

Probability and Statistical Inference
Continuous Assessment Part II
Semester I 2022/2023
Due Date: Friday 6th January 2023

Introduction

For this part of the assignment, you are going to conduct:

- a dimension reduction,
- linear regression
- and logistic regression.

For the dimension reduction you will use a new set of data designed for this purpose. The data to be used is available as part of the assignment in Brightspace.

For the linear and logistic regression, you can use some of the same variables you used for part I of the CA or new variables from the dataset used in part I you have chosen for the purposes of this part of the CA. If you choose the same variables then you do not have to repeat your assessments simply include relevant descriptive statistics and report relevant test findings as evidence for inclusion of variables as predictors. If you choose new variables you will need to present an assessment and conduct any relevant tests to generate evidence for inclusion of variables as predictors.

You are required to develop clear hypotheses which you then test using the multivariate techniques proposed.

This part of the assignment is worth 60% of the module marks.

Probability and Statistical Inference
Continuous Assessment Part II
Semester I 2022/2023
Due Date: Friday 6th January 2023

Task Overview

You are expected to :

- Discuss the data used using appropriate statistical illustration addressing relevant issues which impact statistical analysis (e.g. statistical summaries, comments on normality, reflection on outliers and missing data, statistical evidence of relationships and differences);
- Use appropriate statistical techniques to achieve your goals;
- Present and interpret the findings;
- Briefly draw conclusions discussing your findings in terms of other related work and any implications for future work;
- 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 generate and correctly state a hypothesis or hypotheses that is/are theoretically-informed;
 - 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 masters level academic work;
 - The ability to construct a report on a statistical inquiry.
-

Probability and Statistical Inference
Continuous Assessment Part II
Semester I 2022/2023
Due Date: Friday 6th January 2023

Task Details:

1. Dimension Reduction
 - Develop hypotheses testable by dimension reduction using 30 of the variables in the dataset provided.
 - Present appropriate statistical summaries of the variables chosen.
 - Assess the suitability of the dataset for dimension reduction.
 - Conduct the dimension reduction
 - Present and discuss the outcomes of dimension reduction.
 - Assess the effectiveness of the dimension reduction.
 - Present conclusions from your analysis.
2. Linear Regression
 - Develop hypotheses testable by building a linear regression model.
 - Your hypotheses must involve three concepts as predictors
 - Present appropriate statistical summaries and evidence for inclusion of your predictors.
 - Build a regression model (including at least two predictors)
 - Assess its fit and usefulness.
 - Test the assumptions of the approach.
 - Illustrate your findings using appropriate examples from your data.
 - Present conclusions from your analysis.
3. Logistic Regression
 - Develop a hypothesis testable by building a linear regression model.
 - Your hypothesis must involve at least two concepts as predictors.
 - Present appropriate statistical summaries and evidence for inclusion of your predictors
 - Build a binary logistic 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.
 - Present conclusions from your analysis.
4. Model Comparison
 - For either the linear or logistic regression model developed
 - Build one additional model which extends your baseline model (the first model developed) either adding or removing predictors relevant to your hypotheses (this must have at least two predictors)
 - Assess its fit and usefulness.
 - Test the assumptions of the approach.
 - Compare the fit and usefulness of this model with the first model created.
 - Illustrate your findings using appropriate examples from your data.

Note 1: At least one of your hypotheses for regression must involve a differential effect.

Note 2: All models must have at least two predictors and at least one of your models must include a nominal variable.

Probability and Statistical Inference
Continuous Assessment Part II
Semester I 2022/2023
Due Date: Friday 6th January 2023

DELIVERABLES

- You are required to address two aspects in your submission:
 - Report on your analysis, adhering to APA guidelines for reporting relevant test results, addressing the items outlined in the overview.
 - An RMD file containing the R commands plus an file containing the 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.

SUBMISSION

- All required documents should be submitted using the **Continuous Assessment Part II** assignment in Brightspace in **one single zip file** containing all of your files named with your student number e.g. D123456.zip.
 - You must include the following information at the start of all files submitted:
 - Student Number: <<your student number>>
 - Student Name: <<your name>>
 - Programme Code: <<programme code>>
 - The version of R used.
 - The R packages needed for your code to execute successfully.
 - All files must include your student number at the start of the file name e.g. D123456.rmd, D123455.nb.html.
 - Only .rmd files will be accepted.
-

Probability and Statistical Inference
Continuous Assessment Part II
Semester I 2022/2023
Due Date: Friday 6th January 2023

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. You are required to adopt the APA guidelines for reporting statistics and for citation <https://apastyle.apa.org/> and report your tests adhering to APA conventions (this style guide should provide you with the information you need <http://spss.allenandunwin.com.s3-website-ap-southeast-2.amazonaws.com/Files/APAStyle.pdf>)
 4. Assignments must be submitted via Brightspace through the assignment section. Email submissions will be ignored.
 5. Extensions due to acceptable personal circumstances must be requested by email in advance of the deadline.
 6. For late submissions (i.e. without an agreed extension), a penalty of 5% will be applied for every day a submission is late.
 7. No submissions will be accepted after Friday January 13th @ 23:39 unless an extension has been agreed.
NB: Anything submitted later than this date without agreement will be ignored.
 8. Assignments which do not adhere to the requirements or which are submitted incorrectly will attract a penalty of up to 10%.
 9. No resubmission of assignments after feedback is given is allowed.
-

Probability and Statistical Inference
Continuous Assessment Part II
Semester I 2022/2023
Due Date: Friday 6th January 2023

Marking Scheme

Hypotheses	
The ability to correctly develop and present hypotheses suitable for testing using dimension reduction and either linear and logistic regression. ¹	10
Dimension Reduction	
Description of the conduct and outcomes of the dimension reduction technique	10
Critical discussion of effectiveness of the dimension reduction technique including description of the outcomes.	12
Linear Regression	
The ability to correctly prepare, present, analyse and critically assess the dataset used for the purposes of building the proposed linear regression.	10
Model– description of conduct and outcomes, assessment of fit and usefulness, illustration of usefulness.	10
Critical discussion of model and overall conclusion.	6
Logistic Regression	
The ability to correctly prepare, present, analyse and critically assess the dataset used for the purposes of building the proposed logistic regression.	10
Model – description of conduct and outcomes, assessment of fit and usefulness, illustration of usefulness.	10
Critical discussion of model including overall conclusion.	6
Comparison model	
Model - description conduct and outcomes, assessment of fit and usefulness, illustration of usefulness.	10
Critical discussion of model including relevant statistical comparison with baseline model and overall conclusion	6
	100

¹ At least one of your hypotheses for regression must involve a differential effect and therefore at least one of your models needs to include a nominal variable as a predictor.