Introduction to R

TASK 1 - Using R: File `diet.dta` is a Stata database including information about several diseases and confounding variables (columns 1-16), nutrients (columns 17-26) and food consumption (columns 27-48).

1. Load data into R and save the information in an object called `diet`.

2. How many samples are this database?

3. Create another database (object called `diet.m`) containing only individuals with 'Bachiller/BUP/COU' studies (variable `estudios`)

4. Print the table `diet.m` for the rows 9,10,11 and 12

5. Which is the median weight (variable `peso`) of all samples?

6. Which is the mean weight of each type of cancer (variable `tipocancer`). HINT: type `help(aggregate)` and investigate how to do this task.

7. Create a boxplot describing the variable `t_zinc` accross the different types of cancer (variable `tipocancer`)

TASK 2 - Data analysis with R: Let us imagine that researchers are interested in determining those nutrients and foods that are associated with colorectal and breast cancer (variable `tipocancer`). Let us perform such analysis using `compareGroups` package:

1. First, create another database called `diet.cc` containing control individuals and those being diagnosed with colorectal and breast cancer (`Control, Colorrectal, Mama`)

2. Create a table describing whether patient's characteristics are comparable among cases and controls (variables `edad, sexo, estudios, peso, altura, mets_10a, mets_5a, Diabetes, Hipertensio, Colesterol`)

3. Create a table computing OR for cases vs colorectal cancer for all variables but: `id, casoc, casom, casop, casoe` and p-values for association and trend (NOTE: use subset argument in `compareGroups` function to avoid creating more datasets)

4. Create the same table for breast cancer analysis