

Introducción a R

Manejo de objetos, lectura de datos y medidas descriptivas

Rivera Palacio, Juan Camilo
j.c.rivera@cgiar.org

Dorado Betancourt, Hugo Andres
h.a.dorado@cgiar.org

July 4, 2018

Ejercicio 1. Si x , y son vectores [1]. ¿Cual será el resultado de ejecutar las siguientes instrucciones?

```
x = c(1,3,4,5,7,9)
y = c(2,3,5,7,11,13)
```

(1)

- | | | |
|----------------------|---------------------------------------|---|
| • $x + 1$ | • $3 + \text{sqrt}(x)$ | • $y[3]$, $y[-3]$ |
| • $y*2$ | • $\text{sum}(x)$, $\text{sum}(x>5)$ | • $y[x]$ |
| • $\text{length}(x)$ | • $\text{sum}(x>5 \mid x<3)$ | • $x + y[\text{seq}(1:\text{length}(x))]$ |

Ejercicio 2. Para este ejercicio se utilizará la base de datos `mora_toyset.csv`.

1. Lea el archivo y guárdelo en una variable con el nombre `datos_mora`.
2. ¿Que tipo de clase es `datos_mora` y como se accede a las variables?
3. ¿Cuántas y de que tipo son las variables de `datos_mora`? Convierta las variables `Nar` y `Cal` en variables cuantitativas.
4. Para las siguientes variables `Yield`, `PrecAcc_2` y `trmm_3`. Calcule lo siguiente:
 - Promedio
 - Máximo
 - Mínimo
 - Varianza
 - Desviación Estándar
 - Histograma
 - Boxplot
5. Utilice la función `summary` para las variables anteriores y explique su resultado.

Ejercicio 3. Regresión Lineal Múltiple.

1. Realice un estudio de regresión lineal múltiple donde las variables predictorias sean `AB_Thorn`, `intDrain` y `slope` y la variable dependiente sea `Yield`. (Ayuda: Utilice la función `lm`.)
2. ¿Cuales son los coeficientes del modelo? y ¿Que significa estos modelos?
3. ¿Cuál es el R^2 múltiple?
4. Grafique el modelo.
5. Repita los ejercicios del 1 al 4. Utilizando como variables predictorias **TODAS** las variables.
6. ¿Cuál de los dos modelos tiene mejor resultado?. Explique

Ejercicio 4. Regresión Lineal

1. Descargue la base de datos.
http://www.uam.es/personal_pdi/ciencias/joser/datos/EdadPesoGrasas.txt
(Ayuda: `read.table`).
2. Grafique la relación entre variables usando la función `pairs`.
3. Encuentre las covarianza entre las variables. Ayuda: (`cor`).
4. Crear un modelo lineal y utilice el resumen para conocer los parámetros.
5. Hacer las predicciones de diez años. Ayuda: (`predict`).

Ejercicio 5. El paquete `MASS` contiene las bases de datos `UScereal` con información de los cereales

1. Represente cada una de las variables utilizando un barplot y/o boxplot.
2. Estime visualmente las medias, medianas, desviaciones estándar de cada conjunto de datos y a continuación calcule los valores anteriores con las funciones adecuadas. ¿Que gráfico resulta de mayor ayuda para la aproximación?

Ejercicio 6. Random Forest.

Utilice los datos `readingSkills` del paquete `party`.

1. Implemente el modelo random forest para predecir la variable `nativeSpeaker`, con las variables predictorias, `age`, `shoeSize`, `score`.
2. Muestre los resultados del modelo
3. Cuales fueron las variables más importantes del modelo.