

PRÁCTICA 5. Pruebas no paramétricas

Consejos:

Antes de iniciar la práctica recuerda lo siguiente:

1. Crea una carpeta vacía que se llame practica5_MEE
2. Abre Rstudio
3. Crea un nuevo proyecto con el menú de Rstudio

File->Create Project-> Existing Directory

Busca la carpeta creada en el punto 1.

Ya podemos empezar una sesión nueva de R que guardaremos al final en esta carpeta cuando salgamos. La ventaja es que podemos continuar esta sesión cuando queramos accediendo a la carpeta y ejecutando el proyecto.

4. Crea un fichero R script con el menú de Rstudio

File->New File->Rscript

Ponle nombre practica5 su extensión será .R

En este fichero guardaremos todas las órdenes de R de manera que podamos repetirlas o modificarlas cuando lo deseemos.

Aquí tienes alguna de los comandos de R que vas a necesitar:

<p>(b) Procedimiento t para muestras por pares, utilizando los valores de cada par: <code>t.test()</code></p> <pre>t.test(x, y, mu, conf.level, alt, paired=T)</pre> <ul style="list-style-type: none"> • x: vector que contiene los primeros valores de cada par. • y: vector que contiene los segundos valores de cada par. • mu: valor de la diferencia de medias en H_0 (por defecto vale 0) Es decir, por defecto $H_0 : \mu_1 = \mu_2$ • paired: debemos indicar que los datos están apareados ((paired=T). • Si sólo deseamos el intervalo de confianza para la diferencia de medias: <code>t.test(x, y, conf.level=0.95, paired=T)\$conf.int</code> 	
<p>Test de Wilcoxon de suma de rangos: <code>wilcox.test()</code></p> <pre>wilcox.test(x, y, mu, conf.level, alt)</pre> <ul style="list-style-type: none"> • x: vector que contiene los valores de la primera muestra. • y: vector que contiene los valores de la segunda muestra. • mu: valor de la diferencia de medianas en H_0 (por defecto vale 0) Es decir, por defecto $H_0 : mediana_1 = mediana_2$ 	<p>Test de Wilcoxon de suma de rangos de diferencias por pares</p> <pre>wilcox.test(x, y, mu, conf.level, alt, paired=T)</pre> <ul style="list-style-type: none"> • x: vector que contiene los primeros valores de cada par. • y: vector que contiene los segundos valores de cada par. • paired=T debemos indicar el valor de este parámetro porque los datos están apareados. • mu: valor de la diferencia de medianas en H_0 (por defecto vale 0) Es decir, por defecto $H_0 : mediana_1 = mediana_2$

Ejercicio 1

En un estudio para valorar la efectividad de una dieta, se midió el peso inicial y el peso final de 14 hombres que siguieron la dieta durante un mes (los datos se muestran en kilogramos):

Antes	75.3	76.5	82.1	65.4	63.7	78.6	74.2	77.3	70.5	66.9	85.6	66.3	65.1	68.2
Después	73.2	76.2	80.4	65.5	64.0	75.5	71.6	75.8	69.1	63.8	86.0	67.1	64.5	62.0

Se desea saber si la dieta es efectiva para perder peso. Queremos extraer conclusiones no sólo para los individuos de la muestra, sino para todos los hombres que sigan la dieta.

(a) Identifique los siguientes elementos:

Muestra/-as:

Población/-es:

Variable(s):

Hipótesis:

H_0 :

En palabras:

H_1 :

En palabras:

(b) ¿Qué test resulta apropiado para estos datos y este objetivo?

(c) Justifique si se dan las condiciones necesarias para aplicarlo.

(d) Resolver el problema utilizando un procedimiento t .

Estadístico t :

Valor p :

Obtenga también un intervalo de confianza al 95% para la pérdida media de peso tras un mes de dieta:

Explica la conclusión que se puede extraer de estos cálculos:

(e) Compara los resultados con aquellos que se obtienen al aplicar el test de Wilcoxon que corresponda. ¿Se obtiene la misma conclusión?

Valor p :

Explica de forma clara y concisa la conclusión del análisis en el contexto del problema:

Ejercicio 2

Una librería exhibe en lugares estratégicos de su tienda las novelas finalistas del Premio Planeta. Las ventas de estas novelas (A y B) en dos muestras aleatorias de 11 días han sido las siguientes:

Novel·la A	280	222	217	331	181	203	223	199	220	149	197
Novel·la B	180	222	200	220	220	290	200	245	283	391	211

NOTA: Los datos NO son pareados, porque se han obtenido 2 muestras en días diferentes.

La pregunta final que queremos contestar es: En previsión de futuras ventas, ¿muestran estos datos suficiente evidencia para justificar que, en el próximo pedido de estas novelas en la editorial, el número de ejemplares solicitados de A y B debe ser diferente?

(a) Defina las variables implicadas en el estudio:

Muestra/-as:

Población/-ones:

Variable(s):

Introduzca los datos en R y realice un análisis descriptivo de cada muestra.

(b) No podemos utilizar la distribución t de Student para realizar una prueba de comparación de medias, ya que se incumple una de las condiciones necesarias. ¿Cuál es esta condición?

(c) Haciendo uso de R, realice un contraste de suma de rangos de Wilcoxon para decidir si existen diferencias significativas entre las ventas de ambas novelas (considere $\alpha = 5\%$). Complete:

Hipótesis:

H_0 : En palabras:

H_1 : En palabras:

Estadístico W : Valor p :

¿Podemos rechazar la hipótesis nula?

Explica de forma clara y concisa la conclusión en el contexto del problema:

Imagínese ahora que los datos están apareados, porque hemos estudiado las ventas de las novelas A y B durante los mismos días.

(d) ¿Se obtendría la misma conclusión con estos datos? Utilice el test de Wilcoxon apropiado para responder esta pregunta.

Valor p :

Conclusión en el contexto del problema: