Exam Template: Statistical Inference

23069818

Semester 1 Sep2023 group

# Instructions to students

Save this template as your studentID.Rmd; you will upload this file as part of your submission. Change the author information on line 3 of this file to your **student ID**. Do not change the authorship to your name.

Your should knit this file to a document **Word** format. The Word document is what will be marked!

Any changes that you make to the data (e.g. variable name changes) should be made entirely within R.

The subsubsections labelled **Answer:** indicate where you should put in your written Answers. The template also provides blank code chunks for you to complete your Answers; you may choose to add additional chunks if required.

This is an individual assessment: do not work with any other person during this exam. Text-matching software will be used on all submissions.

# Instructions for submission

You must submit your assignment before the stated deadline by electronic submission through Blackboard.

* It is a good idea to save your work early and frequently to ensure you have no issues with the submission portal. Multiple submissions can be made to the portal, but only the final one will be accepted.
* It is your responsibility to submit the exam in a format stipulated above. Your marks may be affected if your tutor cannot open or properly view your submission.
* Do not leave submission to the very last minute. Always allow time in case of technical issues.
* The date and time of your submission is taken from the Blackboard server and is recorded when your submission is complete, not when you click Submit.
* It is essential that you check that you have submitted the correct file(s), and that each complete file was received. Submission receipts are accessed from the Coursework tab.

There is no late submission permitted on this timed assessment. Ensure that you submit your submission in good time. Neither the module leader nor module team can accept late assessments, do not ask them to do so.

# Background to the research

The head of school for a four year degree course has provided you with some data based on student demographics, marks and graduate outcomes.

They have asked you if the data could reveal findings that may be relevant for monitoring student performance and outcomes.

# Data instructions

Your individual data set is accessed via Blackboard >>> Assessments >>> Dewis Data For Exam.

You must only analyse the specified data. No other data is to be used for this assessment.

All data manipulation and analyses must be done within R.

# Data structure

The variables collected for each student are:

studentID – a unique student identifier issued to each student at the start of the course

outcome – employment status one year after finishing the course (E1 = employed in a graduate role, E2 =Employed in a non-graduate role, Education = in full time further education, Unemployed = not yet employed)

age – age at start of course

gender – gender at start of course

language – score given for student level of English proficiency determined as part of the application process for the course (minimum 0, maximum 10)

feedback – score given by student for their satisfaction of the course when asked at the end of Year 4 (minimum 0, maximum 10)

Mark1 - Mark for Year 1 (out of 100)

Mark2 - Mark for Year 2 (out of 100)

Mark3 - Mark for Year 3 (out of 100)

Mark4 – Mark for Year 4 (out of 100)

# QUESTIONS START HERE

# Question 1: Data Preparation

1. Ensure you have prepared your knitted Word document as per Instructions to Students
2. You should load the data in R, describe and perform any actions with respect to:

-any manipulation of the data structure

-missing values

-ensuring data is valid

**(10 marks)**

### Answer:

# load the dataset here

# further data preparation here

# Question 2

A colleague suggests the following research question,

“do students perform differently in their final year relative to their performance at the start?”

To assess this research question:

* create a new variable for the difference between Year 4 mark and Year 1 Mark.
* show and interpret a confidence interval for the mean difference, in context of the research question.

**(12 marks)**

### Answer:

# Question 3

Another research question is suggested,

“is there a relationship between student marks across each of the years?”

* Assess this research question by showing and interpreting the linear correlations between the marks for each of the four years.

Marks are awarded for well-designed output, and the interpretation of the output.

**(12 marks)**

### Answer:

# Question 4

A further research question states,

“can the final year mark be predicted based on one mark for a previous year?”

Produce simple linear regression with Year 4 mark as the dependent variable, and only one independent variable.

Your answer should include:

* justification for the choice of explanatory variable, including any additional supporting exploratory data analyses used to make the choice;
* interpretation of the slope (gradient) coefficient;
* comment on the r-square value, and the validity of model assumptions.

**(23 marks)**

### Answer:

# Question 5: Report

Clearly state one **alternative new** research question based on the full original data set supplied to you. Explain why this is a worthwhile research question to consider.

You are required to write a short report for the client showing some analyses based only on the research question you have selected.

In your report you may wish to include a number of the following: exploratory data analyses; a hypothesis test; data modelling; discussion of limitations; how you could extend the research if given more time.

To clarify, your answer to this question must be a report based on **your analyses of your own research question** arising from the data, which is not addressed in the questions above. This report should contain a maximum 5 outputs (i.e. graphics + tables) and a maximum of 500 words.

**(35 marks)**

### Answer:

# Question 6

Describe how you have applied principles of reproducible research in this submission (maximum 100 words).

Marks are awarded for identification of appropriate reproducible research principles, only if also evidenced throughout your submission that they have been applied.

**(8 marks)**

### Answer:

# End matter - Session Information

Do not edit this part. Make sure that you compile your document so that the information about your session (including software / package versions) is included in your submission.

sessionInfo()

R version 4.3.2 (2023-10-31)  
Platform: aarch64-apple-darwin20 (64-bit)  
Running under: macOS Sonoma 14.2.1  
  
Matrix products: default  
BLAS: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib   
LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib; LAPACK version 3.11.0  
  
locale:  
[1] en\_US.UTF-8/en\_US.UTF-8/en\_US.UTF-8/C/en\_US.UTF-8/en\_US.UTF-8  
  
time zone: Asia/Dhaka  
tzcode source: internal  
  
attached base packages:  
[1] stats graphics grDevices utils datasets methods base   
  
loaded via a namespace (and not attached):  
 [1] compiler\_4.3.2 fastmap\_1.1.1 cli\_3.6.2 tools\_4.3.2   
 [5] htmltools\_0.5.7 yaml\_2.3.8 rmarkdown\_2.25 knitr\_1.45   
 [9] xfun\_0.41 digest\_0.6.34 rlang\_1.1.3 evaluate\_0.23