

Tarea-Examen, versión B. Pruebas no paramétricas.

Gonzalo Pérez, Sheyla Barradas y Luis Genaro Coria

Semestre 2024-2

La tarea-examen se deberá subir al classroom antes de las 11:59 PM del 4 de junio de 2024. La pregunta 1 vale 2 puntos, mientras que el resto de preguntas valen 1.4 puntos cada una.

Favor de argumentar con detalle las respuestas. Cuando no se especifique, considere una significancia de $\alpha = .05$ y que no es necesario verificar los supuestos del modelo o prueba.

NOTA. En caso de que se identifiquen respuestas iguales en otras tareas-exámenes, se procederá a la anulación de las tareas-exámenes involucradas. Incluir el(los) nombre(s) completo(s) de la(s) persona(s) que está(n) resolviendo los ejercicios. Equipos de máximo cuatro integrantes.

1.

Los datos de la siguiente tabla muestran las mediciones de “low density lipid (LDL) cholesterol”, también conocido como colesterol malo por las consecuencias asociadas a tener altos niveles de este colesterol. Cuatro tratamientos (treat:1,2,3,4) se probaron con la finalidad de reducir los niveles de LDL, los datos de 39 observaciones aleatorias se muestran en la tabla siguiente (datos “quail” en paquete “Rfit”).

Cuadro 1: Low density lipid (LDL) cholesterol para 39 observaciones independientes

<i>n=39</i>							
treat	ldl	treat	ldl	treat	ldl	treat	ldl
1	52	2	36	3	52	4	62
1	67	2	34	3	55	4	71
1	54	2	47	3	66	4	41
1	69	2	125	3	50	4	118
1	116	2	30	3	58	4	48
1	79	2	31	3	176	4	82
1	68	2	30	3	91	4	65
1	47	2	59	3	66	4	72
1	120	2	33	3	61	4	49
1	73	2	98	3	63		

- Presente el boxplot de la variable LDL para cada tratamiento (grupo) y comente.
- Indique si se puede asumir que los datos en cada grupo provienen de una distribución Normal.
- Dependiendo de la respuesta anterior, realice una prueba adecuada para indicar si es plausible asumir que la varianza es similar en los cuatro grupos.
- ¿Los cuatro tratamientos proporcionan los mismos valores de LDL? Realice la prueba adecuada con $\alpha = .1$.
- Si la respuesta a la pregunta del inciso iv) es negativa, indique si el tratamiento 2 reduce más los niveles de colesterol en comparación con el resto de tratamientos. Use $\alpha = .1$.

2.

Los datos de la tabla de abajo muestran los resultados en la escala de depresión “Hamilton Depression Scale Factor IV” para nueve pacientes con depresión, antes de recibir un tratamiento (x) y después de recibir el tratamiento (y). A menor valor en la escala se observa una mejoría. En este sentido ¿se puede concluir que el tratamiento ha tenido éxito? Realice.

- I. La prueba paramétrica asociada.
- II. La prueba no paramétrica asociada.

Paciente	1	2	3	4	5	6	7	8	9
x	1.53	1.68	1.88	1.55	3.06	1.3	0.5	1.62	2.48
y	0.578	1.06	1.29	1.06	3.14	1.29	0.647	0.598	2.05

3.

Considere los datos de la tabla de abajo.

x	3.9	7.9	4.1	8.8	9.4	0.46	5.3	8.92	5.5	4.6	47.9	23.2	34.2	29.1	6.0	45.1	13.1	3.1	17.1	47.8
y	2.3	2.4	1.8	2.3	1.7	0.10	2.2	0.22	2.4	1.0	2.8	2.9	2.5	2.6	1.2	2.1	3.4	1.3	1.7	1.6

Calcule las medidas de asociación

- I. Coeficiente de correlación de Pearson.
- II. Coeficiente tau b de Kendall, τ_b .
- III. Coeficiente rho de Spearman, r_s .

Además, presente la gráfica de dispersión de las variables x y y . ¿Se puede indicar que las variables son independientes? Argumente.

Nota. El coeficiente de correlación de *Pearson* se interpreta bajo el **supuesto de normalidad bivariada** en términos de la asociación lineal entre las dos variables; en ocasiones su medición se ve afectada de forma considerable ante presencia de observaciones extremas. Por otro lado, los coeficientes *Kendall's tau* y *Spearman's rho* se interpretan en términos de **asociaciones monótonas** entre dos variables y son más robustos ante la presencia de observaciones extremas, además de que se pueden usar para variables **continuas u ordinales**.

4.

Los datos en el archivo *Ejercicio4Ex3.csv* corresponden a una encuesta aleatoria levantada a 1000 personas. En esta encuesta se les preguntó el nivel de escolaridad (NivEdu) y además se analizó el impacto de un conjunto de Fake News actuales (FakeNews), clasificando éste en cuatro niveles: Muy Poco, Poco, Regular y Mucho. ¿Se puede decir que a mayor nivel educativo hay menor impacto por las Fake News?

Los primeros datos del archivo son los siguientes.

```
##           NivEdu FakeNews
## 1  Secundaria Muy Poco
## 2  Secundaria Muy Poco
## 3  Profesional Muy Poco
## 4  Bachillerato Poco
## 5  Bachillerato Muy Poco
## 6  Profesional Regular
```

5.

La sustancia Sulphinpyrazone fue usada para la reducción de muerte cardíaca después de un infarto. Para saber si realmente funcionaba, investigadores recolectaron información de 1475 pacientes que sufrieron un paro cardíaco. A cada paciente, de forma aleatoria, se le recetó tomar tabletas con Sulphinpyrazone o bien un placebo por un período de dos años (733 lo recibieron y 742 no). Durante este tiempo se tomó registro de los pacientes que habían sobrevivido y de los que habían muerto. El número de pacientes (Frec) de acuerdo a si tomó o no el medicamento y su condición después del paro se presenta en la siguiente tabla.

Frec	Tratamiento	Vivo
692	Sulphinpyrazone	Sí
41	Sulphinpyrazone	No
682	Placebo	Sí
60	Placebo	No

Realice una prueba de hipótesis para indicar si la condición de muerte después de un paro cardíaco es diferente de acuerdo a si se recibió o no el tratamiento con Sulphinpyrazone. Considere $\alpha = .01$.

6.

Se reportan 40 números aleatorios presentados en orden creciente.

0.0023, 0.0150, 0.0298, 0.0337, 0.0729, 0.0943, 0.0950, 0.1080, 0.1180, 0.1300, 0.1500, 0.1592, 0.1617, 0.2016, 0.2083, 0.2316, 0.2403, 0.2863, 0.3427, 0.3766, 0.4384, 0.4715, 0.4895, 0.5544, 0.5575, 0.5910, 0.5960, 0.6224, 0.6517, 0.6602, 0.7197, 0.7317, 0.7687, 0.8212, 0.9439, 1.1242, 1.2681, 1.2885, 2.3626, 2.6055

Suponga que se desea contrastar las hipótesis:

H_0 : los datos provienen de la distribución con función de densidad $f(x) = 2e^{-2x}$, $x > 0$

vs

H_a : los datos no provienen de esa distribución.

Usando la prueba de bondad de ajuste ji-cuadrada, con $\alpha = .05$, realice la prueba de hipótesis usando $k = 4$, donde las clases están determinadas por: $(0, 0.3]$, $(0.3, 0.7]$, $(0.7, 1.1]$, $(1.1, \infty)$.

7.

Con los datos del ejercicio 6 y usando la prueba adecuada tipo Kolmogorov–Smirnov realice el siguiente contraste de hipótesis:

H_0 : los datos provienen de una distribución $Exp(\lambda)$

vs

H_a : los datos no provienen de una distribución $Exp(\lambda)$