

Introduction to R software

Session I

Teaching Assistant: Manasa Shanta Yerramalla - manasa-shanta.yerramalla@inserm.fr
Coordination : Nolwenn Le Meur – nolwenn.lemeur@ehesp.fr

Objectives: The objectives of this lab are to get familiar with the RStudio interface, to read data in R, and compute basic summary descriptive statistics.

Sleep quality among children in France in 2016

The present study deals with the quality of sleep among children. Sleep quality is of major importance for children as it seems to play a significant role in concentration, attention and watchfulness during the day and especially at school. A poor sleep quality may partly alter the learning process and be detrimental for the child.

The main objective of this study is to explore the factors associated with the sleep quality in a sample of children aged from 4 to 12 years old in France in 2016.

Children have been asked to evaluate their quality of sleep each day during a week using a visual analog scale from 1 (poor sleep quality) to 10 (very good sleep quality). An average score over the week has been computed (**sleep_quality**). Several individual characteristics have been collected such as gender and age. Other characteristics have been recorded from the parents: time spent in front of the TV after 6 p.m., number of siblings, presence or absence of story-telling before sleeping, presence of at least another child in the same bedroom, number of wakes up during the night (average score computed over the week), type of residence area, bedtime before or after 9 p.m.

Part A – Reading and formatting data using R

1. Creating a R script

1. After launching the software, use the drop-down menu to create a new script.
2. Verify where, in your computer folder architecture, R has opened, using **getwd()**
(To keep in mind: the command in R is equivalent to **pwd** in STATA)

3. Specify the working directory using the command **setwd()**
4. Save the script using the drop-down menu, creating a *.R file.

2. Open the dataset

1. Have a look at the help page for the **read.csv()** function typing `?read.csv` in the console ,
2. Using the **read.csv()** function read the data file "SleepQualityData.csv" in R into an object named *SleepQualityData* (Note that the data file has a header and data are separated by ";")
3. Have a look at the **tail()**, the **head()** of the dataset
4. Use the **View()** function to display the data in a tabular spreadsheet (be careful with capital letter!).
5. Use the **summary()** function to obtain summarized information concerning the variables in the dataset
6. Is the output consistent with what you expected in terms of the variables' types?

3. Data type conversion

1. Transform into character the ID variable
2. Transform into factor the following categorical variables:
 - **gender** variable with labels **boy** and **girl** for categorical levels **0** and **1**, respectively
 - **sibling** variable with **0="Single_child"**, **1="OneSibling"**, **2="TwoSiblings"**, **3="ThreeSiblings"** and **4="FourSiblings"**
 - **singleroom** variable with labels **No** and **Yes** for categorical levels **0** and **1**, respectively
 - **story** variable with labels **No** and **Yes** for categorical levels **0** and **1**, respectively

Part B - Descriptive Analysis – univariate analysis

1. Use again the **summary()** function to get descriptive statistics of the data.
Comment on the effect of variable transformation
2. Describe the **age** variable (mean, standard deviation, median, mode, quartiles,...).
Use the *help window* to find the appropriate commands.
3. Draw the **age** distribution using two different types of plot. Do not forget to give a title and proper axis labels to your plot. Interpret it.

4. Describe the **gender** variable using the **table()** command. To compute proportions, use **prop.table()** (pay attention to the margin option).
5. Draw a bar chart for the **story** variable
6. Build a two-way contingency table for the **gender** variable together with the **story** variable.