

CA1__2023

August 23, 2023

1 Data Analysis - Coursework 1 (35%)

Short Style Data Analysis Questions Deadline Thursday week 6, 2pm.

INSTRUCTIONS

This coursework assesses learning outcomes from Chapters 1 - 4 of the course.

These assessments are equivalent to an exam: - Submit your work via Turn-It-In on Learning Central. **Note that you will need to upload your final notebook exported as a pdf.** Don't forget to execute all of your cells before you export the notebook to pdf. You can constantly resubmit your turnitin document until the deadline. - Don't worry about how your code looks - marks are not given for pretty code, but rather for the approach used in solving the problem, your reasoning, explanation and answer. - Please also take note of the University's policy on **plagiarism**, which is outlined in your student handbook.

AIMS OF THE ASSESSMENT

The assessment will evidence your ability to synthesise the tools provided during weeks 1-5 and how you communicate it to the audience.

1. Ability to identify the correct statistical method with sound reasoning.
2. Evidence of knowledge of the assumptions made or inferred from the question.
3. Application of the statistical method, including correct treatment of the given / inferred data.
4. Critical explanations and interpretation of the results, ability to turn raw data into a statistically correct statement

Specific requirements and guidance

- For short-form questions (for example question 1 on CA 1), marks are given in steps for the correct answer and reasoning.
- In longer form questions there are a small proportion of marks allocated for additional work beyond the basic principles outlined in the question. This is intended to recognise student work that goes beyond the basic investigation and shows evidence of insight and use of appropriate tools.

Major lapses

- Full marks cannot be awarded if there is no explanation or justification of the method/approach/equations/statistical tool used.

- If the question is answered correctly but there is no explanation of the reasoning being used to answer the question, the maximum mark that can be awarded for that (sub)question is 60%.
 - If the question is correct with explanations and steps explained but there are no details of the code to get that answer, the maximum mark that can be awarded for that (sub)question is 50%.
-

1.1 QUESTION 1

A student is taking three modules, and the probability that they pass any individual module depends on the fraction of the weekly live sessions they attend f so that the probability of passing the module is $0.8f$.

- a.) If the student attends all the weekly live sessions, ($f = 1$), calculate the probability that they will
- (i) pass all three modules,
 - (ii) fail one module and pass the other two
 - (iii) pass only one module,
 - (iv) pass no modules.
- b.) Show that the sum of these four probabilities is 1.
- c.) Calculate the four probabilities on the assumption that the student attends only half of the weekly live sessions.
- d.) What fraction of weekly live sessions must the student attend to have a 50% chance of passing all three modules?
- e.) Is this a realistic way to model the probability that the student will pass the modules?

[10 marks]

1.2 QUESTION 2

A group researching cancer have previously found that the genetic marker D3 is a useful indication that a person will develop the more aggressive form of melanoma skin cancer, in that D3 is present in 65% of the aggressive cases. However the test is expensive. A rival group claim that the marker M23 is more sensitive than D3, and works out considerably cheaper to test for. The rival research team manage to get DNA samples from 7 patients with the aggressive form of the disease, all of whom test positive for the genetic marker M23. Based on these results, is M23 a better marker for the disease than D3?

Give full mathematical working for your reasoning, and show labeled plots of the underlying functions.

[20 marks]

Answer:

1.3 QUESTION 3

A computer chip manufacturer suspects that roughly half of its latest batch of CPUs contains a flaw. The accounts department are clearly concerned, and are trying to predict how the fault will affect the number of customers returning products. How many CPUs from the batch would they need to examine to know the probability that any given CPU is faulty to better than 2.5%?

Tip: think carefully about what you are trying to estimate here, you want the *error in your success probability* to be less than 2.5%.

[15 marks]

1.4 QUESTION 4

The state of Florida is thinking of relaxing its policy on alcohol sales, to allow supermarkets to sell hard alcohol, since the police predict that this can reduce violence. After some extensive polling, they find that only 35% and 10% of Republican and Independent voters are, respectively, behind the change in the law, while 80% of the Democrat voters are in favour. You are visiting the state, and ask a Police Officer what she thinks of the idea. They says they're against the change to the law. What is the probability that they votes Democrat?

You may assume that voting in the Florida polls that year was split in the following way: 40% Democrat and 36% Republican. You can also assume that Independent covers everything that is not Republican or Democrat.

[25 marks]

1.5 QUESTION 5

Ten new recruits for a basketball team are timed (in secs) in running the 100 meters and 1,500m races to determine how fast they can run. The following results were obtained,

Distance	1	2	3	4	5	6	7	8	9	10
100m:	11	12	12	13	13	15	11	16	11	12
1500m:	270	300	230	260	270	230	260	240	270	260

What trends do we see in the data above? Are they significant? Use appropriate tests to answer this question.

Please **create your own** statistical functions when answering this question. Please include any sources you have used to answer this question.

[30 marks]