Problem set 1
Econometrics 2018

Boyko Amarov

February 25, 2019

Problem 1 [**Arithmetics, Assignment**]

a) Create a new script tile called `Problem_1.R` in R-Studio. Use it to store the code for this problem.

b) Calculate the sum of 8 and 31 and assign it to a variable named `x`.

c) Calculate 2 to the power of 8 and assign it to a variable named `y`.

d) Calculate the ratio of `x` and `y`.

e) Run the code `X + y` and *read* the error message.

f) Assign the string literal `"six"` to a variable called `x_char`.

g) Try to calculate the sum of `x_char` and `y` and *read* the error message.

h) Assign `TRUE` to a variable named `x_true` and `FALSE` to a variable named `x_false`. Calculate the sum of these variables and inspect the result.

Problem 2 [**Vectors**]

A vector is an ordered collection of values. Each vector can hold an arbitrary number of values of a *single* type. The primary function used to create vectors is `c` (concatenate). Sometimes we need to create vectors with special structures, e.g. the numbers from 1 to 100, the even numbers from 2 to 20, etc. R provides the built-in functions `seq` (sequence), `rep` (repeat) and `:` to help in these cases.

a) Create a new script tile called `Problem_2.R` in R-Studio.

b) Use the `c` function to create a numeric vector with values: 1, 4, 5.8 and assign it to an object called `x`.

c) Use the `c` function to create a numeric vector with values 2, 3, 8 and assign it to an object called `y`.

d) Use the `c` function to concatenate `x` and `y` to a vector called `z`.

e) Create a vectors with the numbers from 1 to 10 using both `seq` and `:`.

f) Create a vector of length 10 with the number 1 at each index. *Hint*: use the `rep` function.

g) Create the vector 1, 2, 3, 1, 2, 3, 1, 2, 3 using `rep`.

h) Create the vector 1, 1, 1, 2, 2, 2 using `rep`.

i) Use the assignment operator to change the first element of the last vector you created from 1 to 3.

j) Use the assignment operator to change the last element to 9.

k) Use the subset operator `[]` to select the first and the last elements.

l) Use the subset operator to select only those elements that are equal to 1.

Problem 3 [Data frame, subsets, descriptive statistics]

The dataset `mall_customers` contains data collected on mall customers described by the following variables:

Id : Customer Id

Gender (character) : Customer gender (male/female)

Age (numeric) : Age in years

Income (numeric) : Annual income (in 1,000 USD)

Score (numeric) : Customer purchasing score computed by the mall.

a) Copy paste the content of `Problem_3.R` located in the github repository into a script in your R-Studio.

b) Run the code that downloads and reads the data file.

c) Examine the result using the `str` function.

d) How many customers are included in the dataset?

e) What is the age of the first customer in the dataset?

f) What is the gender of the last customer in the dataset?

g) Compute the average customer age in this dataset using the `mean` function.

h) Compute the minium, maximum, and median age and interpret the results.

i) Compute the 0.05, 0.2, 0.5, 0.75 and 0,95 empirical quantiles of age and interpret the result.

j) How many customers were younger than the average? *Hint* create a `logical` vector and compute its sum using `sum`.

k) How many customers were male/female? Compute the result using `summary` and `table`.

l) Is there a difference (on average) between the age of female and male customers?

m) Use a box plot to compare the distribution of age by gender.

n) Use a scatter plot to examine how Score varies with age. Interpret the result.

o) Create a new variable in the dataset `mall` that is `TRUE` if age is less than 30 years and `FALSE` otherwise. Choose an informative name for the new variable.

p) Compare the distributions of Score by age group.