Métodos No Paramétricos. II-2023
Taller 3. Ejercicios Pruebas Clásicas

Ramón Giraldo
Departamento de Estadística
Universidad Nacional de Colombia

1. Intervalos de confianza clásicos

1. Sean $Y_1, \cdots, Y_n$, variables aleatorias *iid* con $Y_i \sim N\left(\mu_0, \sigma^2\right), \forall i = 1, \ldots, n, n = 100$.

   a) Asigne valores a μ_0 y σ^2

   b) Usando `Monte Carlo` simule en `R` 1000 intervalos de confianza para μ_0, a través de $\left[\bar{y}_j \pm 1{,}96 \frac{\sigma}{\sqrt{n}}\right], j = 1, \ldots, 1000$,
   con $\bar{y}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$ (media muestral calculada en la j-ésima simulación)

   c) ¿Cuántos intervalos contienen a μ_0?. Es éste resultado coherente con lo esperado?.

2. Sean $Y_1, \cdots, Y_n$, variables aleatorias *iid* con $Y_i \sim N\left(\mu_0, \sigma^2\right), \forall i = 1, \ldots, n$, $n = 100$.

   a) Asigne valores a μ_0 y σ^2.

   b) Simule 1000 intervalos de confianza para μ, usando $\left[\bar{y}_j \pm t_{(0{,}975,99)} \frac{s_j}{\sqrt{n}}\right]$, $j = 1, \ldots, 1000$. $t_{(0{,}975,99)}$ corresponde al percentil 97.5 de una distribución T-student con 99 grados de libertad. s_j corresponde a la desviación estándar muestral de la j-ésima simulació

   - ¿Cuántos intervalos contienen a μ?. ¿Es este resultado coherente con lo esperado?.
   - Con cada simulación haga la prueba de hipótesis $H_0 : \mu = \mu_0$ vs $H_0 : \mu \neq \mu_0$. Ayuda: Con los datos de cada simulación use la función `t.test` de `R` y almacene en un archivo los valores-p de cada prueba. Después cuente cuantos valores-p son menores del 5 %. Esto dará el número total de rechazos.
   - Compruebe gráficamente que la distribución del valor-p es uniforme (0,1).

2. Pruebas No Paramétricas

3. Haga un cuadro resumen de las pruebas no paramétricas estudiadas. Debe incluir los supuestos, las hipótesis, la estadística de prueba y su distribución (incluyendo la asintótica).

4. Ejercicios libro Hollander y Wolfe. Para cada conjunto de ejercicios (de cada página) realice 2.

   a) Test de Wilcoxon: Página 50. Ejercicios 1-17

   b) Test de Signo: Página 70. Ejercicios 43-57.

   c) Test del Signo muestras pareadas. Página 82. Ejercicios 87-93

   d) Test de Wilcoxon dos muestras. Páginas 123-125. Ejercicios 1-14

   e) Test de Kruskal-Wallis. Página 199. Ejercicios 1-12

2.1. Una muestra

5. Los tiempos de sobrevivencia (en años) de 12 personas que se han sometido a un trasplante de corazón son: 3.1, 0.9, 2.8, 4.3, 0.6, 1.4, 5.8, 9.9, 6.3, 10.4, 0, 11.5. Con los datos anteriores, encuentre intervalos de confianza del 95 % para la mediana (usando `Bootstrap`) y para el coeficiente de variación (usando `Jacknife`).

6. Suponga que tiene una muestra aleatoria de $X \sim N(\theta, \sigma^2)$. Quiere probar $H_0 : \theta = 0$ vs $H_a : \theta > 0$. Dispone de una muestra de tamaño 20 ($n = 20$).

   - Encuentre una expresión para hallar el valor crítico K del test del `Signo` usando la distribución asintótica de la estadística de prueba.
   - Encuentre una expresión, basada en la distribución asintótica de la estadística del `Signo`, para calcular la potencia
   - Calcule la potencia del test del `Signo` para $p = 0{,}51, 0{,}52, \cdots, 0{,}99$ usando la expresión del item anterior.
   - Grafique la función de potencia
   - Compare la función de potencia del `Signo` con la función de potencia del test basado en la distribución `T-student`

7. Suponga que quiere probar $H_0 : \theta = 0$ vs $H_a : \theta > 0$ y que dispone de una muestra de tamaño 20 ($n = 20$).

   - Encuentre una expresión para hallar el valor crítico K del test de `Wilcoxon` usando la distribución asintótica de la estadística de prueba.
   - Encuentre una expresión, basada en la distribución asintótica de la estadística de `Wilcoxon`, para calcular la potencia de este test.
   - Usando simulación de una distribución normal $\mu = 0$, compare la potencia de los tests del `Signo`, `Wilcoxon` y `T-student`.

2.2. Dos muestra pareadas

8. Se llevó a cabo un estudio para determinar el grado en el cual el alcohol entorpece la habilidad de pensamiento para realizar una tarea. Se seleccionaron 10 personas de distintas características. Cada persona hizo la actividad sin nada de alcohol y después de haber consumido alcohol (hasta tener 0.1 %). Los tiempos antes y después (en minutos), se dan a continuación.

Antes	Después
28	35
22	40
55	25
45	37
32	30
32	30
39	58
45	51
67	34
61	48
46	30

   a) El tiempo antes es menor que el tiempo después por más de 10 minutos?. Emplee los tests del `Signo` y `Wilcoxon`. Haga las pruebas a mano y usando `R`.

   b) Realice la comparación usando un test de `permutaciones`

2.3. Dos Muestras Independientes

9. Defina el test de `Ansary-Bradley` de igualdad de varianzas (ver capítulo 5 del libro de Hollander y Wolfe).

   a) Aplíquelo a un conjunto de datos reales o simulados usando R (ver función `ansari.test`).

   b) Haga un test de `permutaciones` para comparar las varianzas de las dos poblaciones

2.4. K-Muestras

10. Use R para los siguientes ejercicios:

    - Genere muestras aleatorias de tres distribuciones normales con medias y varianzas iguales. Pruebe la hipótesis de normalidad de cada muestra usando el test de `Lilliefors`. Pruebe la hipótesis de igualdad de medianas usando la estadística de `Kruskal-Wallis`. Compruebe previamente que las varianzas son iguales (usando un test de `Bartlett`).

- Genere muestras aleatorias de tres distribuciones normales con medias distintas y varianzas iguales. Pruebe la hipótesis de igualdad de medianas usando la estadística de `Kruskal-Wallis`. Pruebe previamente que las varianzas son iguales. Haga un test de `comparación múltiple` no paramétrico
- Genere muestras de tres distribuciones normales (dos con medias iguales y una con media distinta) de varianzas iguales. Pruebe la hipótesis de igualdad de medianas usando la estadística de `Kruskal-Wallis`. Pruebe previamente que las varianzas son iguales.
- Use un test de `comparaciones múltiples` no paramétrico (si se requiere) para determinar en el ejercicio anterior entre cuales poblaciones hay diferencias.
- Simule un conjunto de datos para el cual sea válido usar el test de `Friedman` asumiendo cierta la hipótesis nula y haga la correspondiente prueba.
- Simule un conjunto de datos para el cual sea válido usar el test de `Friedman` asumiendo falsa la hipótesis nula y haga la correspondiente prueba.

11. ¿Cuándo es apropiado usar el test de `Kruskal-Wallis` y cuándo el de `Jonckheere-Tersptra` (no visto en el curso)?. ¿Cuál es el criterio para escoger uno u otro?. Aplique los tests de `Kruskal-Wallis` y `Jonckheere -Tersptra` al ejemplo dado en la tabla 6.2 del libro de Hollander y Wolfe.

3. Maestría

12. Determine un intervalo de confianza del $100(1-\alpha)\,\%$ para θ usando la estadística de `Wilcoxon` (ver página 56 Hollander y Wolfe)

13. Determine un intervalo de confianza del $100(1-\alpha)\,\%$ para θ usando la estadística de `Signo` (ver página 76 Hollander y Wolfe)

14. Determine un intervalo de confianza del $100(1-\alpha)\,\%$ para la diferencia de medianas usando la estadística de `Wilcoxon` (ver página 132 Hollander y Wolfe)

15. Demuestre (en el test de `Kruskal-Wallis`) que $H \sim \chi_{k-1}$ (ver página 197 Hollander y Wolfe)

16. Sea

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{NMV}\left(\begin{pmatrix} \mu \\ \eta \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix} \right)$$

a) Asigne valores a $\mu, \sigma, \sigma^2, \tau^2$ y $\rho \neq 0$. Genere una muestra bivariada tamaño 30 y pruebe la hipótesis de independencia (correlación igual cero) usando las estadísticas de `Pearson`, `Spearman` y `Kendall`

b) Con base en simulaciones de la distribución bivariada (variando ρ y dejando fijos los otros parámetros) compare la potencia de los tests de independencia basados en las estadísticas de `Pearson`, `Spearman` y `Kendall`.