Prueba de Rango con Signo Wilcoxon para dos Muestras Dependientes

Esta lección se ocupa de la prueba de rangos con signo de Wilcoxon que se usa con datos muestrales apareados. Esta prueba se utiliza con la hipótesis nula de que la población de diferencias de los datos apareados tiene una mediana igual a cero.

Definición

La prueba de rangos con signo de Wilcoxon es una prueba no paramétrica que utiliza rangos ordenados de datos muestrales que consisten en datos apareados. Se usa para probar la hipótesis nula de que la población de diferencias tiene una mediana de cero, de manera que las hipótesis nula y alternativa son las siguientes:

H0: Los datos apareados tienen diferencias que provienen de una población con una mediana igual a cero.

H1: Los datos apareados tienen diferencias que provienen de una población con una mediana diferente de cero.

Requisitos

1. Los datos consisten en datos apareados que se seleccionaron aleatoriamente.

2. La población de las diferencias (calculadas a partir de los pares de datos) tiene una distribución que es aproximadamente simétrica, lo que quiere decir que la mitad izquierda de su histograma es aproximadamente una imagen de espejo de la mitad derecha (no existe el requisito de que los datos tengan una distribución normal).

Notación

T= la más pequeña de las siguientes dos sumas:

1. La suma de los valores absolutos de los rangos negativos de las diferencias d que no sean cero.

2. La suma de los rangos positivos de las diferencias d que no sean cero.

Estadístico de prueba

Si n ≤ 30, el estadístico de prueba es T.

Si n > 30, el estadístico de prueba es

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Valores críticos

1. Si n ≤ 30, el valor crítico T se encuentra en la tabla a A-8. (Dar clic para visualizar la tabla)

2. Si n > 30, los valores críticos z se encuentran en la tabla A-2. (Dar clic para visualizar la tabla)

Procedimiento de la prueba de rangos con signo de Wilcoxon

Paso 1: Para cada par de datos, calcule la diferencia d restando el segundo valor del primero. Mantenga los signos, pero descarte cualquier par para el que d =0.

Paso 2: Ignore los signos de las diferencias, luego acomode las diferencias de la menor a la mayor y reemplácelas por el valor del rango correspondiente. Cuando las diferencias tengan el mismo valor numérico, asígneles la media de los rangos implicados en el empate.

Paso 3: Agregue a cada rango el signo de la diferencia de la que provino. Esto es, inserte aquellos signos que se ignoraron en el paso 2.

Paso 4: Calcule la suma de los valores absolutos de los rangos negativos. También calcule la suma de los rangos positivos.

Paso 5: Permita que T sea la más pequeña de las dos sumas calculadas en el paso 4. Podría utilizarse cualquier suma, pero para simplificar el procedimiento seleccionamos arbitrariamente la más pequeña de las dos sumas.

Paso 6: Permita que n sea el número de pares de datos para los que la diferencia d no es 0.

Paso 7: Determine el estadístico de prueba y los valores críticos con base en el tamaño muestral.

Paso 8: Cuando plantee la conclusión, rechace la hipótesis nula si los datos muestrales le llevan a un estadístico de prueba que se ubica en la región crítica, esto es, cuando el estadístico de prueba sea menor o igual que el valor (o los valores) crítico(s). De otra forma, no rechace la hipótesis nula.

EJEMPLO

En 1908 William Gosset publicó el artículo "The Probable Error of a Mean" bajo el seudónimo de "Student". Él incluyó los datos que se listan en la tabla 2

Tabla 1. Cosechas de maíz de diferentes semillas

Normales	1903	1935	1910	2496	2108	1961	2060	1444	1612	1316	1511
Secadas en horno	2009	1915	2011	2463	2180	1925	2122	1482	1542	1443	1535
Diferencia d	-106	20	-101	33	-72	36	-62	-38	70	-127	-24
Rangos de |diferencias|	10	1	9	3	8	4	6	5	7	11	2
Rango con signo	-10	1	-9	3	-8	4	-6	-5	7	-11	-2

para dos tipos diferentes de semillas de maíz (normales y secadas en horno), que se utilizaron en parcelas de tierra *adyacentes*. Los valores corresponden a las cosechas de cabezas de maíz (o mazorcas) en libras por acre. Utilice la prueba de rangos con signos de Wilcoxon, con un nivel de significancia de 0.05, para probar la aseveración de que no hay diferencia entre las cosechas de las semillas normales y de las semillas secadas en horno.

Solución.

Debemos tener datos apareados seleccionados al azar. Los datos están apareados y, dado el diseño de este experimento, es razonable suponer que los datos apareados fueron elegidos al azar.

Las hipótesis nula y alternativa son como sigue:

$H0$: Las cosechas de las semillas normales y de las semillas secadas en horno son tales que la mediana de la población de las diferencias es igual a cero.

$H1$: La mediana de la población de diferencias no es igual a cero.

Prueba de Rango con Signo Wilcoxon para dos Muestras Dependientes

El nivel de significancia es $\alpha = 0.05$. Estamos utilizando el procedimiento de la prueba de rangos con signo de Wilcoxon, por lo que el estadístico de prueba se calcula aplicando el procedimiento de ocho pasos presentado anteriormente.

Paso 1: En la tabla 2 el renglón de diferencias se obtiene calculando esta diferencia para cada par de datos:

d= cosecha de las semillas normales - cosecha de las semillas secadas en horno

Paso 2: Ignorando sus signos, ordenamos los rangos de las diferencias absolutas de la menor a la mayor (si existiera algún empate en los rangos tendríamos que manejarlos asignando la media de los rangos implicados a cada uno de los valores empatados. Además, tendríamos que descartar cualquier diferencia de 0).

Paso 3: El renglón inferior de la tabla 2 se crea insertando a cada rango el signo de la diferencia correspondiente. Si en realidad no existe diferencia entre las cosechas de los dos tipos de semillas (como en la hipótesis nula), esperamos que la suma de los rangos positivos sea aproximadamente igual a la suma de los valores absolutos de los rangos negativos.

Paso 4: Ahora calculamos la suma de los valores absolutos de los rangos negativos y también calculamos la suma de los rangos positivos. Suma de los valores absolutos de los rangos negativos: 51 (de 10+9+8+6+5+11+2). Suma de los rangos positivos: 15 (de 1+3+4+7)

Paso 5: Permitiendo que T sea la menor de las dos sumas calculadas en el paso 4, encontramos que T =15.

Prueba de Rango con Signo Wilcoxon para dos Muestras Dependientes

Paso 6: Permitiendo que n sea el número de pares de datos para los que la diferencia d no es 0, tenemos n=11.

Paso 7: Puesto que n=11, tenemos que $n \leq 30$, por lo cual utilizamos un estadístico de prueba de T=15 (y no calculamos un estadístico de prueba z). Además, puesto que $n \leq 30$, utilizamos la tabla de A-8 para encontrar el valor crítico de 11 (utilizando n=11 y $\alpha = 0.05$ en dos colas).

Paso 8: El estadístico de prueba T=15 no es menor o igual que el valor crítico de 11, por lo que no rechazamos la hipótesis nula. Parece que no hay una diferencia entre las cosechas de las semillas normales y de las semillas secadas en horno.

Referencia:

Triola, M., (2013). Estadística. Decimoprimera edición. Pearson educación. México