Pruebas no paramétricas

Aleksander Dietrichson, PhD

10 de junio 2021

Agenda

1. Prueba U de Mann-Whitney
2. Prueba de rangos con signos de Wilcoxon
3. Prueba de signos
4. Tareas
5. Ejemplos de Modelaje

Prueba U de Mann-Whitney

La prueba U de Mann-Whitney también se conoce con otros nombres: *Mann–Whitney–Wilcoxon*, *Wilcoxon rank-sum test* y *Wilcoxon–Mann–Whitney*. Por ello está disponible en R por medio de la función `wilcox.test`.

Se basa en rangos. De manera que:

```
my_sample <- my_data %>%
  sample_n(10) %>%
  select(sexo, ingreso) %>%
  mutate(rango = rank(ingreso)) %>%
  arrange(rango) %>%
  select(rango,everything())
```

Prueba U de Mann-Whitney

Da:

```
my_sample %>%
  knitr::kable(booktabs=TRUE)
```

rango	sexo	ingreso
1	varón	0
2	mujer	7500
3	varón	8000
4	varón	12000
5	varón	15000
6	varón	18000
7	mujer	30000
8	mujer	40000
9	varón	45000
10	varón	60000

Prueba U de Mann-Whitney

$H_0 : P(x_i > y_j) = \frac{1}{2}$

$H_1 : P(x_i > y_j) \neq \frac{1}{2}$

```
mujeres <- my_sample %>% filter(sexo == "mujer") %>% pull(i
hombres <- my_sample %>% filter(sexo != "mujer") %>% pull(i
wilcox.test(hombres,mujeres)
```

```
    Wilcoxon rank sum test

data:  hombres and mujeres
W = 10, p-value = 1
alternative hypothesis: true location shift is not equal to
```

Prueba de rangos con signos de Wilcoxon

Table 2: Eficiencia de dos medicamentos reportada por los pacientes.

Paciente	Droga.A	Droga.B
1	2,0	3,5
2	3,6	5,7
3	2,6	2,9
4	2,7	2,4
5	7,3	9,9
6	3,4	3,3
7	14,9	16,7
8	6,6	6,0
9	2,3	3,8
10	2,1	4,0
11	6,8	9,1
12	8,5	20,9

Prueba de rangos con signos de Wilcoxon

```
shapiro.test(datos$Droga.A)
```

```

    Shapiro-Wilk normality test

data:  datos$Droga.A
W = 0.80692, p-value = 0.01124
```

```
shapiro.test(datos$Droga.B)
```

```

    Shapiro-Wilk normality test

data:  datos$Droga.B
W = 0.7883, p-value = 0.006919
```

Prueba de rangos con signos de Wilcoxon

```
wilcox.test(datos$Droga.A, datos$Droga.B, paired = TRUE)
```

```
    Wilcoxon signed rank test with continuity correction

data:  datos$Droga.A and datos$Droga.B
V = 8, p-value = 0.01669
alternative hypothesis: true location shift is not equal to
```

Test de T con los mismos datos

```
t.test(datos$Droga.A,datos$Droga.B,paired = TRUE)
```

```
    Paired t-test

data:  datos$Droga.A and datos$Droga.B
t = -2.1465, df = 11, p-value = 0.05498
alternative hypothesis: true difference in means is not equ
95 percent confidence interval:
 -4.28706458  0.05373125
sample estimates:
mean of the differences
              -2.116667
```

Prueba de signos

- Si las variables son solo ordinales

$$z = \frac{N - 2 \times W - 1}{\sqrt{N}}$$

Prueba de signos

Ejemplo:

39 prefieren Martínez 61 prefieren Habana. 50 No tienen preferencia.

$$z = \frac{N - 2 \times W - 1}{\sqrt{N}} = \frac{100 - 2 \times 39 - 1}{\sqrt{100}} = \frac{21}{10} = 2,1$$

Implementación en R

```
library(BSDA)
martinez <- c(rep(5,39),rep(1,61),rep(3,50))
habana <- c(rep(1,39),rep(5,61),rep(3,50))
SIGN.test(martinez,habana)
```

```
    Dependent-samples Sign-Test

data:  martinez and habana
S = 39, p-value = 0.0352
alternative hypothesis: true median difference is not equal
95 percent confidence interval:
 0 0
sample estimates:
median of x-y
            0

Achieved and Interpolated Confidence Intervals:

                  Conf.Level L.E.pt U.E.pt
```

Otro uso del test de signos

- Si se conoce el mediano de la población, se puede usar para determinar si una variable x es significativamente distinto a ese.

```
Mediano <- median(my_data$ingreso)
SIGN.test(my_sample$ingreso, md=Mediano)
```

```
    One-sample Sign-Test

data:  my_sample$ingreso
s = 5, p-value = 1
alternative hypothesis: true median is not equal to 16000
95 percent confidence interval:
  7662.222 43377.778
sample estimates:
median of x
      16500

Achieved and Interpolated Confidence Intervals:
```

Otro uso del test de signos

Ejemplo

```r
Mediano <- median(my_data$ingreso)
ingresos_observados <- c(rep(10000,20),300000,300000)
SIGN.test(ingresos_observados, md=Mediano)
```

```
    One-sample Sign-Test

data:  ingresos_observados
s = 2, p-value = 0.0001211
alternative hypothesis: true median is not equal to 16000
95 percent confidence interval:
 10000 10000
sample estimates:
median of x
      10000

Achieved and Interpolated Confidence Intervals:
```

Tareas

- ¿Cuál es la distribución de sus datos?
- ¿Qué impacto tiene a la hora de seleccionar metodos estadísticos?