Lab Exercises

1. Install and the load the package bootsrap.
2. The volume of a sphere with radius R is given by $\frac{4}{3}\pi R^3$. Using the for loop in R find the volumes of the spheres with radii 3, 4, 5,..., 20. Repeat the exercise by using vectors in order to avoid the usage of the for loop.
3. Multiply all the numbers from 1 to 200 using the for loop. Repeat the exercise by using the function prod.
4. Load in R the following data as a numerical vector

   2, 3, 1, 5, 6, 3, 2, 5, 7, 9, 10, 11, 9, 8.

   Find the length of the vector, the minimum and the maximum. Sort the vector and then find the ranks of its values using all available methods to treat the ties.
5. Create a function in R that has an argument a numerical vector x. The function should sort the values of the vector and then return the median.
6. Try in R to make the calculation $\frac{22^{300}}{21^{250}}$. What do you notice? Using the log function repeat the calculation.
7. Read the data from http://www.math.ntua.gr/~fouskakis/data1-da.txt in R and create a matrix X with 6 columns. Find the dimension of the matrix. Then export only the 2nd and 3rd columns of the matrix to a file. Using the function apply() add by row all the elements of X. Convert the matrix to a data frame and give the following names to the columns: ID, AGE, FEV, HEIGHT, SEX, SMOKING. Finally add a column of 1s to the data frame.
8. Create a function in R that takes as an argument a square matrix X. If the matrix is not squared the function should return an error message. The function should also return an error message if the matrix is not invertible. In all other cases the function should return the determinant of the matrix and its inverse.
9. Using R do the following calculations:

a) If $X \sim N(0, 1)$, find $P(X > 1)$.
b) If $X \sim N(0, 1)$, find y: $P(X < y) = 0.8$.
c) If $X \sim N(\mu = 10, \sigma^2 = 4)$, find $P(X < 11)$.
d) If $X \sim N(\mu = 10, \sigma^2 = 4)$, find y: $P(X > y) = 0.6$.
e) If $X\sim Beta(2, 2)$ find $P(X \leq 3)$.
f) If $X\sim Gamma(4, 2)$ find $P(1 < X < 2)$.
g) If $X\sim Bin(10, 0.7)$, find the value of the probability mass function (p.m.f.) for every possible value of the random variable X.
h) If $X\sim Nbin(3, 0.5)$, find the value of the p.m.f. for the values 3,…,10.
i) Find the median of the Snedecor distribution with parameters 3 and 5.
j) Find the value of the probability density function (p.d.f.) of the chisquare distribution with parameter 10 for x = 1, 3 and 5.
k) Simulate 100 variates from the hypergeometric distribution with parameters m = 10, n = 7 and k = 8.

10. Make the plot of the following p.d.f.s (or p.m.f.s).
a) Snedecor(10, 15).
b) Weibull(2, 4).
c) NegBin(2, 0.7).

11. In an environmental study we wish to check if the height of a particular plant is affected by its geographic position. In the following table we have data from 28 different areas in Greece and we present the height (in meters) of the plant and the geographic position (1: North Greece, 2: South Greece).

Height	Geographic Position	Height	Geographic Position
3.33	1	0.84	1
5.87	2	3.95	1
6.31	2	4.16	1
1.98	2	2.99	1
4.86	1	2.15	2
2.65	2	6.78	2
2.78	1	2.84	1
2.21	1	4.23	2
0.45	2	2.89	1
1.51	1	2.56	1
0.56	2	2.65	2
3.68	1	1.54	2
2.16	1	0.98	2
3.15	1	0.87	2

a) Generate a data frame in R that contains the above data and name appropriately its columns.
b) For the variable Height calculate the mean, median, s.d. and the 1st and 3rd quartiles. Also find the interquartile range.

c) Using different number of classes plot make the histogram of the values of the variable Height using frequencies and relative frequencies.
d) In the same graph present the boxplots of the plant Heights for the two different Geographical Positions. Give appropriate names to each boxplot.
e) Make a plot of all the values of the variable Height in your sample and then replace the plotting symbol with the characters ‘1’ and ‘2’ for the north and south Geographical Positions respectively. Give a title to your graph and an appropriate legend describing the meaning of the two different plotting symbols.
f) Calculate the median of the Height of the plants for each Geographical Position separately.
g) Create a new variable “Height_Cat” that takes the value 1 if the plants have Height $\leq$ 1 meter, the value 2 if the plants have Height between 1 and 3 meters and the value 3 for plants with Heights > 3 meters. Using the command levels() name the categories of the new variable as "Short", "Medium" and "Tall". Create the contingency table for the variables Height_Cat and Geographical Position.
h) For the previous contingency table find the relative frequencies for each cell as well as relative frequencies for rows and columns. Make stacked and grouped barplots.

12. Create a 3×2 graphical window in R. By using the curve() command, depict the plots of $f(x) = x^i$, $i = 1, \ldots, 5$ in the interval [-70, 70] as well as the plot of the function $g(x) = |x|$. Give different color as well as different line type and line width in each graph. *Hint:* the for loop and the paste() command may be useful in terms of code optimization.

13. Consider the following p.d.f. of a random variable X coming from a Gamma distribution with parameters a and b:

$$f(x) = \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)}, \quad x > 0,\ a > 0,\ b > 0.$$

a) Construct a contour plot of the given p.d.f. Assume that X takes values in the interval [1, 5], the shape parameter a takes values in the interval [0.01, 10] and the rate parameter is constant (b = 2).
b) Regarding the same p.d.f., construct the corresponding colored grid using the image() command and the rainbow(30) color palette.

14. Consider the mtcars dataset in R. The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and various aspects of automobile design and performance for 32 automobiles. We are interested in the relationship between car weight (wt) and the consumption index in miles per gallon (mpg). Use appropriate values in the plt argument of the par() command in order to properly formulate the space of the graphical window which will follow. Construct the scatterplot of car weight (x axis) against the fuel consumption index (y axis) adding suitable axis labels. Next, after using the plt and the add = T arguments of the par() command, add a horizontal boxplot of car weight on the upper margin of the graphical window. In a similar manner, add a vertical boxplot of the fuel consumption index on the

right margin of the current graphical window. The boxplots should have different colors and no axes. Finally, give a title to the overall graph using the mtext() command.