

# Examen 3, versión A. Pruebas no paramétricas.

Gonzalo Pérez, Leonardo de la Cruz, José Ángel Román y David González

Semestre 2023-1

El examen se deberá subir al classroom antes de las 11:59 PM del 14 de diciembre de 2022. 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 otros exámenes, se procederá a la anulación de los exámenes involucrados. Incluir el(los) nombre(s) completo(s) de la(s) persona(s) que está(n) resolviendo los ejercicios. Equipos de máximo tres 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  $\alpha = .1$ .

- v. Si la respuesta a la pregunta del inciso iv) es negativa, indique si hay algún grupo que reduzca 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	1.83	3.06	1.3	0.5	1.62	2.48	1.68	1.88	1.55
y	0.878	3.14	1.29	0.647	0.598	2.05	1.06	1.29	1.06

## 3.

Considere los datos de la tabla de abajo. Calcule las medidas de asociación

x	1.1	6.2	6.09	6.2	8.6	6.4	0.09	2.3	6.66	5.1	41.9	33	17.7	55.5	18.25	50.40	17.9	16.74	12.0	14.70
y	-1.3	-1.2	-0.52	2.0	5.0	1.5	-3.89	-1.9	-0.62	8.9	4.1	2	1.6	5.4	0.82	-0.42	4.6	-0.25	2.4	-0.12

- I. Coeficiente de correlación de Pearson.
- II. Coeficiente tau b de Kendall,  $\tau_b$ .
- 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 tabla de abajo muestra el número de individuos (Frec) de acuerdo con su comportamiento después de ser liberados de la cárcel: adecuado y no adecuado, y su raza o grupo étnico: Asiático, Negro, Hispano y Blanco.

Frec	GrupoEt	Comportamiento
6	Asiático	Adecuado
14	Negro	Adecuado
29	Hispano	Adecuado
12	Blanco	Adecuado
4	Asiático	No Adecuado
7	Negro	No Adecuado
18	Hispano	No Adecuado
10	Blanco	No Adecuado

Pruebe si el comportamiento de los individuos al salir de la cárcel depende o no de los diferentes grupos étnicos. Use  $\alpha = .01$ .

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 una distribución  $Exp(\lambda)$

vs

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

Usando la prueba de bondad de ajuste ji-cuadrada, con  $\alpha = .05$ , pruebe la hipótesis nula usando  $k = 4$ , donde las clases están determinadas por:  $(0, 0.3]$ ,  $(0.3, 0.6]$ ,  $(0.6, 1.0]$ ,  $(1.0, \infty)$ .

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

vs

$H_a$  : los datos no provienen de esa distribución.