

Unit: Advanced Mathematics and Statistics for Data Science and AI
Assessment Element2: Practical Examination

Housekeeping Information

Before you start, I would recommend you please read the whole document once including the caution note, before you dive in straight.

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 2: Practical Examination

Please find below the task for **Element 2 Practical Examination**

Plagiarism: [Avoid Plagiarism Essentials](#)

Preferred Language: Python

Platform/IDE: Any

- 1) With a simple python code to demonstrate the following using this simple dataset:

Memory (Negative Memory Bias)	Anxiety	Depression	Self-Esteem
5	20	0	16
5	21	0	15
6	24	0	19
6	32	1	18
7	32	1	17
7	21	1	18
7	45	3	16
8	45	3	10
9	31	5	15
9	22	8	15
10	44	6	14
4	45	6	13
11	46	7	15
13	49	8	12
13	45	10	10
13	60	16	9
16	56	20	2
16	57	13	2
19	48	28	12
20	41	24	1

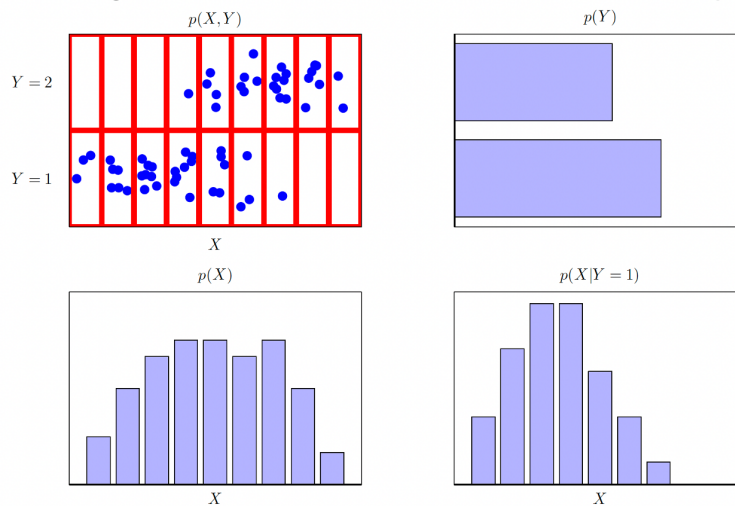
- (a) Perform a series of correlations on the above (fictitious) data.
- (b) Demonstrate through multiple regression to examine the contribution of each independent variable to the prediction of Memory Bias. Also report how much of the variance is accounted for by the regression equation?

- (c) Based on the above analysis, what would be the predicted value of Memory Bias for a person with an Anxiety score of 44, a Depression score of 13 and a Self-Esteem score of 12?
 - (d) Using the same data, perform a multiple regression to determine the best predictor of Memory Bias.
 - (e) Using the same data, perform a multiple regression to test the idea that Anxiety is the salient predictor of Memory Bias. Enter Anxiety on the first step, and Depression and Self-Esteem on the second.
- 2) With a simple python code to demonstrate the following using this simple dataset:

This is a 2x3 between-subjects design with 2 levels of therapy (new vs. old) and 3 levels duration (short- vs. medium- vs. long-term improvement). The dependent variable is an improvement index where the higher the score, the greater the improvement.

		Therapy (A)	
		New (A ₁)	Old (A ₂)
Duration (B)	Short-term (B ₁)	6	10
		9	11
		8	5
		5	9
	Medium-term (B ₂)	15	9
		18	12
		16	13
		20	9
	Long-term (B ₃)	10	8
		12	7
		9	9
		13	12

- (a) Generate a table of means and SDs.
 - (b) Perform an ANOVA using General Linear Model, Univariate, and report the significant effects.
 - (c) Plot the interaction in two ways:
 - i. Duration x Therapy
 - ii. Therapy x Duration
 - (d) Perform the simple effects analyses of the following and comment on the significance of the following simple effects:
 - i. Duration at New Therapy (B at A₁)
 - ii. Therapy at Mid-term (A at B₂)
 - iii. Therapy at Long-term (A at B₃)
 - (e) Provide the conclusions drawn from the simple effects analyses?
- 3) Create a Python code to generate similar output:



*This is a finite sample of $N = 60$ data points drawn from the joint distribution (**top left**)*

*In the **top right** is a histogram of the fractions of data points having each of the two values of Y .*

- 4) For this task you need to choose any two appropriate datasets from the Dataset folder available in Moodle (In Class Assignment Element 1 datasets), Write Python code to perform the following (on both the chosen datasets) :
 - (a) Generate classifier objects (with default hyperparameters) using the following:
 - a. LogisticRegression
 - b. LinearSVC
 - c. SVC
 - d. KNeighborsClassifier
 - e. Bayesian Logistic Regression
 - (b) Fit each of the classifiers on the respective data
 - (c) Provide plots to demonstrate the decision boundaries
 - (d) Comment on each of their performance on the given dataset

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.