DOCIMAS PARA DOS MUESTRAS RELACIONADAS

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

PRUEBA DE SIGNOS

Prueba aplicable a experimentos en muestras pareadas o muestras relacionadas, contrasta la hipótesis de que ambos tratamientos no difieren o que pertenecen a poblaciones idénticas.

Consiste en datos u observaciones de m.a. bivariadas esto es "n" pares de observaciones, cada par (x_i, y_i) es comparado colocándole signos:

+ si
$$x_i > y_i$$
 entonces $x_i - y_i > 0$

0 si
$$x_i = y_i$$
 entonces $xi - yi = 0$

- si
$$x_i < y_i$$
 entonces $x_i - y_i < 0$

La escala de medida es la ordinal.

ESTADISTICA NO PARAMETRICA

a)Unilateral Izquierda (1 cola):

 H_0 : La mediana de las diferencias ≥ 0 ó $P(+) \geq P(-)$

 H_1 : La mediana de las diferencias < 0 ó P(+) < P(-)

b) Unilateral Derecha (1 cola):

 H_0 : La mediana de las diferencias ≤ 0 ó $P(+) \leq P(-)$

 H_1 : La mediana de las diferencias > 0 ó P(+) > P(-)

c) Bilateral (2 colas):

 H_0 : La mediana de las diferencias = 0 ó P(+) = P(-)

 H_1 : La mediana de las diferencias (\neq) 0 ó P(+) (\neq) P(-)

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Estadístico de contraste o prueba:

a) El estadístico de prueba para muestras pequeñas (N < 20), se basa en un ajuste de las diferencias observadas a una distribución binomial con probabilidad constante 0.5, y se calcula :

Unilateral:

$$p = P(X \le x) = \sum_{i=0}^{x} {n \choose i} (0.5)^{N}$$

Donde:

N=número de pares cuyas diferencias son diferentes de cero(0).

p = probabilidad asociada con la ocurrencia de H_0 de valores tan pequeños como "x".

x = frecuencia del signo menos frecuente.

ESTADISTICA NO PARAMETRICA

Bilateral:

$$p = P(X \le x_1) + P(X \ge x_2)$$

$$p = \sum_{i=0}^{x_1} {\binom{N}{i}} (0.5)^N + \sum_{i=x_2}^{N} {\binom{N}{i}} (0.5)^N$$

Donde:

N=número de pares cuyas diferencias son diferentes de cero(0).

p = probabilidad asociada con la ocurrencia de H₀ de valores tan pequeños como "x".

 X_1 = frecuencia del signo menos frecuente.

 x_2 = frecuencia del signo de mayor frecuencia.

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

b) Cuando el numero de muestras es grandes ($N \ge 20$), las diferencias observadas son aproximadas de una distribución binomial a una distribución normal, entonces:

$$Media = \mu_x = \frac{N}{2}$$

$$Varianza = \sigma_x^2 = \frac{N}{4}$$

Donde:

N=número de pares cuyas diferencias son diferentes de cero(0).

x = frecuencia del signo menos frecuente.

$$Z = \frac{x - \mu_X}{\sigma_X} = \frac{x - \frac{N}{2}}{\frac{\sqrt{N}}{2}} \approx N(0,1)$$

La aproximación se hace excelente cuando se aplica una corrección de continuidad.

ESTADISTICA NO PARAMETRICA

La corrección se hace reduciendo la diferencia entre el número de observados de más (o menos) conforme a $\rm H_0$, en 0.5, con la corrección de continuidad tendremos que:

$$Z = \frac{(x \pm 0.5) - \frac{N}{2}}{\frac{\sqrt{N}}{2}}$$
 Donde

Para el caso de considerar el signo menos frecuente:

$$x + 0.5$$
 para $x < \frac{N}{2}$

Para el caso de considerar el signo de mayor frecuencia:

$$x$$
 - 0.5 para $x > \frac{N}{2}$

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Regla de Decisión:

Se rechaza H_0 , si: p-valor $\leq \alpha$ ó Si:

> Unilateral:

 \checkmark Unilateral Derecha: $Z_c > Z_{tabla,(lpha)}$

 \checkmark Unilateral Izquierda: $Z_c < -Z_{tabla,(lpha)}$

> Bilateral:

$$Z_c \le -Z_{tabla,\frac{\alpha}{2}}$$
 o $Z_c \ge Z_{tabla,\frac{\alpha}{2}}$

También:

Si p-valor $\leq \alpha/2$ rechazamos H_0 .

ESTADISTICA NO PARAMETRICA

Aplicativo 1:

En un estudio diseñado para determinar la eficacia de una nueva dieta, una compañía de seguros selecciona una muestra de 12 sujetos con sobre peso entre 40 y 50 años. Toma la medida de los pesos antes de iniciar la dieta y 60 días después.

Formular y contrastar la hipótesis nula apropiada con α=0.05

Individuo	1	2	3	4	5	6	7	8	9	10	11	12
Antes	90	105	76	71	82	93	92	85	84	84	78	81
Después	84	98	77	71	80	87	91	85	82	83	76	82

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Solución:

Planteamos las hipótesis:

 H_0 : La dieta no disminuye el sobre peso, $P(+) \le P(-)$

 H_1 : La dieta disminuye el sobre peso, P(+) > P(-)

se evaluará con un nivel de significancia de α=0.05

Comparando los resultados antes y después de la aplicación de la nueva dieta, se asignará los signos más y menos según corresponda:

Individuo	1	2	3	4	5	6	7	8	9	10	11	12
Antes	90	105	76	71	82	93	92	85	84	84	78	81
Después	84	108	74	71	80	93	92	85	82	84	76	82
	+	-	+	0	+	0	0	0	+	0	+	-

Entonces tenemos N=7 (no se consideran los ceros) y X=2 (frecuencia del signo menos frecuente)

ESTADISTICA NO PARAMETRICA

Calculamos el estadístico de prueba:

como N<20, calculamos la probabilidad a través del ajuste a la distribución binomial:

$$p = P(X \le 2) = \sum_{i=0}^{2} {7 \choose i} (\frac{1}{2})^{7} = 0.23$$

Luego p> 0.05, no rechazamos H_0 y concluimos que con un riesgo del 5% no existe evidencia suficiente para decir que la dieta disminuye de peso.

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Aplicativo 2:

Honda probó la resistencia al uso de dos tipos de bandas de rodamiento de los neumáticos en su motocicleta Night. Se seleccionaron aleatoriamente una muestra de 40 motos. Los mecánicos montaron los neumáticos con un tipo de banda en el frente y la otra banda de rodamiento atrás.

Después de manejar las motocicletas un número de millas especificando bajo las condiciones establecidas, se obtuvo en comparación entre el tipo de banda 1 y tipo de banda 2 : 8 signos más, 28 signos menos y 4 diferencias cero.

El analista de Honda desea probar la hipótesis de que no hay diferencia en las clasificaciones de desgaste al nivel del 10%.

ESTADISTICA NO PARAMETRICA

Solución:

Planteamos las hipótesis:

 H_0 : No existe diferencia en el desgaste de las bandas, P(+) = P(-)

 H_1 : Existe diferencia entre los desgastes de las 2 bandas, $P(+) \neq P(-)$

se evaluará con un nivel de significancia de α=0.10

Como n>20, utilizaremos la aproximación normal.

Entonces tenemos N=36 (no se consideran los 4 ceros) y X=8 (frecuencia del signo menos frecuente), calculando el estadístico de prueba considerando el signo menos frecuente:

$$Z_c = \frac{(x+0.5) - \frac{N}{2}}{\frac{\sqrt{N}}{2}} = \frac{(8+0.05) - 0.5(36)}{0.5(\sqrt{36})} = -3.17$$

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

$$Z_c = -3.17$$

Regla de decisión, rechazamos H_0 si :

$$Z_c \le -Z_{tabla\frac{\alpha}{2}}$$
 o $Z_c \ge Z_{tabla\frac{\alpha}{2}}$, $Z_{tabla0.05} = -1.64$

Entonces: -3.17< -1.64, se rechaza H_0

Concluimos que con un riesgo del 10% existe diferencia significativa entre los desgastes de los dos tipos de bandas de neumáticos de motocicletas en estudio.

ESTADISTICA NO PARAMETRICA

Prueba de McNemar

La prueba de McNemar se utiliza para decidir si puede o no aceptarse que determinado "tratamiento" induce un cambio en la respuesta de los elementos sometidos al mismo, y es aplicable a los diseños del tipo "antes-después" en los que cada elemento actúa como su propio control.

Consiste en n observaciones de una v.a. bidimensional (X_i, Y_i) para i=1,...,n.

La escala de medición para X e Y es nominal con 2 categorías tales como positivo o negativo ó hembra, macho ó presencia, ausencia, que se puede denominar "0" y "1".

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

En la dócima de McNemar, los resultados se presentan en una tabla 2x2 en la siguiente forma:

		Clasific	cación Y _i
		(+) Y _i = 0	(-) Y _i = 1
	(+)	A	B
Clasificación X _i	$X_i = 0$	(0,0)	(0,1)
·	(-)	C	D
	$X_i = 1$	(1,0)	(1,1)

En el cuadro de entrada de frecuencias se presenta al primero y al segundo conjunto de respuestas de los mismos individuos.

Obs.: En la tabla se usan + y – para simbolizar respuestas diferentes.

ESTADISTICA NO PARAMETRICA

			cación Y _i
		(+)	(-)
Clasificación X _i	(+)	A	В
	(-)	С	D

Los casos que muestran cambios entre la primera y segunda respuesta aparecen en las celdillas B y C. Un individuo es clasificado en la celdilla B si cambió de + a - . Es clasificado en la celdilla C si cambió de – a +. Si no es observado ningún cambio, va a la celdilla A (respuestas de + antes y después) o a la celdilla D (respuestas de – antes y después).

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Hipótesis:

- H₀ :EI "tratamiento" no induce cambios significativos en las respuestas, es decir, los cambios observados en la muestra se deben al azar
- H₁: El "tratamiento" induce cambios significativos en las respuestas, es decir, los cambios observados en la muestra no se deben al azar

En esta prueba para la significación de cambios solamente nos va a interesar conocer las celdas que presentan cambios ("B" y "C") y puesto que B+C es el número de personas que cambiaron, se espera de acuerdo a la hipótesis planteada H₀ que (B+C)/2 casos cambien en una dirección y (B+C)/2 casos a otra dirección.

ESTADISTICA NO PARAMETRICA

Sea: B+C=n

Estadístico de prueba:

1) Si n < 20 la estadística de contraste es:

$$T_1 = B$$

Regla de Decisión:

Para esta estadística de contraste, rechazamos H₀, si:

$$T_1 \ge \chi^2_{\alpha(1g.l.)}$$

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Sea: B+C=n

2) Si n ≥ 20 entonces la estadística de contraste es:

$$T_1 = \frac{(B-C)^2}{B+C} \approx \chi_{(1)}^2$$

La aproximación es más precisa si se realiza la corrección de continuidad de yates, quedando el estadístico:

$$T_1 = \frac{(|B-C|-1)^2}{B+C} \approx \chi_{(1)}^2$$

ESTADISTICA NO PARAMETRICA

Regla de Decisión:

Para esta estadística de contraste, rechazamos H₀, si:

$$T_1 \ge \chi^2_{\alpha(1g.l.)}$$

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Aplicación 1:

En un debate televisivo a nivel nacional entre 2 candidatos presidenciales, se toma una muestra de 100 personas a quienes se les pregunta sobre sus candidatos antes del debate de los cuales 84 resultaron convencidos por el candidato demócrata y 16 al candidato republicano. Después del debate las mismas 100 personas son vueltas a preguntar acerca de sus preferencias, y del grupo de personas que favorecieron al demócrata cambian exactamente ¼ de ellos su manera de pensar y también ¼ de los republicanos.

ESTADISTICA NO PARAMETRICA

Solución:

Planteamos las hipótesis

 H_0 : Los votantes no han cambiado su manera de pensar

H₁: Los votantes han cambiado su manera de pensar debido al debate

		Después	del debate	
		Demócratas	Republicanos	
Antes del	Demócratas	A 63	B 21	84
debate	Republicanos	C 4	D 12	<u>16</u>
	_	_		100

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Calculamos el estadístico de prueba:

$$T_1 = \frac{(|B-C|-1)^2}{B+C} = \frac{(|21-4|-1)^2}{25} = 10.24$$

$$\chi^2_{\alpha=0.05,(1g.l)} = 3.84$$

Por lo tanto H₀ es rechazada.

<u>Concluimos</u> con un riesgo del 5%, existe evidencia suficiente para afirmar que los votantes han cambiado su manera de pensar debido al debate.

ESTADISTICA NO PARAMETRICA

Aplicación 2:

Se requiere saber si en los hospitales de la Solidaridad (ambulatorio) la asistencia de salud es del mismo tipo que en los Centros de Salud. Para ello se elije al azar a 125 individuos para hacer el seguimiento de la atención en ambos establecimientos tomando la opinión de cada uno de ellos respecto al tipo de asistencia recibida en el Centro de Salud y posteriormente, con la recibida en el Hospital de la Solidaridad.

Para dicho fin, se nos proporciona la siguiente información para determinar en base a un α =0.05, si existe alguna diferencia entre la asistencia de salud brindada por los 2 establecimientos.

CUADRO 1

CUADRO I			
	OPI	NIÓN	Total
	SI NO		Iotai
Hospital de la Solidaridad	62	63	125
Centro de Salud	70	55	125
Total	132	118	250

CUADRO 2

	Hospital de l		
Centro de Salud	SI	Total	
SI	27	43	70
NO	35	20	55
Total	62	63	125

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Solución:

Planteamos las hipótesis

- ${
 m H_0}$: No existe diferencia entre la atención brindada entre el Hospital de la Solidaridad y los Centros de Salud.
- H₁: Las atenciones brindadas entre el Hospital de la Solidaridad y los Centros de Salud difieren.

CUADRO 2

	Hospital d	Total	
Centro de Salud	SI	iotai	
	A	В	
SI	27	43	70
NO	35 C	20	55
Total	62	63	125

N=B+C=43+35=78

ESTADISTICA NO PARAMETRICA

Calculamos el estadístico de prueba:

$$T_1 = \frac{(|B - C| - 1)^2}{B + C} = \frac{(|43 - 35| - 1)^2}{43 + 35} = 0.63$$

$$\chi^2_{\alpha=0.05,(1g.l)} = 3.84$$

Por lo tanto H_0 No se rechaza.

Por lo tanto, con un riesgo del 5%, existe evidencia suficiente para afirmar que las atenciones brindadas por el Hospital de la Solidaridad y los Centros de Salud son iguales.

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

La Prueba de Rangos Señalados y Pares Igualados de Wilcoxon para Dos Muestras Relacionadas

Sean $(X_1, Y_1),..., (X_n, Y_n)$ observaciones independientes.

Donde X_i es el resultado correspondiente al control, e Y_i el correspondiente al tratamiento. Sea:

$$D_i=X_i-Y_i$$
 Para todo $i=1,...,n$

La prueba de Wilcoxon para datos pareados o relacionados consiste en la aplicación de la prueba a las diferencias " D_i ", suponiendo que la distribución de las diferencias es simétrica.

ESTADISTICA NO PARAMETRICA

Hipótesis a contrastar:

H₀: No existe diferencia entre las observaciones pareadas o relacionadas.

No hay efecto del tratamiento.

H₁: Existe diferencia significativa entre las observaciones pareadas o relacionadas.

El efecto del tratamiento produce valores grandes

El efecto del tratamiento produce valores pequeños

Nivel de significación : a

Para el caso de Wilcoxon para una muestra se incluía la magnitud de la mediana, para el caso de muestras relacionadas o apareadas se incluye la magnitud de las diferencias.

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Estadístico de Prueba:

Para calcular el estadístico de prueba, tomaremos las diferencias " $d_i = x_i - y_i$ " y ordenarlas de menor a mayor prescindiendo del signo y asignarles rango desde 1, 2, . . . ,n. Los valores iguales a cero se ignoran y si varias " d_i " son iguales se les asigna el rango promedio de los valores empatados.

a)Para N<15, N=número de pares cuyas diferencias son diferentes de cero(0) :

Consideramos el estadístico:

 $T = min(T_+, T_-)$.

Siendo:

 T_{+} = Suma de los rangos correspondientes a las diferencias positivas.

 T_{-} = Suma de los rangos correspondientes a las diferencias negativas.

ESTADISTICA NO PARAMETRICA

Regla de Decisión:

- Para una prueba Unilateral:

 $T \leq W_{N,\alpha}$



Se rechaza H₀

se puede contrastar también con el p-valor (p):

p ≤ α



Se rechaza H₀, si:

- Prueba Bilateral:

 $T \le W_{N,\alpha/2}$



Se rechaza H₀, si:

se puede contrastar también con el p-valor (p):

 $p \le \alpha/2$



Se rechaza H₀, si:

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

b) Para N ≥ 15, se puede aproximar T (bajo la hipótesis nula) a una normal con :

$$\mu_T = \frac{N(N+1)}{4}$$

$$Varianza = \sigma_T^2 = \frac{N(N+1)(2N+1)}{24}$$

$$Z = \frac{T - \mu_T}{\sigma_T} \approx N(0,1)$$

N=número de pares cuyas diferencias son diferentes de cero(0):

ESTADISTICA NO PARAMETRICA

Regla de Decisión:

Se rechaza H₀, si:

> Unilateral:

 \checkmark Unilateral Derecha: $Z_c > Z_{tabla,(\alpha)}$

 \checkmark Unilateral Izquierda: $Z_c < -Z_{tabla,(lpha)}$

También Si : p-valor $\leq \alpha$

➤ Bilateral:

$$Z_c \le -Z_{tabla,\frac{\alpha}{2}}$$
 o $Z_c \ge Z_{tabla,\frac{\alpha}{2}}$

También:

Si p-valor $\leq \alpha/2$ rechazamos H_0 .

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Aplicativo 1:

Se realizó un estudio comparativo en el cual se evaluó la efectividad de dos métodos, uno tradicional y uno moderno de enseñanza del álgebra. En ese estudio 14 individuos fueron extraídos al azar de la población de interés y se formaron 7 pares en base a su coeficiente intelectual. Los miembros de cada par fueron asignados al azar a uno de los dos métodos de enseñanza, y posteriormente ambos grupos fueron instruidos durante 3 semanas. Todos los estudiantes rindieron el mismo examen al final del periodo de instrucción y los resultados obtenidos fueron los siguientes:

Par	Tradicional	Moderno
1	36	31
2	38	42
3	33	44
4	36	48
5	53	51
6	49	57
7	52	62

ESTADISTICA NO PARAMETRICA

Solución:

Planteamos las hipótesis:

H₀: Los dos métodos tienen la misma efectividad.

H₁: El método moderno es mucho mas efectivo que el tradicional.

Establecemos el nivel de significación que se usará en la prueba:α=0.05

Calculamos del estadístico de prueba

ESTADISTICA NO PARAMETRICA

	Par	Tradicion	nal N	Moderno	D _i		
	1	36		31	-5		
	2	42		38	-4		
	3	33		44	11		
	4	36		48	12		
	5	53		51	-2		
	6	49		57	8		
	7	62		52	-10		
$ D_i $	2	4	5	8	10	11	12
ID _i I						11	
Rango	1	2	3	4	5	6	7
Considerando el signo	-2	-4	-5	8	-10	11	12
Rango con signos	-1	-2	-3	4	-5	6	7

Luego: N=7

 $T = min(T_+, T_-)$: estadístico de prueba para N<25

Donde:

$$T_{+}=4+6+7=17$$
 $T=Min(11,17) \longrightarrow T=11$

T₌1+2+3+5=11

Tabla XI Test signo-rango de Wilcoxon

Una cola	Dos colas	n = 5	n = 6	n = 7	n = 8	n = 9	n = 10
P = .05 $P = .025$ $P = .01$ $P = .005$	P = .10 P = .05 P = .02 P = .01	1	2 1	4	6 4 2	8 6 3	11 8 5
Line cole	Dog colog	9 = 11	2 - 12	- 13	2 - 14	- 15	- 16

Decisión:

T=11> $W_{7,0.05}$ =4, por lo tanto con un nivel de significancia del 5% No rechazamos H_0 .

Conclusión: podemos afirmar con un riesgo del 5% que el método moderno no es más efectivo que el método tradicional.

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Aplicativo 2:

Para proteger contra las verrugas se ensayaron 2 cremas A y B, aplicadas en dos zonas diferentes de cada paciente.

Posteriormente se midió el diámetro en mm de la zona afectada por verrugas, donde hemos aplicado la crema A y donde hemos aplicado la crema B. los resultados a las medidas de una muestra de 8 pacientes son los siguientes:

Α	4	3	6	7	9	4	2	5
В	3	4	6	6	5	4	7	3

Determinar si existen diferencias entre los resultados de las cremas A y B. α =0,05

ESTADISTICA NO PARAMETRICA

Solución:

Planteamos las hipótesis:

 $\mbox{H}_{\mbox{\scriptsize 0}}$: Las diferencias entre las zonas tratadas por las cremas A y B son nulas.

H₁: Las diferencias entre las zonas tratadas por las cremas A y B son significativas.

ESTADISTICA NO PARAMETRICA

LIC. RITA GUZMAN LOPEZ

Α	4	3	6	7	9	4	2	5
В	3	4	6	6	5	4	7	3
А-В	1	-1	0	1	4	0	-5	2

$ D_i $	1	1	1	2	4	5
Rango	2	2	2	4	5	6
Considerando el signo	1	-1	1	2	4	-5
Rango con signos	2	-2	2	4	5	-6

Luego: N=6

Donde:

$$T_{+}=2+2+4+5=13$$

 $T_{-}=2+6=8$
 $T=Min(13,8) \longrightarrow T=8$

ESTADISTICA NO PARAMETRICA

Tabla XI Test signo-rango de Wilcoxon

Una cola	Dos colas	n = 5	n = 6	n = 7	n = 8	n = 9	n = 10
P = .05	P = .10	1	2	4	6	8	11
P = .025	P = .05	22 W	1	2.	100 4 1	6	8
P = .01	P = .02	i .	30.11	0	2	. 3	5
P = .005	P = .01	Record	G 1 .	F 17 1-12	. 0	2	3
Una cola	Dos colas	n = 11	n = 12	n = 13	n = 14	n = 15	n = 16

Decisión:

Dado que T=8> $W_{6,0.025}$ =1, con un nivel de significancia del 5% No rechazamos $H_{\rm o}$.

Conclusión: podemos afirmar con un riesgo del 5% que no se detectan diferencias estadísticamente significativas entre las zonas protegidas por la crema A y la crema B.

ESTADISTICA NO PARAMETRICA