CARRERA DE INGENIERÍA DE SISTEMAS

Pruebas No Paramétricas

ESTADÍSTICA APLICADA
PROF. MSC. VILMA S. ROMERO ROMERO
5TO CICLO | 14-06-2022

UNIVERSIDAD DE LIMA

Introducción

- Hasta el momento hemos estudiado pruebas de hipótesis que dependen de un conjunto de supuestos.
  - Si alguno de los supuestos no se cumple, las conclusiones pueden ser inválidas.
- La mayoría de procedimientos estadísticos se basan en el supuesto de normalidad. Las técnicas estadísticas basadas en este supuesto se conocen como **pruebas paramétricas**.
- Existen técnicas estadísticas que requieren pocos supuestos sobre la población de donde se extrajo la muestra, llamadas **pruebas no paramétricas**.

1. Prueba de Rangos de Signos de Wilcoxon para una población

Prueba no paramétrica utilizada para evaluar la hipótesis sobre la mediana de la población.

- Dado que la mediana es el punto medio de una población, se espera que aproximadamente la mitad de los datos de una muestra aleatoria se encuentren por debajo de la mediana hipotética y la otra mitad por encima.
- Se rechaza la hipótesis nula si la distribución real de los datos muestra una desviación demasiada grande de lo esperado.

Ejemplo 1

Se tiene información sobre las ventas semanales, en soles, de cierto producto. Los datos de las ventas semanales de una muestra aleatoria de 15 tiendas se encuentran en el archivo "`VentasSemanales.csv`". La gerencia ha solicitado la siguiente prueba de hipótesis sobre la mediana de las ventas semanales de la población:

$$H_0 : \text{Mediana} = 460$$
$$H_1 : \text{Mediana} \neq 460$$

Considerar un nivel de significación del 5%.

Prueba de Rangos de Signos de Wilcoxon en R

```r
# Lectura de datos
datos <- read.csv("VentasSemanales.csv")
```

```r
wilcox.test(datos$Ventas, mu = 460, alternative = "two.sided")

##
##  Wilcoxon signed rank exact test
##
## data:  datos$Ventas
## V = 35, p-value = 0.1688
## alternative hypothesis: true location is not equal to 460
```

Como p-value $= 0.1688 > \alpha = 0.05$, entonces no se rechaza la H_0. Por lo tanto, no existe suficiente evidencia para sugerir que la mediana de las ventas semanales del producto sea diferente de 460 soles, al 5% de significación.

UNIVERSIDAD DE LIMA

2. Prueba U de Mann Whitney

- Prueba no paramétrica para evaluar la diferencia entre dos poblaciones basado en dos muestras independientes.
- Es una alternativa a la prueba t de muestras independientes.
- No requiere el supuesto de que las poblaciones se distribuyan de manera normal.

Ejemplo 2

Los siguientes datos se han recopilado de dos muestras independientes de dos poblaciones. Evaluar la afirmación de que la mediana de la segunda población excederá la mediana de la primera población. Considerar un nivel de significación del 5%.

Muestra 1	13	21	15	10	11	14	12	8
Muestra 2	9	18	16	17	20	7	22	19

Solución:

1) Se formulan las hipótesis:

$$H_0 : \text{Mediana}_{\text{Muestra 2}} \leq \text{Mediana}_{\text{muestra 1}}$$
$$H_1 : \text{Mediana}_{\text{Muestra 2}} > \text{Mediana}_{\text{muestra 1}}$$

2) $\alpha = 0.05$

```
# Ingreso de datos
muestra1 <- c(13,21,15,10,11,14,12,8)
muestra2 <- c(9,18,16,17,20,7,22,19)
```

UNIVERSIDAD DE LIMA

Prueba U de Mann Whitney en R

3) Prueba U de Mann Whitney

```
wilcox.test(muestra2, muestra1, alternative = "greater")
##
##  Wilcoxon rank sum exact test
##
## data:  muestra2 and muestra1
## W = 44, p-value = 0.1172
## alternative hypothesis: true location shift is greater than (
```

4) Como p-value $= 0.1172 > \alpha = 0.05$, entonces no se rechaza la H_0. Por lo tanto, no existe suficiente evidencia para sugerir que los valores de la población 2 tienen una mediana superior a los de la población 1, al 5% de significación.

UNIVERSIDAD DE LIMA

3. Prueba U de Mann Whitney

- Prueba no paramétrica para muestras relacionadas cuando no se quiere o no se puede (debido a limitaciones de los datos) usar la prueba t de muestras relacionadas.
- La prueba utiliza valores cuantitativos pero no requiere el supuesto de que las diferencias entre las observaciones pareadas están normalmente distribuidas.
  - Solo requiere el supuesto de que las diferencias entre las observaciones pareadas tengan una distribución simétrica alrededor de su mediana.

Ejemplo 3

Una determinada compañía de software desea evaluar si su nueva aplicación de análisis de datos reduce el tiempo en ciertos procesos analíticos. Para esto selecciona un focus group de 7 personas dedicadas a la ciencia de datos, a las cuales les asigna una tarea analítica específica antes y de manera similar después del entrenamiento de uso de la aplicación. Los tiempos registrados, en minutos, se presentan a continuación:

Persona	1	2	3	4	5	6	7
Antes	24	20	19	20	13	28	15
Después	11	18	23	15	16	22	8

Realizar la prueba de hipótesis respectiva empleando en su análisis un nivel de significación del 5%.

Solución:

1) Se formulan las hipótesis:

H_0 : $\text{Mediana}_{\text{Antes}} \leq \text{Mediana}_{\text{Después}}$
H_1 : $\text{Mediana}_{\text{Antes}} > \text{Mediana}_{\text{Después}}$
(Mediana de los tiempos será inferior después del entrenamiento)

Alternativamente,

H_0 : (La mediana de los tiempos se mantiene)
H_1 : (La mediana de los tiempos será inferior después del entrenamiento)

2) $\alpha = 0.05$

```r
# Ingreso de datos
antes <- c(24,20,19,20,13,28,15)
despues <- c(11,18,23,15,16,22,8)
```

3) Prueba de Rangos de Signos de Wilcoxon

```
wilcox.test(antes, despues, paired = TRUE,
            alternative = "greater")
```

```
##
##  Wilcoxon signed rank exact test
##
## data:  antes and despues
## V = 23, p-value = 0.07813
## alternative hypothesis: true location shift is greater than (
```

4) Como p-value $= 0.07813 > \alpha = 0.05$, entonces no se rechaza la H_0. Por lo tanto, según estos datos muestrales, la compañía de software no tiene suficiente evidencia estadística para afirmar que su aplicación reducirá el tiempo mediano necesario para realizar tareas analíticas, al 5% de significación.

UNIVERSIDAD DE LIMA

Recursos Adicionales I

Devore, J. (2019). *Introducción a la probabilidad y estadística para ingeniería y ciencias*. Cengage, 1 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_CENGAGE.

Johnson, R. A. (2012). *Probabilidad y estadística para ingenieros*. Pearson Educación, 8 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_PEARSON.

Kokoska, S. (2015). *Introductory Statistics*. W. H. Freeman and Company, 2 edition.

Mendenhall, W., Beaver, R. J., and Beaver, B. M. (2015). *Introducción a la probabilidad y estadística*. Cengage, 14 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_CENGAGE.

UNIVERSIDAD DE LIMA

Recursos Adicionales II

Millones, R., Barreno, E., Vásquez, F., and Castillo, C. (2017). *Estadística Descriptiva y Probabilidades: Aplicaciones en la ingeniería y los negocios*. Lima: Fondo Editorial de la Universidad de Lima, 1 edition. Código Biblioteca U.Lima: 519.53 E.

Triola, M. (2018). *Estadística*. Pearson Educación, 12 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_PEARSON.

UNIVERSIDAD DE LIMA