

Taller primer corte

Santiago Enrique Lozano González

Programación en R

Universidad Piloto de Colombia Seccional Alto Magdalena

santiago-lozano@unipiloto.edu.co

19 de marzo de 2020

Este taller se debe realizar en formato R Markdown que ofrece Rstudio como lo explicaré en la clase del viernes 20 de marzo de 2020, en grupos de 3 personas, así los grupos deben escoger entre realizar los puntos pares o los puntos impares para entregar en el taller, con el objetivo de tener total conocimiento de los integrantes de los grupos, deben entrar al excel de one drive para especificar los integrantes y que grupo de ejercicios se realizarán, los cupos para cada grupo de pares o impares son limitados, solo habrán 3 cupos para ejercicios pares y 2 para ejercicios impares, así que examinen los punto y decidan

Los datos como JTRAIN2 y BWGHT deben ser importados y mtcars, iris y Titanic se guardan con la codificación **data(mtcars)**, por ejemplo

1. La base de datos de JTRAIN2.txt proviene de un experimento de capacitación para el trabajo realizado para hombres con bajos ingresos durante 1976-1977; véase Lalonde (1986).
 - a) Emplee la variable indicadora train para determinar la proporción de hombres a los que se les dio capacitación para el trabajo.
 - b) La variable re78 es ingresos desde 1978, dados en dólares de 1982. Determine el promedio de re78 para la muestra de hombres a los que se les dio capacitación laboral y para la muestra de hombres a los que no se les dio. ¿Es esta diferencia económicamente grande?
 - c) La variable unem78 indica si un hombre estuvo desempleado o no en 1978. ¿Qué proporción de los hombres a los que se les dio capacitación para el trabajo están desempleados? ¿Y de aquellos a los que no se les dio capacitación laboral? Comente la diferencia.
 - d) Con base en los incisos (a) y (bi), ¿parece haber sido efectivo el programa de capacitación laboral?

2. Para responder estas preguntas emplee la base de datos BWGHT.txt.
 - a) ¿Cuántas mujeres hay en la muestra (*male* = 0) y cuántas de las informantes fumaron durante un embarazo?
 - b) ¿Cuál es la cantidad promedio de cigarros consumidos por día (*cigs*)?
 - c) Entre las mujeres que fumaron durante el embarazo, ¿cuál es la cantidad promedio de cigarros consumidos por día?
 - d) Determine el promedio de *fatheduc* (años de educación del padre) en la muestra. ¿Por qué se emplean sólo 1 192 observaciones para calcular este promedio?
 - e) Dé el ingreso familiar promedio (*famine*) y su desviación estándar en dólares.

3. Para la base de Datos mtcars extraída de 1974 Motor Trend US magazine que contiene información sobre 32 modelos de automóviles
 - a) Esciba la codificación que me permite saber si en la variable *am* ¿Cuáles tipos de carros hay más, si con transmisión automática *am* = 0 o manual *am* = 1
 - b) Haga un diagrama de dispersión entre *hp* vs. *weight*

4. Para la base de datos iris la cual da la medidas en centímetros del largo de los sépalos (Sepal.Length) y el ancho de los mismos (Sepal.Width), también el largo de los pétalos (Petal.Length) y el ancho de estos (Petal.Width), de 50 flores de 3 especies de la flor iris (Iris setosa, versicolor y virginica)
 - a) Cree un nuevo data frame solo con flores de tipo versicolor
 - b) Para el data frame creado en el inciso anterior cree un vector denominado Sepal.Diff con la diferencial entre el ancho y el largo de los sépalos
 - c) Cree una nueva variable en el dataframe del punto (a) que tenga los valores del punto (b)

5. Usando los datos de mtcars
 - a) Mediante la función *sapply* vea una forma de ver el tipo de vector de cada variable
 - b) cambie las variables 'am', 'cyl' y 'vs' to integer and y cree un nuevo data frame con este cambio
 - c) Redondee todos los datos en el dataframe de inciso anterior a un decimal después del punto

6. Usando iris

- a) extraiga del data frame todos los datos de la especie virginica con Sepal.Width mayores a 3.7
- b) Cómo extraería de la base de datos iris, de todos los datos de la especie virginica aquellas que tengan Sepal.Width mayores a 3.5 pero sin mostrar la última columna

7. Tomando iris

- a) Repita cada valor de Sepal.Length dos veces y todo el vector repítalo dos veces, es decir. De $x < -c(1, 2, 3)$, debe obtener el vector $c(1,1,2,2,3,3,1,1,2,2,3,3)$
- b) Obtenga un nuevo vector que extraiga las ubicaciones impares de Sepal.Length
- c) obtenga un vector aplicando el mismo procedimiento de (a) al vector que resultó de (b)
- d) Reemplace el vector de inciso anterior por el que está ubicado en la columna Sepal.Length de iris

8. Use mtcars

- a) Use la función `lapply()` para obtener los valores mínimos de cada columna de mtcars
- b) Use la función `sapply()` para obtener los valores mínimos de cada columna de mtcars
- c) guarde las acciones de los incisos anteriores en variables
- d) Cree una lista donde almacene los dos resultados
- e) mediante la función `sapply()` diga que clase de objeto es cada elemento de la lista

9. Usando la base de datos Titanic. Este conjunto de datos proporciona información sobre el destino de los pasajeros en el fatal viaje inaugural del transatlántico "Titanic", resumido de acuerdo con el estado económico (clase), sexo, edad y supervivencia.

- a) Use de manera apropiada la función `apply()` para obtenerla cantidad de hombres vs. la cantidad de mujeres

Survived		
Sex	No	Yes
Male	1364	367
Female	126	344

b) obtenga una tabla cruzada de sobrevivientes por sexo

	Sex	
Age	Male	Female
Child	64	45
Adult	1667	425

c) obtenga una tabla cruzada de la cantidad de pasajeros sexo vs. edad

10. cree las siguientes matrices y listas

```
first = matrix(38:66, 3)
second = matrix(56:91, 3)
third = matrix(82:145, 3)
fourth = matrix(46:93, 5)
listobj = list(first, second, third, fourth)
```

a) Extraiga la segunda columna de la lista de matrices

b) Extraiga la tercera fila de la lista de matrices

Sugerencia: Use la función `lapply()`