

Clase 06 - Programación con Rmarkdown: Reportes dinámicos de pruebas no paramétricas.

Curso Análisis de datos con R para Biociencias

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

24 January 2022

PLAN DE CLASE

1). Introducción

- ▶ Pruebas no paramétricas
- ▶ ¿Qué es Rmarkdown y para que sirve?
- ▶ ¿Cómo funciona?

2). Práctica con R y Rstudio cloud

- ▶ Escribir un código de programación con Rmarkdown.
- ▶ Elaborar reportes dinámicos de análisis de datos con Rmarkdown.

MÉTODOS NO PARAMÉTRICOS

Conjunto diverso de pruebas estadísticas.

El concepto de “no paramétrico” a veces es confuso, pues los métodos no paramétricos si estiman y someten a prueban hipótesis usando parámetros, pero no los de distribución normal.

Se aplican usualmente para variables cuantitativas que no cumplen con el supuesto de normalidad y para variables cualitativas.

SUPUESTOS: MÉTODOS NO PARAMÉTRICOS

El principal supuesto de los métodos no paramétricos es que las variables aleatorias son independientes y con idéntica distribución.

Usualmente no tienen supuestos acerca de que tipo de distribución tiene la variable, alternativamente se conocen como métodos de distribución libre.

PRUEBA DE CORRELACIÓN NO PARAMÉTRICA - SPEARMAN

¿Para que sirve?

Para estudiar asociación de dos variables, cuando no se cumple uno o varios supuestos de la correlación paramétrica:

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

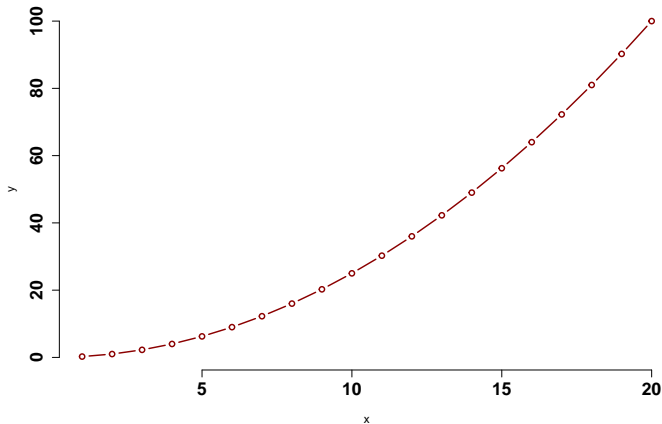
Hipótesis	Verdadera cuando
H_0 : X e Y mutuamente independientes	$\rho = 0$
H_1 : X e Y no son mutuamente independientes	$\rho \neq 0$

Coeficiente de correlación de Spearman ($\rho = rho$)

EJEMPLO FUNCIÓN MONÓTONA

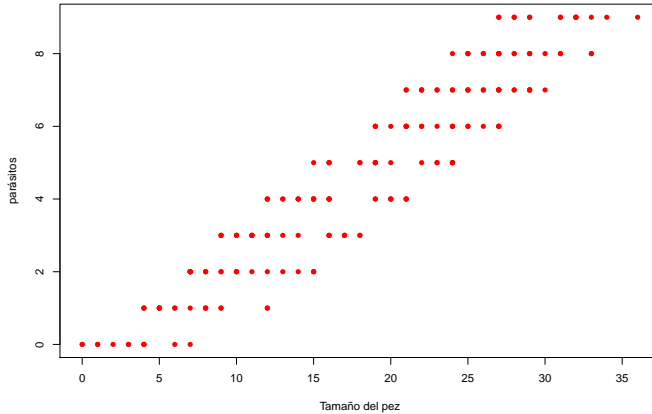
¿Cuál es el supuesto que no se cumple?

Relación no lineal



EJEMPLO VARIABLES DISCRETAS U ORDINALES

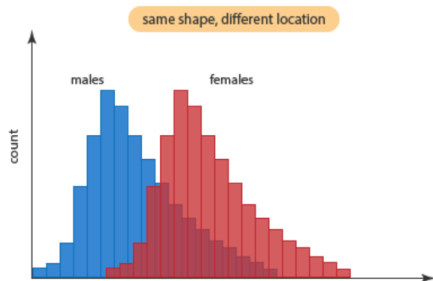
¿Cuál es el supuesto que no se cumple?



PRUEBA DE WILCOXON PARA DOS MUESTRAS INDEPENDIENTES

¿Para qué sirve?

Para comparar dos muestras con idéntica distribución, con diferentes medianas y sin normalidad. Usualmente para variables discretas.



Hipótesis

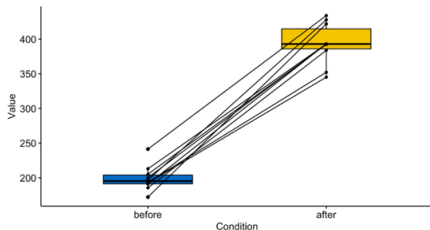
H_0 : Tratamiento = Control

H_1 : Tratamiento > Control

PRUEBA DE WILCOXON PARA DOS MUESTRAS PAREADAS

¿Para que sirve?

Para comparar dos muestras *pareadas* con idéntica distribución, con diferentes medianas y sin normalidad.



Hipótesis

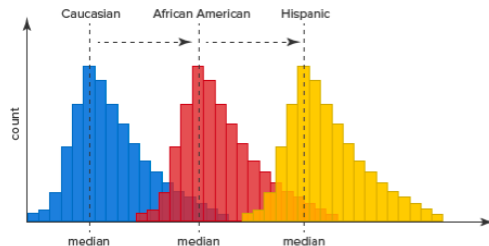
$$H_0: d = 0$$

$$H_1: d > 0$$

PRUEBA DE KRUSKAL-WALLIS PARA MÚLTIPLES MUESTRAS INDEPENDIENTES

¿Para que sirve?

Para comparar múltiples muestras con idéntica distribución, con diferentes medianas y sin normalidad.



Hipótesis

H_0 : La distribución de los k grupos son iguales.

H_1 : Al menos 2 grupos son distintos.

EXISTEN DISTINTAS FORMAS DE CREAR REPORTES

- ▶ **What You See Is What You Get**

Microsoft Word, Microsoft Power Point, formato de documento portátil (pdf)

- ▶ **What You See Is What You Mean.**

Latex, Markdown, HTML, Rmarkdown



¿QUÉ ES R MARKDOWN?

Rmarkdown es un procesador de texto que permite la creación de reportes de alta calidad.



VENTAJAS DE USAR RMARKDOWN?

- ▶ Es un software libre y de código abierto, por lo que podemos usarlo sin necesidad de comprar una licencia.
- ▶ Permite trabajar bajo el paradigma de la investigación reproducible (texto sin formato).
- ▶ Cualquiera puede crear reportes, documentos, presentaciones y libros de alta calidad, con poco esfuerzo.
- ▶ Uno de los mejores sistemas para crear reportes colaborativos y mejorar el rendimiento del trabajo de los analistas de datos.

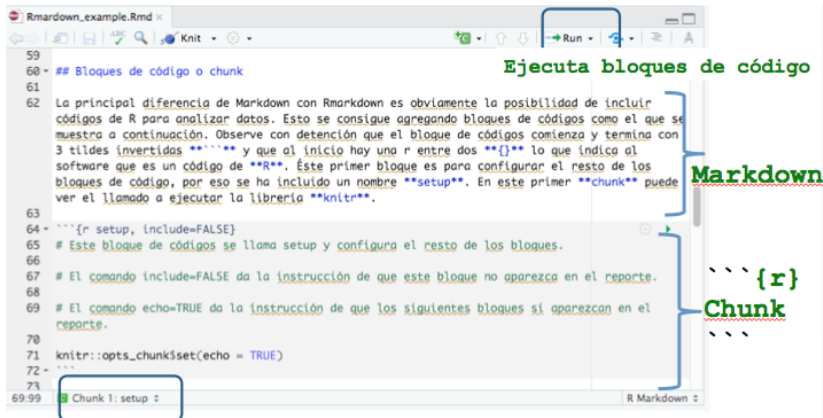
¿CÓMO FUNCIONA R MARKDOWN? 1

The screenshot shows the RStudio interface with the following elements:

- File Name:** Rmarkdown_example.Rmd
- Active Tab:** Reporte
- Other Tab:** Crea bloques de código
- Buttons:** Knit, Run, and a button to create code blocks.
- Code Editor Content:**

```
1 ---
2 title: "Ejemplo reporte pdf con Rmarkdown"
3 author: "Dr. José A. Gallardo | Pontificia Universidad Católica de Valparaíso |
4 <jose.gallardo@pucv.cl>"
5 date: "`r format(Sys.time(), '%d %B %Y')`"
6 output:
7   pdf_document: default
8   html_document:
9     df_print: paged
10   word_document: default
11 subtitle: "OCE 386 - Introducción al análisis de datos con R"
12 ---
13 ## Metadata
14
15 Este script contiene toda la información necesaria para generar un reporte
16 dinámico en formato pdf. Al inicio de este documento usted puede observar la
17 **metadata** en formato **YAML**. Note que hemos agregado un *subtítulo*, un
18 formato de fecha dinámica que establece la fecha del día que se imprime el reporte
19 y el formato pdf con el comando **output**.
```
- Annotations:**
 - A bracket on the right side of the code editor groups lines 2 through 11, labeled **Metadata en formato YAML**.
 - A bracket on the right side of the code editor groups lines 13 through 19, labeled **Markdown**.

¿CÓMO FUNCIONA R MARKDOWN? 2



The screenshot shows the RStudio interface with a file named 'Rmarkdown_example.Rmd'. The editor contains R Markdown code with several annotations:

- A blue box highlights the 'Run' button in the top toolbar, with a green arrow pointing to it and the text 'Ejecuta bloques de código' (Executes code blocks).
- A blue bracket on the right side groups the first two code blocks (lines 60-62 and 64-72), with the label 'Markdown' in green text.
- A blue bracket on the right side groups the third code block (lines 74-76), with the label 'Chunk' in green text. Above this label is the R code syntax for a chunk: `{r}`.
- A blue box highlights the status bar at the bottom left, which displays 'Chunk 1: setup', with the text 'Nombre del bloque de códigos' (Code block name) in green.

```
59
60 ## Bloques de código o chunk
61
62 La principal diferencia de Markdown con Rmarkdown es obviamente la posibilidad de incluir
63 códigos de R para analizar datos. Esto se consigue agregando bloques de códigos como el que se
64 muestra a continuación. Observe con detenimiento que el bloque de códigos comienza y termina con
65 3 tildes invertidas ```` y que al inicio hay una r entre dos **{} lo que indica al
66 software que es un código de **R**. Éste primer bloque es para configurar el resto de los
67 bloques de código, por eso se ha incluido un nombre **setup**. En este primer **chunk** puede
68 ver el llamado a ejecutar la librería **knitr**.
69
70 ```{r setup, include=FALSE}
71 # Este bloque de códigos se llama setup y configura el resto de los bloques.
72 # El comando include=FALSE da la instrucción de que este bloque no aparezca en el reporte.
73 # El comando echo=TRUE da la instrucción de que los siguientes bloques sí aparezcan en el
74 reporte.
75 knitr::opts_chunk$set(echo = TRUE)
76 ```
77
78 69:99 Chunk 1: setup R Markdown
```

Nombre del bloque de códigos

PRÁCTICA CON RMARKDOWN

- 1). Guía de trabajo programación con Rmarkdown disponible en drive.
- 2). La tarea se realiza en Rstudio.cloud.

RESUMEN DE LA CLASE

- ▶ Comunicar resultados de forma efectiva.
- ▶ Ventajas de la elaborar reportes dinámicos con Rmarkdown.
- ▶ Escribir un código de programación con Rmarkdown.
- ▶ Elaborar diferentes reportes dinámicos.