Guía Ayudantía 3, parte 2

Viernes 22 de abril de 2022

Ayudante Santiago Ortúzar

Prof. Alejandro González

Estadística descriptiva-R, FEN-UAH

En esta guía, usted trabajará con el data frame ‘swiss’ de R base.

Para usar este dataframe, siga los siguientes pasos:

* Limpie su environment:

**rm(list=ls())**

* Fije su directorio con el comando **setwd()** (puede usar **getwd()** para buscar la ruta de forma más simple).
* Cargue el data frame ‘swiss’:

**data(swiss)**

* Puede visualizar el dataframe usando **head(swiss)**, **View(swiss)** o **print(swiss)**.

En términos globales, el objeto general de este ejercicio es que ustedes observen tres distribuciones de datos bastante distintas entre sí, y puedan entender por qué necesitamos distintos estadísticos, así como la observación de las distribuciones (vía histogramas y curvas de densidad) para describirlas apropiadamente.

1. Usando el comando ‘help(swiss)’, describa con sus palabras en qué consiste el data frame ‘swiss’: de dónde proviene la información, qué significa cada variable, cuántas observaciones tiene la base de datos y cómo está medida cada variable.
2. Trabaje usando la variable Fertility:
   1. Grafique el histograma para la variable Fertility, utilizando breaks=10.
   2. Indique la media, la mediana y la moda de Fertility. Para la moda, puede utilizar una función obtenida de una librería de R o usar el código entregado en clases.
   3. Calcule el rango de Fertility. ¿Por qué podría ser útil conocer el rango de una determinada distribución?
   4. Divida la distribución en cuartiles. Indique cuál es el valor del primer cuartil (el valor que toma la variable en el 25% de la distribución), el segundo cuartil (50%, lo mismo que la mediana) y el tercer cuartil (75%).
   5. Utilizando los estadísticos calculados entre 2.a y 2.d, describa lo que observa en el histograma.
   6. Grafique el histograma de Fertility nuevamente, pero ahora con 5 breaks, 10 breaks y 25 breaks.
   7. Ahora grafique el density plot de Fertility (sin histograma).
   8. Calcule nuevamente los histogramas de 2.f, agregando ahora la curva de densidad. Recuerde que ahora al código del histograma debe agregar el comando ‘freq=FALSE’.
3. Para la variable Education:
   1. Grafique el histograma (con curva de densidad), utilizando 8 breaks.
   2. Utilizando el rango, la media, la mediana, la moda, cuartiles y/o deciles, y cualquier otro estadístico que usted considere pertinente, además del histograma creado en 3.a, ¿qué diría usted sobre la distribución de educación en nuestra muestra?
4. Repita el ejercicio anterior para la variable Catholic. Grafique el histograma con breaks=14.