

Reporte pruebas no paramétricas con Rmarkdown

Curso Análisis de datos con R para Biociencias

Dr. José A. Gallardo | Pontificia Universidad Católica de Valparaíso | jose.gallardo@pucv.cl

24 January 2022

INTRODUCCIÓN

Este script contiene toda la información necesaria para generar un reporte dinámico en formato pdf. Al inicio de este documento usted puede observar la **metadata** en formato **YAML**. Note que hemos agregado un título al reporte con *title*, un autor con *author*, un *subtítulo* con el nombre del curso, un formato de fecha dinámica *date* que establece la fecha del día que se imprime el reporte y el formato con el que se imprimirá el reporte a pdf con el comando *output*.

La principal diferencia entre un script de .R y un script de Rmarkdown .Rmd es que el código está separado del texto.

Markdown

Este texto que usted lee ahora se denomina markdown y puede modificarlo a voluntad cuando escriba un reporte, agregando todo tipo de formatos, así como tablas, listas, párrafos o referencias.

Bloques de código o chunk

Los códigos se agregan en bloques llamados también *chunk*. Observe con detención que el bloque de códigos comienza y termina con 3 tildes invertidas “`” y que al inicio hay una r entre dos `{ }` lo que indica al software que es un código de **R**.

Éste primer bloque es para configurar el resto de los bloques de código, por eso se ha incluido un nombre **setup**. En este primer **chunk** puede ver el llamado a ejecutar las librerías **knitr** y **tdyer**. Mientras que para ejecutar un bloque de código, solo debes presionar **la punta de flecha verde o play** que se observa en la parte superior derecha del mismo bloque de códigos.

PRUEBAS ESTADÍSTICAS NO PARAMÉTRICAS

Las pruebas estadísticas no paramétricas son un conjunto diverso de pruebas estadísticas que se aplican usualmente para variables cuantitativas que no cumplen con el supuesto de normalidad y para variables cualitativas. El concepto de “no paramétrico” a veces es confuso, pues los métodos no paramétricos si estiman y someten a prueba hipótesis usando parámetros, pero no los de distribución normal.

PRUEBA DE CORRELACIÓN DE SPEARMAN

La prueba de correlación de Spearman se utiliza para estudiar la asociación de dos variables, cuando no se cumple uno o varios supuestos de la correlación paramétrica:

- 1 Las variables X e Y no son continuas.
- 2 No existe relación lineal.
- 3 La distribución conjunta de (X, Y) no es una distribución Bivariable normal.

En el siguiente ejemplo se realizará una prueba no paramétrica de Spearman al set de datos de la siguiente tabla.

```
# Este bloque de códigos se llama resumen Spearman test.

# Crea objetos X e Y
X <- c(942,101,313,800)
Y <- c(13,14,18,10)
cor_data <- data.frame(X,Y)

# Imprime una tabla con knitr

kable(cor_data, caption = "Set de datos de correlación", col.names = c("weight","lice"))
```

Table 1: Set de datos de correlación

weight	lice
942	13
101	14
313	18
800	10

La hipótesis que se someterá a prueba es la siguiente.

H_0 : weight y lice son mutuamente independientes, $\rho = 0$

H_1 : weight y lice no son mutuamente independientes, $\rho \neq 0$

```
# Realiza test de correlación
cor_test <- cor.test(X,Y, method = "spearman",
  alternative = "two.sided")
```

La correlación de spearman entre **weight** y **lice** toma un valor de $\rho = -0.6$, no obstante aquello, no es posible rechazar la hipótesis nula pues el valor de **p**= **0.42**.

COMPARACIÓN DE DOS O MÁS MUESTRAS

2 Muestras independientes

La prueba de Mann-Whitney o Wilcoxon sirve para comparar dos muestras independientes con idéntica distribución, con diferentes medianas y cuando no se cumple el supuesto de normalidad o de variable continua.

```
# Minutos de conducta agresiva (t= solo machos; c= machos y hembras)

t <- c(9, 12, 13)
c <- c(0, 4, 6)
m_data <- data.frame(t,c)

# Imprime una tabla con knitr
kable(m_data, caption = "Set de datos de correlación",
      col.names = c("tratamiento", "control"))
```

Table 2: Set de datos de correlación

tratamiento	control
9	0
12	4
13	6

```
# Realiza prueba de Mann-Whitney
wil_test <- wilcox.test(t, c, alternative = "g",
                       paired = FALSE)

wil_test

##
## Wilcoxon rank sum exact test
##
## data: t and c
## W = 9, p-value = 0.05
## alternative hypothesis: true location shift is greater than 0
```

2 Muestras pareadas

La prueba de Wilcoxon también sirve para comparar dos muestras pareadas.

```
# Nivel de hormona pre y post maduración
pre <- c(45, 41, 47, 52)
post <- c(49, 50, 52, 50)

# Realiza prueba de Wilcoxon
wilcox.test(post - pre, alternative = "greater")

##
## Wilcoxon signed rank exact test
##
## data: post - pre
## V = 9, p-value = 0.125
## alternative hypothesis: true location is greater than 0
```

```
# no es necesario indicar muestras pareadas
# pues estamos haciendo la resta en la función.
```

Mas de 2 muestras independientes

La prueba de Kruskal - wallis es el gold standar para comparar múltiples muestras con idéntica distribución, con diferentes medianas y sin normalidad.

```
# Score de calidad de producto congelado 1 (peor) - 9 (mejor)
```

```
d0 <- c(9,8,9,8) # day0
d4 <- c(7,7,6,8) # day4
d8 <- c(6,5,5,6) # day8

calidad <- data.frame(d0, d4, d8)

kable(calidad, caption = "Score de calidad por día",
      col.names = c("día 0", "día 4", "día 8"))
```

Table 3: Score de calidad por día

día 0	día 4	día 8
9	7	6
8	7	5
9	6	5
8	8	6

```
# Este bloque de códigos se llama datos calidad.
```

```
grafica <- gather(data = calidad, key = "Dias", value = "Score de calidad", 1:3)
```

```
kable(grafica, caption = "Score de calidad por día")
```

Table 4: Score de calidad por día

Dias	Score de calidad
d0	9
d0	8
d0	9
d0	8
d4	7
d4	7
d4	6
d4	8
d8	6
d8	5
d8	5
d8	6

También puedes incluir y dar nombre a las gráficas en el mismo bloque de código usando el comando `fig.cap=" "`.

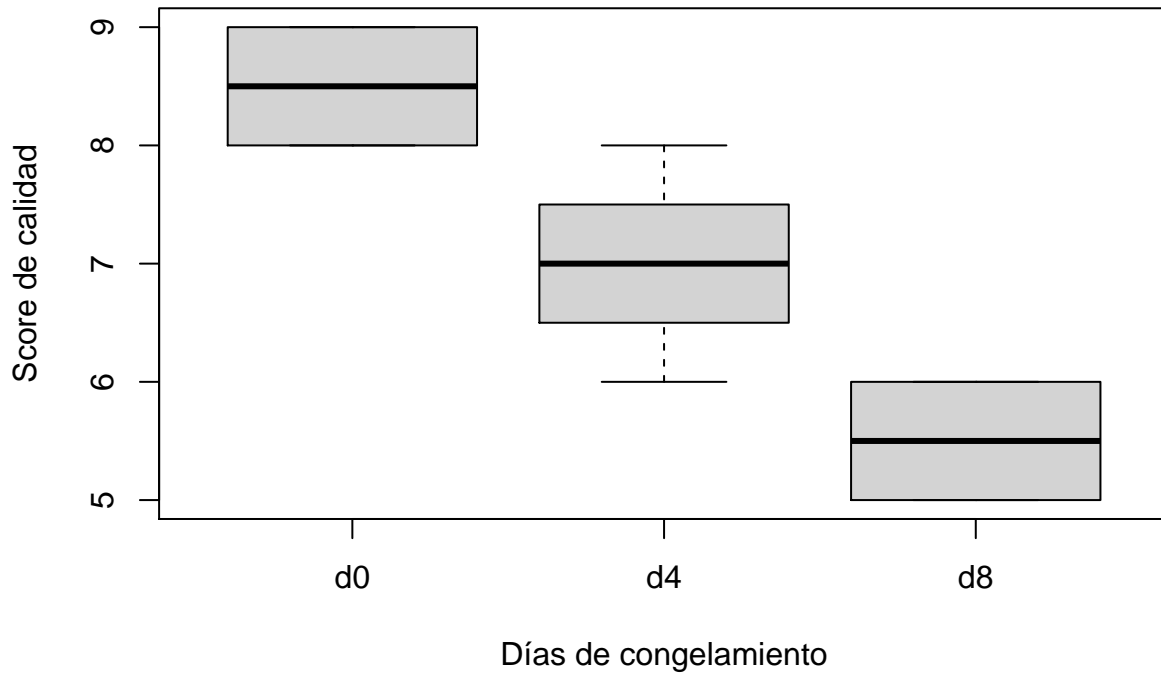


Figure 1: Calidad de producto a diferentes tiempos de congelación

```
kruskal.test(list(d0, d4, d8))

##
##  Kruskal-Wallis rank sum test
##
## data:  list(d0, d4, d8)
## Kruskal-Wallis chi-squared = 9, df = 2, p-value = 0.01111
```

Imprimir el reporte

Finalmente, para poder generar un reporte debes presionar el **triángulo negro** que está junto a la palabra **Knit** en la parte superior de este documento. Si presionas **Knit to pdf**, esto indicará a la librería **knitr** que debe leer el presente documento, interpretar y ejecutar el metadata, el texto de bajo nivel y los diferentes bloques de código para elaborar un reporte pdf.

Para finalizar, puedes generar dos reportes adicionales uno en formato html y otro en formato word. Compare los documentos e identifique sus diferencias.

Para más detalles de como trabajar con Rmarkdown investigue el siguiente **enlace**.