

BIOE70037 - Computational and Statistical Methods for Research 2024-2025

Jupyter Notebook Assignment

This is the last assessable component of the course. It will be marked out of 50 marks as detailed below.

Dataset for you

There are 25 distinct datasets in the folder and the data set you will work on corresponds to the remainder of the division of your CID number by 25. For example, if you CID number is 12345678, then your dataset is **3** (because $12345678 = 493827 \times 25 + 3$).

Data visualization (15 marks)

In this section students are expected to produce a figure (or figures) that fully represent the data they are given. This plot (or plots) should clearly show the shape of the data and the relationship between the independent and dependent data. In addition to the figure/s, students must write a paragraph describing the data and explaining their choice of display figures. This paragraph should clearly highlight any interesting properties of the data such as shape, trend, etc. (10 marks for figures, 5 marks for explanatory paragraph)

Statistical summary (20 marks)

In this section the students must use both descriptive statistics and inferential statistics to produce a figure (or figures) clearly representing the comparison of data and any statistical significance between groups. This plot (or plots) should clearly show appropriate descriptive statistics for the groups as well as any significant differences demonstrated by inferential analysis. In addition to the figure/s, students must write a paragraph summarizing how the data looks from a statistical standpoint as well as providing a clear justification of any statistics used (i.e. ANOVA vs. independent/dependent t-test, mean vs. median, standard deviation vs. standard error). (10 marks for figures, 10 marks for paragraph)

Discussion of results (15 marks)

In this section students must write a discussion (500 words maximum) summarizing the findings of the data analysis as a whole. The need to clearly describe the relationships that can be inferred from the statistical analysis and the importance of these findings relative to the initial aim of the study (the aim or objective will be noted in the specific data file allotted to each student). The students must also comment on the experimental design challenges or issues they can identify from the data analysis, and how to improve the study/analysis of the study. Specific statements on study design related to aims, powering and the level of significance should be included. Finally, specify an experimental approach including variables, sample numbers and powering that would enable a superior study to address your stated aim. (5 marks for summary, 5 marks for comment on study design and 5 marks for proposed new design)

Submission will be via Blackboard and the Jupyter Notebook file will be directly uploaded. The assignment is due on Friday, 13th December 2023, by 3pm.