

# Práctica 4

Trabajaremos con el archivo `Mundo95.RData`, que se debe descargar de Moodle. Dicho archivo contiene datos de 109 países. Concretamente, contiene las siguientes variables:

- `país`: nombre del país al que pertenece el caso estudiado.
- `poblac`: población del país, en miles de habitantes.
- `densidad`: densidad de población (habitantes/Km<sup>2</sup>).
- `urbana`: porcentaje de la población que habita en ciudades.
- `espvidaf`: esperanza de vida femenina, en años.
- `espvidam`: esperanza de vida masculina, en años.
- `alfabet`: tasa de alfabetización de la población.
- `pib_cap`: producto interior bruto per cápita.
- `región`: región mundial a la que pertenece el país (OCDE, Europa Oriental, Asia/Pacífico, África, Oriente Medio, América Latina).
- `calorías`: ingesta diaria media de calorías por habitante.
- `nac_def`: tasa nacimientos/defunciones.
- `fertilid`: tasa de fertilidad (número de nacimientos por cada mujer).
- `religión`: religión mayoritaria en cada país.
- `región_cod`: región codificada (OCDE: 1, Europa Oriental: 2, Asia/Pacífico:3, África:4, Oriente Medio:5, América Latina:6).

Realizar un estudio descriptivo sobre el conjunto de datos, respondiendo a las siguientes cuestiones:

- a) Calcular e interpretar las siguientes medidas descriptivas para la tasa de alfabetización: mínimo, máximo, media, percentil 30, coeficiente de asimetría y coeficiente de apuntamiento.
- b) ¿Qué datos presentan una mayor dispersión relativa, los correspondientes a la esperanza de vida femenina o los datos de la tasa de alfabetización?
- c) Recodificar la variable *urbana* en otra variable de nombre *urbana\_rec*, en cuatro intervalos igual amplitud, comenzando en 0 y terminando en 100. Construir tablas de contingencia adecuadas, para las variables *urbana\_rec* y *región* para responder a las siguientes preguntas:
  - c1) ¿Qué porcentaje de países, de los que tienen un porcentaje en ciudades entre el 25 % y el 50 %, se encuentran en América Latina?
  - c2) ¿Cuál es el porcentaje de países de Oriente medio en los que más del 50 % de la población vive en ciudades?
  - c3) ¿Qué porcentaje de países pertenecen a la OCDE y tienen un porcentaje de habitantes que habitan en ciudades entre el 25 % y el 75 %?
- d) Construir tablas de estadísticas adecuadas, que tengan como factores las variables *región* y *urbana\_rec*, que permitan responder a las siguientes cuestiones sobre la ingesta diaria de calorías:
  - d1) ¿Cuál es la media de las ingestas diarias de calorías por habitante para los países de la OCDE en los que más del 75 % de los habitantes vive en ciudades?
  - d2) Si consideramos los países en los que el porcentaje de habitantes en ciudades se encuentra entre el 25 % y el 50 %, cuál es la región en la que la ingesta diaria por habitante presenta una mayor variabilidad?
- e) Construir un gráfico de sectores para la variable *religión*. Determinar el porcentaje de países que corresponde a cada sector.
- f) Construir un gráfico de barras que muestre el número de países en los que cada religión es mayoritaria, agrupándolos según las distintas regiones. Elegir en la pestaña **Opciones**, de la ventana **Gráfica de barras**, la opción **Lado a lado** y comparar los resultados que se obtienen para los distintos grupos.

- g) Construir histogramas que permitan estudiar el porcentaje de habitantes en ciudades, agrupando los datos según las distintas regiones. Utilizar cuatro intervalos para la construcción de los histogramas que comiencen en 0 y terminen en 100.

**Sugerencias:** Construir el histograma usando el correspondiente menú de R-Commander y modificar la instrucción que aparece en la ventana **R Script**, indicando los intervalos deseados. En la pestaña **Opciones** del menú **Histograma**, seleccionar la opción porcentajes para que sea posible realizar la comparación entre los distintos histogramas.

- h) Construir una variable de nombre *dif\_ev* como la diferencia entre las variables *espvidam* y *espvidaf*. Representar e interpretar gráficos de cajas para la nueva variable, según las distintas regiones.
- i) Construir un gráfico XY para estudiar la esperanza de vida femenina en función de la tasa de fertilidad, eligiendo como condición el valor de la variable *región*. Interpretar los resultados obtenidos.