Tutorial 2: Importing, viewing and screening imported data

Student 1 name and ID:

Student 2 name and ID:

To do before the tutorial starts

Install and load the package psych and forcats

install.packages(“psych”)

library(“psych”)

1. Refresh your memory

During the first online session you learn some commands to obtain basic information on an object such as the mean, the median, the standard deviation.

During the first online session you learn how to load an existing dataset in R

Exercize 1:

1. Load the dataset USArrests
2. Complete the following table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Murder | Assault | Urban\_Pop | Rap |
| Mean |  |  |  |  |
| Median |  |  |  |  |
| Standard dev. |  |  |  |  |

1. Create a R script with the commands that you use for this exercize
2. Use the command describe to analyze the variable Murder. What is the advantage to use describe compare with the command summary?

Exercize 2:

1. Load the UCBAdmissions data from R
2. Which command can help you to explore those data

Explore the other commands to get ride of your data

The command **summary** is particularly interesting for knowing basic information about the distribution of continuous variable which are named numeric in R. However this command does not work for factor and character variable. Different command are necessary to explore different type of variable

1st Possibility:

When the variable is **numeric, integer, factor or character** the command str can be use

str(object)

2nd Possibility:

When you have a **numeric or integer** the command describe can be used

describe(objet)

3rd Possibility

When the variable is **factor**, I can know the possible “label” of the answers using the command levels(object).

4th Possibility

When the variable is **factor**, you can use the command table to explore the variable

table(object)

Exercize 3:

1. Load the gss\_cat data from R
2. Use the str command and tell how many levels do you have in the variable rincome?
3. Use the command levels to know what is the ninth level of the variable rincom?
4. Is rincome a well ordered variable? Why?
5. In the gss\_cat dataset how many people are widowed? divorced?
6. What is the proportion of Buddhist?
7. What is the 9th label in the vector “relig”

Exercize 4:

1. Load the dataset Golf\_trainer\_short.Rdata
2. Which variables contain levels arranged in a messy order?
3. What can you say about the variable Year\_birth?
4. How many character variable do you have in the dataset?
5. For which variable can we find a level named “missing” or “missing value”
6. To sum-up what do you need to before analyzing those data?

Recap:

summary

describe

str

levels

table