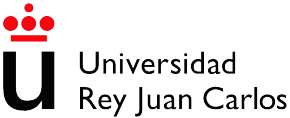
Inferencia no paramétrica

Víctor Aceña - Isaac Martín - Carmen Lancho

DSLAB

2025-09-24



Contrastes de hipótesis no paramétricos

¿Qué son? Los contrastes no paramétricos son métodos estadísticos utilizados para probar hipótesis cuando no se cumplen los supuestos necesarios para los contrastes paramétricos (por ejemplo, la normalidad)

- **Características Principales:**
  - **Flexibilidad**: No requieren que los datos sigan una distribución específica (se usa la Función de Distribución Empírica (EDF, por sus siglas en inglés)).
  - **Robustez**: Menos sensibles a valores atípicos y a desviaciones de supuestos de normalidad.
  - **Usos frecuentes**: Cuando las muestras son pequeñas, los datos son ordinales o tienen distribuciones asimétricas.

Contrastes de hipótesis no paramétricos

- **Aplicaciones**:
  - Comparación de medianas entre grupos.
  - Evaluación de relaciones de orden entre variables.
  - Pruebas de independencia entre variables categóricas.
- **Ejemplos**:
  - Prueba de Kolmogorov-Smirnov: Compara la EDF de dos muestras o de una muestra con una distribución teórica.
  - Prueba de Mann-Whitney: Utiliza la posición o el orden de los datos
  - Prueba de Kruskal-Wallis: Comparación de varias muestras
  - Prueba de Chi-cuadrado: Análisis de independencia para variables categóricas.

Prueba Chi-cuadrado de independencia

- La prueba chi-cuadrado de independencia se utiliza para determinar si hay una asociación significativa entre dos variables categóricas X con categorías $X_1, X_2, \dots, X_r$ e Y con categorías $Y_1, Y_2, \dots, Y_c$
- Esta prueba compara las frecuencias observadas en la tabla de contingencia con las frecuencias esperadas bajo la hipótesis de independencia

	Y_1	$\dots$	Y_j	$\dots$	Y_c	
X_1	$n_{1,1}$	$\dots$	$n_{1,j}$	$\dots$	$n_{1,c}$	$n_{1.}$
$\dots$						$\dots$
X_i	$n_{i,1}$	$\dots$	$n_{i,j}$	$\dots$	$n_{i,c}$	$n_{i.}$
$\dots$						$\dots$
X_r	$n_{r,1}$	$\dots$	$n_{r,j}$	$\dots$	$n_{r,c}$	$n_{r.}$
	$n_{.1}$		$n_{.j}$		$n_{.c}$	$n_{..}$

Prueba Chi-cuadrado de independencia

La hipótesis nula H_0 de esta prueba es que no hay asociación entre las variables, esto es, que las variables implicadas son independientes:

- **Hipótesis nula** H_0: No hay asociación entre las variables (son independientes)
- **Hipótesis alternativa** H_1: Hay una asociación entre las variables (son dependientes)

Las frecuencias esperadas se calculan como sigue:

$$E_{ij} = \frac{(n_{i.} \times n_{.j})}{N}$$

donde:

- E_{ij} es la frecuencia esperada en la celda (i, j)
- $n_{i.}$ es el total de la fila i
- $n_{.j}$ es el total de la columna j
- $N = n_{..}$ es el total general

Prueba Chi-cuadrado de independencia

Ahora, comparamos las frecuencias esperadas con las frecuencias observadas, definiendo con ellos el estadístico chi-cuadrado:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

donde O_{ij} es la frecuencia observada en la celda (i, j).

Bajo la hipótesis nula, el estadístico de prueba sigue una distribución chi-cuadrado con $(r-1)(c-1)$ grados de libertad, donde r es el número de filas y c es el número de columnas. Podemos calcular el $p - valor$ como:

$$p - valor = P(\chi^2_{(r-1)(c-1)} \geq \chi^2_{observado})$$

Como ocurría en los contrastes de hipótesis paramétricos, comparamos el $p - valor$ con el nivel de significancia α, generalmente 0.05. Si $p - valor < \alpha$, se rechaza la hipótesis nula.

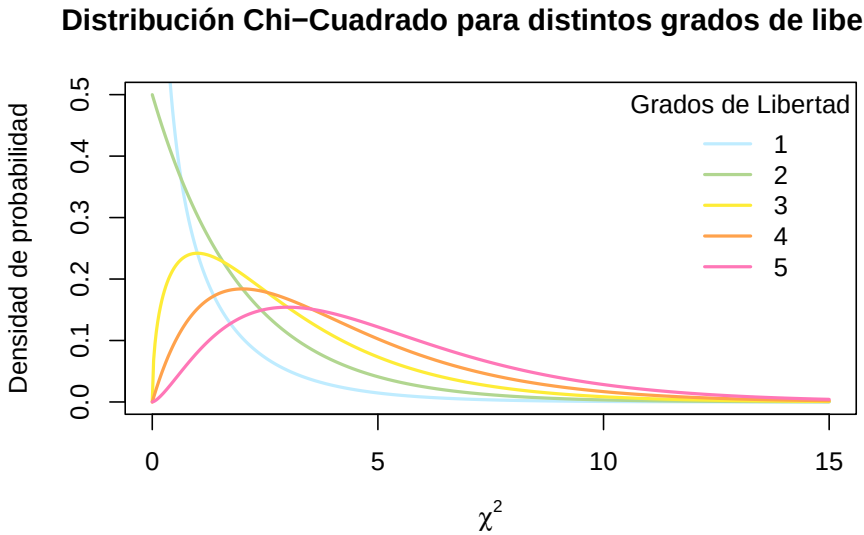
Distribución Chi-cuadrado

- La distribución Chi-cuadrado χ^2 es una distribución de probabilidad que surge especialmente en pruebas de hipótesis
- Se define como la distribución de la suma de los cuadrados de k variables aleatorias independientes que siguen una distribución normal estándar $N(0,1)$:

$$\chi^2_k = \sum_{i=1}^{k} Z_i^2$$

siendo $Z_i \sim N(0,1)$.

- Los grados de libertad k determinan la forma de la distribución. A medida que k aumenta, la distribución es más simétrica.
- $E(\chi^2_k) = k$ y $V(\chi^2_k) = 2k$
- Se trata de un caso especial de la distribución Gamma.


Distribución Chi−Cuadrado para distintos grados de libertad
Densidad de probabilidad
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
χ^2
Grados de Libertad
1
2
3
4
5

Ejemplo

Supongamos que un investigador desea determinar si hay una asociación entre el tipo de dispositivo usado (Laptop, Tablet, Smartphone) y la satisfacción del usuario (Satisfecho, No Satisfecho).

Recolectamos datos de una muestra de 150 usuarios y construimos la siguiente tabla de contingencia:

	Satisfecho	No Satisfecho	Total
Laptop	30	10	40
Tablet	20	20	40
Smartphone	50	20	70
Total	100	50	150

¿Existe asociación entre las variables?

Prueba Chi-cuadrado de independencia

Cuando la tabla de contingencia es pequeña (2x2), el estadístico chi-cuadrado puede devolver valores demasiados pequeños (lo que aumenta la probabilidad de error de tipo I).

Soluciones planteadas:

- Corrección de continuidad de Yates

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij} - 0.5)^2}{E_{ij}}$$

- Test exacto de Fisher (tabla 2x2 y alguna celda con menos de 5 obs)

Prueba de bondad de ajuste

Esta prueba se utiliza para determinar si una distribución de frecuencias observadas $F(x)$ sigue una distribución teórica esperada $F_0(x)$ (Normal, exponencial, Poisson, etc.). Este tipo de pruebas se llaman pruebas de bondad de ajuste (test of goodness of fit) y contrastan:

$$H_0 : F(x) = F_0(x) \quad \forall x$$

$$H_1 : F(x) \neq F_0(x)$$

En particular, veremos la prueba Chi-cuadrado de bondad de ajuste

Prueba Chi-cuadrado de bondad de ajuste

Dada la muestra aleatoria simple $X_1, \ldots, X_n$ de n observaciones se pretende analizar si concuerdan con una distribución específica conocida $F_0(x)$:

- Hipótesis nula H_0: Las frecuencias observadas siguen la distribución esperada $F_0(x)$
- Hipótesis alternativa H_1: Las frecuencias observadas no siguen la distribución esperada

Para ello, vamos a dividir el dominio en k trocitos y vamos a comparar las frecuencias de ambas distribuciones en ellos. Por ejemplo, en el caso real, se divide la recta real en: $(-\infty, b_1], (b_1, b_2], \dots, (b_{k-1}, \infty)$

Como la distribución $F_0(x)$ es conocida, sabemos cuáles son las probabilidades de dichos intervalos: $p_1, \dots, p_k$, $\sum_{i=1}^{k} p_i = 1$. Por tanto, según la distribución teórica, el número esperado de observaciones en cada intervalo es $E_i = np_i$, $\forall i = 1, \dots, k$.

Por otro lado, sabemos las observaciones O_i de la muestra que caen en cada intervalo (siendo $\sum_{i=1}^{k} O_i = n$).

Finalmente, se comparan las frecuencias observadas O_i con las frecuencias esperadas $E_i = np_i$ con el estadístico Chi-cuadrado

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-1,1-\alpha}$$

Ejemplo

Supongamos que un investigador quiere determinar si los resultados de un dado son uniformemente distribuidos. El dado se lanza 60 veces y los resultados son los siguientes:

- 1: 8 veces
- 2: 10 veces
- 3: 9 veces
- 4: 11 veces
- 5: 12 veces
- 6: 10 veces

Queremos comprobar si estos resultados siguen una distribución uniforme, es decir, cada número tiene la misma probabilidad de $1/6$.

Pruebas de homogeneidad

- La prueba no paramétrica de homogeneidad se utiliza para determinar si dos o más muestras independientes provienen de la misma distribución o de distribuciones similares
- Ejemplo de pruebas no paramétricas de homogeneidad:
  - **Prueba de Kolmogorov-Smirnov** para dos muestras: Compara dos muestras para verificar si provienen de la misma distribución
  - **Prueba de Mann-Whitney U** (o Wilcoxon Rank-Sum Test): Compara dos muestras independientes para determinar si tienen la misma distribución
  - **Prueba de Kruskal-Wallis**: Extiende la prueba de Mann-Whitney U a más de dos muestras independientes

Concepto previo

- La **función de distribución empírica** (EDF, por sus siglas en inglés) es una función de distribución de probabilidad utilizada para estimar la distribución subyacente de un conjunto de datos observados.
- Es una herramienta no paramétrica que proporciona una estimación de la función de distribución acumulada de una muestra de datos

Función de distribución empírica

- Dada una muestra de datos $(X_1, X_2, \ldots, X_n)$, la función de distribución empírica $F_n(x)$ se define como:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$$

donde $I(X_i \leq x)$ es una función indicadora que toma el valor 1 si $X_i \leq x$ y 0 en caso contrario

- En otras palabras, $F_n(x)$ es la proporción de valores en la muestra que son menores o iguales a x

Propiedades de la Función de Distribución Empírica

1. Escalonada: La EDF es una función escalonada que incrementa en pasos de $1/n$ en cada punto de datos
2. No decreciente: La EDF nunca disminuye a medida que x aumenta.
3. Límites: La EDF varía entre 0 y 1. Específicamente, $F_n(x) = 0$ para x menor que el valor mínimo de la muestra y $F_n(x) = 1$ para x mayor que el valor máximo de la muestra

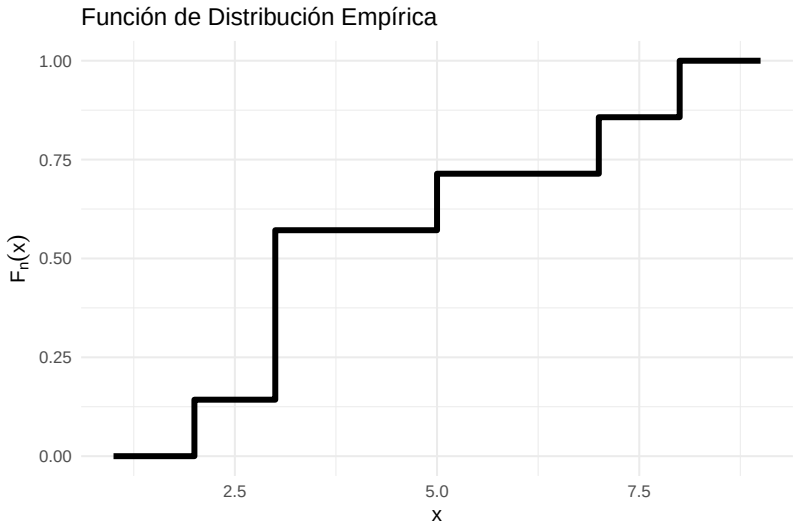
Función de distribución empírica

Calculemos la función de distribución empírica de los datos $X = \{8, 3, 5, 3, 7, 3, 2\}$.

1. Ordenar los datos: $\{2, 3, 3, 3, 5, 7, 8\}$
2. Calcular la función de distribución empírica $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$ para nuestras $n = 7$ observaciones

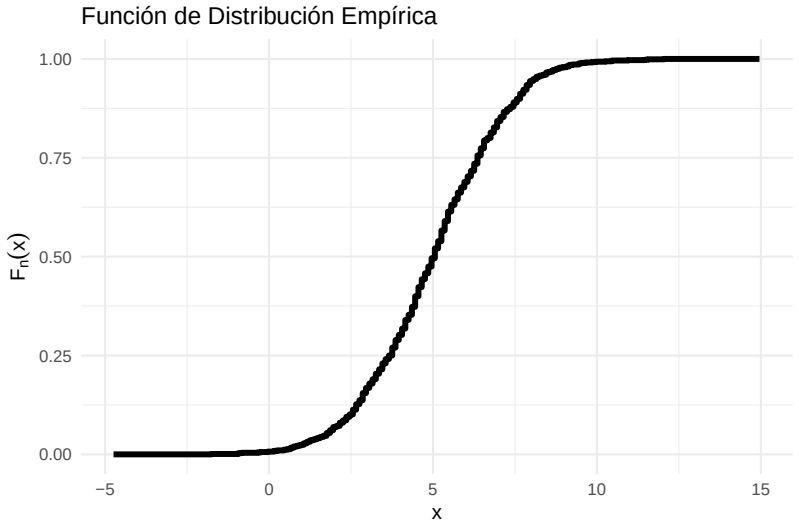
$$F_n(x) = \begin{cases} 0 & \text{si } x < 2 \\ \frac{1}{7} = 0.143 & \text{si } 2 \leq x < 3 \\ \frac{4}{7} = 0.571 & \text{si } 3 \leq x < 5 \\ \frac{5}{7} = 0.714 & \text{si } 5 \leq x < 7 \\ \frac{6}{7} = 0.857 & \text{si } 7 \leq x < 8 \\ 1 & \text{si } x \geq 8 \end{cases}$$

Función de distribución empírica



Función de distribución empírica

Con muchos datos:



Prueba de Kolmogorov-Smirnov para dos muestras

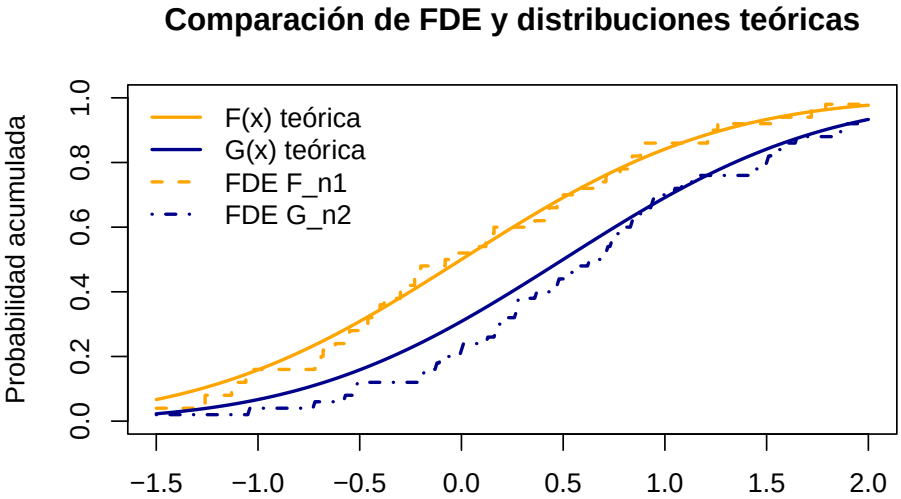
- La prueba de Kolmogorov-Smirnov (K-S) para dos muestras independientes es una prueba no paramétrica utilizada para determinar si dos muestras independientes provienen de la misma distribución
- A diferencia de otras pruebas que se centran en comparar medias o varianzas, la prueba K-S compara las distribuciones acumuladas de dos muestras
- Las hipótesis de la prueba son:
  - H_0: Las dos muestras provienen de la misma distribución
  - H_1: Las dos muestras provienen de distribuciones diferentes

Prueba de Kolmogorov-Smirnov para dos muestras

- Sea $X_1, \dots, X_{n_1}$ una m.a.s de una población 1 donde las X_i son independientes e idénticamente distribuidas y sea $Y_1, \dots, Y_{n_2}$ otra m.a.s de una población 2 donde las Y_j son independientes e idénticamente distribuidas
- Para cada muestra, se construyen las funciones de distribución empírica (EDF)
- Calculamos el estadístico D de la prueba K-S que es la máxima diferencia absoluta entre las dos funciones de distribución empírica:

$$D = \sup_x |F_{n_1}(x) - F_{n_2}(x)|$$

donde, $F_{n_1}(x)$ y $F_{n_2}(x)$ son las funciones de distribución empírica de las dos muestras.


Comparación de FDE y distribuciones teóricas
Probabilidad acumulada
0.0
0.2
0.4
0.6
0.8
1.0
F(x) teórica
G(x) teórica
FDE F_n1
FDE G_n2
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

Prueba de Kolmogorov-Smirnov para dos muestras

- El $p - valor$ se determina utilizando la distribución del estadístico D bajo la hipótesis nula de que ambas muestras provienen de la misma distribución
- El cálculo exacto del $p - valor$ para la prueba de K-S no es trivial y generalmente se realiza mediante métodos numéricos o tablas pre-calculadas. Sin embargo, se puede aproximar utilizando la distribución asintótica del estadístico D

Prueba de Kolmogorov-Smirnov para dos muestras

- Para muestras grandes, el $p - valor$ se puede aproximar usando la fórmula:

$$p \approx Q_{KS}(\sqrt{n}D)$$

donde:

  - $n = \frac{n_1 \cdot n_2}{n_1 + n_2}$ es el número efectivo de muestras
  - D es el valor del estadístico K-S
  - Q_{KS} es una función que representa la cola superior de la distribución de Kolmogorov-Smirnov

- La función Q_{KS} para grandes valores de n se puede aproximar usando la siguiente fórmula:

$$Q_{KS}(\lambda) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2\lambda^2}$$

donde $\lambda = \sqrt{n}D$

Prueba de Mann-Whitney U

- La prueba de Mann-Whitney U, también conocida como prueba de Wilcoxon para muestras independientes, es una prueba no paramétrica que se utiliza para contrastar si dos muestras independientes provienen de la misma distribución
- Es una alternativa a la prueba t de Student cuando no se cumplen los supuestos de normalidad. En lugar de trabajar con los valores originales, la prueba utiliza los rangos de los datos
- Esta prueba se basa en combinar y ordenar juntas ambas muestras. Si en dicha ordenación:
  1. Los valores de ambas muestras se mezclan de forma aleatoria –> entenderemos que las muestras no son distintas
  2. Los valores de cada muestra quedan claramente agrupados –> las muestras son distintas

Prueba de Mann-Whitney U

Queremos contrastar si el tratamiento A y el tratamiento B tienen el mismo efecto

Tx	A	A	A	A	B	A	B	B	B	B
Rango	1	2	3	4	5	6	7	8	9	10

(a)

Tx	B	A	A	B	B	A	B	A	B	A
Rango	1	2	3	4	5	6	7	8	9	10

(b)

Prueba de Mann-Whitney U

- En primer lugar se combinan los datos de ambas muestras y se ordena el total de los valores de menor a mayor
- A continuación se asignan rangos a estos valores. Los empates se gestionan otorgando a los valores iguales el rango promedio
- Se calcula el estadístico del contraste tal y como sigue para cada muestra ($i = 1, 2$):

$$U_i = n_1 n_2 + \frac{n_i(n_i + 1)}{2} - R_i$$

donde:

  - n_1 y n_2 son los tamaños de las dos muestras
  - R_1 y R_2 son la suma de los rangos de las muestras 1 y 2, respectivamente

- El estadístico U final es $U = min(U_1, U_2)$

Prueba de Mann-Whitney U

1. Para Muestras Pequeñas (Método Exacto)

- El **p-valor** se obtiene comparando el estadístico **U** con los valores críticos de una **tabla de referencia de Mann-Whitney** o mediante software estadístico.

2. Para Muestras Grandes (Aproximación a la Normal)

- Si las muestras son grandes (ej. $n_1, n_2 > 20$), la distribución de U se aproxima a una normal.
- Se estandariza el estadístico U calculando el valor Z_{calc}:

$$Z_{calc} = \frac{U - E(U)}{\sqrt{Var(U)}}$$

- Donde la media y la varianza esperadas son:

$$E(U) = \frac{n_1 n_2}{2} \quad ; \quad Var(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

3. Regla de Decisión

- El p-valor obtenido (ya sea por el método exacto o por aproximación) se compara con un **nivel de significancia** α, usualmente 0.05.
- Si **p-valor** $< \alpha$ **–> Se rechaza la hipótesis nula (**H_0**)**.
- **En la práctica**: Rechazar H_0 significa que existe evidencia estadística para afirmar que las dos muestras provienen de poblaciones diferentes (es decir, hay una diferencia significativa).

Ejemplo

Supongamos que queremos comparar los tiempos de entrega (en días) de dos proveedores distintos:

- Proveedor A: $2, 3, 5, 6, 8$
- Proveedor B: $1, 4, 4, 7, 9$

Prueba de Kruskal-Wallis

- La prueba de Kruskal-Wallis es una prueba no paramétrica utilizada para comparar tres o más muestras independientes para determinar si provienen de la misma distribución
- Es una extensión de la prueba de Mann-Whitney U a más de dos grupos y una alternativa robusta a la ANOVA cuando no se cumplen los supuestos de normalidad y homogeneidad de varianzas
- Dado que es una prueba no paramétrica, no requiere que los datos provengan de una distribución normal. Al igual que la prueba de Mann-Whitney, la prueba de Kruskal-Wallis trabaja con los rangos de los datos en lugar de los valores originales

Las hipótesis son:

- H_0: Todas las muestras provienen de la misma distribución
- H_1: Al menos una de las muestras proviene de una distribución diferente

Prueba de Kruskal-Wallis

- Se comienza combinando los datos de todas las muestras y se ordenan los valores de menor a mayor
- A continuación se asignan rangos R_i a estos valores. Los empates se gestionan otorgando a los valores iguales el rango promedio
- El estadístico Kruskal-Wallis H se define como:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

donde N es el tamaño total de la muestra (suma de los tamaños de las k muestras), R_i es la suma de los rangos de la i-ésima muestra, n_i es el tamaño de la i-ésima muestra y k es el número de muestras

Prueba de Kruskal-Wallis

- El $p - valor$ se determina comparando el estadístico H con una distribución de referencia para H (tabla de Kruskal-Wallis)
- Para tamaños de muestra relativamente grandes el estadístico H sigue una distribución χ^2 con $k - 1$ grados de libertad. El $p - valor$ se determina como

$$p - valor = P(H > h)$$

siendo h el valor que toma H en las muestras en cuestión

Ejemplo

Unos investigadores estaban interesados en estudiar la interacción social de distintos adultos para estudiar si la interacción social puede vincularse a la confianza en uno mismo

Para ello, clasificaron a 17 participantes en tres grupos en función de la interacción social exhibida: alta, media, baja

Después de clasificar a los participantes en los tres grupos, se les pidió que completaran una autoevaluación de la autoconfianza en una escala de 25 puntos:

Alta	Media	Baja
21, 23, 18, 12, 19, 20	19, 5, 10, 11, 9	7, 8, 15, 3, 6, 4

Se quiere determinan si existe diferencia entre alguno de los 3 grupos.

Prueba sobre muestras pareadas

- **Muestras pareadas**: las muestras no son independientes, están relacionadas, emparejadas entre sí
- Ejemplos:
  - Nivel de colesterol en un grupo de personas antes y después de tomar un medicamento
  - Medidas tomadas en los dos pies
  - Distintas versiones de un modelo de Machine Learning en el mismo conjunto de datasets
- Contrastes:
  - Prueba del signo (sign test)
  - Prueba de rangos de signos de Wilcoxon (Wilcoxon signed-rank test)

Prueba del signo

- Prueba no paramétrica utilizada para comparar dos muestras relacionadas o emparejadas
- Se emplea cuando se tienen dos conjuntos de datos dependientes y se desea determinar si hay una diferencia significativa en sus medianas
- Alternativa a la prueba de la t-Student cuando no se cumplen las hipótesis
- Simplifica la comparación entre las muestras mediante una binarización de la misma: convierte los resultados en "+" y "-" y lo compara en esa versión

Prueba del signo

- Sean dos muestras relacionadas $X_1 \dots, X_n$ e $Y_1, \dots, Y_n$ de tamaño n
- Comenzamos calculando las diferencias entre pares: Para cada par (X_i, Y_i), calcular la diferencia $D_i = X_i - Y_i$
- Contar los signos
  - S_+ es el número de diferencias positivas $D_i > 0$
  - S_- es el número de diferencias negativas $D_i < 0$
  - Ignorar las diferencias que son cero $D_i = 0$
- Estadístico de Prueba: $S = min(S_+, S_-)$

Prueba del signo

- Determinar el valor crítico. Consultar una tabla de la distribución binomial para obtener el valor crítico correspondiente al nivel de significancia α y el tamaño de la muestra efectiva n (número de pares no nulos). Podemos calcular el valor crítico como sigue:

$$p - valor = P(S \leq s)$$

donde $S \sim Binomial(n, p = 0.5)$

- Decisión:
  - Rechazar H_0 si el estadístico de la prueba es menor o igual al valor crítico
  - No rechazar H_0 si el estadístico de la prueba es mayor que el valor crítico

- Si $S_+ + S_- \geq 25$, se puede usar la siguiente formula

$$z = \frac{max(S_+, S_-) - 1/2(S_+ + S_-) - 1/2}{1/2\sqrt{S_+ + S_-}} \sim N(0,1)$$

Así, $p - value = P(Z \leq z)$ en el contraste unilateral y se multiplicará por 2 en el bilateral.

Prueba de rangos de signos de Wilcoxon

- La prueba de Wilcoxon de rangos de signos se utiliza para comparar muestras pareadas
- Es una alternativa no paramétrica a la prueba t de muestras pareadas. En lugar de comparar medias, esta prueba compara las medianas de las diferencias entre las dos muestras pareadas
- Es una prueba ideal para muestras pequeñas cuando la normalidad no puede ser asumida

Prueba de rangos de signos de Wilcoxon

1. **Calcular las diferencias** entre cada par de observaciones
2. **Ordenar las diferencias** en valor absoluto y asignarles rangos, ignorando las diferencias que sean cero
3. **Asignar signos** a los rangos de acuerdo con el signo de las diferencias originales
4. **Calcular la suma de los rangos positivos** y la suma de los rangos negativos
5. **Determinar el estadístico de prueba**: El estadístico de Wilcoxon es el menor de las dos sumas de rangos: $T = min(\sum R_+, \sum R_-)$

siendo R_+ la suma de los rangos con diferencias positivas y R_- la suma de los rangos con diferencias negativas

6. **Significatividad.** Se compara el estadístico de prueba T con los valores críticos de la tabla de Wilcoxon para determinar la significancia estadística

Si el tamaño muestral es grande, se puede determinal la significatividad mediante la aproximación normal para grandes tamaños de muestra:

$$Z = \frac{T - E(T)}{Var(T)} - \sim N(0,1)$$

siendo

$$E(T) = \frac{n(n+1)}{4} \qquad Var(T) = \frac{n(n+1)(2n+1)}{24}$$

con n el número de muestras pareadas.

Con ello, $p - valor = P(Z > z_{\alpha/2})$

Referencias

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons.

Deshpande, J. V., Naik-Nimbalkar, U., & Dewan, I. (2017). *Nonparametric statistics: theory and methods*. World Scientific.

Gomez Villegas, M. A. (2005). *Inferencia estadística*. Ediciones Díaz de Santos.

Canavos, G. C., & Medal, E. G. U. (1987). *Probabilidad y estadística* (p. 651). México: McGraw Hill.