Probability and Statistical Inference
Continuous Assessment
Due Date: Saturday December 12th 2020 @ 23:59

OVERVIEW

You are required create a digital portfolio to demonstrate:

- Your knowledge and understanding of key concepts related to preparing and conducting a statistical analysis;
- Your ability to apply knowledge and understanding to preparing and conducting key aspects of a statistical analysis;
- Your ability to communicate your knowledge and understanding and findings from the application of your knowledge and understanding in an acceptable manner.

Where will I create it?

You are free to choose how/where you create your portfolio, creating a free website/blog/wiki .Some possible tools to use:

- WordPress.com (https://wordpress.com/)
- Blogger.com (https://www.blogger.com/)
- Google Sites (https://sites.google.com/)
- PBWorks (https://www.pbworks.com/education.html)).

However, your digital portfolio **must be accessible** for review by the lecturer and external examiners.

What dataset will I use?

- You will use the dataset on student exam performance to illustrate your portfolio. The dataset is available for download from the UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/student+performance) where you will find a description. It is also used in the following paper which also provides a dataset descriptor:

P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7. (https://repositorium.sdum.uminho.pt/bitstream/1822/8024/1/student.pdf)

NOTE: Please ensure that you include this citation in your portfolio.

- An additional dataset will be provided later in the module to support dimension reduction.

- For the purposes of the CA for this module you should consider that this is training data only. The dataset does not contain data from sufficient years to be able to fully fit a model, but it does contain enough for use to build and assess the fit of an initial model.

How many marks can I get?

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

Probability and Statistical Inference

Continuous Assessment

Due Date: Saturday December 12th 2020 @ 23:59

What do I include in the portfolio?

The format and structure you use is largely at your discretion, however you need to include the following:

- A **cover introduction post/page** which includes the following information:
  - Student Number: <<your student number>>
  - Student Name: <<your student name>>
  - Programme Code: <<your programme code>>
  - The version of R used.
  - The R packages needed for your code to execute successfully.

- A **clickable index/table of contents** which facilitates access to all components of your portfolio and also includes any guidance needed to successfully navigate your portfolio and download and run the R code provided.

- You need to include sections to demonstrate the following:

Probability and Statistical Inference 2020/2021

Digital Portfolio Requirements

REQUIREMENT	PREPARE	EXPLORE
Knowledge and understanding	• Populations and samples • Key issues related to using a sample for the purposes of statistical analysis	• Statistical measures • Common issues which could impact a statistical analysis
Ability to apply knowledge and understanding	• Choose an appropriate sample for the purposes of a statistical analysis • Frame a research question so that it is suitable to use a statistical analysis to investigate • State hypotheses related to research questions	• Select/create appropriate measures and justify your choices • Inspect data for common issues which could impact a statistical analysis • Select appropriate tests to test hypotheses • Interpret test findings and draw conclusions
Ability to Communicate	• Appropriate details about a population and sample • A research question and hypotheses	• Your understanding of a dataset using appropriate statistics and visualisations • The implications of your inspection for a proposed statistical analysis

REQUIREMENT	ANALYSE	MODEL
Knowledge and understanding	• A range of statistical tests and their use • The implications of test outcomes for a statistical analysis	• Assumptions of the technique, when/how to use the technique • Common issues and challenges that may occur when using the technique and how to address these • The statistics, visualisations and other information needed to be able to interpret the outcomes of using the technique
Ability to apply knowledge and understanding	• Select appropriate statistical tests to test hypotheses with cognisance of the outcomes of a data inspection • Conduct appropriate statistical tests to test hypotheses • Interpret test findings and draw conclusions for a statistical analysis	• Apply the technique as part of a statistical analysis taking cognisance of the outcomes of data exploration and analysis • Address the common issues and challenges that may occur when using the technique • Interpret the findings to draw conclusions for proposed hypotheses
Ability to Communicate	• Reasons for test selection, interpretation, and conclusions using appropriate statistics and language	• The outcomes of the application of the technique using appropriate statistics, language and visualisation • Your interpretation of the outcomes of applying the technique for proposed hypotheses • The implications of the application of the technique for your research question

Probability and Statistical Inference

Continuous Assessment

Due Date: Saturday December 12th 2020 @ 23:59

REQUIREMENT	REPORT	USING R

Apply knowledge and understanding	• 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
	  ○ You are required to adopt the APA guidelines for reporting statistics and for citation https://apastyle.apa.org/ and adhere 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)
	• Using appropriate statistics, visualisations and language.
		• Create R Code to
		  ○ Inspect your dataset, create any additional measures, generate descriptive statistics and visualisations needed
		  ○ Execute appropriate statistical tests and generate statistics needed to test hypotheses
		  ○ Build models using logistic and linear regression and dimension reduction, generating appropriate statistics and visualisations to interpret outcomes and test assumptions
		  ○ **NOTE**: This R Code should be accessible in a manner so that it is easy to download and run without extensive setup.

NOTE: The organisation/presentation is at your discretion but **you must provide a guide** to demonstrate how/where you have addressed these (this can be the contents/index/introduction page).

NOTE: The requirements are available in a separate document in Brightspace under the Content Section called Continuous Assessment.

Probability and Statistical Inference

Continuous Assessment

Due Date: Saturday December 12th 2020 @ 23:59

What do I submit?

- You are required to submit a URI to your digital portfolio including 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 Digital Portfolio.

BASIC MARKING SCHEME

PREPARE	10
EXPLORE	15
ANALYSE	20
MODEL	30
REPORT	15
USING R	10
	100

* A detailed rubric will be released after week 6.

Probability and Statistical Inference

Continuous Assessment

Due Date: Saturday December 12th 2020 @ 23:59

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.
    a. Email submissions will be ignored.
4. Late Submission:
    a. Extensions due to acceptable personal circumstances must be requested by email in advance of the deadline.
    b. For late submissions without an agreed extension, a penalty of 5% will be applied for every day a submission is late.
    c. 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.