ESTADÍSTICA NO PARAMÉTRICA

PRUEBAS

DE HIPÓTESIS

MET. Alejandra Cerda

Métodos no paramétricos

En éstos métodos no paramétricos se encuentran aquellas pruebas cuyas hipótesis no corresponden a una afirmación sobre un parámetro, y las pruebas de libre distribución donde su aplicación no depende de la distribución de la variable de interés en la población de estudio.

Cuando la hipótesis corresponde a un parámetro debemos evaluar sobre usar métodos paramétricos (revisados en inferencia estadística) o métodos no paramétricos. Mientras que si la hipótesis a probar corresponde a una afirmación que no es la relacionada con un parámetro los métodos no paramétricos representan la única alternativa.

Algunas ventajas

- Son más rápidos y fáciles de aplicar
- Con frecuencia, más fáciles de entender
- Relativamente insensibles a datos atípicos
- Los supuestos requeridos son, en general, más fáciles de cumplir
- Se pueden aplicar en muestras pequeñas donde no se pueden verificar los supuestos de la estadística inferencial clásica (métodos paramétricos)

Algunas pruebas no paramétricas

- Prueba del signo
- Prueba de rangos con signo de Wilcoxón
- Prueba de suma de rangos de Wilcoxón
- Prueba U de Mann-Whitney
- Prueba de rachas de Wald-Wolfowitz
- Prueba de McNemar
- Prueba de la mediana
- Prueba Kruskall-Wallis
- Prueba de Friedman
- Coeficiente de correlación de rangos de Spearman-ρ
- Prueba de homogeneidad
- Prueba de bondad de ajuste

PRUEBA DEL SIGNO

Prueba del signo

Se usa como alternativa a la prueba t, suponemos que la población es continua y simétrica. Reemplazamos cada valor que excede a $\mu 0$ con un signo "+" y con un signo "-" cada valor menos que $\mu 0$ descartando la observación que cumpla con la igualdad.

Aquí probamos la hipótesis nula del número de signos "+" es un valor de una v.a. con distribución binomial con n= numero total de signos "+" y "-" asignados y $\theta=1/2$. Así:

H_0	$H_0\colon \mu = \mu_0$ *equivale a* $H_0\colon \theta = {}^1\!/_2$	
H_1	$H_1\colon \mu > \mu_0$ *equivale a* $\theta > {}^1\!/_2$	$H_1\colon \mu < \mu_0$ *equivale a* $\theta < {}^1\!/_2$
Estadístico de prueba	X el número de signos "+"	
Región de Rechazo	$X \geq K_\alpha$ donde K_α es el entero más pequeño para el cual $\sum_{y=K_\alpha}^{n} bin(y;n,1/2) \leq \alpha$	$X \leq K'_\alpha$ donde K'_α es el entero más grande para el cual $\sum_{y=0}^{K'_\alpha} bin(y;n,1/2) \leq \alpha$

Prueba del signo para observaciones pareadas

Suponga que tenemos n pares (x_i, y_i) y que deseamos comparar la distribución de las X y Y respecto a la igualdad o diferencia de su ubicación. Consideremos las diferencias

$D_i = x_i - y_i$ que no muestran empate

H_0	$p = {}^1/_2$ (las distribuciones de X y Y no difieren en localización)	
H_1	$p > {}^1/_2$ (la distribución de X se localiza a la derecha de distribución de Y)	$p < {}^1/_2$ (la distribución de X se localiza a la izquierda de distribución de Y)
Estadístico de prueba	M= número de diferencias positivas	
Región de Rechazo	$M \geq K_\alpha$ donde K_α es el entero más pequeño para el cual $\sum_{y=K_\alpha}^{n} bin(y; n, 1/2) \leq \alpha$	$M \leq K'_\alpha$ donde K'_α es el entero más grande para el cual $\sum_{y=0}^{K'_\alpha} bin(y; n, 1/2) \leq \alpha$

La prueba involucra muestra(s) grande(s)

De lo anterior, la prueba puede realizarse al comparar una muestra respecto a un valor específico µ0 o respecto a la comparación con otra muestra. Si la muestra o muestras involucradas son suficientemente grandes (mayor a 30) se puede hacer uso de la aproximación de la distribución binomial a la distribución normal.

Así por ejemplos para comparar una muestra con el valor μ_0 se tiene:

H_0	$H_0: \mu = \mu_0$	
H_1	$H_1: \mu > \mu_0$	$H_1: \mu < \mu_0$
Estadístico de prueba	$Z = \frac{X-n\theta}{\sqrt{n\theta(1-\theta)}}$ donde $\theta = 0.5$ y X es el número de signos "+"	
Región de Rechazo	$Z \geq Z_\alpha$	$Z \leq -Z_\alpha$

PRUEBA DE RANGOS CON SIGNO DE WILCOXÓN

Prueba de rangos con signo

Suponga que tenemos n pares (x_i, y_i) y que deseamos comparar la distribución de las X y Y respecto a la igualdad o diferencia de su ubicación. Consideremos las diferencias

$D_i = x_i - y_i$ que no muestran empate

H_0	Las distribuciones poblacionales de X y Y son idénticas		
H_1	La distribución de X se localiza a la derecha de distribución de Y	La distribución de X se localiza a la izquierda de distribución de Y	Las distribuciones poblacionales difieren en localización
Estadístico de prueba	T^-= suma de rangos de las diferencias "-"	T^+= suma de rangos de las diferencias "+"	$\mathrm{T} = \min(T^+, T^-)$
Región de Rechazo	$T^- \leq T_0$ el valor crítico para prueba unilateral	$T^+ \leq T_0$ el valor crítico para prueba unilateral	$\mathrm{T} \leq T_0$ el valor critico para prueba de dos colas

¿Cómo asignar el rango?

- Considerando las observaciones ordenadas de menor a mayor
- Los rangos se asignan de 1 en adelante
- Considerando observaciones con empate se asigna el promedio de los rangos involucrados. Por ejemplo, si los siguientes rangos son 3 y 4 y es empate a ambos se les asigna rango 3.5
- Se continua con el rango siguiente a los rangos involucrados en el empate(el cual no necesariamente es el rango asignado)
- No necesariamente la observación mayor tendrá signado el rango n

orden	observación	rango
1	1.5	1
2	2	2
3	2.3	3.5
4	2.3	3.5
5	2.9	5
6	3	6.5
7	3	6.5

Variaciones de la prueba

La prueba se puede llevar a cabo al considerar las magnitudes de las diferencias de la observación de una muestra y µ0 . Continuando con la asignación del rango como ya se ha explicado.

Si la(s) muestra(s) son consideradas grandes, T^+ es una variable aleatoria con distribución Z con parámetros dados por

$$\mu = \frac{n(n+1)}{4} \qquad \text{y} \qquad \sigma^2 = \frac{n(n+1)(2n+1)}{24}$$

por simetría los resultados son validos de igual forma si sustituimos T^- en lugar de T^+ . Así se tiene una distribución normal y se pude utilizar el estadístico de prueba Z

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

PRUEBA U DE MANN-WHITNEY

Prueba U

Considere dos muestras aleatorias e independientes cuyas hipótesis a contrastar son:

H_0	H_0: las distribuciones de las poblaciones I y II son idénticas		
H_1	las distribuciones tienen localizaciones diferentes	la distribución de la población I se desplaza a la derecha de la distribución de población II	la distribución de la población I se desplaza a la izquierda de la distribución de población II
Estadístico de prueba	$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - w$ donde n_1 es el número de observaciones de la muestra I, la muestra más pequeña w es la suma de rangos de la muestra I, donde el rango se obtiene al ordenar las observaciones de las dos muestras		
Región de Rechazo	$U \leq U_0$ ó $U \geq n_1 n_2 - U_0$ donde $P(U \leq U_0) = \propto/2$	$U \leq U_0$ donde $P(U \leq U_0) = \propto$	$U \geq n_1 n_2 - U_0$ donde $P(U \leq U_0) = \propto$

Uso de la distribución Z

Se puede demostrar que el estadístico U, para distribuciones poblacionales idénticas, tiene

$$E(U) = \frac{n_1 n_2}{2} \qquad \text{y} \qquad Var(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Cuando $n_1, n_2 > 10$ se tiene una distribución aproximadamente normal , así se pude utilizar el estadístico de prueba Z

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Uso de la distribución Z

Así la prueba U simplificada para muestras grandes esta dada por :

H_0	H_0 :las distribuciones de las poblaciones I y II son idénticas		
H_1	las distribuciones tienen localizaciones diferentes	la distribución de la población I se desplaza a la derecha de la distribución de población II	la distribución de la población I se desplaza a la izquierda de la distribución de población II
Estadístico de prueba	$Z = \dfrac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$		
Región de Rechazo	$Z \leq -Z_{\propto/2}$ ó $Z \geq Z_{\propto/2}$	$Z \leq -Z_\alpha$	$Z \geq Z_\alpha$

PRUEBA KRUSKAL-WALLIS

Prueba Kruskall-Wallis

Esta prueba no paramétrica permite probar si las poblaciones difieren en localización, bajo la suposición de que las muestras son aleatorias e independientes, tomadas de k poblaciones que difieren sólo quizá en localización, no es necesario suponer que las poblaciones poseen distribuciones normales.

La prueba Kruskall-Wallis considera el estadístico H para compara k distribuciones poblacionales basadas en muestras grandes (aquí "grande" es considerar cinco o más mediciones en cada muestra)

Prueba Kruskall-Wallis

Hipótesis a contrastar

H0: Las k distribuciones poblacionales son idénticas

Ha: Al menos dos de las k distribuciones poblacionales difieren en localización

Estadístico de Prueba

$$H = \frac{12}{n(n+1)} \sum \frac{{R_i}^2}{n_i} - 3(n+1)$$

donde n_i es el número de observaciones en la muestra de la población i

R_i es la suma de rangos para la muestra i, donde el rango se obtiene al ordenar las observaciones de todas las muestras

Rechazando H0 si $\mathrm{H} \geq X^2_{\alpha,k-1}$ donde k es el número de poblaciones a comparar

Consideraciones finales

- El termino de muestra “grande” no siempre es considerado de igual forma.
- La asignación de rangos al utilizar diferencias no considera diferencia 0.
- Para las pruebas anteriormente mencionadas usan el mismo criterio de asignación del rango promedio al tener diferencias con empate.
- Los rangos no siempre consideran todos los datos, en ocasiones se usa la asignación de rangos para las diferencias.
- La prueba Kruskall-Wallis se puede usar al comparar a partir de 2 poblaciones.
- $X^2_{\alpha,k-1}$ se calcula mediante el uso de la cola derecha.
- La prueba de rangos con signo y la prueba U tienen sus correspondientes tablas estadísticas.