

Tema 10. Contrastes no paramétricos en dos o más poblaciones

Resumen del tema

10.1. Contraste de homogeneidad con dos o más muestras independientes

Test de Kruskal-Wallis	
contraste	H_0 : las r poblaciones son homogéneas H_1 : las r poblaciones no son homogéneas
condiciones	La variable aleatoria observada en las r poblaciones no es cualitativa nominal. Se extraen r muestras aleatorias simples (una de cada población) independientes de tamaños n_1, n_2, \dots, n_r , respectivamente.
estadístico	$H = \frac{1}{C} \left(\frac{12}{N(N+1)} \sum_{i=1}^r \frac{R_i^2}{n_i} - 3(N+1) \right), \quad \text{donde}$ $R_i = \text{suma de los rangos de la muestra } i\text{-ésima,}$ $N = \sum n_i,$ $C = \begin{cases} 1 & \text{si no hay empates en los rangos,} \\ 1 - \frac{\sum T_i}{N^3 - N} & \text{si hay empates en los rangos,} \end{cases}$ <p style="margin-left: 40px;">con $T_i = t_i^3 - t_i$,</p> <p style="margin-left: 40px;">$t_i = \text{número de empatados del grupo } i\text{-ésimo de empates.}$</p>
región crítica	Si $r = 3, 4, 5$ y n_i pequeño, rechazamos H_0 si $H \geq h$ de la tabla de los puntos críticos del test de Kruskal-Wallis.
	Si $r > 5$ ó n_i grande, rechazamos H_0 si $H \geq \chi_{r-1, 1-\alpha}^2$.

10.2. Contraste de homogeneidad con dos o más muestras apareadas

Test de Friedman	
contraste	H_0 : las r poblaciones son homogéneas H_1 : las r poblaciones no son homogéneas
condiciones	La variable aleatoria observada en las r poblaciones no es cualitativa nominal. Se extraen r muestras aleatorias simples (una de cada población) apareadas de tamaño n .
estadístico	$V = \frac{1}{C} \left(\frac{12}{nr(r+1)} \sum_{i=1}^r R_i^2 - 3n(r+1) \right), \quad \text{donde}$ $R_i = \text{suma de los rangos de la muestra } i\text{-ésima,}$ $C = \begin{cases} 1 & \text{si no hay empates en los rangos,} \\ 1 - \frac{\sum \sum T_{ij}}{(r^3 - r)n} & \text{si hay empates en los rangos,} \end{cases}$ <p>con $T_{ij} = t_{ij}^3 - t_{ij}$,</p> <p>$t_{ij}$=número de empatados del grupo j en el bloque i.</p>
región crítica	<p>Si $r = 3, 4, 5, 6$ y n pequeño, rechazamos H_0 si $V \geq v$ de la tabla de los puntos críticos del test de Friedman.</p> <p>Si $r > 6$ ó n grande, rechazamos H_0 si $V \geq \chi_{r-1, 1-\alpha}^2$.</p>

10.3. Contraste chi-cuadrado sobre homogeneidad de dos o más poblaciones

Test χ^2 de Pearson sobre homogeneidad	
contraste	H_0 : las r poblaciones son homogéneas H_1 : las r poblaciones no son homogéneas
estadístico	$\chi_{exp}^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(f_{ij} - e_{ij})^2}{e_{ij}}, \quad \text{donde}$ <p> k = número de clases, f_{ij} = frecuencia absoluta observada en la muestra i para la clase j, f_{i*} = tamaño de la muestra i, f_{*j} = número de individuos en la clase j, $e_{ij} = \frac{f_{i*} \cdot f_{*j}}{n}$ = frecuencia esperada bajo H_0. </p>
condiciones	<p>Las r muestras aleatorias son independientes.</p> <p>$e_{ij} \geq 1$ para todas las clases.</p> <p>$e_{ij} \geq 5$, salvo para un 20% de las clases como máximo.</p>
región crítica	$\chi_{exp}^2 \geq \chi_{(r-1)(k-1), 1-\alpha}^2$

Problemas propuestos

Problema 10.1. El número de libros juveniles prestados en 15 días elegidos al azar en los meses de verano (V) e invierno (I) ha sido:

V	54	61	44	50	50	54	59	54	22	58	45	30	25	29	24
I	61	46	50	17	45	31	20	54	37	38	30	42	58	44	58

¿Hay la misma demanda de libros en verano que en invierno?

Problema 10.2. Una colección de libros ha sido incluida en un índice de dos formas distintas: a) fichero de entrada simple e índice en cadena, y b) fichero de entrada múltiple e índice simple por orden alfabético. El número de entradas en los dos ficheros para una muestra aleatoria de 12 documentos ha sido:

a)	4	3	4	4	5	4	3	3	3	5	5	2
b)	4	3	6	4	6	6	4	3	4	6	6	2

¿El número de entradas por documento depende del tipo de fichero?

Problema 10.3. Se selecciona una muestra aleatoria simple de 10 bibliotecas y se observa el número de items (libros, artículos, revistas, ...) obtenidos y el número de items pedidos por el servicio de préstamo interbibliotecario de cada una de ellas en el último año. Los resultados son los siguientes:

obtenidos	920	1.274	768	608	776	874	744	484	826	2.174
pedidos	874	489	1.175	1.034	1.752	588	670	622	747	1.793

¿En toda la población de bibliotecas, el número medio de items obtenidos es igual al número medio de items pedidos?

Problema 10.4. Se eligen aleatoria e independientemente 15 alumnos del primer curso de bachillerato y 12 alumnos del segundo curso de bachillerato, y se observa el número de libros distintos que han pedido prestados en la biblioteca de su instituto durante un curso académico determinado. Los resultados son los siguientes:

1º	2	7	5	9	7	10	8	6	4	3	1	6	9	10	11
2º	10	12	3	7	9	11	7	12	14	9	8	10			

¿Son iguales las medias del número de libros que los alumnos de 1º y 2º han pedido prestados a la biblioteca del instituto durante el curso?

Problema 10.5. En un volumen de libros para jóvenes se observa que, para edades comprendidas entre 9 y 11 años, 68 libros fueron escritos por hombres y 94 por mujeres; y para edades comprendidas entre 12 y 14 años, 116 libros fueron escritos por hombres y 28 por mujeres. ¿Hay diferencia significativa entre los dos grupos de edades respecto de la variable sexo de la persona que escribe los libros?

Problema 10.6. En un experimento se encuentra que en el año 1980 el número de citas en sociología fue 330 y el número de citas en economía fue 299. En 1990, el número de citas en sociología fue 414 y en economía fue 393. ¿Hay diferencia entre los dos años investigados respecto del número de citas en sociología y economía?

Problema 10.7. Los siguientes datos corresponden al número de libros científicos y de ficción prestados a adultos residentes en dos áreas de una determinada ciudad:

	científicos	de ficción
área A	870	745
área B	304	251

¿Hay diferencia significativa entre las dos áreas respecto del tipo de libro demandado?

Problema 10.8. Los resúmenes de *Economics Abstracts* se escriben en inglés, francés y alemán. Se extraen muestras aleatorias independientes de 8 resúmenes escritos en cada uno de los tres idiomas mencionados, observando el número de palabras por resumen, siendo los resultados los siguientes:

	inglés	francés	alemán
	71	111	67
	118	113	75
	52	84	61
	47	84	99
	59	84	58
	65	94	107
	84	90	113
	111	90	95

¿La extensión de los resúmenes es la misma para los tres idiomas?

Problema 10.9. En una investigación sobre la transferencia de la información se recogieron los siguientes datos:

grupo de trabajo	en persona	por teléfono	otras
A	1.008	269	708
B	409	194	497
C	2.252	544	1.524

¿Hay diferencia entre los grupos de trabajo A, B y C en cuanto a los métodos empleados para transmitir la información?

Problema 10.10. Se pregunta a una muestra aleatoria de alumnos de 3º de una facultad de documentación, de cuatro cursos académicos distintos, si conocen los registros MARC de la *British Library*, y los resultados son los siguientes:

	No	Sí	No responde
1994-95	37	56	24
1995-96	24	44	30
1996-97	14	34	41
1997-98	28	54	15

¿Hay diferencia significativa entre los cuatro cursos académicos con respecto a la respuesta dada?

Problema 10.11. Se eligen aleatoria e independientemente 10 estantes con libros de geografía, 10 con libros de derecho, 10 con libros de matemáticas y 10 con libros de filosofía, y se cuenta el número de libros por estante. Los resultados son los siguientes:

Geografía	Derecho	Matemáticas	Filosofía
25	21	36	25
30	21	32	27
30	33	30	26
29	23	30	26
25	16	32	21
23	26	33	28
28	26	33	30
33	28	28	31
25	26	39	28
25	21	43	32

¿El número medio de libros por estante es igual para las cuatro materias?

Problema 10.12. En una muestra aleatoria simple de 12 días se observa el número de libros prestados en diferentes materias (científicos, novelas, ensayos, arte, música) siendo los resultados los siguientes:

científicos	novelas	ensayos	arte	música
24	40	19	23	21
29	39	17	15	17
33	45	15	13	20
30	38	10	19	16
36	33	12	17	14
27	30	15	20	12
24	25	20	21	11
19	38	23	9	8
16	27	25	23	21
35	39	14	17	14
37	41	21	19	12
32	47	11	14	17

¿El número medio de libros prestados diariamente es igual en las cinco materias?

Problema 10.13. En una investigación sobre el uso que los profesores de distintos departamentos hacen de las revistas científicas, se encontró que 34 de los 50 profesores del departamento A, 22 de los 40 profesores del departamento B y 15 de los 35 profesores del departamento C, utilizan las revistas como ayuda en su trabajo académico (y el resto no). ¿Hay diferencia significativa entre los tres departamentos respecto del uso que hacen de las revistas científicas?