**Student Number:** C15311966

**Name:** Maksymilian Drzezdzon

**Programme Code:** TU060/1

# Section 1 - Research Question(s)

State your research question(s).

In this section you should state the main research question/questions you are hoping to investigate by using linear/logistic regression or dimension reduction.

There is no need to state sub-questions which relate to establishing evidence for including predictors (e.g. establishing correlation or difference).

# Section 2 - Dataset

In this section you should present a relevant analysis of the dataset (which you are treating as a single sample). You should consider

* Representativeness
* The variables of interest to your research questions and their statistical measurement type
* Provide relevant descriptive statistics and visuals
* Assess normality where relevant
* Missingness

Report your findings as we discussed through the lectures.

For this part of the assignment you will be using a dataset of academic performance evolution for engineering students. The dataset contains the results in national assessments for secondary and university education in engineering students and contains academic, social, economic information for 12,411 students. The data was collected as part of the Master's Degree in Engineering project of the Technological University of Bolívar (UTB) titled Academic Efficiency Analysis in Engineering students

A full descriptor is available at: <https://www.sciencedirect.com/science/article/pii/S2352340920304315#utbl0001>

The dataset is available for download at <https://data.mendeley.com/datasets/83tcx8psxv/1>

Delahoz-Dominguez, Enrique, Rohemi Zuluaga, and Tomas Fontalvo-Herrera. "Dataset of academic performance evolution for engineering students." *Data in Brief* (2020): 105537.

# Section 3 - Results

In this section you are going to present the results of your tests and models. You should adhere to the sections for Option A or Option B as relevant (refer to the assignment specification).

## Option B

## Section 3.1 Dimension Reduction

In this section you should present your assessment of the data for dimension reduction and the outcomes. Present the findings as discussed in lectures and include relevant illustrations.

You should report (which includes commenting on):

* Correlation matrix
* Bartlett’s Test of Sphericity
* KMO
* Loadings
* Communalities
* Variance Explained
* Factors/Components extracted

## Section 3.2 – Model

### Section 3.2.1 Statistical Evidence

Of particular importance for Option B is to describe the factors/components extracted. You should include the relevant descriptive statistics and visualisations. If you are unable to use the factors/components (or none were suitable) then outline the proxy variables you intend to use in the model for the concept of interest and describe these if you haven’t previously done so.

In this section you should present the results of your correlation and difference tests conducted to justify inclusion of predictors in your model/describe the manifest variables you intend to use in your dimension reduction.

### Section 3.2.2 Model

In this section you should present the findings from the first model. Present the findings as discussed in lectures and include relevant illustrations. You should:

* Report the model as discussed during lectures
* Make sure you comment on the fit and usefulness of the model in your report
* Make sure you comment on the predictors and their significance
* Make sure you assess the model against assumptions
* Make sure you illustrate how the model works for the dataset e.g. state the regression equation and illustrate how it works by using mean/median values for continuous data and the differences by looking at the equation for all values of the differential effects.

## Section 3.n Model n

Please add additional sections if you wish to include more than 1 model.

# Section 4 – Discussion/Conclusion

In this section you should reflect on your results from the perspective of your research question(s). You should also suggest some changes/additional research that could be conducted to try to better answer this research question.