

Unit: Advanced Mathematics and Statistics for Data Science and AI
Assessment Element 1: In-Class Assignment

Housekeeping Information

Assessments: There are two Elements where Element 1 is In Class Assignments and Element 2 is Practical Exam

Element 1: In Class Assignments

A collection of assignments applying concepts and skills from the unit, as specified in the Assessment Brief (50%).

Element 2:

Practical Exam You will be individually asked to generate a report integrating text, code and visualisations (e.g., in the form of a Jupyter notebook) detailing statistical insights given a previously unseen dataset (50%).

Submission Date:

Element 1: In Class Assignments:

By 6pm (1800) GMT on Monday 4th December 2023

Element 2: Practical Exam:

By 6pm (1800) GMT on Friday 8th December 2023. Open 7 days prior to deadline; open from Friday 1st December 2023.

Submission Method:

Element 1: In Class Assignments:

A Zip folder (100MB max) containing:

1. a *link to your GIT code repository*,
2. a link to a video (with voice over/detailed captions) demonstration of your work and
3. *Readme* file to contextualise the overall design and development that includes a 300-word mini project description written in your style highlighting on
 - a. What is the task about?
 - i. Understanding of the bigger picture and each sub task,
 - b. What Maths and Statistics involved?
 - i. Identifying and applying the appropriate concepts learnt
 - c. How did you implement?
 - i. Possible libraries, packages involved.
 - ii. Logic used and why?
 - iii. Cite the references used
 - d. Outcomes of each sub task
 - i. Interpret and provide detailed explanation about the arrived results at each step
 - e. Challenges and How you resolved?
 - i. Challenges if any and how it was overcome
 - f. References
 - i. Acknowledge by listing all of the resources that helped to complete the task
4. The GIT repository must include
 - a. the programming tasks,

- b. dataset used,
- c. a link to a video demonstration of your work, and
- d. Readme file to contextualise the design and development.

Element 2: Practical Exam: Zip folder via Moodle:

A Zip folder (100MB max) containing your portfolio of works, along with a link to your implementation notebook with necessary comments, exported as PDF and submitted in the Zip.

Element 1: In Class Assignments

Please find below the task for Element 1 In-Class Assignment 1. This assessment will have a series of smaller sub-tasks to be completed.

Dataset: [Click here for the dataset for this Task](#)

Plagiarism: [Avoid Plagiarism Essentials](#)

Preferred Language: Python

Platform/IDE: Any

Overall Task brief:

For this In-Class Assignment, you will be using the dataset provided above and you are to complete each item below:

Bigger Picture of the Task: We wanted to study if users with high mobile usage find a given visual search target faster than their counterparts and also to assess if there is any significant interaction with other IVs (eg: Gender, Age etc.,)?

1. Check Data quality:
 - a. Create a Data Pre-processing pipeline with only the necessary tasks such a way that you confirm the current dataset is all set for next steps.
 - b. Record the data shape at this stage
Note: This to be supported with necessary comments and justifications throughout the process.
2. Data Relationship/Distribution:
 - a. Provide a **Frequency table** and an appropriate plot to visualize Pickup counts split by gender.
 - b. Provide a Frequency table and at least 2 different types of plots to visualize the distribution of Daily average minutes.
 - c. Provide at least 2 appropriate plots (for each) to check the relationship between the
 - i. Participant's age and their Response time on singleton visual search
 - ii. Participant's gender and their Response time on conjunction visual search
3. Correlation Check: Produce a **bivariate correlation table** between Age, STAI, BRIEF_Total, DailyAvgMins and VS_RT_correct_Single.
4. Linear Regression: Perform a **linear regression** to see if DailyAvgMins predicts VS_RT_correct_Single
5. Multiple Regression: Add **predictors** Age, GenderNum, STAI, BRIEF_Total and DailyAvgPickups to the multiple regression model. Does the amount of variance

accounted for in the outcome increase? Is DailyAvgMins a significant predictor of the outcome?

6. Scenario 1:

- a. Participants including male, female and non-binary from various age groups took part in a study which ran for a week.
- b. They were grouped under three Age Groups a ≤ 30 years, b: 31 to 49 years, and c ≥ 50 years
- c. And then their mobile usage was monitored, and the daily average minutes of usage was recorded.
- d. And the usage (average minutes per day) was grouped under three Usage Groups a ≤ 300 minutes, b: 301 to 899 minutes, or c ≥ 900 minutes
- e. Then they were asked to locate target red apple amongst distractors of only blue apples (known as Singleton search) and Mean time to identify was recorded (reaction time; RT)

Perform the following based on the Scenario 1:

- f. Create any groups described above and choose appropriate Omnibus test statistic for Scenario 1 to test the hypothesis stated below:
 - i. Hypothesis: The researcher hypothesised that as participant mobile phone use increased, the time taken for them to find the search target would decrease in singleton search, with the greatest decrease in participants less than 30 years-of-age.
 - ii. Justify reasons for choosing the test.
 - iii. List the details of assumptions and the corresponding statistic test deployed to achieve it (in a markdown cell) (Example assumptions (but not limited to) may be Normality, Variance, Linearity etc.,)
 - iv. Check validity of those assumptions and provide appropriate visual charts.
Note: If there are deviations, apply methods of rectification before the test is applied on the dataset
 - v. Use appropriate follow-on tests to identify where the specific effects lie (e.g. between which IV levels or combination of levels).

7. Scenario 2:

Instruction: For Scenario 2 we are going to transform the existing data to fake a different set of empirical data. Sort the existing dataset by VS_RT_correct_Single (Descending order) and rename “VS_RT_correct_Conjunction” → “VS_RT_Before” and “VS_RT_correct_Single” → “VS_RT_After”

- a. Participants including male, female and non-binary (i.e. Gender) took part in a study which ran a week span.
- b. Then they were asked to locate target red apple amongst distractors of both blue apples and red slices and Mean time to identify was recorded (reaction time; VS_RT_Before).
- c. Create a new variable “Brain Training” and randomly allocate half the participants to a Trained and the other half to a non-trained group.
- d. Their mobile usage was monitored, and the daily average minutes of usage was recorded. And the usage (average minutes per day) was grouped under three

UsageGroup categories a =< 300 minutes, b= b: 301 to 899 minutes, or c >= 900 minutes

- e. They were finally asked to locate target red apple amongst distractors of only blue apples (Singleton Search) and Mean time to identify was recorded (reaction time; `VS_RT_After`).

Perform the following based on the Scenario 2:

- f. Create any groups described above and choose appropriate **Omnibus test statistic** for Scenario 2 to test the hypothesis stated below:
 - i. Hypothesis: The researcher hypothesised that as participant mobile phone use increased, the time taken for them to find the search target would decrease, with the greatest decrease in female participants after playing brain games.
 - ii. Justify reasons for choosing the test.
 - iii. And WRT to the test you have finalized list the details of assumptions and the corresponding statistic test deployed to achieve it (in a markdown cell) (Example assumptions (but not limited to) may be Normality, Variance, Linearity etc.,)
 - iv. Check validity of those assumptions and provide appropriate visual charts.
Note: If there are deviations, apply methods of rectification before the test is applied on the dataset
 - v. Use appropriate follow-on tests to identify where the specific effects lie (e.g. between which IV levels or combination of levels).

Caution Note: Should provide sufficient comments throughout the Assignment. Should credit every reference used. Having said that too much of external code use may lead to loss of grade, care to be taken to show case originality of work. You are allowed to develop the necessary logics to restructure the dataset to complete task without distorting the meaning of the dataset (eg: Adding necessary columns, Transposing, Grouping etc.,). If this step is carried out upload this dataset to the git else the original would do. But should be supported with appropriate comments and justification, else it will be considered deviation from the expected task leading to loss of grade.