

**ESTADÍSTICA II**  
**Grado en Matemáticas (2022/23)**

**Tema 2: CONTRASTES NO PARAMÉTRICOS**

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

**2.2.** Se desea contrastar la hipótesis nula de que una única observación  $x$  procede de una distribución  $N(0, 1)$ . Si se utiliza para ello el contraste de Kolmogorov-Smirnov, determina para qué valores de  $x$  se rechaza la hipótesis nula a nivel  $\alpha = 0.05$ .

**2.3.** Da una demostración directa para el caso  $k = 2$  de que la distribución del estadístico del contraste  $\chi^2$  de bondad de ajuste converge a una distribución  $\chi_1^2$ , es decir,

$$T = \frac{(O_1 - e_1)^2}{e_1} + \frac{(O_2 - e_2)^2}{e_2} \rightarrow_d \chi_1^2, \text{ si } n \rightarrow \infty.$$

[INDICACIÓN: Hay que demostrar que  $T = X_n^2$ , donde  $X_n \rightarrow_d N(0, 1)$ . Para reducir los dos sumandos a uno, utilizar la relación existente entre  $O_1$ ,  $e_1$  y  $O_2$ ,  $e_2$ .]

**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  $\chi^2$ , 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).

**2.5.** Se tira un par de dados 500 veces. Las frecuencias de la suma de las dos caras resultantes aparecen en la siguiente tabla:

Suma	2	3	4	5	6	7	8	9	10	11	12
Frecuencia	12	26	36	58	62	83	102	33	20	9	59

¿Hay evidencia de que alguno de los dados no sea equilibrado?

**2.6.** 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?

2.7. Contrastar a nivel  $\alpha = 0.1$  si la muestra

$$0.14 \quad -1.12 \quad -0.81 \quad 1.81$$

procede de una distribución  $N(0,1)$ .

2.8. La “medicina alternativa” es toda práctica que afirma tener un efecto sanador sin justificar esta afirmación mediante pruebas científicas, por lo que su efectividad puede ser debida simplemente al efecto placebo. *The Lancet* (2005) publicó los resultados de un estudio en que se comprobaban la efectividad de dos tipos de tratamientos alternativos: música, imágenes y tacto (MIT) frente a oración terapéutica (OT), en enfermos cardíacos que se recuperaban de una intervención. Se asignó aleatoriamente a los pacientes a uno de los cuatro siguientes tipos de tratamiento: (1) OT, (2) MIT, (3) OT y MIT y (4) cuidados estándar (no OT y no MIT). Seis meses después del comienzo del estudio, se evaluó si los pacientes habían sufrido algún episodio cardiovascular grave (por ejemplo, un ataque al corazón). Los resultados (en frecuencias absolutas de pacientes) aparecen en la siguiente tabla:

Terapia	Episodio cardiovascular grave	Sin eventos reseñables	Total
OT	43	139	182
MIT	47	138	185
OT y MIT	39	150	189
Estándar	50	142	192
Total	179	569	748

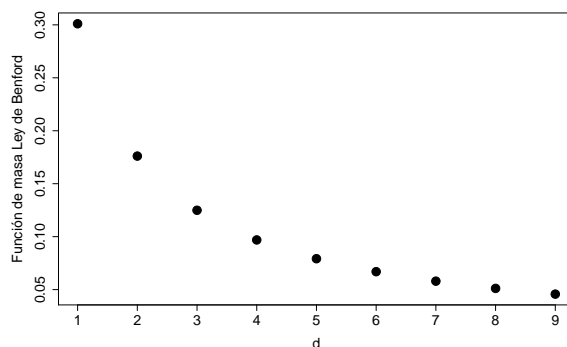
- Identifica las dos variables cualitativas (y sus posibles valores) que intervienen en el estudio.
- A nivel  $\alpha = 0.05$ , ¿hay relación entre el tipo de terapia y el hecho de que sufra o no un episodio cardiovascular de gravedad?
- Consideremos sólo los pacientes que sí sufrieron episodios cardiovasculares graves. A nivel  $\alpha = 0.01$ , ¿hay evidencia muestral de que no se asignaran uniformemente a los tipos de terapia?

2.9. 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:

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

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*. Con el siguiente código R dibujamos la función de masa de esta distribución:

```
d = 1:9 # Numeros del 1 al 9
PBenford = log10((d+1)/d) # Fn de masa Ley de Benford
plot(d,PBenford,pch=19,cex=2,ylab="Función de masa Ley de Benford",
     cex.axis=1.5,cex.lab=1.5, xaxt="n")
axis(1,at=d,labels=d,cex.axis=1.5)
```



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`.

**2.10.** 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  $\chi^2$  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  $\alpha$  para los que la hipótesis de independencia se puede rechazar con el contraste del apartado anterior.

**2.11.** Sea  $X_1, \dots, X_n$  una muestra de una distribución  $\text{Bin}(1, p)$ . 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  $\chi^2$  de bondad de ajuste con  $k = 2$  clases. ¿Cuál es la relación entre ambos contrastes?

**2.12.** En Smith (1990)<sup>1</sup> se tomó una muestra de mujeres de entre 20 y 44 años en la ciudad de Toronto. Smith defendía que dicha muestra representaba bien la distribución de la población total de mujeres de ese rango de edad en Toronto. Para determinar la veracidad de esta afirmación se ofrecen a continuación los recuentos de la muestra de Smith y las proporciones del censo municipal:

Edad	Recuento muestral	Porcentaje en censo
20-24	103	18
25-34	216	50
35-44	171	32
Total	490	100

A nivel 0.05, ¿hay alguna evidencia en contra de la afirmación de Smith?

<sup>1</sup>Smith, M.D. (1990). Sociodemographic risk factors in wife abuse: Results from a survey of Toronto women. *Canadian Journal of Sociology*, 15, 47.

**2.13.** Se toma una muestra de 300 personas, que se clasifican primero según el tipo de sangre (O, A, B, AB) y segundo por el Rh (positivo o negativo), obteniéndose la siguiente tabla de contingencia:

	O	A	B	AB
Rh positivo	82	89	54	19
Rh negativo	13	27	7	9

- Al nivel  $\alpha = 0.05$ , contrastar si los dos tipos de clasificaciones de sangre son independientes entre sí.
- Al nivel  $\alpha = 0.05$ , contrastar si la distribución del tipo de sangre (O, A, B, AB) es uniforme.
- Explicar qué hace el siguiente código R y completar los resultados borrados (sustituidos por !!)

```
Tabla0 = matrix(c(82, 89, 54, 19, 13, 27, 7, 9),nrow=2,byrow=TRUE)
chisq.test(Tabla0)
```

Pearson's Chi-squared test

```
data: Tabla0
X-squared = !!, df = !!, p-value = 0.03505
```

**2.14.** Se tiene la muestra de 25 datos (ordenados de menor a mayor por filas) de la Tabla 1.

0.01	0.06	0.08	0.09	0.11
0.16	0.22	0.23	0.29	0.30
0.35	0.38	0.40	0.41	0.42
0.48	0.57	0.66	0.71	0.75
0.78	0.79	0.82	0.88	0.90

Tabla 1

- Utiliza el estadístico de Kolmogorov-Smirnov para contrastar, a nivel  $\alpha = 0.05$ , si la población que generó los datos es uniforme en el intervalo  $[0,1]$ .
- A nivel  $\alpha = 0.05$ , determina si la población que generó los datos es uniforme en el intervalo  $[0,1]$  utilizando un contraste  $\chi^2$ . (Sugerencia: Utilizar la partición de  $[0,1]$  dada por  $[0,1/3]$ ,  $[1/3,2/3]$ ,  $[2/3,1]$ ).
- Compara los resultados de (a) y (b).