ESTADÍSTICA NO PARAMÉTRICA

Introducción a la unidad

En la unidad 4 correspondiente a “Pruebas de Hipótesis”, se estudiaron pruebas tanto para las medias poblacionales como para las proporciones poblacionales. En algunos casos el tamaño de la muestra era mayor que 30, mientras que en otras la muestra era pequeña.

Sin embargo, todas estas situaciones de pruebas presentaron una característica común: necesitaban de ciertos supuestos respecto a la población. Por ejemplo, las pruebas “t” y las pruebas “F” requerían el supuesto de que la población estuviese distribuida normalmente. Debido a que tales pruebas dependen de postulados sobre la población y sus parámetros, se denominan **pruebas paramétricas.**

En la práctica, surgen muchas situaciones en las cuales simplemente no es posible hacer de forma segura ningún supuesto sobre el valor de un parámetro o sobre la forma de la distribución poblacional, por lo que la mayoría de las pruebas descritas en los capítulos anteriores no son aplicables. Más bien se deben utilizar otras pruebas que no dependan de un solo tipo de distribución o de valores de parámetros específicos; estas pruebas se denominan pruebas no paramétricas (o libres de distribución).

Objetivo particular de la unidad

Analizar los fundamentos de la estadística no paramétrica, su importancia, desarrollo y evolución, así como su aplicación en las áreas económico-administrativas.

Lo que sé

Elige la respuesta correcta a la siguiente pregunta:

1. La fórmula del estadístico “z” es:

❍ a) $z = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n$

❍ b) $z = \frac{x-\mu}{\sigma}$

❍ c) $z = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}$

Temas de la unidad V

1. Características de las pruebas no paramétricas
2. Pruebas de bondad de ajuste
3. Tablas de contingencia
4. Prueba de los signos de Wilcoxon
5. Prueba de rachas
6. Otras pruebas

Resumen de la unidad

En esta unidad estudiamos de los métodos estadísticos conocidos como pruebas no parámetricas. Todos sabemos que los procedimientos paramétricos operan bajo algunos supuestos con respecto a la distribución de la población de la cual se obtiene la muestra para trabajar. Las pruebas no paramétricas no necesitan de éstos supuestos para su operación, y esto las convierte en procedimientos estadísticos de gran aplicación, además estas pruebas no paramétricas se pueden aplicar a datos nominales y ordinales.

La prueba de los signos es uno de estos procedimientos no paramétricos que permite identificar posibles diferencias entre dos poblaciones cuando se cuenta únicamente con datos nominales. Aplicando cuando la muestra es pequeña la distribución de probabilidad binomial para determinar los valores críticos de la prueba de los signos. Y cuando la muestra es grande se puede usar la distribución normal como aproximación.

La prueba de los signos de Wilcoxon es una prueba no paramétrica que analiza datos pareados de muestras cuando se dispone de datos con escala de intervalo o de razón para cada una de las parejas formadas; es claro que no requiere de supuestos acerca de la distribución de probabilidad de la población de la que se obtienen las muestras y en términos generales, la prueba de Wilcoxon maneja la hipótesis de que las dos poblaciones son idénticas.

La prueba de Mann-Whitney-Wilcoxon comprueba si hay una diferencia entre dos poblaciones, para lo cual se basa en dos muestras aleatorias independientes. La prueba de Kruskal-Wallis amplia la de Mann-Whitney-Wilcoxon al caso de tres poblaciones o más.

Tema 1. Características de las pruebas no paramétricas

Objetivo del tema

Distinguir las características de la pruebas no paramétricas y su utilidad en las áreas económico administrativas.

Desarrollo

Las pruebas no paramétricas son útiles sobre todo cuando no se conoce la distribución del cual provienen los datos y, por tanto, no se conoce la distribución del estadístico para hacer una estimación por intervalos de confianza o una prueba de hipótesis. Estas pruebas son útiles por ejemplo cuando el tipo de datos es nominal u ordinal.

Generalmente son más fáciles de realizar y comprender ya que no requieren cálculos laboriosos ni el ordenamiento o clasificación formal de datos o mediciones más exactas de parámetros poblacionales.

ACTIVIDAD 1

Con los autores Berenson, Levin y Mason, has un cuadro comparativo de las característica que hacen útiles las pruebas no paramétricas en las áreas económico administrativas.

	CARACTERISTICAS
BERENSON	
LEVIN	
MASON	

Descarga el siguiente cuadro para completarlo, una vez que lo tengas listo presiona el botón **Examinar**. Localiza el archivo, ya seleccionado, presiona **Subir este archivo** para guardarlo en la plataforma.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Tema 2. Pruebas de bondad de ajuste

Objetivos del tema

Aplicar las pruebas de bondad de ajuste en las áreas económico administrativas.

Desarrollo

Pruebas de bondad de ajuste. Medidas sobre qué tan cerca se ajustan los datos muestrales observados a una forma de distribución particular planteada como hipótesis.

Con frecuencia, las decisiones en los negocios requieren que se pruebe alguna hipótesis sobre la distribución poblacional desconocida. Por ejemplo, se puede plantear la hipótesis que la distribución poblacional es uniforme y que todos los valores posibles tienen la misma probabilidad de ocurrir.

Las hipótesis que se probarían son las siguientes:

H0: La distribución poblacional es uniforme.
H1: La distribución poblacional no es uniforme.

La prueba de bondad de ajuste se utiliza entonces para determinar si la distribución de los valores en la población se ajusta a una forma en particular planteada como hipótesis —en este caso, una distribución uniforme. De la misma manera que con todas las pruebas estadísticas de esta naturaleza, los datos muestrales se toman de la población y éstos constituyen la base de los hallazgos.

Si existe gran diferencia entre lo que realmente se observa en la muestra y lo que se esperaría observar si la hipótesis nula fuera correcta, es menos probable que la hipótesis nula sea verdadera. Es decir, la hipótesis nula debe rechazarse cuando las observaciones obtenidas en la muestra tienen diferencias significativas del patrón que se espera que ocurra en la distribución planteada como hipótesis.

Por ejemplo, si se hace rodar un dado “bueno”, es razonable plantear como hipótesis un patrón de resultados tal que cada resultado (números del 1 al 6) ocurra aproximadamente un sexto de las veces. Sin embargo, si un porcentaje significativamente grande o pequeño de número pares ocurre, puede concluirse que el dado no está balanceado adecuadamente y que la hipótesis es falsa. Es decir, si la diferencia entre los patrones de eventos que en realidad se observaron y el patrón de eventos que se espera que ocurra si la hipótesis nula es correcta, prueba ser demasiado grande como para atribuirlo a un error de muestreo debe concluirse entonces que la población presenta una distribución distinta de la especificada en la hipótesis nula.

Para **contrastar la hipótesis relativa** a una distribución poblacional, se debe analizar la diferencia entre las expectativas con base en la distribución planteada como hipótesis y los datos reales que aparecen en la muestra.

Para lo anterior, se utiliza la distribución χ^2 (Chi-cuadrada) como prueba estadística de bondad de ajuste y se utiliza alguna de las siguientes fórmulas:

$$\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} \quad \text{o} \quad \chi_e^2 = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n$$

En donde:

χ_e^2 Es el estadístico de prueba.

f_o Es la frecuencia de los eventos observados en los datos muestrales

f_e Es la frecuencia de los eventos esperados si la hipótesis nula es correcta

k Es el número de categorías o clases.

n Es el número de datos.

La prueba tiene K-m-1 grados de libertad, en donde “m” es el número de parámetros por estimar.

Por ejemplo si se desconoce la media o varianza de la población y se tienen que “estimar” cada uno representa un grado menos de libertad.

En la fórmula podemos observar que el numerador mide la diferencia entre las frecuencias de los eventos observados y las frecuencias de los eventos esperados.

Tipo de pruebas no paramétricas

Paso 1. Establecer la hipótesis nula (H_o) y la hipótesis alternativa (H_1).

La H_o indica que no hay diferencias significativas entre las frecuencias observadas y las frecuencias esperadas. Cualquier diferencia puede atribuirse al muestreo o a la casualidad. La H_i indica por lo tanto que si hay diferencias significativas entre una distribución esperada y la estimada para la población.

Paso 2. Elegir un nivel de significación (α).

Paso 3. Elegir y calcular el estadístico de prueba χ_e^2

Paso 4. Establecer la regla de decisión.

Paso 5. Calcular el valor de Chi-cuadrada crítica (χ_c^2) y tomar la decisión.

Tipos de pruebas de bondad de ajuste

1. Prueba para un ajuste uniforme

Se pretende probar que la distribución de datos es uniforme.

Ejemplo de aplicación; un nuevo director de mercadotecnia tiene la responsabilidad de controlar el nivel de existencias para 4 tipos (A, B, C, D) de automóviles vendidos por su empresa de distribución. Le han informado que la demanda de cada tipo de automóviles es la misma. Para probar esta hipótesis se selecciona una muestra aleatoria de 100 automóviles vendidos en los últimos meses. Se requiere un nivel de significación del 10%.

Se cuenta con la siguiente información:

Tipo Automóvil	Ventas Observadas
A	32
B	21
C	19
D	28

Solución:

1. H_o = La demanda es uniforme para los 4 tipos de automóviles.

   H_1 = La demanda no es uniforme para los 4 tipos de automóviles.

2. $\alpha = 0.10$

3. Se elegirá el estadístico de prueba: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}$ y se comprueba con:

$$\chi_e^2 = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n$$

Tabla de frecuencias observadas y esperadas

Tipo Automóvil	Ventas Observadas	Ventas Esperadas	$\frac{(f_o - f_e)^2}{f_e'}$	$\frac{f_o^2}{f_e}$
A	32	25	1.96	40.96
B	21	25	0.64	17.64
C	19	25	1.44	14.44
D	28	25	0.36	31.36
Suma	100	100	4.40	104.40

Por lo tanto: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} = 4.40$; utilizando la otra fórmula, comprobamos:

$$\chi_e^2 = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n = 104.4 - 100 = 4.40$$

4. Regla de decisión: Si χ_e^2 es ≤ que χ_c^2 no se rechaza la H_o. En caso contrario rechazar la H_o.

5. En la tabla de la distribución χ^2:, si se tienen *gl=k-m-1 = 4-1=3* y el nivel de significación es de *0.10*, se observa:

$$\chi^2_{c,0.10,3} = 6.251$$

Por lo tanto como χ^2_e < 6.251, la hipótesis nula de que la demanda es uniforme, no se rechaza. Las diferencias no son lo suficientemente grandes para refutar la hipótesis nula; las diferencias no son significativas y pueden atribuirse simplemente a un error de muestreo.

Veamos otro ejemplo, una tienda vende 6 tipos de tarjetas de onomástico y se quiere saber si todas se venden en las mismas cantidades. Si en el siguiente día se vendieron 120 tarjetas, se esperaría que se vendieran 20 de cada una. Sin embargo, el número de tarjetas que se vendieron de cada tipo fueron:

A – 13; B – 33; C – 14; D – 14; E – 36; F – 17.

Con esta información, probar que no hay diferencias significativas en el número de ventas de las tarjetas en estudio a un nivel de significación del 5%.

Solución:

1. H_o = Las tarjetas se venden en la misma cantidad.

   H_1 = Las tarjetas no se venden en la misma cantidad.

2. $\alpha = 0.05$

3. Se elegirá el estadístico de prueba: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}$ y se comprueba con:

$$\chi_e^2 = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n$$

Tipo Tarjeta	Ventas Observadas	Ventas Esperadas	$\frac{(f_o - f_e)^2}{f_e'}$	$\frac{f_o^2}{f_e}$
A	33	20	2.45	8.45
B	13	20	8.45	54.45
C	14	20	1.80	9.80
D	7	20	8.45	2.45
E	36	20	12.80	64.80
F	17	20	0.45	14.45
Suma	120	120	34.40	154.40

Tabla de frecuencias observadas y esperadas

Por lo tanto: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} = 34.40$; utilizando la otra fórmula, comprobamos:

$$\chi_e^2 = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n = 154.40 - 120 = 104.40$$

4. Regla de decisión: Si χ_e^2 es $\leq$ que χ_c^2 no se rechaza la H_o. En caso contrario rechazar la H_o.

5. En la tabla de la distribución χ^2:, si se tienen *gl=k-m-1 = 6-153* y el nivel de significación es de *0.10*, se observa:

$$\chi^2_{c,0.05,5} = 11.070$$

Por lo tanto como χ^2_e > 11.070, se encuentra en la zona de rechazo. Las diferencias son lo suficientemente grandes para considerarlas significativas. Se concluye que es improbable que todas las tarjetas se vendan en el mismo número.

2. Prueba de ajuste a un patrón específico

Existen muchos casos en los cuales las frecuencias se prueban contra un patrón determinado en las que las frecuencias esperadas no son todas iguales.
Las frecuencias esperadas se calculan con datos porcentuales de la siguiente forma: $f_e = np_i$; en donde

n= Tamaño de la muestra

p_i= Probabilidad de cada categoría como se especifica en la hipótesis nula.

Ejemplo de aplicación; un director de un banco trata de seguir una política de extender un 35% de sus créditos a empresas industriales; un 20% a empresas comerciales; un 18% a empresas de servicios; un 25% a empresas maquiladoras; y un 5% a empresas extranjeras.

Para demostrar que la política se está siguiendo, se seleccionaron 113 créditos que se aprobaron recientemente. Se encontró que 28 créditos se otorgaron a empresas industriales; 22 a comerciales; 25 a empresas de servicios; 30 a maquiladoras; y 8 a empresas extranjeras. Probar esta hipótesis a un nivel de significación del 20%.

Solución:

1. H_o = Se mantuvo el patrón deseado.

   H_1 = No se mantuvo el patrón deseado.

2. $\alpha = 0.20$

3. Se elegirá el estadístico de prueba: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}$ y se comprueba con:

$$\chi_e^2 = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n$$

Tabla de frecuencias observadas y esperadas:

Tipo de Empresa	Frecuencias Observadas	Frecuencias Esperadas	$\frac{(f_o - f_e)^2}{f_e'}$	$\frac{f_o^2}{f_e}$
Industrial	28	39.55	3.37	19.82
Comercial	22	22.60	0.02	21.42
De servicios	25	20.34	1.07	30.73
Maquiladoras	30	24.86	1.06	36.20
Extranjeras	08	05.65	0.98	11.33
Suma	113	113.00	6.50	119.50

Por lo tanto: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} = 6.50$; utilizando la otra fórmula, comprobamos:

$$\chi_e^2 = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n = 154.40 - 120 = 104.40$$

4. Regla de decisión: Si χ_e^2 es ≤ que χ_c^2 no se rechaza la H_o. En caso contrario rechazar la H_o.

5. En la tabla de la distribución χ^2:, si se tienen *gl=k-m-1 = 6-153* y el nivel de significación es de *0.10*, se observa:

$$\chi_{c,0.05,5}^2 = 11.070$$

Por lo tanto como χ_e^2 > 11.070, se encuentra en la zona de rechazo. Las diferencias son lo suficientemente grandes para considerarlas significativas. Se concluye que es improbable que todas las tarjetas se vendan en el mismo número.

Otro ejemplo sería el de tres monedas fueron lanzadas 80 veces y se registró el número de veces que salieron “águilas”:

x	0	1	2	3
f	20	38	18	4

Siendo *“x”* el “lado águila” y *“f”* el número de veces que salió “águila”.

Con esta información, poner a prueba la hipótesis nula de que “x” es binomial con *n=3 y p= 0.5.* Usar un nivel de significación del 5%.

Solución:

1. H_o = *“x”* sigue una distribución binomial.

   H_1 = *“x”* no sigue una distribución binomial.

2. $\alpha = 0.05$

3. Se elegirá el estadístico de prueba: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}$ y se comprueba con:

$$\chi_e^2 = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n$$

Se calculan las probabilidades de éxito binomiales para 0, 1, 2, y 3.

Fórmula: $P(x) = C_x^n \cdot p^x \cdot q^{n-x}$

$$P(0) = \frac{3!}{0!(3-0)!} 0.5^0 \cdot 0.5^3 = 0.125$$

$$P(1)=\frac{3!}{1!(3-1)!}0.5^1\cdot 0.5^2=0.375$$

$$P(2)=\frac{3!}{2!(3-2)!}0.5^2\cdot 0.5^1=0.0.375$$

$$P(3)=\frac{3!}{3!(3-3)!}0.5^3\cdot 0.5^0=0.125$$

Tabla de frecuencias observadas y esperadas:

x	f_o	f_e	$\frac{(f_o-f_e)^2}{f_e'}$	$\frac{f_o^2}{f_e}$
0	20	10	10.00	40.00
1	35	30	02.13	48.13
2	18	30	04.80	10.80
3	04	10	03.60	01.60
Suma	80	80	20.53	100.53

Solución:

Paso 1. H_o = Los datos siguen una distribución normal

H_1 = Los datos no siguen una distribución normal

Paso 2. $\alpha=0.05$

Paso 3. Se elegirá el estadístico de prueba: $\chi_e^2=\sum_{i=1}^{k}\frac{(f_o-f_e)^2}{f_e}$ Se calculan las probabilidades de éxito de una distribución normal para cada evento.

Fórmula: $z = \dfrac{x - \mu}{\sigma}$

Para el evento A: $P(A) = 0.5000 - P(z_1)$

$$z_1 = \frac{580 - 600}{10} = -2.0$$

En la tabla de distribución normal se encuentra: $P(z_1) = 0.4772$

Por lo tanto : $P(x \leq 580) = 0.5000 - 0.4772$

Para el evento B: $P(B) = P(\ddot{}z_1) - P(z_2)$

$$z_2 = \frac{590 - 600}{10} = -1.0$$

En la tabla de distribución normal se encuentra: $P(z_2) = 0.3413$

Por lo tanto: $P(580 \leq x \leq 590) = 0.4772 - 0.3413 = 0.1359$

Para el evento C: $P(C) = P(z_2)$

$$z_2 = \frac{590 - 600}{10} = -1.0$$

En la tabla de distribución normal se encuentra: $P(z_2) = 0.3413$

Por lo tanto : $P(590 \leq x \leq 600) = 0.3413$

Tabla de frecuencias observadas y esperadas:

Evento	f_o	$P(x)$	f_e	$\frac{(f_o - f_e)^2}{f_e'}$
A	020	0.0228	0022.8	0.344
B	142	0.1359	0135.9	0.274
C	310	0.3413	0341.3	2.870
D	370	0.3413	0341.3	2.413
E	128	0.1359	0135.9	0.459
F	030	0.0228	0022.8	2.274
Suma	1000	1.0000	1000.0	8.634

Por lo tanto: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} = 8.634$;

Paso 4. Regla de decisión: Si χ_e^2 es $\leq$ que χ_c^2 no se rechaza la H_o. En caso contrario rechazar la H_o.

Paso 5. En la tabla de la distribución χ^2:, si se tienen *gl=k-m-1 = 6-0-1=5* y el nivel de significación es de *0.05*, se observa: $\chi^2_{c,0.05,5} = 11.070$

Por lo tanto como χ_e^2 < 11.070, se encuentra en la zona de aceptación. En consecuencia se concluye que los datos de la población siguen una distribución normal.

Se presenta un siguiente ejemplo; los fabricantes de una marca de computadoras reportan en su publicidad que su vida media útil es de 6 años con una desviación estándar de 1.4 años. En una muestra de 90 computadoras vendidas hace 10 años se encontraron los siguientes tiempos de vida útil:

Evento	Tiempo de vida (años)	Frecuencia
A	Hasta 4	07
B	De 4 a 5	14
C	De 5 a 6	25
D	De 6 a 7	22
E	De 7 a 8	16
F	8 o más	06

Con esta información, ¿puede concluir el fabricante, con un nivel de significación del 5% que la vida útil de las computadoras tiene una distribución normal?

Solución:

1. H_o = La vida útil de las computadoras sigue una distribución normal.

H_1 = La vida útil de las computadoras no sigue una distribución normal

2. $\alpha = 0.05$

3. Se elegirá el estadístico de prueba: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}$. Se calculan las probabilidades de éxito de una distribución normal para cada evento.

Fórmula: $z = \frac{x - \mu}{\sigma}$

Para el evento A: $P(A) = 0.5000 - P(z_1)$

$$z_1 = \frac{4-6}{1.4} = -1.43$$

En la tabla de distribución normal se encuentra: $P(z_1) = 0.4236$

Por lo tanto : $P(x \leq 4) = 0.5000 - 0.4236 = 0.0764$

Para el evento B: $P(B) = P(z_1) - P(z_2)$

$$z_2 = \frac{5-6}{1.4} = -0.71$$

En la tabla de distribución normal se encuentra: $P(z_2) = 0.2611$

Por lo tanto: $P = 0.4236 - 0.2611 = 0.1625$

Para el evento C: $P(C) = P(z_2)$

$$z_2 = \frac{5-6}{1.4} = -0.71$$

En la tabla de distribución normal se encuentra: $P(z_2) = 0.2611$

Por lo tanto: $P(5 \leq x \leq 6) = 0.2611$

Por lo tanto: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} = 20.53$; utilizando la otra fórmula, comprobamos:

$$\chi_e^2 = \frac{\sum_{i=1}^{k} f_o^2}{f_e} - n = 100.53 - 80 = 20.53$$

4. Regla de decisión: Si χ_e^2 es ≤ que χ_c^2 no se rechaza la H_o. En caso contrario rechazar la H_o.

5. En la tabla de la distribución χ^2:, si se tienen *gl=k-m-1 = 4-0-1=3* y el nivel de significación es de *0.20*, se observa: $\chi^2_{c,0.05,3} = 7.815$

Por lo tanto como χ_e^2 > 7.815, se encuentra en la zona de rechazo. En consecuencia se concluye que *"x"* no sigue una distribución binomial con *n=3* y *p=0.50*.

3. Pruebas de normalidad

Se requiere probar que una serie de elementos de una población sigue una distribución normal por medio de una muestra.

Ejercicio de aplicación; en clases de buceo, los tanques de inversión se llenan a una presión promedio de 600 libras por pulgada cúbica (psi). Se permite una desviación estándar de 10 psi. Las especificaciones de seguridad permiten una distribución normal en los niveles de llenado. Probar la hipótesis a un nivel de significación del 5% si en una muestra se miden 1,000 tanques con los siguientes resultados:

Tabla de frecuencias observadas y esperadas:

Evento	f_o	$P(x)$	f_e	$\frac{(f_o - f_e)^2}{f_e'}$
A	07	0.0764	06.876	0.0022
B	14	0.1625	14.625	0.0267
C	25	0.2611	23.499	0.0959
D	22	0.3413	23.499	0.0959
E	16	0.1359	14.625	0.1293
F	06	0.0228	06.876	0.1116
Suma	90	1.0000	90.000	0.4613

Por lo tanto: $\chi_e^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} = 0.4613$;

4. Regla de decisión: Si χ_e^2 es $\leq$ que χ_c^2 no se rechaza la H_o. En caso contrario rechazar la H_o.

5. En la tabla de la distribución χ^2 :, si se tienen *gl=k-m-1 = 6-0-1=5* y el nivel de significación es de *0.05*, se observa: $\chi_{c,0.05,5}^2 = 11.070$

Por lo tanto como χ_e^2 < 11.070, se encuentra en la zona de aceptación. En consecuencia se concluye que los datos de la población siguen una distribución normal.

ACTIVIDAD 1

1. La crisis mundial actual ha provocado no solo despidos masivos, también ha provocado que las ventas de autos disminuya drásticamente; considere una agencia automotriz que desea controlar las existencias de sus tres versiones del auto compacto menos solicitado.

Si consideramos que las ventas de las tres versiones es la misma, ayude a decidir las hipótesis central, si al tomar una muestra aleatoria de 50 autos. ompactos vendidos últimamente, los datos obtenidos se muestran en la siguiente tabla.

Versiones del automóvil compacto	Ventas observadas en una muestra de 50 autos
M	21
K	19
L	10

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar**, localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

2. Una empresa de neumáticos está probando un nuevo modelo, que requiere ser llenado a una presión promedio de 35 libras con una desviación estándar de 3 libras. Si las especificaciones de seguridad requieren una distribución normal en la presión de las llantas. Pruebe la hipótesis para un nivel de significancia de 3% si una muestra de 500 llantas ofreció los siguientes resultados.

Presión de la llanta	frecuencia
Menor de 35 libras	70
De 36 a 40 libras	320
De 41 a 45 libras	85
De 46 a 50 libras	25

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar,** localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Selecciona si las siguientes aseveraciones son verdaderas (V) o falsas (F). Una vez que concluyas, obtendrás tu calificación de manera automática.

	Verdadera	Falsa
1. Las pruebas de bondad de ajuste son medidas sobre qué tan cerca se ajustan los datos muestrales observados a una forma de distribución particular planteada como hipótesis.	( )	( )
2. Las pruebas de bondad de ajuste son muy importantes pues en los negocios frecuentemente se requiere probar alguna hipótesis sobre una distribución poblacional desconocida.	( )	( )
3. En una prueba para un ajuste uniforme se requiere probar que una serie de elementos de una población sigue una distribución normal por medio de una muestra.	( )	( )
4. En las pruebas de normalidad se pretende probar que la distribución de datos es uniforme	( )	( )
5. En estadística no paramétrica, una muestra es grande cuando su tamaño es mayor de 20.	( )	( )

Tema 3. Tablas de contingencia

Objetivos del tema

Identificar si dos variables están relacionadas o no, utilizando como herramienta las tablas de contingencia.

Desarrollo

En aplicaciones estadísticas es frecuente interesarse en calcular si 2 variables de clasificación, cuantitativas o cualitativas, son independientes o si están relacionadas.

Las hipótesis son:

H_o: Las variables de clasificación son independientes.

H_1: Las variables de clasificación son dependientes.

Estos modelos se basan también en la prueba JI-cuadrada por lo que se procede a comparar las frecuencias esperadas con las observadas para determinar que tan grande debe ser el alejamiento permitido para que la hipótesis de independencia pueda rechazarse.

Si el valor del estadístico de prueba JI-cuadrada es mayor que el valor critico, no se puede suponer que las 2 variables de clasificación sean independientes.

La fórmula del estadístico de prueba es la siguiente:

$$\chi_e^2 = \sum_{1}^{rc} \frac{\left(f_o - f_e\right)^2}{f_e}$$

en donde:

r es el N° de renglones de una tabla de contingencia.

C es el N° de columnas de una tabla de contingencia.

Los grados de libertad serán: $gl = (r-1)(c-1)$

Ejemplo; un director de investigación de productos debe determinar si existe alguna relación entre la clasificación de efectividad que los consumidores asignan a un nuevo insecticida y el sitio (urbano o rural) en los cuales se utilizan. De los 120 consumidores de la encuesta, 90 viven en zonas urbanas y 30 en rurales. El nivel de significación es del 1%.

En la siguiente tabla de contingencia se muestran las clasificaciones.

Atributo “A”: Clasificación	Atributo “B”: Urbano	Ubicación Rural
Por encima del promedio	24	13
En el promedio	48	10
Por debajo del promedio	18	07

Probar esta hipótesis con un nivel de significación es del 1%.

Solución

Paso 1. H_o = La clasificación y la ubicación son independientes.

H_1 = La clasificación y la ubicación no son independientes.

Paso 2. $\alpha = 0.01$

Paso 3. Se elegirá el estadístico de prueba: $\chi_e^2 = \sum_1^{rc} \frac{(f_o - f_e)^2}{f_e}$

Tabla de contingencia

Atributo "A": Clasificación	Atributo "B": Urbano	Ubicación Rural	Total
Por encima del promedio	24	13	037
En el promedio	48	10	058
Por debajo del promedio	18	07	025
Total	90	30	120

Se calculan las frecuencias esperadas relacionando las 2 variables. En un esquema matricial se utiliza su nomenclatura:

Calculo de:

$$f_{e_{11}} = \frac{f_{t_{13}}}{f_{t_{43}}} \cdot f_{t_{41}} = \frac{37}{120} \cdot 90 = 27.75 \quad f_{e_{12}} = \frac{f_{t_{13}}}{f_{t_{33}}} \cdot f_{t_{42}} = \frac{37}{120} \cdot 30 = 9.25$$

$$f_{e_{21}} = \frac{f_{t_{32}}}{f_{t_{43}}} \cdot f_{t_{41}} = \frac{58}{120} \cdot 90 = 43.50 \quad f_{e_{22}} = \frac{f_{t_{23}}}{f_{t43}} \cdot f_{t_{42}} = \frac{58}{120} \cdot 30 = 14.5$$

$$f_{e_{31}} = \frac{f_{t_{33}}}{f_{t_{43}}} \cdot f_{t_{33}} = \frac{25}{120} \cdot 90 = 18.75 \quad f_{e_{32}} = \frac{f_{t_{33}}}{f_{t_{43}}} \cdot f_{t_{42}} = \frac{25}{120} \cdot 30 = 6.25$$

Tabla de frecuencias de clasificación

Atributo "A": Clasificación	Atributo "B": Urbano		Ubicación Rural	
	f_o	f_e	f_o	f_e
Por encima del promedio	24	24.75	13	09.25
En el promedio	48	43.50	10	14.50
Por debajo del promedio	18	18.75	07	06.25

Por lo tanto utilizando la fórmula del estadístico de prueba:

$$\chi_e^2 = \sum_1^{rc} \frac{(f_o - f_e)^2}{f_e} = \frac{(24-27.75)^2}{27.75} + \frac{(13-9.25)^2}{9.25} + \frac{(48-43.5)^2}{43.5} + \ldots\ldots + \frac{(7-6.25)^2}{6.25} = 4.009$$

Paso 4. Regla de decisión: Si χ_e^2 es $\leq$ que χ_c^2 no se rechaza la H_o. En caso contrario rechazar la H_o.

Paso 5. En la tabla de la distribución χ^2:, si se tienen $gl = (r-1)(c-1) = (3-1)(2-1) = 2$ y el nivel de significación es de *0.01*, se observa:

$$\chi_{c,0.01,2}^2 = 9.210$$

Por lo tanto como χ_e^2 < 9.210, la hipótesis nula es aceptada y se concluye que tanto la clasificación como la ubicación son factores independientes.

Un siguiente ejemplo, es el director de Marketing de un diario metropolitano de gran circulación, estudia la relación entre el tipo de actividad y la sección del periódico que es de su preferencia. De una muestra de lectores se obtuvo la siguiente información:

Tabla de frecuencias de clasificación

Lectores	Noticias Nacionales	Sociales	Deportes
Profesionistas	170	124	090
Estudiantes	112	100	120
Otros	130	088	090

¿Podemos concluir con un nivel de significación del 10% que si hay relación entre el tipo de actividad y la sección del periódico de su preferencia?

Solución:

Paso 1. H_o = No hay relación entre el tipo de actividad y la sección de su preferencia.

H_1 = Si hay relación entre el tipo de actividad y la sección de su preferencia.

Paso 2. $\alpha = 0.10$

Paso 3. Se elegirá el estadístico de prueba: $$\chi_e^2 = \sum_{1}^{rc} \frac{\left(f_o - f_e\right)^2}{f_e}$$

Tabla de contingencia

Lectores	Noticias Nacionales	Sociales	Deportes	Total
Profesionistas	170	124	090	0384
Estudiantes	112	100	120	0332
Otros	130	088	090	0308
Total	412	312	300	1,024

Se calculan las frecuencias esperadas relacionando las 2 variables. En un esquema matricial se utiliza su nomenclatura:

Calculo de:

$$f_{e_{11}} = \frac{f_{t_{13}}}{f_{t_{43}}} \cdot f_{t_{41}} = \frac{384}{1024} \cdot 412 = 154.5 \quad f_{e_{12}} = \frac{f_{t_{13}}}{f_{t_{33}}} \cdot f_{t_{42}} = \frac{384}{1024} \cdot 312 = 117.0$$

$$f_{e_{21}} = \frac{f_{t_{32}}}{f_{t_{43}}} \cdot f_{t_{41}} = \frac{332}{1024} \cdot 412 = 133.6 \quad f_{e_{22}} = \frac{f_{t_{23}}}{f_{t43}} \cdot f_{t_{42}} = \frac{332}{1024} \cdot 312 = 101.2$$

$$f_{e_{31}} = \frac{f_{t_{33}}}{f_{t_{43}}} \cdot f_{t_{33}} = \frac{308}{1024} \cdot 412 = 123.9 \quad f_{e_{32}} = \frac{f_{t_{33}}}{f_{t_{43}}} \cdot f_{t_{42}} = \frac{308}{1024} \cdot 312 = 93.8$$

Se realizan cálculos similares para la sección de deportes.

Lectores	Noticias nacionales		Sociales		Deportes	
	f_o	f_e	f_o	f_e	f_o	f_e
Profesionistas	170	154.5	124	117.0	090	112.5
Estudiantes	112	133.6	100	101.2	120	097.2
Otros	130	123.9	088	93.8	090	090.3

Por lo tanto utilizando la fórmula del estadístico de prueba:

$$\chi_e^2 = \sum_1^{rc} \frac{(f_o - f_e)^2}{f_e} = \frac{(170-154.5)^2}{154.5} + \frac{(112-133.6)^2}{133.6} + \frac{(130-123.9)^2}{123.9} + \ldots\ldots + \frac{(90-90.3)^2}{90.3} = 15.936$$

Paso 4. Regla de decisión: Si χ_e^2 es $\leq$ que χ_c^2 no se rechaza la H_o. En caso contrario rechazar la H_o.

Paso 5. En la tabla de la distribución χ^2, si se tienen $gl = (r-1)(c-1) = (3-1)(3-1) = 4$ y el nivel de significación es de *0.10*, se observa: $\chi^2_{c,0.10,4} = 7.779$

Por lo tanto como χ_e^2 > 7.779, se rechaza la hipótesis nula por lo que se puede afirmar que si hay una relación entre la actividad de los lectores y la sección del periódico de su preferencia.

ACTIVIDAD 1

1. Una agencia de marketing que trabaja para el sector automotriz, estudia la relación entre la clase social de los jóvenes y su preferencia por tres marcas específicas de automóviles. De una muestra aleatoria obtuvo los siguientes datos.

Usuario	Volvo	Toyota	Chevrolet
Jóvenes hijos de padres ricos	150	110	70
Jóvenes que trabajan (proletariados)	05	70	120
Otros	40	130	90

Concluye si para un nivel de significancia del 5% hay relación entre la clase social y la marca automotriz de preferencia.

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar**, localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

ACTIVIDAD 2

Discute en el **Foro Tablas de contingencia**, las ventajas de poder determinar mediante una tabla de contingencia la dependencia o independencia de dos variables de tipo cualitativo.

Pulsa el botón **Colocar un nuevo tema de discusión aquí**; pon en el apartado Asunto el título de tu aportación, redacta tu comentario en el área de texto y haz clic en **Enviar al foro**.

Autoevaluación

Selecciona si las siguientes aseveraciones son verdaderas (V) o falsas (F). Una vez que concluyas, obtendrás tu calificación de manera automática.

	Verdadera	Falsa
1. Las pruebas de bondad de ajuste son medidas sobre qué tan cerca se ajustan los datos muestrales observados a una forma de distribución particular planteada como hipótesis.	( )	( )
2. Las pruebas de bondad de ajuste son muy importantes pues en los negocios frecuentemente se requiere probar alguna hipótesis sobre una distribución poblacional desconocida.	( )	( )
3. En una prueba para un ajuste uniforme se requiere probar que una serie de elementos de una población sigue una distribución normal por medio de una muestra.	( )	( )
4. En las pruebas de normalidad se pretende probar que la distribución de datos es uniforme	( )	( )

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Tema 4. Prueba de los signos de Wilcoxon

Objetivos del tema

Aplicar la prueba de los signos de Wilcoxon como una alternativa no paramétrica para comparar datos obtenidos mediante una muestra pareada.

Desarrollo

Se utiliza como una **alternativa no paramétrica** cuando se trata de comparar los datos de 2 poblaciones o de una misma población mediante una muestra apareada en la que cada unidad experimental genera 2 observaciones pareadas o ajustadas, una de la población 1 y una de la población 2. Las diferencias entre las observaciones pareadas permiten tener una buena perspectiva respecto de la diferencia entre las 2 poblaciones.

La metodología del análisis paramétrico de una muestra pareada requiere de datos de intervalo y de la suposición de que la población de las diferencias entre los pares de observaciones tenga una distribución normal. Con este supuesto se puede usar la distribución "t" para probar la hipótesis nula es decir que no hay diferencias entre las medias poblacionales. Si no es así se debe utilizar la prueba de rango con signo de Wilcoxon.

La prueba de los rangos con signo usa los rangos de los valores absolutos de las diferencias pareadas, asignando el rango 1 a la diferencia con valor absoluto mínimo, el rango 2 a la siguiente diferencia con menor valor absoluto y así se procede sucesivamente. Se deben descartar los rangos con diferencias de cero y en caso de valores absolutos repetidos, a cada uno de ellos se les otorga el valor promedio de los rangos ocupados por los valores repetidos. A cada uno de los rangos positivos o negativos, se les asocia el signo correspondiente.

La suma de los rangos positivos se indica por T^{+}, la suma de los rangos negativos se denota por T^{-} y el máximo valor entre estos 2 valores se escribe solamente *"T"* y se utiliza generalmente como estadístico de prueba. Si el número de diferencias es igual o mayor de 15 entonces la distribución muestral de *"T"* es aproximadamente normal por lo que se utilizará la variable parametrizada "z". Si es menor se deberán utilizar tablas especiales que proporcionan los valores críticos para la prueba de rangos con signo.

La suma de los rangos es: $S = \frac{n(n+1)}{2}$ y deberá ser igual a $T^{+} + T^{-}$

Las fórmulas de la media y desviación estándar de la distribución muestral *"T"* son las siguientes:

Media: $$\mu_T = \frac{n(n+1)}{4}$$

Desviación estándar: $$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

y el estadístico de prueba es: $$z = \frac{T - \mu_T}{\sigma_T}$$

Ejemplo de aplicación; se desea saber si un programa de capacitación en cómputo en una empresa especializada, mejoró las habilidades de los empleados en dicha materia. Por ello se observa el nivel de habilidades antes del programa y después del programa en una muestra de 22 empleados, obteniéndose los siguientes resultados y probar la hipótesis a un nivel de significación del 1%.

					Diferencias	Rango	Rangos
Número	Puntaje		Diferencia		absolutas		con signos
Empleado	Antes (a)	Después (b)	b-a		ordenadas		correctos
1	18	15	-3		2	1	1
2	60	70	10		3	2	-2
3	81	75	-6		4	3	-3
4	15	20	5		5	4	4.5
5	20	50	30		5	5	4.5
6	17	40	23		6	6	-6
7	26	50	24		8	7	-7.5
8	11	30	19		8	8	7.5
9	20	40	20		9	9	-9
10	38	30	-8		10	10	10.5
11	80	85	5		10	11	10.5
12	59	86	27		11	12	12
13	12	72	60		19	13	13
14	87	98	11		20	14	15
15	88	79	-9		20	15	15
16	64	88	24		20	16	15
17	88	90	2		23	17	17
18	76	96	20		24	18	18.5
19	43	39	-4		24	19	18.5
20	90	98	8		27	20	20
21	40	60	20		30	21	21
22	50	60	10		60	22	22

Se obtienen las diferencias de los puntajes antes y después, sus diferencias, las diferencias absolutas ordenadas, sus rangos y los rangos con signos correctos.

La suma de rangos positivos es: $T^{+} = 225.5$

La suma de rangos negativos es: $T^{+} = 27.5$

Comprobación: $$S = T^{+} + T^{-} = \frac{n(n+1)}{2} = \frac{22(22+1)}{2} = 253.0$$

Por lo tanto $T = 225.5$

La hipótesis por probar son:

Ho: No hay diferencia significativa debido al tratamiento.

Ha: Hay diferencia significativa por el tratamiento

La columna de rangos con signos correctos se determinó mediante el promedio de rangos, si la diferencia absoluta se repite y los rangos son signos correctos se preserva el signo de la diferencia que le dio origen. Por ejemplo, para el rango 4 y 5 se promedio (4+5)/2=4.5 y como el rango 4 corresponde a una diferencia 5 positiva entonces se le asigna 4.5 positivo, lo mismo para el rango 5. En el caso de los rangos 7 y 8 (correspondientes a una diferencia de 8), el promedio es 7.5 y como la diferencia de 8 corresponde a un valor negativo y otro positivo, entonces se le asigna un rango con signo correcto de -7.5 y 7.5.

Estadístico de prueba: $$z = \frac{T - \mu_T}{\sigma_T}$$

La media es: $$\mu_T = \frac{n(n+1)}{4} = \frac{22 \cdot 23}{4} = 126.5$$

La desviación estándar: $$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{22 \cdot 23 \cdot 43}{24}} = 30.1$$

Por lo tanto: $$z = \frac{T - \mu_T}{\sigma_T} = \frac{225.5 - 126.5}{30.1} = 3.29$$

Nivel de significación: $\alpha = 0.01$ por lo que $z_c = 2.33$

Como $z > z_c$ cae en la zona de rechazo, se puede concluir que el programa de capacitación de computo en esta empresa si mejoró las habilidades del personal.

ACTIVIDAD 1

1. Una manufacturera automotriz desea conocer la preferencia de los clientes por los colores ocre o índigo del modelo de lujo, pues sólo uno saldrá al mercado. Se invitó a los 20 mejores vendedores para que opinaran y se encontró que doce prefirieron el color ocre, siete el índigo y uno indeciso. En un nivel del 10% probar si:

H0: Cualquier color gustará por igual a los clientes

H1: Hay preferencia por alguno de los colores de los clientes

Para enviar tu actividad, pulsa **Editar mi envío** y se mostrará un editor de texto en el que deberás redactar tu información. Cuando termines, guarda tu tarea haciendo clic en **Guardar cambios.**

2. Para el aniversario de la empresa se organizó una convención y se dio a escoger entre el menú tradicional o uno especial. La muestra fue de 81 clientes de los cuales 42 prefirieron el especial. Utilizando la prueba del signo y un nivel de 0.02, pruebe si a los clientes les gustó más el menú especial que el tradicional:

H0: Ambos menús gustaron por igual ($p=0.50$)

H1: Gustó más el menú especial ($p>0.50$)

Para enviar tu actividad, pulsa **Editar mi envío** y se mostrará un editor de texto en el que deberás redactar tu información. Cuando termines, guarda tu tarea haciendo clic en **Guardar cambios**.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Elige la respuesta correcta a las siguientes preguntas, una vez que concluyas, obtendrás de manera automática tu calificación.

1. La prueba de los signos de Wilcoxon:

- a) Se utiliza para averiguar el signo de una prueba no paramétrica.
- b) Se utiliza como una alternativa no paramétrica para comparar datos obtenidos mediante una muestra pareada.
- c) Se utiliza para probar la varianza de dos poblaciones utilizando la distribución F.

2. En una prueba de signo:

- a) Se manejan sólo valores absolutos
- b) Se ignora la magnitud de la diferencia
- c) Es importante la dispersión de los datos

3. En un grupo piloto de 14 competidores del equipo olímpico mexicano se efectuó una prueba de “confianza en sí mismo” antes de cursar un seminario. La autoconfianza se clasificó como negativa, baja, alta y muy alta. Usando un nivel de significación del 5% y analizando la tabla diga si el seminario ayuda a mejorar la autoconfianza:

H_0: p=0.50

H_1: p>0.50

Nombre	Antes del curso	Después del curso
Elizabeth	Negativa	Alta
Luisa	Baja	Muy alta
Mario	Baja	Alta
René	Negativa	Baja
Cristina	Baja	Alta
Eloísa	Negativa	Baja
Arturo	Baja	Alta
Luis	Baja	Muy alta
Xóchitl	Negativa	Baja
Mónica	Negativa	Negativa
Jaime	Baja	Alta
Soledad	Muy alta	Baja
Estrella	Baja	Alta
Francisco	Baja	Alta

- a) Se rechaza H0 pues hubo 12 signos “+”
- b) Se rechaza H_0 pues hubo 10 signos “+”
- c) Se rechaza H_0 pues hubo 11 signos “+”

4. El sindicato de taxistas afirma que la mediana del recorrido semestral de cada unidad es de 40,000 km. Sin embargo, esto es rebatido por algunos dirigentes que afirman que la mediana es mayor; por lo tanto, se tomó una muestra aleatoria de 205 taxis y se encontró que 170 recorren una mayor distancia; 5 recorren 40,000 km. y los restantes menos de 40,000 km. Utilizando la prueba del signo y un nivel de 0.05, pruebe si:

H_0: La mediana semestral es igual a 40,000 km.

H_1: La mediana semestral es mayor de 40,000 km

- a) Se acepta H_0 pues la Zc es 1.509
- b) Se rechaza H_0 pues la Zc es 3.748
- c) Se rechaza H_0 pues la Zc es 10.540

5. En los botes pesqueros se desea probar la efectividad de una nueva técnica. Así que al tomar una prueba con 11 botes los resultados fueron los siguientes:

No.	Normal	Nueva
1	15	25
2	10	28
3	16	16
4	10	22
5	20	19
6	17	20
7	24	30
8	23	26
9	17	18
10	21	23
11	25	22

Aplicando la prueba de Wilcoxon de rangos con signo y un nivel de significación del 5% probar si la nueva técnica incrementa o no la pesca:

H_0: Los volúmenes de pesca no cambian con la nueva técnica

H_1: Los volúmenes de pesca mejoran con la nueva técnica

- a) Se acepta H_0 pues la T calculada es 18.1
- b) Se acepta H_1 pues la T calculada es 6.5
- c) Se acepta H_0 pues la T calculada es 2.7

Tema 5. Prueba de rachas

Objetivos del tema

Comprobar la aleatoriedad de una muestra utilizando la prueba de rachas.

Desarrollo

Es una prueba que se utiliza para comprobar la aleatoriedad de muestras. Es muy importante demostrar la aleatoriedad de las muestras en los estudios estadísticos. Si no es así se crea una gran desconfianza en los procesos de muestreo.

En una prueba de rachas, se asigna a todas las observaciones de la muestra uno o dos símbolos. Una racha se designa como una secuencia de uno o más símbolos similares y también se expresa como una serie continua de uno o más símbolos. Si el número de rachas es menor de 20, se utilizan tablas específicas en donde se muestran valores críticos mínimos y máximos por lo que si el número de rachas (r) es menor o excede de esos valores críticos, se indica una ausencia de aleatoriedad.

Si se tienen 2 categorías y los datos muestrales no caen en alguna de ellas, se puede utilizar la mediana como valor de referencia. Una importante aplicación de la prueba de rachas es en el método de mínimos cuadrados en el análisis de regresión. Una propiedad básica en estos modelos de regresión es que los errores son aleatorios.

Las hipótesis para probar son:

$H_o:$ Existe aleatoriedad en las muestras.

$H_1:$ No existe aleatoriedad en las muestras.

Si el número de datos en 2 categorías n_1 y n_2 son mayores a 20, la distribución de muestreo para "*r*" se aproxima a una distribución normal.

Las fórmulas son:

Media de la distribución muestral del número de rachas:

$$\mu_r = \frac{2n_1n_2}{n_1+n_2}+1$$

Desviación estándar: $$\sigma_r = \sqrt{\frac{2n_1n_2\left(2n_1n_2-n_1-n_2\right)}{\left(n_1+n_2\right)^2\left(n_1+n_2-1\right)}}$$

Estadístico de prueba: $$z=\frac{r-\mu_r}{\sigma_r}$$

Ejemplo de aplicación; en una campaña a 100 posibles compradores de un producto especializado, se realizaron 52 ventas, 48 no ventas y 40 rachas. A un nivel de significación del 1% probar la hipótesis que la muestra es aleatoria.

Las hipótesis son:

H_o : La muestra es aleatoria.

H_1 : La muestra no es aleatoria.

Estadístico de prueba: $$z=\frac{r-\mu_r}{\sigma_r}$$

La media es: $$\mu_r = \frac{2n_1n_2}{n_1+n_2}+1 = \frac{2\cdot 52\cdot 48}{52+48}+1 = 50.92$$

La desviación estándar:

$$\sigma_r = \sqrt{\frac{2n_1n_2\left(2n_1n_2-n_1-n_2\right)}{\left(n_1+n_2\right)^2\left(n_1+n_2-1\right)}} = \sqrt{\frac{2\cdot 52\cdot 48\left(2\cdot 52\cdot 48-52-48\right)}{\left(52+48\right)^2\left(52+48-1\right)}} = \sqrt{24.67} = 4.97$$

Por lo tanto: $$z = \frac{r-\mu_r}{\sigma_r} = \frac{40-50.92}{4.97} = -2.20$$

Nivel de significación: $\alpha = 0.01$ por lo que $z_c = \pm 2.58$ ya que es una prueba de 2 colas. Como $z < z_c$ cae en la zona de aceptación se puede concluir que no hay evidencia suficiente para rechazar la hipótesis nula, por lo que se puede indicar que la muestra es aleatoria.

ACTIVIDAD 1

Consulta en tu bibliografía y en sitios de Internet, acerca de las pruebas de rachas y elabora un resumen completando la información que se expuso en el tema. Recuerda citar tus fuentes.

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar**, localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Responde si son verdaderas (V) o falsas (F) las siguientes aseveraciones. Una vez que concluyas, obtendrás tu calificación de manera automática.

	Verdadera	Falsa
1. La prueba de las rachas se utiliza para comprobar la aleatoriedad de una muestra.	( )	( )
2. En una prueba de rachas se asigna a todas las observaciones de la muestra una función matemática representativa de la muestra.	( )	( )
3. Una racha se designa como una secuencia de uno o más símbolos similares o como una serie continua de uno o más símbolos.	( )	( )
4. En el análisis de regresión lineal no es posible aplicar la prueba de rachas.		
5. En la prueba de rachas, la hipótesis nula siempre dice que existe aleatoriedad en la muestra.	( )	( )

Tema 6. Otras pruebas

Objetivos del tema

Utilizar la Prueba U de Mann-Whitney para la toma de decisiones.

Desarrollo

Prueba U de Mann-Whitney

Esta prueba es útil cuando se seleccionan dos conjuntos aleatorios independientes y su escala es de tipo ordinal al menos. La prueba consiste en determinar si las dos muestras presentan los mismos promedios poblacionales o no (prueba de medias).

Para esta prueba se considerará que el estadístico de prueba se comportará como una distribución de Mann-Whitney y en ocasiones se prefiere en lugar de la "*t*" de Student debido a que la varianza de las dos poblaciones son independientes o los datos son de tipo ordinal.

Las hipótesis son las siguientes:

$H_o:$ Las medias o medianas son iguales.

$H_1:$ Las medias o medianas no son iguales.

e utilizan los estadísticos U_1 y U_2 para la primera y la segunda muestra respectivamente.

Fórmulas: $$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum Rangos_1 \quad y \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum Rangos_2$$

Media de la distribución muestral "*U*":

$$\mu_u = \frac{n_1 n_2}{2}$$

Desviación estándar.

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Estadístico de prueba:

$$z = \frac{U_i - \mu_u}{\sigma_u}$$

Veamos un ejemplo; en una empresa se está efectuando una prueba de aptitud mecánica en la línea de producción y se desea saber si la aptitud mecánica de los hombres es la misma que la de las mujeres o son distintas (prueba de dos colas). Para ello se extrae una muestra de nueve hombres y cinco mujeres y se les calificó en puntos el nivel de aptitud; este último varia en un rango de 600 a 1600 puntos y a cada puntuación se le asigna un rango del 1 al 14 (rango 1=mayor puntuación y rango 14=menor puntuación), obteniéndose los siguientes resultados:

Puntuaciones y rangos de hombres y mujeres en la prueba de aptitudes mecánicas

HOMBRES			MUJERES	
Puntuación	Rango		Puntuación	Rango
1 500	2			
1 600	1			
670	13		1400	3
800 *	10.5		1200	6
1 100	8		780	12
800 *	10.5		1350	4
1 320	5		890	9
1 150	7			
600	14		TOTAL	34
TOTAL	71			

Nota: * El caso de empate se resolvió asignando el promedio de los rangos que le corresponderían y que serían el rango 10 y rango 11.

Realizar una prueba de hipótesis a un nivel de significación del 10%.

Cálculo de U_1 y U_2:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum Rangos_1 \quad y \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum Rangos_2$$

que en este caso es igual a:

$$U_1 = (9)(5) + \frac{9(10)}{2} - 71 = 19 \quad y \quad U_2 = (9)(5) + \frac{5(6)}{2} - 34 = 26$$

Media de la distribución: $\mu_u = \frac{n_1 n_2}{2} = \frac{9 \cdot 5}{2} = 22.5$

Desviación estándar.

$$\sigma_U = \sqrt{\frac{n_1 n_2 \left(n_1 + n_2 + 1\right)}{12}} = \sqrt{\frac{9 \cdot 5\left(9+5+1\right)}{12}} = \sqrt{56.25} = 7.50$$

Estadístico de prueba:

$$z_1 = \frac{U_1 - \mu_u}{\sigma_u} = \frac{19 - 22.5}{7.50} = -0.47$$

$$z_2 = \frac{26 - 22.5}{7.50} = 0.47$$

Nivel de significación: $\alpha = 0.10$ por lo que $z_c = \pm 1.96$ ya que es una prueba de 2 colas.

Como $z_1 \ y \ z_2 < z_c$ cae en la zona de aceptación se puede concluir que no hay evidencia suficiente para rechazar la hipótesis nula, por lo que se puede indicar que la aptitud de los hombres es muy similar a las de las mujeres.

Es importante para los alumnos profundizar en el manejo de estas pruebas no paramétricas y estudiar otras más que tienen múltiples aplicaciones en la administración y economía.

ACTIVIDAD 1

1. En la Olimpiada de Matemáticas compitieron 20 ingenieros contra 15 actuarios, las calificaciones fueron las siguientes:

ingenieros: 35, 11, 8, 21, 4, 18, 15, 36, 13, 28, 31, 23, 5, 9, 27, 29, 37, 2, 1, 20, 19, 16, 39, 33 y 7;

actuarios: 12, 34, 32, 25, 22, 3, 40, 17, 30, 24, 14, 10, 38, 6 y 26.

Use la prueba de Mann-Whitney con un nivel de significación de 0.05 y demuestre si hay alguna diferencia en su aptitud hacia las matemáticas de alguno de los grupos.

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar,** localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

2. En dos equipos de vendedores de puerta en puerta se analizaron los resultados de su entrenamiento.

Las puntuaciones del equipo “A” son: 186, 212, 97, 141, 160, 122, 180 y 121;
las del grupo “B” son: 147, 99, 167, 126, 180, 197 y 128.

Como la población de puntuaciones no se distribuye normalmente, use la prueba de Mann-Whitney con un nivel de significación de 5% y compruebe si existe alguna diferencia entre ambos grupos

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar**, localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

Autoevaluación

Responde si son verdaderas (V) o falsas (F) las siguientes aseveraciones. Una vez que concluyas, obtendrás tu calificación de manera automática.

	Verdadera	Falsa
1. La prueba U de Mann-Whitney se aplica únicamente a datos cuantitativos.	( )	( )
2. La prueba U de Mann-Whitney consiste en determinar si las dos muestras analizadas presentan los mismos promedios poblacionales o no.	( )	( )
3. La prueba U de Mann-Whitney es útil cuando se seleccionan dos conjuntos aleatorios independientes y su escala es de tipo ordinal al menos.	( )	( )
4. En la prueba U de Mann-Whitney se considera siempre que el estadístico de prueba se comportará como una distribución de Mann-Whitney.	( )	( )

LO QUE APRENDÍ DE LA UNIDAD

En la actualidad existen diferentes centros educativos que entre sus actividades preparan a los alumnos que desean presentar el examen del CENEVAL. A continuación se enumeran 3 de estos centros y las calificaciones obtenidas por sus alumnos en dicho examen. Utilice la prueba de Kruskal-Wallis con un nivel de significancia de 0.05 con el fin de comprobar si hay diferencias significativas en sus habilidades de enseñanza de tales centros educativos.

Centro educativo No. 1	87	78	82	85	99	99	85	94	
Centro educativo No. 2	88	76	68	82	85	82	84	83	81
Centro educativo No. 3	80	85	56	71	89	87			

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar**, localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

Glosario de la unidad

Métodos no paramétricos

Métodos estadísticos que requieren muy pocos o ningún supuesto acerca de las distribuciones de probabilidad de la población, y acerca del nivel de medición. Estos métodos se pueden aplicar cuando se dispone de datos nominales u ordinales.

Métodos sin distribución

Es otro nombre que se da a los métodos estadísticos no paramétricos, que indica la carencia de supuestos sobre la distribución de probabilidad de la población.

Prueba de signo

Prueba estadística no paramétrica que permite identificar diferencias entre dos poblaciones basándose en el análisis de datos nominales.

Prueba de rango con signo de Wilcoxon

Prueba estadística no paramétrica con la cual se identifican diferencias entre dos poblaciones, basada en el análisis de dos muestras pareadas o ajustadas.

Prueba de Mann-Whitney-Wilcoxon

Prueba estadística no paramétrica con la cual se identifican diferencias entre dos poblaciones basada en el análisis de dos muestras independientes.

Prueba de Kruskal-Wallis

Prueba no paramétrica que permite identificar diferencias entre tres o más poblaciones.

Coeficiente de correlación de rango de Spearman

Es una medida de la correlación basada en datos ordenados para dos variables.

MESOGRAFÍA

Bibliografía básica