

## Introducción a la Estadística y Ciencia de Datos

### ACTIVIDAD - CLASE 2

En el archivo `ENNyS_menorA2.txt` se encuentra parte de la base de datos de la *Encuesta Nacional de Nutrición y Salud 2018-2019* de la República Argentina correspondiente a datos de bebés (o sea, cuya edad es menor a 2 años). Las variables que se seleccionaron para esta actividad son **Sexo**, **Edad**, **Peso**, **Perímetro Cefálico** y **Talla del/ de la bebé registrado/a**, además del **Tipo de embarazo** (simple o múltiple) del que nació. **Sexo** y **Tipo de embarazo** son variables categóricas y las restantes son variables continuas.

1. Para cada una de las variables categóricas,
  - a) calcular la tabla de frecuencias relativas. *Sugerencia:* Verificar que la variable en cuestión sea de clase **factor** en R.
  - b) hacer un gráfico de barras.
  - c) hacer un gráfico de tortas.
2. Calcular la tabla de contingencia entre ambas variables categóricas y hacer un gráfico de barras.
3. Para cada una de las variables continuas,
  - a) construir un histograma. ¿La distribución parece ser simétrica? ¿El ajuste normal parece razonable?
  - b) hacer un boxplot. ¿Cómo se compara la información que da este gráfico con la obtenida con el histograma? En base al gráfico obtenido, discutir simetría y presencia de datos anómalos.
  - c) calcular medidas de posición: la media muestral, la mediana muestral y las media  $\alpha$ -podadas para  $\alpha = 0,1; 0,2$ .
  - d) calcular medidas de dispersión: el desvío estándar, la distancia intercuartil y la MAD muestrales.
  - e) obtener los percentiles (muestrales o empíricos) 0, 1; 0, 25; 0, 5; 0, 75 y 0, 9.
  - f) graficar el `qqnorm`. ¿Parece razonable el ajuste normal? En caso negativo y de observar una distribución *famosa* en el histograma correspondiente, graficar un `qq-plot` pertinente.
4. Cargar el paquete `aplpack` y hacer un bagplot entre las variables **Peso** y **Perímetro Cefálico**. *Sugerencia:* Modificar el parámetro `approx.limit` apropiadamente.

5. Hacer boxplots paralelos según Sexo de la variable

- a)* Peso.
- b)* Perímetro Cefálico.
- c)* Talla.