Modelos no paramétricos y de regresión | Estadística II

Tarea 4

Fecha de entrega: 9 de marzo

Prueba de Wilcoxon

1. En el planteamiento de la prueba de Wilcoxon con variables absolutamente continuas, se definió la suma de rangos con signo como $T = \sum_{i=1}^{n} \operatorname{sgn}(D_i) R_i$, donde R_i es el rango de $| D_i$.
   a) Mostrar que $E(T) = 0$.
   b) Mostrar que $V(T) = \frac{n(n+1)(2n+1)}{6}$.
2. Se diseñó un experimento para estudiar el efecto de un nuevo aditivo para gasolina en el rendimiento de los automoviles. Para obtener información se seleccionaron al azar siete vehículos de los 20 más vendidos y de cada uno de ellos se determinó su rendimiento (en millas por galón), con y sin el nuevo aditivo. Los resultados son los siguientes.

Vehículo	1	2	3	4	5	6	7
Sin aditivo	24.2	30.4	32.7	19.8	25.0	24.9	22.2
Con aditivo	23.5	29.6	32.3	17.6	25.3	25.4	20.6

¿Existe evidencia para rechazar que el aditivo no tienen un efecto en el rendimiento de los automoviles? Utilizar un tamaño de prueba $\alpha = 0.01$ y reportar el *p-value* de la prueba (calculado con simulación).

Prueba de Mann-Whitney

3. Sean n y N enteros positivos, con $n < N$. Se define la variable aleatoria T como la suma de n enteros seleccionados al azar (sin reemplazo) del conjunto $\{1, 2, \ldots, N\}$.
   a) Mostrar que $E(T) = \frac{n(N+1)}{2}$.
   b) Mostrar que $V(T) = \frac{n(N+1)(N-n)}{12}$.
4. Se realizó un estudio para aprender cómo cambian las dietas de las mujeres durante el invierno y el verano. Se seleccionó un grupo de 12 mujeres que fueron observadas durante el mes de julio y de cada una se determinó el porcentaje de calorias provenientes de grasas. Se realizaron obervaciones similares de un grupo diferente de 12 mujeres seleccionadas al azar durante el mes de enero. Suponer que los resultados fueron los siguientes.

Julio	32.2	27.4	28.6	32.4	40.5	26.2	29.4	25.8	36.6	30.3	28.5	32.0
Enero	30.5	28.4	40.2	37.6	36.5	38.8	34.7	29.5	29.7	37.2	41.5	37.0

Prueba de Kruskal-Wallis

5. La prueba de Kruskal-Wallis es la alternativa paramétrica a la prueba F del análisis de varianza, o ANOVA, con una fuente de variación. El estadístico utilizado en la prueba F se

calcula como

$$F = \frac{\left(\sum_{i=1}^{k} \frac{X_{i\bullet}^2}{n_i} - \frac{X_{\bullet\bullet}^2}{n}\right)/(k-1)}{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i\bullet}^2}{n_i}\right)/(n-k)},$$

donde $X_{i\bullet} = \sum_{j=1}^{n_i} X_{ij}$ es la suma de las observaciones en la i-ésima muestra, $X_{\bullet\bullet} = \sum_{i=1}^{l}\sum_{j=1}^{n_i} X_{ij}$ es la suma de todas las observaciones y $n = \sum_{i=1}^{k} n_i$ es la suma de los tamaños de muestra. Bajo los supuestos del ANOVA se pude mostrar que la distribución nula de F es $F_{k-1,n-k}$. Mostrar que cuando el estadístico F se calcula con los rangos *combinados* en lugar de con las observaciones, se cumple la siguiente relación

$$F = \frac{T/(k-1)}{(n-1-T)/(n-k)},$$

donde T es el estadístico de Kruskall-Wallis.

6. En un estudio de nutrición se formaron tres grupos con cinco ciclistas cada uno. Los individuos en el primer grupo recibieron suplementos vitamínicos para tomarlos en cada comida durante las siguientes tres semanas. El segundo grupo recibió la indicación de comer cierto tipo de cereal de grano completo con alto contenido de fibra durante las siguientes tres semanas. El tercer grupo recibió la indicación de alimentarse como normalmente lo hacen. Cuando las tres semanas se cumplieron, se requirió a los ciclistas que recorrieran una distancia de 10 km. Se registraron los siguientes tiempos

Grupo control	15.9	17.2	16.6	15.4	16.8
Grupo con vitaminas	15.6	16.5	17.2	15.5	16.4
Grupo con fibra	17.1	16.2	15.8	16.3	16.0

¿Existe evidencia para rechazar que ni las vitaminas ni la fibra tienen algún efecto en la velocidad de los ciclistas? Utilizar un tamaño de prueba $\alpha = 0.01$ y reportar el *p-value* de la prueba (calculado con simulación).