Pruebas no paramétricas U de Mann-Whitney & Kruskall-Wallis

U de Mann-Whitney & Kruskall-Wallis

Mg. RMMB Fisarth[1]

[1]Departamento de matemática
Universidad Nacional de San Cristobal de Huamanga & ESFA FGPA

Sesión ***prueba de hipótesis no paramétrico*** (Estadística inferencial), 2024

Índice

1. Motivación
   - El problema básico que estudiamos
   - Trabajo previo
2. Nuestros resultados/Contribución
   - Prueba U de Mann-Whitney
   - Prueba de Kruskal-Wallis

Índice

1 Motivación
- El problema básico que estudiamos
- Trabajo previo

2 Nuestros resultados/Contribución
- Prueba U de Mann-Whitney
- Prueba de Kruskal-Wallis

Prueba de hipótesis (U de Mann-Whitney & Kruskall Wallis)

- Son pruebas no parámetricas (sin supuesto de normalidad y homocedasticidad)
- Trabaja con variables *ordinales y cuantitativas no normales*
- En estudios transversales de dos o más *grupos no relacionados*

Prueba de hipótesis (U de Mann-Whitney & Kruskall Wallis)

- Son pruebas no parámetricas (sin supuesto de normalidad y homocedasticidad)
- Trabaja con variables *ordinales y cuantitativas no normales*
- En estudios transversales de dos o más *grupos no relacionados*

Índice

1. Motivación
   - El problema básico que estudiamos
   - Trabajo previo
2. Nuestros resultados/Contribución
   - Prueba U de Mann-Whitney
   - Prueba de Kruskal-Wallis

Pruebas no parámetricas

- χ^2
  - χ^2 bondad de ajuste.
  - χ^2 de independencia
  - χ^2 de homogeneidad
- McNemar
- Q de Cochran

Índice

1. Motivación
   - El problema básico que estudiamos
   - Trabajo previo
2. Nuestros resultados/Contribución
   - Prueba U de Mann-Whitney
   - Prueba de Kruskal-Wallis

U de Mann-Whitney

(Dos grupos independientes, datos ordinales o continuos no normales)

Si el tamaño de las muestras es inferior a 20, se compara el valor obtenido de U con los valores de una tabla U de Mann-Whitney. *Si el U es menor que el valor correspondiente en la tabla, la diferencia es significativa* (rechazar h_0).
Si $n_1 > 20$ y $n_2 > 20$ se puede asumir que U se distribuye de forma aproximadamente normal, rechazando h_0 si z calculado es mayor que el valor de z para el α elegido.

U de Mann-Whitney

- Hipótesis. Nula (h_0: La suma de los rangos no difieren en los dos grupos) y alterna (h_1: $\neg h_0$)
- Estadística (n_1 y n_2 mayores a 20 datos)

$$z = \begin{cases} \dfrac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}, & \text{Sin rangos empatados} \\ \dfrac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{n(n-1)}} \sqrt{\frac{n^3-n}{12} - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{12}}}, & \text{Con rangos empatados} \end{cases}$$

  donde n_1 y n_2 son el número de datos de cada grupo, $n = n_1 + n_2$, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$, $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$, R_1 y R_2 son las sumas de los *rangos de cada grupo*. Entonces $U = \min(U_1, U_2)$ además t_i es el número de observaciones repetidas dentro del grupo i.
- Nivel de significancia $\alpha = 0.05$.
- Regla de decisión. Si $|z| < z_\alpha$, se rechaza h_1 (se acepta h_0).

U de Mann-Whitney

- Hipótesis. Nula (h_0: La suma de los rangos no difieren en los dos grupos) y alterna (h_1: $\neg h_0$)
- Estadística (n_1 y n_2 mayores a 20 datos)

$$z = \begin{cases} \dfrac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}, & \text{Sin rangos empatados} \\ \dfrac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{n(n-1)}} \sqrt{\frac{n^3-n}{12} - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{12}}}, & \text{Con rangos empatados} \end{cases}$$

donde n_1 y n_2 son el número de datos de cada grupo, $n = n_1 + n_2$, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$, $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$, R_1 y R_2 son las sumas de los *rangos de cada grupo.* Entonces $U = \min(U_1, U_2)$ además t_i es el número de observaciones repetidas dentro del grupo i.

- Nivel de significancia $\alpha = 0.05$.
- Regla de decisión. Si $|z| < z_\alpha$, se rechaza h_1 (se acepta h_0).

U de Mann-Whitney

- Hipótesis. Nula (h_0: La suma de los rangos no difieren en los dos grupos) y alterna (h_1: $\neg h_0$)
- Estadística (n_1 y n_2 mayores a 20 datos)

$$z = \begin{cases} \dfrac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}, & \text{Sin rangos empatados} \\ \dfrac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{n(n-1)}} \sqrt{\frac{n^3-n}{12} - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{12}}}, & \text{Con rangos empatados} \end{cases}$$

donde n_1 y n_2 son el número de datos de cada grupo, $n = n_1 + n_2$, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$, $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$, R_1 y R_2 son las sumas de los *rangos de cada grupo.* Entonces $U = \min(U_1, U_2)$ además t_i es el número de observaciones repetidas dentro del grupo i.

- Nivel de significancia $\alpha = 0.05$.
- Regla de decisión. Si $|z| < z_\alpha$, se rechaza h_1 (se acepta h_0).

U de Mann-Whitney

- Hipótesis. Nula (h_0: La suma de los rangos no difieren en los dos grupos) y alterna (h_1: $\neg h_0$)
- Estadística (n_1 y n_2 mayores a 20 datos)

$$z = \begin{cases} \dfrac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}, & \text{Sin rangos empatados} \\ \dfrac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{n(n-1)}} \sqrt{\frac{n^3-n}{12} - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{12}}}, & \text{Con rangos empatados} \end{cases}$$

donde n_1 y n_2 son el número de datos de cada grupo, $n = n_1 + n_2$, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$, $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$, R_1 y R_2 son las sumas de los *rangos de cada grupo.* Entonces $U = \min(U_1, U_2)$ además t_i es el número de observaciones repetidas dentro del grupo i.
- Nivel de significancia $\alpha = 0.05$.
- Regla de decisión. Si $|z| < z_\alpha$, se rechaza h_1 (se acepta h_0).

Ejemplo U de Mann-Whitney

Man	Woman
45	34
33	36
35	41
39	43
42	44
	37

h_0: *No hay diferencia (en términos de tendencia central) entre los dos grupos de la población*

h_1: *Existe una diferencia (con respecto a la tendencia central) entre los dos grupos de la población.*

Ejemplo U de Mann-Whitney

- $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 6 \cdot 5 + \frac{6(6+1)}{2} - 37 = 14$
- $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 6 \cdot 5 + \frac{5(5+1)}{2} - 29 = 16$
- $U = \min\{14, 16\} = 14$, $\mu_U = \frac{n_1 n_2}{2} = \frac{6 \cdot 5}{2} = 15$, $\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{30(12)}{12}} = 5.47$
- $\frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} = \frac{14-15}{5.47} = -.1825$

$|z| = |-.1825| < z_{0.05} = 1.96$ por lo tanto (Se acepta h_0)

Ejemplo U de Mann-Whitney

- $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 6 \cdot 5 + \frac{6(6+1)}{2} - 37 = 14$
- $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 6 \cdot 5 + \frac{5(5+1)}{2} - 29 = 16$
- $U = \min\{14, 16\} = 14,\ \mu_U = \frac{n_1 n_2}{2} = \frac{6 \cdot 5}{2} = 15,\ \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{30(12)}{12}} = 5.47$
- $\frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} = \frac{14-15}{5.47} = -.1825$

$|z| = |-.1825| < z_{0.05} = 1.96$ por lo tanto (Se acepta h_0)

Ejemplo U de Mann-Whitney

- $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 6 \cdot 5 + \frac{6(6+1)}{2} - 37 = 14$
- $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 6 \cdot 5 + \frac{5(5+1)}{2} - 29 = 16$
- $U = \min\{14, 16\} = 14,\ \mu_U = \frac{n_1 n_2}{2} = \frac{6 \cdot 5}{2} = 15,\ \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{30(12)}{12}} = 5.47$
- $\frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} = \frac{14-15}{5.47} = -.1825$

$|z| = |-.1825| < z_{0.05} = 1.96$ por lo tanto (Se acepta h_0)

Ejemplo U de Mann-Whitney

- $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 6 \cdot 5 + \frac{6(6+1)}{2} - 37 = 14$
- $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 6 \cdot 5 + \frac{5(5+1)}{2} - 29 = 16$
- $U = \min\{14, 16\} = 14,\ \mu_U = \frac{n_1 n_2}{2} = \frac{6 \cdot 5}{2} = 15,\ \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{30(12)}{12}} = 5.47$
- $\frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} = \frac{14-15}{5.47} = -.1825$

$|z| = |-.1825| < z_{0.05} = 1.96$ por lo tanto (Se acepta h_0)

Índice

1. Motivación
   - El problema básico que estudiamos
   - Trabajo previo
2. Nuestros resultados/Contribución
   - Prueba U de Mann-Whitney
   - Prueba de Kruskal-Wallis

Prueba de Kruskal-Wallis

(Tres o más grupos independientes, datos ordinales o continuos no normales)

Si el tamaño de las muestras es inferior a 5, se compara el valor obtenido de H con los valores de una tabla Kruskal-Wallis. *Si el H es mayor que el valor correspondiente en la tabla, la diferencia es significativa* (rechazar h_0).

Si $n_1 > 5$, $n_2 > 5$, etc, se puede asumir que H tiene distribución χ^2_{g-1}, rechazando h_0 si H calculado es mayor que el valor de $\chi^2_{g-1,\alpha}$.

Prueba de Kruskal-Wallis

- Hipótesis. Nula (h_0 : Los rangos medios de los grupos son los mismos) y alterna (h_1: $\neg h_0$)
- Estadística $(n-1)\frac{\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot}-\bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij}-\bar{r})^2} = \frac{12}{n(n+1)}\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot}-\bar{r})^2$

$$\chi^2_{g-1,\alpha} \sim H = \begin{cases} \frac{12}{n(n+1)}\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot}-\bar{r})^2, & \text{Sin rangos empatados} \\ \frac{\frac{12}{n(n+1)}\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot}-\bar{r})^2}{1-\frac{\sum_{i=1}^{g}(t_i^3-t_i)}{n^3-n}}, & \text{Con rangos empatados} \end{cases}$$

donde n_i es el número de obervaciones en el grupo i, r_{ij} es el rango de la observacion j enel grupo i, n total de observaciones en todos los grupos, g numero de grupos $\bar{r}_{i\cdot} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$, $\bar{r} = \frac{n+1}{2}$ es el promedio de r_{ij}. Además t_i es el número de observaciones repetidas dentro del grupo i.

Prueba de Kruskal-Wallis

- Hipótesis. Nula (h_0 : Los rangos medios de los grupos son los mismos) y alterna (h_1: $\neg h_0$)
- Estadística $(n-1)\frac{\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot}-\bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij}-\bar{r})^2} = \frac{12}{n(n+1)}\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot}-\bar{r})^2$

$$\chi^2_{g-1,\alpha} \sim H = \begin{cases} \frac{12}{n(n+1)}\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot}-\bar{r})^2, & \text{Sin rangos empatados} \\ \frac{\frac{12}{n(n+1)}\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot}-\bar{r})^2}{1-\frac{\sum_{i=1}^{g}(t_i^3-t_i)}{n^3-n}}, & \text{Con rangos empatados} \end{cases}$$

donde n_i es el número de obervaciones en el grupo i, r_{ij} es el rango de la observacion j enel grupo i, n total de observaciones en todos los grupos, g numero de grupos $\bar{r}_{i\cdot} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$, $\bar{r} = \frac{n+1}{2}$ es el promedio de r_{ij}. Además t_i es el número de observaciones repetidas dentro del grupo i.

Prueba de Kruskal-Wallis

- Nivel de significancia $\alpha = 0.05$.
- Regla de decisión. Si $H > \chi^2_{g-1;\alpha}(H \geq K_\alpha)$, se rechaza h_0
- Conclusión. Con un nivel de significancia del 95% se rechaza o acepta la hipótesis nula.

Prueba de Kruskal-Wallis

- Nivel de significancia $\alpha = 0.05$.
- Regla de decisión. Si $H > \chi^2_{g-1;\alpha}(H \geq K_\alpha)$, se rechaza h_0
- Conclusión. Con un nivel de significancia del 95% se rechaza o acepta la hipótesis nula.

Prueba de Kruskal-Wallis

- Nivel de significancia $\alpha = 0.05$.
- Regla de decisión. Si $H > \chi^2_{g-1;\alpha}(H \geq K_\alpha)$, se rechaza h_0
- Conclusión. Con un nivel de significancia del 95% se rechaza o acepta la hipótesis nula.

Ejemplo de Kruskal-Wallis

Supongamos que se ha medido el tiempo de reacción de tres grupos y quieres saber si hay alguna diferencia entre ellos

A	B	C
34	44	35
36	37	39
42	45	42
43	33	46

h_0: *No hay diferencias entre los grupos.*

h_1: *Si hay diferencias entre los grupos.*

Ejemplo de Kruskal-Wallis

$n = 12$, $df = g - 1 = 2$, $E_R = \frac{n+1}{2} = 6.5$, $R_A = \frac{22}{4} = 5.5$, $R_B = \frac{27}{4} = 6.75$ y $R_A = \frac{29}{4} = 7.25$

$$
\begin{aligned}
H &= \frac{12}{n(n+1)} \sum_{i=1}^{g} n_i (\bar{r}_{i\cdot} - \bar{r})^2 \\
&= \frac{12}{12(12+1)} \left(4\,(5.5 - 6.5)^2 + 4\,(6.75 - 6.5)^2 + 4\,(7.25 - 6.5)^2 \right) \\
&= \frac{4}{13} \left((5.5 - 6.5)^2 + (6.75 - 6.5)^2 + (7.25 - 6.5)^2 \right) = .5
\end{aligned}
$$

por lo tanto $.5 = H < \chi^2_{g-1,\alpha} = \chi^2_{2,0.05} = 5.99$ (Se acepta h_0)

Sumario

- U de Mann-Whitney.
- Kruskall-Wallis
- Perspectiva
  - Probar hipótesis asociados con variables ordinales.
  - Realizar investigacion con la prueba de U de Mann-Whitney y Kruskall-Wallis.

Lecturas complementarias I

- Rufino Moya C.
  Probabilidad e Inferencia Estadística.
  2da. Edición, 1990.
- Miguel A. Gómez Villegas
  Inferencia Estadística
  2da. Edición, 2005.