

Assessment 2: Statistical Analysis

Hint: It is very important to comment your code. Some good rules of thumb to follow when commenting your code, is to provide sufficient details such that a) you will understand your coding should you read in 5 years from now, and b) someone foreign could easily follow the logic in your coding. Remember, for the moment, it is just you and the marker reading this code, but in the real world, R code maybe shared by multiple people. Please visit these websites to learn more about effective techniques to comment your code:

<https://swcarpentry.github.io/r-novice-inflammation/06-best-practices-R/>
<https://google.github.io/styleguide/Rguide.xml>

Filename convention:

Does your name appear in the file name? 1 Mark will be deducted from the overall assignment score, if your name is omitted from the file name. For example, if your name is Mary Smith, and you submitted this file as pdf, then you must call your file
MarySmithMA5820Report.pdf

If Mary uses word, then the pdf extension will be replaced with the docx extension. The most important thing is that your name appears in the final name.

Please put your R-code and output in a place that is obvious for the markers in the word (or pdf) file. Please also, submit your code as a dot R file which also has your name in the file name. One R file should be sufficient for all three questions. Please comment your code appropriately.

Question 1:

The lifetime of a particular type of TV follows a normal distribution with $\mu = 4800$ hours, and $\sigma = 400$ hours.

- Find the probability that a single randomly-chosen TV will last less than 4500 hours. Use R to assist with your computations.
- Find the probability that the mean lifetime of a random sample of 16 TVs is less than 4500 hours. Use R to assist with your computations.
- Compare answers from (a) and (b)

Marking Scheme for question 1

Criteria (each worth the same percentage)	0 Marks	1 Mark	2 Marks
1. Knowledge of statistical principles, methods and techniques.	Major errors in application AND/OR Comprehension AND/OR Interpretation AND/OR Errors are major.	Minor errors in application AND/OR Comprehension AND/OR Interpretation AND/OR Errors are minor	Exemplary application of principles, methods and techniques AND demonstrates full comprehension of associated statistical theory, assumptions and limiting factors. AND Student clearly interpreted solutions in a highly articulate statistical and English language
2. Effectively integrate and execute statistical theories and processes in RStudio or RMarkdown.	The strategy in R was difficult to follow AND/OR ineffective AND/OR Failed to demonstrate reasonable mastery of the statistical package.	The strategy in R was recognisable and mostly effective AND Demonstrates reasonable mastery of the statistical package.	The strategy in R follows directly from the theory AND Demonstrates full mastery of the statistical package.
3. Displaying and Summarising Data to support analyses.	Inappropriate choice(s).	Some improvements would enhance the investigation.	Appropriate choice of graphs AND That are well presented with easy to read axes and labels AND Figure number and captions were included and referred to in text.

Question 2:

Dataset: `beta.csv`.

Beta endorphins are morphine like substances produced by the body. They create a sense of well-being. It has been proposed that Beta endorphins increase with exercise. Test this hypothesis using the data in `beta.csv` which has Beta endorphin levels for 10 people measured for each person pre and post exercise. Using this sample, test if Beta endorphins increase with exercise. Adopt a 5% risk of committing a type I error.

1. Enter the data into R
2. Perform some exploratory data analysis procedures
3. Perform an appropriate significance test
4. State any assumptions needed to support the validity of the procedure and where possible comment on the adequacy of these assumptions.
5. If you were conducting this experiment, what would you try to do to minimise confounding?

Hint: use R to assist your calculations.

Marking scheme for question 2 – is the same as the marking scheme for question 1.

Question 3

Dataset: `PlantGrowth`

The dataset `PlantGrowth` is a datafile in the R datasets. If you type `'PlantGrowth'` from the R prompt you will see the data. For more information about this data type `help(PlantGrowth)` from the R prompt. These data are obtained from an experiment to compare plant yields under a control and two other treatments. Test if there is a difference between the control group and treatment 2 on mean plant yields.

1. Check that you have access to the data
2. Perform some exploratory data analysis procedures
3. Perform an appropriate significance test
4. State any assumptions needed to support the validity of the procedure and where possible comment on the adequacy of these assumptions.

Marking scheme for question 2 – is the same as the marking scheme for question 1.