Course outline weeks 7–11
4th May–5th June

BIOSCI220

Students are expected to be able to

- Identify the following types of variables in the dataset
  - discrete
  - continuous
- Write their own R code to summarise each variable using an appropriate plot. Specifically be able to produce the following plots
  - boxplot
  - scatter plot
  - histogram
  - barplot
- Represent their data as a clean data frame in R and manipulate it accordingly. Use of [& : etc.
- Communicate a summary of the dataset accurately and concisely
- Communicate any limitations relevant to the dataset

Week 7 (beginning 4th May): Experimental design

Learning objectives

By the end of this lab student should be able to

- Identify the following experimental variables
  - independent variable
  - dependent variable
- List and describe Fisher's three principals of experimental design
  - Randomization
  - Replication
  - Blocking
- Identify the design concepts used in given case studies
- Discuss the advantages and disadvantages of different designs
- Critique experimental designs

Mini lectures

- Randomization
- Replication
- Local Control (blocking etc.)
- Lab tutorial explained (design your own experiment, peer review)

Assessed quiz on CANVAS worth 1% of your final grade

Lab—worth 6% of your final grade

During this exercise you will design your own experiment (you are not expected to carry this out!).

- concisely summarize what the question is you have you wish to answer during your experiment,
- summarize your experimental design, paying particular attention to Fisher's 3 principles of experimental design.

You will also be expected to peer review 3 other worksheets; this also means that your work will be peer marked by 3 of your classmates. The worksheets you will review will be automatically assigned to you after the assignment due date. When carrying out this peer review please follow the rubric carefully and be mindful that your comments, although anonymous, will be passed on to your peer.

Week 8 (beginning 11th May): Visualising and analysing multivariate data

Learning objectives

By the end of this lab student should be able to

- Discuss the aims and motivations of Multidimensional Scaling (MDS) and its relevance in biology
- Explain the aims and motivation behind Principal Component Analysis (PCA) and its relevance in biology
- Write R code to carry out PCA
- Interpret the effectively communicate the output of PCA

Mini lectures

- Intro to MDS and scaling
- When the distances are Euclidean (PCA & bread)
- Carrying out and drawing inference PCA in 'R'
- Lab tutorial explained (PCA cheatsheet, peer review)

Assessed quiz on CANVAS worth 1% of your final grade

Lab—worth 6% of your final grade

Using a dataset of your choice you should make a cheatsheet illustrating the concepts of PCA and the steps required in the analysis. You may include R code and any plots you feel necessary. The text should be minimal, but easy to follow by your peers.

- Your cheatsheet should briefly
    1. Explain the aims of Principal Component Analysis
    2. Contain reproducible R code to carry out PCA
    3. Explain any inferences drawn
- Be very concise; rely on diagrams where possible.
- Pay attention to the details!
- Code comments inform, but fail to draw the readers attention. It is better to use arrows, speech bubbles, etc. for important information. If it is not important information, leave it out.
- Simple working examples are more helpful than documentation details.
- Add some concise text to help the user make sense of your sections, diagrams and inferences.

You may use whatever software you want to create your cheatsheet, however, you must export the file as a PDF before uploading it to CANVAS. You will also be expected to peer review 3 other cheatsheets; this also means that your work will be peer marked by 4 of your classmates. The cheatsheets you will review will be automatically assigned to you after the assignment due date. When carrying out this peer review please follow the rubric carefully and be mindful that your comments, although anonymous, will be passed on to your peer.

Labs 9–11 will be assessed via a one A4 page (11 pt font; not including plots) Executive Summary.

The aim of this assignment is to give you a chance to demonstrate effective communication skills. There were three main objectives to this section of the course (Module 3); students should be comfortable:

1. recognising which type of analysis is appropriate for a given dataset
2. analysing the data using a the R statistical software, and
3. communicating their findings, not only to the statistically trained, but also to those who have no formal understanding of statistics or its terminology.

This assignment covers the final three weeks of Module 3. Using the statistical skills you've learnt in this and previous labs use the appropriate statistical methods to analyse the dataset you chose at the start of the course (see the Data Sets page). Based on your analysis write an executive summary based on your results using no more than one A4 page (11 pt font; not including plots). Your summary should be aimed at a non-technical audience.

Due to there not being much space for your write-up your report needs to be clearly laid out and in logical order. For example,

- starting with a statement of aims, then
- describing (very briefly) how the data was collected,
- how it was analysed,
- what the main findings are,
- how reliable these are and what should be done next.

There is no need to repeat yourself (e.g., with a long concluding paragraph). Describe the model you have fitted or the hypothesis test you've carried out in detail and interpret any estimated parameters and/or probabilities. You should present the main findings in a quantitative way. For example, quote the average decrease/increase in predicted of the response in relation to the independent variables. It's important also to give a measure of uncertainty on this (e.g., if the model is correct, we can be 95results to too many significant figures (2–3 is enough!). Comment on any plots you make; plots should be effective in communicating the point you are trying to get across. Note: being non-technical should not mean being non-specific!

It will be difficult to convey all the important details of what you've done and what you conclude in such a small amount of space whilst making sure that each piece of technical language is appropriately explained (or is replaced by clear plain English words). However, you should try to find a way to discuss the assumptions and limitations of your analysis.

Your executive summary will be graded based on four main categories:

1. quality of core analysis,
2. quality of presentation,
3. demonstration and understanding,
4. initiative, originality, effort and reproducibility.

This Executive Summary will be worth 18% of your final grade and is in addition to he weekly CANVAS quiz. During lab time you are free to work through the material provided and your final report. You may work in groups (digitally), however, the final report must be your **own** work. Any plagiarism will automatically result in 0% for the report. Your Executive Summary should be no more than one A4 page. In addition to your summary you must upload your R script (as a .r file), which should reproduce your results. A well commented script is highly recommended.

Here are some guidelines to follow when writing your executive summary

- It should **not** contain any statistical terminology that would only be properly understood by a statistician
- Recall there are a set of main messages that you should report from your analysis, the reader doesn't need to know about all the work you carried out
- A brief outline of an Executive Summary should follow the sections listed below
  - Introduction: a one or two sentence description of the data and the purpose of the analysis
  - Methods: important **non technical** information for the reader about the analysis carried out
  - Report findings and the strength of evidence for them
  - Quantification: how reliable/generalisable are those findings
  - Summary: a one or two sentence summary of the major findings

Week 9 (beginning 18th May): Hypothesis testing

Learning objectives

By the end of this lab student should be able to

- List appropriate questions posed by the biological questions and outline an appropriate hypothesis test that would answer it
- Describe the aims of the following hypothesis tests
  - one-sample t-test
  - two-sample t-test
  - randomization test
- List the aims of hypothesis testing and write out the appropriate null and alternative hypothesis using statistical notation
- Write R code to carry out an hypothesis test using the appropriate variables in their dataset. Specifically write R code to carry out
  - one-sample t-test
  - two-sample t-test
  - randomization test
- Interpret and communicate the findings of an hypothesis test accurately and concisely
- List the limitations of the hypothesis in relation to the questions posed by the data

Mini lectures

- Hypotheses, why?
- Differences in mean
- Randomization tests
- Lab tutorial explained (executive summary)

Assessed quiz on CANVAS worth 1% of your final grade

Week 10 (beginning 25th May): Introduction to linear modelling

Learning objectives

By the end of this lab student should be able to

- Develop a biologically relevant question of interest from the dataset and identify the following types of variables in the dataset
  - response variable
  - explanatory variables
- Express their question of interest accurately and concisely
- Carry out and interpret tests for the existence of relationships between explanatory variables and the response in a linear model
- Write R code to fit a linear model with a single continuous explanatory variable
- Write R code to fit a linear model with a continuous explanatory variable and a factor explanatory variable
- Interpret estimated effects with reference to confidence intervals from linear regression models. Specifically the interpretation of
  - the intercept
  - the effect of a factor
  - the effect of a one-unit increase in a numeric variable
  - the effect of an x-unit increase in a numeric variable
- Make a point prediction of the response for a new observation

Mini lectures

- ANOVA $\equiv$ regression
- Explanatory variables and the response
- Fitting and interpreting linear models in R

Assessed quiz on CANVAS worth 1% of your final grade

Week 11 (beginning 1st June): Modelling II

Learning objectives

By the end of this lab student should be able to

- Write R code to fit a linear model with interaction terms in the explanatory variables
- Interpret estimated effects with reference to confidence intervals from linear regression models. Specifically the interpretation of
  - main effects in a model with an interaction
  - the effect of one variable when others are included in the model
- Explain why you may want to include interaction effects in a linear model
- Describe the differences between the operators : and * in R model-fitting formulae
- Critique the fitted model

Mini lectures

- Multiple explanatory variables
- Interactions
- Model diagnostics

Assessed quiz on CANVAS worth 1% of your final grade