

Ejercicios resueltos y propuestos sobre tests no paramétricos

(Unidad 3)

Ejercicio 1 *Los siguientes datos son las edades de una muestra de personas seleccionadas entre los visitantes de un Bingo.*

32, 23, 64, 31, 74, 44, 61, 33, 66, 73,
27, 65, 40, 54, 23, 43, 58, 87, 58, 62,
68, 89, 93, 24, 73, 42, 33, 63, 36, 48,
77, 75, 37, 59, 70, 61, 43, 68, 54, 29,
48, 81, 57, 97, 35, 58, 56, 58, 57, 45

Realiza un test Chi-cuadrado de bondad de ajuste para decidir si puede aceptarse que las edades sigan una distribución normal.

Ordenamos los datos de menor a mayor y realizamos una tabla de frecuencias con 4 clases.

Clase	[20,40)	[40,60)	[60,80)	[80,100)	Total
Frecuencia	12	18	15	5	50

Tenemos que hallar una estimación para la media y la desviación típica. Usamos en esta ocasión la media y la desviación típica de la muestra como estimadores. Para realizar los cálculos, y con el propósito de simplificarlos se han empleado la tabla de datos agrupados en lugar de los datos primitivos, resultando:

$$\hat{\mu} = \bar{x} = 55.2, \hat{\sigma} = S = 18.7$$

Calculamos ahora la probabilidad para cada clase usando la distribución $N(55.2, 18.7)$

La probabilidad que correspondería a las distintas clases si se cumple la hipótesis nula de que los datos siguen una distribución $N(55.2, 18.7)$ es:

$$P(x \leq 40) = \text{NormalDist}(40; 55.2, 18.7) = 0.20816$$

$$P(40 < x \leq 60) = \text{NormalDist}(60; 55.2, 18.7) - \text{NormalDist}(40; 55.2, 18.7) = 0.60129 - 0.20816 = 0.39313$$

$$P(60 < x \leq 80) = \text{NormalDist}(80; 55.2, 18.7) - \text{NormalDist}(60; 55.2, 18.7) = 0.90761 - 0.60129 = 0.30632$$

$$P(80 < x) = 1 - \text{NormalDist}(80; 55.2, 18.7) = 9.2386 \times 10^{-2}$$

Multiplicamos por el número total de datos estas probabilidades para obtener la frecuencia esperada, np_i :

Clase	$(-\infty, 40]$	$(40, 60]$	$(60, 80]$	$(80, -\infty]$	Total
n_i	12	18	15	5	50
p_i	0.21	0.39	0.31	0.09	1
np_i	10.5	19.66	15.32	4.5	50

El valor experimental de Chi es:

$$\chi^2 = \frac{(12-10.5)^2}{10.5} + \frac{(18-19.66)^2}{19.66} + \frac{(15-15.32)^2}{15.32} + \frac{(5-4.5)^2}{4.5} = 0.41669$$

Hallando el valor crítico que corresponde a $\chi^2_{4-2-1, 0.95} = \text{ChiSquareInv}(0.95; 1) = 3.84$, resulta que el intervalo de aceptación es $(0, 3.84)$. Como el valor experimental, 0.41669, pertenece a este intervalo se decide aceptar que los datos siguen una distribución $N(55.2, 18.7)$.

Ejercicio 2 *Se han seleccionado aleatoriamente una muestra de 82 estudiantes de Instituto y otra con 46 estudiantes de centros privados y se ha considerado la nota en Educación Física para cada uno de ellos. Los datos obtenidos vienen resumidos en la siguiente tabla de contingencia*

	<i>Insuf</i>	<i>suf o bien</i>	<i>notable</i>	<i>sobresaliente</i>	<i>Total</i>
<i>centro privado</i>	6	14	17	9	46
<i>Instituto</i>	30	32	17	3	82
	36	46	34	12	128

Se desea contrastar la hipótesis de que la distribución de notas en Educación física es independiente del tipo de centro de Enseñanza.

Para aplicar el test Chi-cuadrado aplicable a tablas de contingencia hallamos, en primer lugar, las frecuencias esperadas caso de cumplirse la hipótesis nula. Se calculan de la forma siguiente:

	Insuf	suf o bien	notable	sobresaliente
c. priv	$128 \cdot \frac{46}{128} \cdot \frac{36}{128}$	$128 \cdot \frac{46}{128} \cdot \frac{46}{128}$	$128 \cdot \frac{46}{128} \cdot \frac{34}{128}$	$128 \cdot \frac{46}{128} \cdot \frac{12}{128}$
Inst.	$128 \cdot \frac{82}{128} \cdot \frac{36}{128}$	$128 \cdot \frac{82}{128} \cdot \frac{46}{128}$	$128 \cdot \frac{82}{128} \cdot \frac{34}{128}$	$128 \cdot \frac{82}{128} \cdot \frac{12}{128}$

Tomando los valores:

	Insuf	suf o bien	notable	sobresaliente	Total
centro privado	12.9	16.5	12.2	4.3	46
Instituto	23.1	29.5	21.8	7.7	82
	36	46	34	12	128

Evaluamos el valor experimental del estadístico de contraste:

$$\chi^2 = \frac{(6-12.9)^2}{12.9} + \frac{(14-16.5)^2}{16.5} + \dots + \frac{(17-21.8)^2}{21.8} + \frac{(3-7.7)^2}{7.7} = 17.3$$

Comparando este valor con el valor crítico de este test que resulta ser $\chi^2_{(4-1)(2-1),0.95} = 7.815$

En este caso el valor experimental queda fuera del intervalo de aceptación, por tanto se rechaza la hipótesis nula y se concluye que el tipo de centro parece tener alguna influencia en la distribución de las notas de Educación Física.

Ejercicio 3 Para comprobar si la velocidad con la que dos ordenadores ejecutan los programas es similar se han seleccionado 9 programas de análoga dificultad, que se han ejecutado en estos ordenadores. Se ha seleccionado aleatoriamente el programa que se va a ejecutar en cada ordenador. Los tiempos de ejecución de estos programas según el ordenador en que se ha ejecutado viene recogido en la siguiente tabla:

Ordenador	B	A	A	B	A	B	B	A	B
Tiempo	8.52	8.00	8.61	8.67	8.23	8.69	8.68	8.12	8.82

¿Se pueden concluir que ambos ordenadores tardan el mismo tiempo en la ejecución de los programas.?

Para tomar una decisión vamos a emplear el test de Wilcoxon para contrastar dos muestras independientes.

Ordenador	B	A	A	B	A	B	B	A	B
Tiempo	8.52	8.00	8.61	8.67	8.23	8.69	8.68	8.12	8.82
Rangos	4	1	5	6	3	8	7	2	9

Considerando la suma de los rangos de la muestra de menor tamaño, $1+5+3+2 = 11$, y consultando en la tabla adecuada resulta que los intervalos de aceptación para la hipótesis nula son (12, 28) al 90% de confianza y (11, 29) al 95% de confianza. Por los tanto en ambos casos se rechaza la hipótesis nula y se concluye que los ordenadores considerados no ejecutan los programas con la misma velocidad.

Ejercicio 4 Considerando los datos siguientes, que por conveniencia se presentan ordenados:

23.8-26.0-26.9-27.4-28.0-30.3-30.7-31.2-31.3-32.8-33.2-33.9-34.3-34.9-35.0
contrastar, al 95% de confianza, la hipótesis de que el valor de su mediana es 30 contra la alternativa de que no vale 30.

En primer lugar restamos a estos valores el valor propuesto para la mediana, resultando:

-6.2, -4.0, -3.1, -2.6, -2.0, 0.3, 0.7, 1.2, 1.3, 2.8, 3.2, 3.9, 4.3, 4.9, 5.0

Ordenandolos por orden de sus valores absolutos resulta:

0.3, 0.7, 1.2, 1.3, -2.0, -2.6, 2.8 -3.1, 3.2, 3.9, -4.0, 4.3, 4.9, 5.0, -6.2

Asignando ahora los rangos resultan

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

Los rangos de los valores negativos son 5, 6, 8, 11, 15. La suma de estos valores, que es el estadístico de contraste resulta:

$$6 + 7 + 8 + 11 + 15 = 47$$

Consultando la tabla adecuada, resulta como región de aceptación (25, 95), que contiene el valor 47, así que no rechaza el valor 30 para la mediana.

Ejercicios propuestos

Ejercicio 5 *Comprobar si los datos siguientes, obtenidos con el generador de números aleatorios de SWP, apoyan la hipótesis de que el generador está bien programado (generan valores de una distribución $U(0,1)$). Emplead cinco clases de igual amplitud.*

0.437 88, 0.773 48, 0.494 65, 0.319 8, 0.169 8, 0.672 27, 0.809 02, 0.515 59,

0.042 26, 0.182 82, 0.348 55, 0.307 36, 0.783 92, 0.323 71, 0.315 60, $3.071 8 \times 10^{-2}$,

0.690 65, 0.607 36, 0.398 60, $2.168 1 \times 10^{-2}$, 0.890 32, 0.983 29, 0.139 6, 0.226 39, 0.660 60, 0.742 93, 0.873 2, 0.519, 0.426 76, 0.734 85, 0.966 90, 0.771 51, 0.808 92, 0.798 18, 0.123 27, 0.556 45, 0.930 97, 0.254 07, 0.615 81, 0.296 01, 0.348 44, $5.374 1 \times 10^{-2}$, 0.641 75, 0.520 51, 0.625 50, 0.675 22, 0.690 63, 0.353 67, 0.109 98, 0.636 73

Ejercicio 6 *Los Estudiantes de una Escuela de Ingeniería proceden de Bachillerato o de Formación Profesional. Se desea saber si este hecho tiene influencia en que el estudiante abandone sus estudios antes de 7 años de cursar la carrera. Con este objeto se ha realizado una encuesta sobre una muestra de ambos grupos de alumnos, obteniéndose los siguientes resultados:*

	<i>Si abandonarán</i>	<i>No abandonaron</i>
<i>Bachillerato</i>	<i>24</i>	<i>50</i>
<i>Form. profesional</i>	<i>6</i>	<i>10</i>

¿Qué conclusión debe adoptarse al 95% de Confianza?

Ejercicio 7 *Se desea comparar la eficacia de dos equipos de trabajo que realizan su labor en dos plantas diferentes. Para ello se ha considerado el*

número medio de artículos por hora que se termina en cada planta en ciertos días elegidos aleatoriamente. Las pruebas se han llevado a cabo durante 10 días, revisando cada día el trabajo terminado por uno sólo de los grupos. Los resultados obtenidos en cada planta han sido.

Planta 1	7.3, 6.9, 7.2, 7.8, 7.2
Planta 2	7.4, 6.8, 6.9, 6.7, 7.1

¿Puede aceptarse que la eficacia de los equipos es similar al 95% de confianza?

Ejercicio 8 Utilizar el test de las rachas para decidir, al 95% de confianza si los valores siguientes pueden considerarse generados aleatoriamente, por no contener pautas ni tendencias: 3, 3, 6, 0, 4, 8, 4, 5, 4, 4

