**
Grado en Matemáticas

Relación 2 de problemas

1. Calcula la distribución exacta bajo la hipótesis nula del estadístico de Kolmogorov-Smirnov para muestras de tamaño 1.

2. Se desea contrastar la hipótesis nula de que una única observación X procede de una distribución $\mathrm{N}(0,1)$. Si se utiliza para ello el contraste de Kolmogorov-Smirnov, determina para qué valores de X se rechaza la hipótesis nula a nivel $\alpha = 0{,}05$.

3. Da una demostración directa para el caso $k = 2$ de que la distribución del estadístico del contraste χ^2 de bondad de ajuste converge a una distribución χ^2_1, es decir,

$$T = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \to_d \chi^2_1, \quad \text{si} \quad n \to \infty.$$

[Indicación: Hay que demostrar que $T = X_n^2$, donde $X_n \to_d \mathrm{N}(0,1)$. Para reducir los dos sumandos a uno, utilizar la relación existente entre O_1, E_1 y O_2, E_2.]

4. El número de asesinatos cometidos en Nueva Jersey cada día de la semana durante el año 2003 se muestra en la tabla siguiente:

Día	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Frecuencia	42	51	45	36	37	65	53

(a) Contrasta a nivel $\alpha = 0{,}05$, mediante un test χ^2, la hipótesis nula de que la probabilidad de que se cometa un asesinato es la misma todos los días de la semana.
(b) ¿Podría utilizarse el test de Kolmogorov-Smirnov para contrastar la misma hipótesis? Si tu respuesta es afirmativa, explica cómo. Si es negativa, explica la razón.
(c) Contrasta la hipótesis nula de que la probabilidad de que se cometa un asesinato es la misma desde el lunes hasta el viernes, y también es la misma los dos días del fin de semana (pero no es necesariamente igual en fin de semana que de lunes a viernes).

5. Para estudiar el número de ejemplares de cierta especie en peligro de extinción que viven en un bosque, se divide el mapa del bosque en nueve zonas y se cuenta el número de ejemplares de cada zona. Se observa que 60 ejemplares viven en el bosque repartidos en las 9 zonas de la siguiente forma:

8	7	3
5	9	11
6	4	7

Mediante un contraste de hipótesis, analiza si estos datos aportan evidencia empírica de que los animales tienen tendencia a ocupar unas zonas del bosque más que otras.

6. Se ha desarrollado un modelo teórico para las diferentes clases de una variedad de moscas. Este modelo nos dice que la mosca puede ser de tipo L con probabilidad p^2, de tipo M con probabilidad q^2 y de tipo N con probabilidad $2pq$ ($p+q=1$). Para confirmar el modelo experimentalmente tomamos una muestra de 100 moscas, obteniendo 10, 50 y 40, respectivamente.
(a) Hallar la estimación de máxima verosimilitud de p con los datos obtenidos.
(b) ¿Se ajustan los datos al modelo teórico, al nivel de significación 0'05?

7. (a) Aplica el test de Kolmogorov-Smirnov, al nivel 0.05, para contrastar si la muestra (3.5, 4, 5, 5.2, 6) procede de la $U(3,8)$.
(b) Aplica el test de Kolmogorov-Smirnov, al nivel 0.05, para contrastar la hipótesis de que la muestra (0, 1.2, 3.6) procede de la distribución $\mathrm{N}(\mu=1;\sigma=5)$.

8. Se ha clasificado una muestra aleatoria de 500 hogares de acuerdo con su situación en la ciudad (Sur o Norte) y su nivel de renta (en miles de euros) con los siguientes resultados:

Renta	**Sur**	**Norte**
0 a 10	42	53
10 a 20	55	90
20 a 30	47	88
más de 30	36	89

(a) A partir de los datos anteriores, contrasta a nivel $\alpha=0{,}05$ la hipótesis nula de que en el sur los hogares se distribuyen uniformemente en los cuatro intervalos de renta considerados.
(b) A partir de los datos anteriores, ¿podemos afirmar a nivel $\alpha=0{,}05$ que la renta de los hogares es independiente de su situación en la ciudad?

9. A finales del siglo XIX el físico norteamericano Newbold descubrió que la proporción de datos que empiezan por una cifra d, $p(d)$, en listas de datos correspondientes a muchos fenómenos naturales y demográficos es aproximadamente:

$$p(d)=\log_{10}\left(\frac{d+1}{d}\right),\quad d=1,2,\ldots,9.$$

Por ejemplo, $p(1)=\log_{10}2\approx 0{,}301030$ es la frecuencia relativa de datos que empiezan por 1. A raíz de un artículo publicado en 1938 por Benford, la fórmula anterior se conoce como *ley de Benford*. El fichero `poblacion.RData` incluye un fichero llamado `poblaciones` con la población total de los municipios españoles, así como su población de hombres y de mujeres.
(a) Contrasta a nivel $\alpha=0{,}05$ la hipótesis nula de que la población total se ajusta a la ley de Benford.
(b) Repite el ejercicio pero considerando sólo los municipios de más de 1000 habitantes.
(c) Considera las poblaciones totales (de los municipios con 10 o más habitantes) y contrasta a nivel $\alpha=0{,}05$ la hipótesis nula de que el primer dígito es independiente del segundo.
(**Indicación**: Puedes utilizar, si te sirven de ayuda, las funciones del fichero `benford.R`).

10. Se ha llevado a cabo una encuesta a 100 hombres y 100 mujeres sobre su intención de voto. De las 100 mujeres, 34 quieren votar al partido A y 66 al partido B. De los 100 hombres, 50 quieren votar al partido A y 50 al partido B.
(a) Utiliza un contraste basado en la distribución χ^2 para determinar si con estos datos se puede

afirmar a nivel $\alpha = 0{,}05$ que el sexo es independiente de la intención de voto.
(b) Determina el intervalo de valores de α para los que la hipótesis de independencia se puede rechazar con el contraste del apartado anterior.

11. Sea $X_1,\ldots,X_n$ una muestra de una distribución $\mathrm{Bin}(1,p)$. Se desea contrastar $H_0 : p = p_0$. Para ello hay dos posibilidades: (a) Un contraste de proporciones basado en la región crítica $R = \{|\hat{p} - p_0| > z_{\alpha/2}\sqrt{p_0(1-p_0)/n}\}$, y (b) un contraste χ^2 de bondad de ajuste con $k = 2$ clases. ¿Cuál es la relación entre ambos contrastes?

12. En un estudio de simulación se han generado 10000 muestras aleatorias de tamaño 10 de una distribución $\mathrm{N}(0,1)$. Para cada una de ellas se ha calculado con R el estadístico de Kolmogorov-Smirnov para contrastar la hipótesis nula de que los datos proceden de una distribución normal estándar, y el correspondiente p-valor.
(a) Determina un valor x tal que la proporción de estadísticos de Kolmogorov-Smirnov mayores que x, entre los 10000 obtenidos, sea aproximadamente igual a 0.05. ¿Cuál es el valor teórico al que se debe aproximar la proporción de p-valores menores que 0.1 entre los 10000 p-valores obtenidos?
(b) ¿Cómo cambian los resultados del apartado anterior si en lugar de considerar la distribución normal estándar se considera una distribución uniforme en el intervalo $(0,1)$?