**Linear Models (20 pts)**

In this section the exercises must have a justification with the model used and the reasons why it was selected: Simple regression, ANOVA, ANCOVA or logistic regression. Statistical tests. The use of the test must be justified. Why did you selected the test to be parametric or non-parametric, one-tailed, or two-tailed test, the p-value threshold to accept or reject the null hypothesis?

The database may have missing values or outliers. Thus, you may have to clean them and filter them. Justify what you did to get the database to be analytical data.

Database: Weather

a) Dependent variable: rain or no rain (PRCP) Independent variables: Month, minimum and maximum temperature. But they can be filtered afterwards.

b) Do a statistical test to see if there is a significant difference in the minimum temperatures of March vs October

**Algorithm Coding. (20 pts)**

All algorithms must be coded by yourself. The only libraries accepted for this part are the ones that already come with the R language. Do not use any package that already have the algorithm coded. Also, the code must be commented with the instructions to use it and at least one example to demonstrate its functionality. It is recommended to do a flow chart to explain the logic of the algorithm. Each algorithm must be coded within a function

TIPs. Use the functions: expression, D, parse and paste within a loop to get the desired output.

a) Linear systems of equations Gauss elimination. The algorithm receives a matrix X with all the equation coefficients. The code must provide as output the solution of the system.

b) Lagrange polynomials. This algorithm receives a nx2 matrix, where the first column represents the x coordinate while the second column represents the y coordinate. The code must provide as output the Lagrange polynomial interpolation expression in terms of “x”.

**Machine Learning  part 1 (20 pts)**

Use python to perform the analysis of the following questions. Code must be commented with the logic of your proposal.

Submit your code (jupyter notebook/ py file)  with the answers to the questions embedded as comments

Perform a PCA analisis using dataset Xab\_180830B.csv , then answer the following questions:  
a) What are the features that best explain the variability of the data?  
b) How much will your data be reduced if you only keep information from the first principal component?

**Machine Learning  part 2 (30 pts)**

**Answer the following questions for WINE dataset**

* Is this a classification or a regression problem? Justify your answer
* Which approaches could we use to work with this data? Depending on your answer, select two methods, perform the analysis in python, and compare the results.
* Do we need all of the features to be used? Justify your answer