Maestría en Computo Estadístico
Inferencia Estadística
Tarea 9
13 de diciembre de 2020
Enrique Santibáñez Cortés
Repositorio de Git: Tarea 9, IE.

Ejercicio de Estadística no parametrica.

Eriksen, Björnstad, y Götestam (1986) estudiaron un programa de entrenamiento de habilidades sociales para alcohólicos. Veinticuatro hombres 'dependientes del alcohol' hospitalizados en un centro de tratamiento de alcohol fueron asignados al azar a dos grupos. A los pacientes del grupo de control se les dio un programa de tratamiento tradicional. A los pacientes del grupo de tratamiento se les dio el programa de tratamiento tradicional más una clase de entrenamiento de habilidades sociales (SST). Después de ser dado de alta del programa, cada paciente informó en intervalos de 2 semanas la cantidad de alcohol consumido, el número de los días anteriores a su primer trago, el número de días sobrio, los días trabajados, las veces admitido en una institución, y las noches dormía en casa. Los informes fueron verificados por otras fuentes (esposas o miembros de la familia). (¡Estos datos pueden ser poco fiables!) Un paciente en el SST grupo, descubierto como adicto a los opiáceos, desapareció después de ser dado de alta y no presentó ningun informe. Los 23 pacientes restantes informaron fielmente durante un año. Los resultados para la ingesta de alcohol se muestran en la siguiente tabla. Consumo de alcohol durante 1 año (Centilitro de alcohol puro):

Control	1042	1617	1180	973	1552	1251	1151	1511	758	1079	951	1319
SST	874	389	612	798	1152	893	541	741	1064	862	213	

Cuadro 1: Resultados del experimento.

1. ¿Qué diferencia existe entre la prueba de Wilcoxon y la prueba de Mann−Whitney?

RESPUESTA
La diferencia que existe en estas pruebas es que **la prueba de Wilconxon considera datos pareados** teniendo como objetivo determinar si existen diferencias en las distribuciones de localización, lo que es equivalente a una prueba clásica de t para datos pareados. Ahora, **la prueba de Mann−Whitney considera que tenemos dos muestras aleatorias independientes dentro y fuera** de la muestra con el objetivo de contrastar un problema de localización.

2. ¿Qué prueba utilizaría para resolver este problema, por qué?

RESPUESTA
Por como se realizo el experimento yo utilizaría la prueba de Mann−Whitney debido a que tenemos dos muestras aleatorias (ya que fueron asignados al azar a dos grupos) independientes dentro y fuera, este supuesto es 'válido' ya que los resultados de cada persona estudiada no dependerán de los resultados de las demás personas y por último la escala de la variable observada es al menos ordinal, por lo que se cumplen todos los supuestos para realizar la prueba.

3. ¿Existe evidencia para afirmar que el grupo con el entrenamiento SST tiene un menor consumo de alcohol comparado con el grupo de control?

RESPUESTA
Para contrastar que el entrenamiento SST tiene un menor consumo de alcohol comparado con el grupo de control ocupemos el siguiente juego de hipótesis:

H_0 las personas que tomaron el programa de tratamiento SST (Y_i) y las personas del tratamiento tradicional (X_i) no tienden a reducir (o aumentar) su consumo de alcohol contra H_A: las personas del entrenamiento

SST tienden a un menor consumo de alcohol comparado con el grupo control (la prueba es de cola inferior). Entonces procedemos a realizar la prueba de Mann−Whitney con ayuda de R.

```r
# datos del problema
control <- c(1042, 1617, 1180, 973, 1552, 1251, 1151 ,1511, 758, 1079, 951, 1319)
sst <- c(874, 389, 612, 798, 1152, 893, 541, 741, 1064, 862, 213)

# tamaños de los datos
n <- length(control)
m <- length(sst)

# mezclamos las m.a.'s
pacientes <- c(sst, control)

# rankeamos los datos
rangos_pacientes <- rank(pacientes)

# calculamos W
W <- sum(rangos_pacientes[1:m])

W
```

```
## [1] 80
```

```r
# calculamos U
U <- W-m*(m+1)/2

# Calculamos el estadístico de prueba
U_2 <- n*m-U
U_2
```

```
## [1] 118
```

```r
# valor crítico
alpha<-0.05
u0<-qwilcox(0.05,m,n, lower.tail = F)

# p.valor
pwilcox(U_2,m,n,lower.tail = F)
```

```
## [1] 0.0002729132
```

Por lo anterior, tenemos que

$$W = \sum_{j=1}^{n} R(Y_i) = 80. \quad \Rightarrow \quad U = W - \frac{n(n+1)}{2} = 14.$$

Y por lo tanto el estadístico de prueba es

$$U' = 118.$$

Entonces con un nivel de significacia del 95 % tenemos que el valor crítico es 93, entonces como $U_2 > 93$ podemos rechazar la hipótesis nula y por lo tanto concluir con 95 % de confianza que el tratamiento SST tuve un efecto en la disminución de consumo de alcohol.

4. Lleve a cabo una aproximación para muestras grandes y compare los resultados con los del apartado anterior, ¿Qué observa?

RESPUESTA
La aproximación para muestras grandes se basa en la normalidad asintótica de W, adecuadamente estandarizada. Dónde el estadístico de prueba es

$$W^* = \frac{W - E_0(W)}{\sqrt{Var_0(W)}} = \frac{W - \frac{n(m+n+1)}{2}}{\sqrt{\frac{mn(n+m+1)}{12}}}.$$

Entonces como tenemos la prueba de cola inferior tenemos que la desición de rechazo es rechazar H_0 si $W^* \leq z_\alpha$. Procedemos a realizar los calculos,

```
W_grande <- (W-(m*(n+m+1)/2))/sqrt(m*n*(m+n+2)/12)

z_alpha <- qnorm(0.95)
```

Entonces tenemos que el estadístico de prueba es

$$W^* = -3{,}135718,$$

por lo tanto como $W^* < 1{,}6448536$ podemos rechazar la hipótesis nula y por lo tanto concluir con un 95 % de confianza que el tratamiento SST tuvo un efecto en la disminución de consumo de alcohol. Observamos que ambas pruebas concluyen lo mismo, algo que tengo duda es cuando se considera ‘una muestra grande’, en este ejercicio los tamaños de cada muestras fueron de 11 y 12 por lo que podremos concluir que ya es apropiado utilizar la prueba de muestras grandes. ■.

Nota: Realizar la prueba en R sin usar la función predeterminada, puede comparar sus resultados con la función de R