



Faculty of Information Technology
FIT4005/5125/5143 Research Methods in IT
Semester 2, 2020

Assignment 3: Weeks 7, 8 and 9 post-workshop exercises

Submission: Post-workshop Exercises for Weeks 7,8 and 9

Value: This assignment is worth 20% of the total marks for FIT4005/FIT5125/FIT5143

Assignment due date: The submission is due: 11:55pm Thursday, Week 10 (22nd of October 2020).

Submission method: Submit to Moodle PDF(s) with your answers to post-workshop exercises for Week 7 (Statistics, see Page 3), Week 8 (Experiments, see Page 4) and Week 9 (Correlation, see Page 5).

This assignment comprises the post-workshop exercises for Weeks 2 to 3.

Weighting of Assessment: 20% of marks for units FIT4005/5125/5143

Weighting of Assessment: 20% of marks for units FIT4005/5125/5143		
Week 7 Statistics Weighting: [20 marks] See page 3.	Week 8 Experiments Weighting: [20 marks] See page 4.	Week 9 Correlation & Regression Weighting: [20 marks] See page 5.

This is an individual assignment; it must be your own work and expressed in your own words.

Assessment criteria:

- (i) The following criteria will be used to determine the score for the assignment:
 - Content and completeness of tasks
 - Clarity and relevance of content
 - Level of critical analysis
 - Logical structure and organization of ideas
 - Use of references (where appropriate)
 - Format, grammar, spelling etc.
- (ii) Note that plagiarism detection procedures may be applied to each submission. See the University rules and regulations regarding plagiarism and resulting penalties. Any case of plagiarism detected will mean automatic failure of the entire assignment.
- (iii) Late submissions will incur a penalty of 5% per day.

Assessed Post-workshop Exercise (Week 7)

1. The aim of this assessment is to evaluate your understanding of basic statistics, and to assess your ability to correctly apply basic statistic computations.
2. This exercise must be completed by the Week 10 submission deadline for Assignment 3 (see Moodle).
3. Submission requirements for Assignment 3 can be found on Moodle and will be a combination of the assessed post-workshop exercise for weeks 7, 8 and 9.
4. This is an individual exercise that forms part of the assessment for the unit, you must therefore work alone and follow Monash University's policies, procedures and regulations relating academic integrity, plagiarism and collusion (see Moodle).
5. Tutor feedback will not be provided for this exercise, but feedback on other workshop activities should be of help to you in completing this exercise.

Assessment (Week 7)

For this assessment you will be using data captured during the Week 6 workshops. As you are aware, all students participated in an online survey, resulting in a large amount of both qualitative and quantitative data. We have selected a sub-set of this data for you to download from the Moodle as an anonymized CSV.

1. Select a single metric from the data (e.g. age)
2. Design a MATLAB script to compute the population standard deviation and the Interquartile Range of this metric. Annotate each line of your script to describe what it does.
3. Imagine that this data is a sample of a wider population (e.g. all students at Monash). Create a script to compute the sample standard deviation. Annotate each line of your script to describe what it does.
4. Choose any data column from the responses (with number values), and use MATLAB to produce a boxplot of this data, you must ensure you label this data appropriately. Annotate your boxplot using MATLAB to indicate what each element means. Annotate each line of your script to describe what it does.

Make sure your submission clearly shows how you got to your results, include:

- a screenshot of each MATLAB script that you created (including annotations)
- an image of the resulting graph including annotations (made using MATLAB).

Assessed Post-workshop Exercise (Week 8)

1. The aim of this assessment is to evaluate your understanding of basic statistics, and to assess your ability to correctly apply basic statistic computations.
2. This exercise must be completed by the Week 10 submission deadline for Assignment 3 (see Moodle).
3. Submission requirements for Assignment 3 can be found on Moodle and will be a combination of the assessed post-workshop exercise for weeks 7, 8 and 9.
4. This is an individual exercise that forms part of the assessment for the unit, you must therefore work alone and follow Monash University's policies, procedures and regulations relating academic integrity, plagiarism and collusion (see Moodle).
5. Tutor feedback will not be provided for this exercise, but feedback on other workshop activities should be of help to you in completing this exercise.

Assessment (Week 8)

This task concerns the design of experiments. In this exercise you will formulate a hypothesis, prepare a plan of your study (including statistical testing) and justify it, including the potential limitations of it.

Consider the topic of the survey that you participated in during the unit, and used for the Week 7 assessment task. Imagine you are asked to develop this research area further.

1. Propose a hypothesis. It should be something you can realistically test using one or more of the statistical tests covered in this course. It can concern any topic or natural phenomena which relates in some way to the survey topic (for example, it might concern user performance on a system in a given scenario and setting) (max. 50 words).
2. Write down the null hypothesis (max. 50 words).
3. Write down the independent and dependent variables as well as at least three confounding variables (max. 50 words).
4. Imagine you had a budget of up to 1000AUD (in addition to up to 100 hours of your time to conduct the study). Explain what data you will collect to investigate this hypothesis and how you would obtain the data in a practical fashion (max. 100 words).
5. What statistical tests do you expect that you may have to conduct. Please explain the circumstances in which you would conduct each test (max 150 words).
6. What are the limitations of your study? Write a paragraph that explains these limitations as well as potential future investigations you might conduct (max 200 words).

Assessed Post-workshop Exercise (Week 9)

1. The aim of this assessment is to evaluate your understanding of basic statistics, and to assess your ability to correctly apply basic statistic computations.
2. This exercise must be completed by the Week 10 submission deadline for Assignment 3 (see Moodle).
3. Submission requirements for Assignment 3 can be found on Moodle and will be a combination of the assessed post-workshop exercise for weeks 7, 8 and 9.
4. This is an individual exercise that forms part of the assessment for the unit, you must therefore work alone and follow Monash University's policies, procedures and regulations relating academic integrity, plagiarism and collusion (see Moodle).
5. Tutor feedback will not be provided for this exercise, but feedback on other workshop activities should be of help to you in completing this exercise.

Assessment (Week 9)

This assessment concerns the evaluation of a machine learning system.

The local Council are considering purchasing a waste sorting system, for the purposes of recycling. This takes a photograph of each object in turn, before applying a computer vision algorithm to sort each object into the relevant category. The machine is capable of sorting into the following categories: *plastic bottles* (B), *tin cans* (T), *glass jars* (J) and *everything else* (X).

To measure the performance of the waste sorting system, the council tested the system with a **sample** of 30 items, which were in turn sorted by the machine and given a Label.

Ground Truth	B	B	B	T	T	J	T	X	X	T	T	B	J	T	J
Label	B	X	B	B	T	J	T	X	T	X	T	B	X	J	J

Ground Truth	X	T	X	T	X	B	B	B	J	B	T	B	B	B	B
Label	X	T	X	T	X	X	B	B	J	B	X	B	T	X	B

By hand and showing your working:

- (a) Compute the **Confusion Matrix** for the Machine, based upon the sample above
- (b) Using the **Confusion Matrix**, or otherwise, compute the **Accuracy** of the waste sorting system, based on the sample above.
- (c) The Council is considering using the **Accuracy** figure based on the **sample** as a means for deciding if it should purchase the waste sorting system. Explain why this might not be appropriate (max 200 words).

(d) The Council considers that they will use the system purely as a machine for identifying and extracting plastic bottles (and which will reject everything else). On that basis, compute the **Precision**, **Recall** and **F1-Score** using the **Confusion Matrix**, or otherwise.

Assignment criteria	N - three or more of the following:	P - mostly the following:	C - mostly the following:	D – mostly the following:	HD – mainly D plus two or more areas at higher standard:
Statistics					
Calculations	No calculations or annotations present.	Calculations need further improvement. Annotations are not clear.	Calculations are partially coherent, elements are missing. Annotations are somewhat clear.	Calculations are clear, but some elements are missing. Annotations are clear.	Calculations are clear and easily understood. Annotations are clear.
MATLAB	No commands present by MATLAB	very few commands shown, appear not to generate the correct features.	Commands are not clear, there are missing elements.	Commands are clear, some further work is needed to be in the right order.	Commands used appear legitimate, in the right order
Experiments					
Hypothesis and Statistical test	Unclear hypothesis, or not framed as such. No statistical tests are identified.	Hypothesis does not clearly relate back to the scenario. Some elements are unclear or not well defined. Statistical test is not clearly outlined with the experiment.	Hypothesis relates to the scenario. Some elements are unclear or not well defined. Limited investigation on statistical test.	Good outline of hypothesis quality shows that a measurable phenomenon can be tested. Good outline of statistical test and description to why it should be applied.	Excellent hypothesis quality and, clearly bounded. Identifies a measurable phenomenon. Excellent outline of statistical test and rich detail relating to why it should be applied.
Variables and data collection	No phenomena or variables identified. Data collection not clear	Further work is needed on variables outlined. Data collection is not connected to the test.	Some valid variables are described, others may be unclear Data collection lacks clarity.	Good outline of variables. Data collection is clear but further details can be included.	Excellent outline of variables. Data collection is relevant to the hypothesis
Limitations and future investigations	Limitations not present. Future investigations not present.	Limitations are unclear and do not relate to the study. Future investigation also needs further reasoning.	Some limitations are discussed but lacking detail. Some mention of further investigation but not clear.	Good outline of limitations for the study. Brief outline of future research.	Excellent outline of limitations for the study. Clear direction for future research.
Correlation					
Working out, correctness and justification.	Incorrect answer and no working out. Justification not present.	Working out is not clear. Aspects of the formula is missing. Limited justification outlined.	Working out needs further detail, formulas present. Limited justification outlined.	Working out is partially coherent, good use of formulas. Good justification outlined.	Working out is excellent, easily understood, formulas are clear. Excellent justification outlined.
Overall					
Quality of writing	Major revision is required, and spelling, grammar and writing must be revise. Difficult to understand sentences.	Significant spelling errors. Most spelling, punctuation, and grammar correct allowing reader to progress though work	Few spelling errors present. Punctuation, grammatical errors, and few fragments.	Quality of writing is at high standard; no distracting spelling errors.	Excellent standard and quality of writing. Quality of the writing is outstanding and engaging. No spelling errors or distractions present.