Probability and Statistical Inference

Continuous Assessment Part II

Semester I 2020/2021

DATASET

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.

MARKS

This part of the assignment is worth 40% of the module marks.

Probability and Statistical Inference
Continuous Assessment Part II
Semester I 2020/2021

WHAT DO I NEED TO DO?

Choose either Option A or Option B.

Option A:

- State (a) research question(s) which can be investigated using statistical analysis.
- Develop hypotheses testable by building a linear or a logistic regression model.
- Derive and present appropriate statistical evidence.
- Using either linear regression or logistic regression.
    - Build a baseline regression model with at least 2 predictors.
        - Assess its fit and usefulness.
        - Test the assumptions of the approach.
        - Illustrate your findings using appropriate examples from your data.
    - Build at least one additional model which extends this baseline either adding or removing predictors relevant to your hypotheses.
        - Note: your model must have at least 2 predictors.
        - Assess its fit and usefulness.
        - Test the assumptions of the approach.
        - Illustrate your findings using appropriate examples from your data.
    - Compare the fit and usefulness of your successive models.
- **Note**: At least one of your models should investigate a differential effect.

Option B:

- State (a) research question(s) which can be investigated using statistical analysis.
- Develop hypotheses testable by dimension reduction followed by either linear or logistic regression.
- Derive and present appropriate statistical evidence.
- Conduct a dimension reduction using either principal components analysis or exploratory factor analysis.
    - Assess the suitability of the dataset for dimension reduction.
    - Demonstrate the outcomes.
    - Assess the effectiveness of the dimension reduction.
    - Illustrate your findings using appropriate examples from your data.
- Using either linear regression or logistic regression.
    - Build a baseline regression model with at least 2 predictors.
        - Assess its fit and usefulness.
        - Test the assumptions of the approach.
        - Illustrate your findings using appropriate examples from your data.
- **Note**: The objective is to use either components or factors derived from the dimension reduction as part of a linear regression or logistic regression model. If the dimension reduction does not succeed, identify an alternate mechanism to derive a measure for concept for which you conducted dimension reduction.

Probability and Statistical Inference

Continuous Assessment Part II

Semester I 2020/2021

For Each Option:

You are expected to :

- Present a summary of the variables representing the concepts of interest, critically discussing relevant issues which impact statistical analysis;
    - Include statistical summaries of the variables of interest and evidence to support relationships or difference to justify their inclusion in a dimension reduction or regression model.
- Use appropriate statistical techniques to achieve your goals.
- Present and interpret the findings;
- Briefly draw conclusions discussing your findings;
- Adopt the APA guidelines for reporting statistical analysis using APA citation and referencing.
- You must use R to conduct your analysis;
- You should cite appropriate sources (which are accessible) in order to support the guidelines you adopt in your decision making and interpretation of findings.

Your will need to demonstrate:

- An ability to state a research question suitable for a statistical analysis;
- Generate and correctly state a hypothesis or hypotheses;
- The ability to correctly prepare, present, analyse and critically assess the dataset used from the perspective of statistical analysis;
- The ability to correctly execute, present and interpret appropriate statistical tests using statistical software;
- The ability to analyse and present the findings gained from your statistical analysis in a clear and accurate way to a standard expected of postgraduate level academic work;
- The ability to construct a report on a statistical inquiry.

Probability and Statistical Inference

Continuous Assessment Part II

Semester I 2020/2021

What Do I Submit?

- You are required to create two artefacts:
  - A report constructed adhering to APA guidelines which addresses the items outlined in the overview.
    - A guide to writing a paper APA style is available at https://www.verywellmind.com/how-to-write-an-apa-format-paper-2794838
      - Note: You should follow this in terms of layout but you are not required to research the topic beyond describing the dataset.
  - The R commands plus output generated from this to support the statistics and information included in your report is required. It should be possible to execute the R commands to verify the statistics you have included.
- You are required to submit a URI to the section of your digital portfolio where the artefacts can be located - include any guidance you feel is needed in order to navigate the portfolio.

How Do I Submit?

You will submit via Brightspace via the Assignments section in the assignment called Independent Project.

Basic Marking Scheme

Option A	
Ability to correctly develop and present a research question and hypotheses suitable for testing using linear or logistic regression.	10
Ability to correctly prepare, present, analyse and critically assess the dataset used for the purposes of building the proposed regression models.	20
Model 1 – description, assessment, illustration	25
Model 2 -description, assessment, illustration	20
Comparison of successive models using appropriate statistics.	15
Ability to interpret the findings from your data within the context of your question and draw conclusions from this.	10
Option B	
Ability to correctly develop and present hypotheses suitable for testing using dimension reduction and either linear or logistic regression.	10
Assessing the suitability of the dataset for the purposes of the dimension reduction technique chosen.	15
Description and interpretation of the outcomes and effectiveness of the dimension reduction technique	25
The ability to correctly prepare, present, analyse and critically assess the dataset used for the purposes of building the proposed regression model.	15
Model – description, assessment, illustration	25
Ability to interpret the findings from your data within the context of your question and draw conclusions from this.	10

Probability and Statistical Inference

Continuous Assessment Part II

Semester I 2020/2021

NOTES

1. Unfair practice is a very serious offence in TU Dublin and you must acknowledge any material used by including a referenced bibliography in your report. Any issues will be investigated and those considered serious will be handled via the TU Dublin Plagiarism policy (details are available in the General Assessment Regulations).
2. You are required to treat the dataset provided ethically and conduct your statistical analysis ethically. As such you should adopt the guidelines for ethical statistical practice provided by the American Statistical Association https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx
3. Assignments must be submitted via Brightspace through the assignment section.
    - Email submissions will be ignored.
4. Late Submission:
    - Extensions due to acceptable personal circumstances must be requested by email in advance of the deadline.
    - For late submissions without an agreed extension, a penalty of 5% will be applied for every day a submission is late.
    - No submissions will be accepted after Saturday 19th December @ 23:59 unless an extension has been agreed.

      NB: Anything submitted later than this date without agreement will be ignored.
5. Assignments which do not adhere to the requirements will attract a penalty of up to 10%.
6. No resubmission of assignments after feedback is given is allowed.