Assignment 1 - November 8, 2021 - Advanced Multivariate Statistics
Instructor: Giancarlo Manzi
Master in Data Science & Economics

NOT MANDATORY!

Must be done alone. Maximum score: 3 points.
If you have questions, please contact the instructor for assistance.

Issued on Monday, November 8th, 2021 **Due by** midnight on Friday, November 12th, 2021

Part a: (***Solutions should be found using*** `R`).

****

Consider the following function:

$$F(x) = \begin{cases} 1 - \exp(-\frac{1}{8}x^2) & x > 0 \\ 0 & x \leq 0. \end{cases}$$

(i) Check whether it is a proper CDF using the limit properties only.

- If it is a proper CDF:

(ii) Plot it.

– If X is the corresponding r.v.:

(iii) plot its density for $X > 0$;

(iv) compute $P(0 < X \leq 2.3)$;

(v) compute $E[X]$;

(vi) compute $Var[X]$;

(vii) compute the median.

Hints: for (i) you might use the `Ryacas R` library; for (ii) you might use the `makeFun()` function in the `mosaic R` package (for piecewise functions) and the `curve() R` function; for (iii) you might use the `D()` function in the `stats` library.
****

Part b: (***Solutions should be found using*** `R`).

****

Consider the `CustomerCare.csv` data set on rating about a new product (columns `Rating_price` and `Rating_quality`). After considering the variable `Education` as group variable:

(i) Perform a MANOVA test to determine if there are differences in the mean vectors of the ratings in the education level groups.

(ii) Perform univariate ANOVA tests on each variable.

(iii) Compare and comment on results.

Hints: for (i) you might use the `manova R` function; for (ii) you might use the `aov()` function.
****

Part c: (***Solutions should be found using*** `R`).
****
Consider the `CustomerCare.csv` data set on rating about a new product (column `Rating_price`).
Implement the classical bootstrap to estimate the (95%) confidence intervals of the population Pearson's correlation coefficient between rating and `Age`. Compare with the Fisher transform method. Implement an iterative procedure and a package procedure (for example using `boot`).
Plot the confidence intervals to visually compare the results.
Let Y be the `Rating_price` and X the `Age`. Compute the (95%) boostrap confidence interval for $\frac{Y}{X}$. Comment on results.
NOTE: This part is similar to an example given in a previous class and should be intended as an exercise in which you can freely add extra material on bootstrap. *For example, you may extend that example including the use of other bootstrap* `R` *packages you may find.*
****

A script containing the `R` code and a MS word file with solutions (and comments) should be sent to **giancarlo.manzi@unimi.it**.
You can also use an `R` markdown script file containing both code and comments.
Please give your student ID number in your message.