BIOL2022

Biology experimental design and analysis

Unit manual

Semester 2, 2024
School of Life and Environmental Sciences

If found, please return to
SOLES Education Support
Room 519, Carslaw Building (F07)
The University of Sydney NSW 2006

Table of contents

Preface **1**
Your lecturers . . . 2

Outline **3**
Structure . . . 3
Lectures, Labs & drop-in sessions . . . 3
Assessements . . . 4
Late penalties . . . 4
Generative AI . . . 4

Policies **6**
Academic integrity . . . 6
Simple extensions . . . 7
Special consideration . . . 7

Studying for BEDA **9**
Important tips . . . 9
Study plan . . . 9
Assignments . . . 10
Cheatsheets . . . 10

Module 1 (Weeks 1 – 3) **11**
Preamble . . . 11
What this module is about . . . 11
Outline . . . 11

Week 1 – Introduction **12**
Welcome . . . 12
Learning outcomes . . . 12
What to submit at the end of the practical . . . 12
Workshop . . . 13
Exercise 1 – cheatsheets . . . 13
Background . . . 13
Try out a cheatsheet . . . 13
Exercise 2 – introduction to models . . . 14
Background . . . 14
Tasks . . . 17
End of practical . . . 18

Week 2 - Study design **19**
Welcome . . . 19

Learning objectives . . . 19
What to submit at the end of the practical . . . 19
Workshop . . . 19
Exercise 1 – Modelling from observations . . . 19
Background . . . 20
Getting started . . . 20
Study design . . . 20
Task . . . 21
Exercise 2 – data entry . . . 23
Submit your data . . . 24
End of practical . . . 24

Week 3 – Data analysis **25**
Welcome . . . 25
Learning objectives . . . 25
What to submit at the end of the practical . . . 25
Workshop . . . 25
Lab activity . . . 25
Exercise 1 – Preparing your data for analysis . . . 26
Background . . . 26
Missing data, outliers and data entry errors . . . 26
Exercise 2 – Data analysis . . . 26
Fitting a model to data . . . 26
Checking assumptions . . . 27
Statistical testing . . . 27
Interpretation . . . 28
Submit your results . . . 28
End . . . 28

Module 2 (Weeks 4 – 8) **29**
Preamble . . . 29
About this Module . . . 29
What you need to submit for this practical component . . . 30
How you will get feedback . . . 30

Report 1 – Projects **36**
Animal Ethics requirements for Projects on birds . . . 40
Conditions of bird project approval: . . . 40

Report 1 – Instructions **42**
About . . . 42
A note about plagiarism . . . 42
Submitting your individual report . . . 42
Requirements for your report . . . 43

Clare's tips on writing your report . . . 43
Abstract . . . 44
Introduction . . . 44
Materials and Methods . . . 44
Results . . . 45
Discussion . . . 45
General . . . 45
More tips on writing your report . . . 46
Structuring Research Papers: by Clare's Lab Group . . . 46
General notes . . . 46

Module 3 (Weeks 9 – 13) 50
Preamble . . . 50
Background . . . 50
Analyses and programs we'll use . . . 51
Timeline . . . 51

Week 9 – Part 1 52
Outline . . . 52
What you need to submit in Module 3 . . . 52
Analyses covered in this practical . . . 53
Part 1: Principal components analysis – "perceptions of biology lecturers" 53
1. Running the PCA . . . 53
2. Interpreting the PCA . . . 54
3. After the PCA . . . 54
Part 2: designing your multivariate experiment . . . 54
The data and your model system . . . 54
An example of a non-biological model system . . . 55
Groups . . . 55
Assessment . . . 55
Template . . . 56
References . . . 56

Week 10 – Part 2 66
Outline . . . 66
Analyses covered in practical 2 . . . 66
Community structure/Assemblage structure . . . 67
Using PRIMER (R instructions are given on Canvas within an R Studio files) 67
PERMANOVA, SIMPER and MDS analysis with PRIMER . . . 67
ANOSIM . . . 68
For next week . . . 69
1. Your habitat data for PCA . . . 69
2. Organising and entering assemblage data for PRIMER . . . 69
3. Generating summary statistics for additional (univariate) analyses 70

Week 11 – Part 3 **73**
Outline 73
Your own data 74
Tips for PRIMER 74

Week 12 – Support session **75**
Outline 75
Marking scheme 75

Preface

Welcome to Biology Experimental Design and analysis, also known as **BEDA**. This book contains all of the information you need to navigate your way around the practicals for this unit. It also contains some additional 'housekeeping' information that you may find useful.

Many students think of BEDA as *yet another statistics unit*. This is not quite true. Unlike most other statistical units, BEDA focuses on the **practical** aspects of data analysis. We assume that you already have a basic understanding of statistical concepts, and we build on that knowledge, focusing on the **how** and **why** of statistical methods:

- *How* do we design a good study?
- *Why* does study design make a difference in the results we get?
- *Why* is it important to understand the context in which we use statistical tests?

Importantly, **attending labs is crucial to your learning** as most of the data we use are collected by **you** – please do not skip any of them!

Please read all the sections before the practicals as they contain important housekeeping information.

Your lecturers

Dr Januar Harianto
Module 1 Lecturer
Rm 364, C81 – Biomedical Building
Unit Coordinator

Prof. Clare McArthur
Module 2 Lecturer
Room 303, A08 – Heydon-Laurence Building

Prof. Mat Crowther
Module 3 Lecturer
Room 225c, A08 – Heydon-Laurence Building

Outline

BEDA is taught under three modules, each led by a different academic, from lectures to practicals. Because of this, you will notice that the delivery differs between modules, and some content may be *repeated*. This is intentional. We believe that repetition is key to learning some of the most important concepts in study design and statistics. Some techniques *need* to be applied in different situations to be fully understood.

Structure

Module 1: led by Januar Harianto, covers the fundamentals. We expect most of you to be familiar with the basic statistical techniques, but we will go through the standard concepts and unify them under a linear modelling framework. In addition we introduce to you the concept of study design and how it influences the results of your analysis.

Module 2: led by Clare McArthur, critically assesses study designs that have already been done. Using real data examples, we will look at how variations in study design can influence the results. You *will* notice that the concepts from Module 1 are applicable to more *complex* models. You will also start working on Report 1 which requires you to design and analyse an experiment.

Module 3: led by Mathew Crowther, focuses on the application of statistical techniques to real-world data that is often complex and messy – and how to deal with it using multivariate techniques. You will start working on Report 2 which requires you to collect and analyse data from your environment, rather than a controlled experiment.

Lectures, Labs & drop-in sessions

Lectures are *compulsory* and are held on Tuesdays and Wednesdays at different locations:

- Tuesdays: 10am-11am – The Quad General Lecture Theatre
- Wednesdays: 10am-11am – Carslaw Lecture Theatre 159-259

Labs are held on Tuesdays, Wednesdays and Fridays. You will be assigned to a lab and you *must* attend the lab session you are assigned to – attendance is recorded. Labs are also *compulsory* at either 2pm-4pm or 4pm-6pm. Check your timetable for your assigned lab time.

All labs are held in Carslaw Lab 307.

Drop-in sessions are held weekly on Zoom. These are optional sessions where you can ask questions about the content covered in the lectures and labs and get help with your assessments. The schedule for these sessions will be posted on Ed, as the times may vary depending on the time of the semester.

Assessements

The assessments for this unit are outlined in Table 1. Compulsory assessments are marked in **bold** and must be attempted and submitted to prevent a fail grade.

Table 1: Assessments for BIOL2022. For detailed information, see Canvas.

Assessment	Type	Mode	Due	Weight
Quiz 1	Ind	Online	Week 1	0%
Quiz 2	Ind	Online	Week 2	0%
Quiz 3	Ind	Online	Week 3	0%
Evaluation Quiz	**Ind**	**Online**	**Week 4**	15%
Report 1	**Ind**	**Submitted work**	**Week 9**	25%
Report 2 – data upload	Grp	Submitted work	Week 11	5%
Report 2	**Ind**	**Submitted work**	**Week 13**	15%
Final Exam	**Ind**	**In-person**	**Exam period**	40%

Late penalties

- Failure to complete the Quiz on time will result in a 0 mark.
- A penalty of 5% per day will be applied for late submissions of all assessments, except where otherwise stated (e.g. an extension has been granted).

Generative AI

There's no escaping generative AI (GenAI) – so let's use it responsibly. The University maintains a Canvas site AI in Education which provides information on the use of AI tools in your studies. Some useful links include:

- Different generative AI options
- Acknowledging and referencing the use of AI
- the Guidelines for the use of AI in your studies

In most cases, we're happy for you to use it **as long as you're transparent about using it** and adhere to the academic integrity guidelines while doing so. Remember that you are here to learn – GenAI is in fact an excellent tool to help you understand the concepts better, but using it to do your work defeats that purpose.

Your report guidelines will outline our expectations on the use of *any* AI tool. If you're unsure, ask us!

Policies

Academic integrity

All students are expected to be familiar and act in compliance with the relevant University policies, procedures and codes. These include the following:

- Academic Integrity Policy (pdf)
- Academic Integrity Procedures (pdf)
- The Student Charter
- Student Responsibilities
- University of Sydney (Student Discipline) Rule (pdf)

Below are the types of academic integrity breaches that are monitored and enforced by the University, which students should be aware of:

1. **Recycling**: Resubmitting work that is the same or substantially similar to work previously submitted for assessment without permission.
2. **Plagiarism**: Presenting another person's work as one's own without appropriate acknowledgment of the source. This includes using phrases, sentences, or longer extracts from published or unpublished work without proper attribution.
3. **Collusion**: Presenting work as independent when it has been produced in whole or part with others, and with the knowledge of the parties involved.
4. **Contract cheating**: Engaging in various forms of outsourcing work to third parties, such as having someone else complete an assessment, purchasing completed assessments, or submitting work generated by artificial intelligence without permission.
5. **Exam cheating**: Attempting to gain an unfair advantage in an exam through prohibited materials, communicating with others, copying, or using electronic devices inappropriately.

Other breaches include fabricating data, inappropriately sharing assessments online, and using digital tools like paraphrasing software to disguise plagiarism or contract cheating. The severity of these breaches is determined by factors like the student's level of study, prior academic integrity breaches, and the extent and type of the breach. Breaches are classified as minor, major, or student misconduct, each carrying different penalties.

Simple extensions

Simple extensions are intended as an informal, streamlined way for students to get short extensions for legitimate reasons, without impacting their special consideration rights for more significant circumstances. According to the Coursework Policy 2021:

The special considerations unit may permit a student to submit an assignment task **up to five calendar days** after the due date with **no penalty** on certain grounds, including:

- illness;
- injury;
- misadventure;
- death of relative, close personal friend or pet; and
- carer responsibilities.

Simple extension must be applied for **before the due date** of the assessment task. **The student must apply for the extension through the Special Consideration and Arrangements portal.**

> ❗ Important
>
> Simple extensions are **not** applicable under the following circumstances:
>
> - Group work assignments;
> - where the assignment release time is 10 working days or less;
> - paid or volunteer employment;
> - clash with other assessments i.e. balancing workload;
> - computer/software malfunctions, unless it could not be prevented.

Students needing an extension **exceeding five calendar days** can apply through the special consideration portal with necessary documentation. For details, see below.

Special consideration

For more significant, longer-term impacts that are not sufficiently addressed by simple extension, students may apply for special consideration due to the following (clause 82):

- **Attendance at a funeral** of a family member or close friend.

- **Death** of a family member or close friend.
- **Illness** or **injury**.
- **Misadventure**.
- **Technology-related problem** which **could not** have been prevented, avoided, or minimised by your reasonable diligence, e.g. an extended power failure due to a storm.

Each of the above reasons **must** be supported by appropriate documentation, outlined under Eligibility in the special consideration Portal.

All applications for special consideration should be submitted **no later than 3 working days after the due date or examination date**, although late applications may be considered in exceptional circumstances. Applications for special consideration should be submitted via the Special Consideration and Arrangements portal on the course website.

Special consideration is not guaranteed and may result in alternative outcomes like alternative assessments or adjusted assessment weightings.

> **Note**
>
> Disability-related special consideration is not covered here but may factor into special consideration applications. For more information, please refer to the Inclusion and Disability and Disability Inclusion Action Plan webpages.

Studying for BEDA

Important tips

BEDA can be challenging, but with the right approach, mastering our content is achievable. Here are some tips to help you study effectively.

- **Keep up with lectures.** Attend lectures in-person or watch the recordings *as soon as possible*. It is important to go through the material regularly and understand it before the practicals. This is especially important for those of you who have practicals on Tuesday, as the Lectures are in the morning.
- **Regular practice is key.** Statistics requires regular practice. Make sure to allocate dedicated time to study and understand the concepts. Skipping content or cramming at the last minute will hinder your understanding and ability to apply the concepts to different scenarios.
- **No question is silly.** Don't hesitate to ask questions! There are multiple channels available for you to seek help and clarify any doubts you may have. You can ask questions during lectures, practicals, or in the Ed discussion forums. If you prefer anonymity/privacy, you can also ask questions anonymously on Ed or request a consultation with the lecturer. Additionally, Januar (your coordinator) holds weekly drop-in sessions where you can address any queries you may have.
- **Master the GLM.** Module 1 covers the general linear model (GLM) approach to study design, which forms the basis of all the statistical tests in this unit. Understanding the GLM well will simplify the rationale behind all the statistics you learn. Modules 2 and 3 will build on this foundation, but only mention it briefly. **If you have a strong grasp of the GLM approach, you will find these modules much easier to understand.**

Study plan

For BEDA, we recommend that you spend an hour after each lecture to revise the content and write your notes. Additionally, spend an hour at the end of the week to consolidate both the lectures and practicals. A futher minimum of two hours per week should be spent on reading up and understanding the content – **independent study is key to mastering the concepts in this unit**.

Assignments

The above excludes the time you will spend on assignments and preparing for the exams. We recommend that you **start your assignments early and not leave them to the last minute**. This will give you time to ask questions if you are stuck.

Cheatsheets

A cheatsheets section is available on the online version of this manual, updated frequently based on the questions you ask on Ed. This section will contain cheatsheets which will help you with data analysis. We hope that this will aid you in your understanding of the content and prepare you for the assessments.

Module 1 (Weeks 1 – 3)

Preamble

Science is a process of asking the right questions in the right way, such that we can draw meaningful conclusions when we collect data to answer them. This is not as easy as it sounds, but it remains a fundamental skill in scientific enquiry. We want to be able to ask questions in a quantitative manner to draw distinctions between different explanations for the phenomena we observe.

Before we can start performing t-tests, regressions and ANOVAs; before we can start implementing *fancy* machine learning algorithms; before we can start using the *latest* and *greatest* tools for data analysis, **we need to be able to ask the right questions**.

So let's (re)-start with the fundamentals.

What this module is about

Module 1 is designed to help you develop the fundamental skills in scientific enquiry. Together, we will cover experimental design, model formulation and sampling design. Each week, we start off with a short **30-minute workshop** to introduce you to the concepts and techniques. You will then work through the practical.

Outline

Week 1 Introduction – We will introduce you to the practical structure, the software you will be using throughout the unit and also get you started on some modelling exercises.

Week 2 Data sampling – Explore the concept of sampling and how to design a study that can draw meaningful conclusions – by collecting data from images. More modelling exercises will be provided.

Week 3 Data analysis – Practice data analysis and diagnostics using the data you have collected in the previous week, and discuss what worked and what didn't.

Week 1 – Introduction

Welcome

Welcome to the first practical of BEDA. In this practical we will introduce you to the software you will be using throughout the unit.

The only expectation for this practical is that you have attended the first lecture. If you were not able to attend, please watch the recording before you come into the laboratory!

In Module 1, all practicals begin with a **workshop**. The workshop is a short class that goes through a key concept. The workshop is followed by **exercises** that you will complete in the laboratory. The exercises are designed to help you understand the concepts covered in the lectures and gives you a chance to practice study design and data analysis.

Learning outcomes

By the end of this practical, you should be able to:

1. Identify the software you will be using in this unit.
2. Be able to use cheatsheets to help you with your data analysis.
3. Have a basic understanding of models and how they can be used in study design.
4. Be able to formulate a model with study design in mind.
5. Be able to interpret plots as models.

What to submit at the end of the practical

If possible, send us some quick **feedback on cheatsheets.** Your feedback will help us improve the cheatsheets for future students. Otherwise no other submissions are required since this is the first practical, but think about how you would show your work to your peers and demonstrators as you work on the exercises.

Workshop

In today's workshop we will go through:

- About us
- Introduction to R, SPSS and Jamovi

The workshop should take no more than 30 minutes (**unless Januar talks too much**). Workshop slides are available on Canvas.

Exercise 1 – cheatsheets

Background

In this short exercise we will introduce you to **cheatsheets**, which are quick reference guides that help you remember important functions and commands. Choose a cheatsheet and try to complete the task(s) within.

Try out a cheatsheet

You will need to use a web browser to access the cheatsheets. Check with your demonstrators if you are unsure how to do this. For those of you who are already reading this on the web, you can simply click on the **Cheatsheets** link in the menu to the left.

1. In the **Cheatsheets** section, you will find a list of cheatsheets. More will come soon depending on feedback and discussions on Ed.
2. **Choose a cheatsheet** that you would like to try out – it could be something that you are *not* familiar with, or you could also choose one that you have some experience with to see if it aids your understanding.
3. **Follow the steps on the cheatsheet**. If you get stuck, don't worry! You can ask your peers or the demonstrators for help, and you may provide feedback so that it can be improved.

Once you are done, please provide feedback on the cheatsheet(s) you have tried out. You can do this by submitting a comment on the form provided to you on the cheatsheets page. The form will ask you to rate the cheatsheet, and should take no more than a minute to complete, so please help us out!

Your feedback will help us improve the cheatsheets for future students.

> **i** Contribute to the cheatsheets (optional)
>
> If you would like to take one step further and *contribute* to the cheatsheets, we are more than happy to accept your suggestions. You can do this in several ways:
>
> - **Email Januar with your suggestions.** This is absolutely fine!
> - **Submit a comment on Ed.** We will discuss, and update the cheatsheets accordingly.
> - **Submit a pull request on GitHub.** This requires technical know-how, but some of you may be able to do this. You can find the link to the GitHub repository on the cheatsheet.
>
> If your changes are accepted, you will be credited and given co-authorship of the cheatsheet. Your work will be part of the unit's resources. You can also refer to your contribution in your CV or portfolio! For more information, please check with Januar.
> **Suggestions and contributions can be provided throughout the semester, so don't worry if you don't have time to contribute now. We will be updating the cheatsheets regularly.**

Exercise 2 – introduction to models

Background

One important aspect of study design is the ability to model data. One way to do it is by using graphical plots, which essentially capture the essence of a study design in a visual form. In this exercise, we will introduce you to the concept of modeling data and get you thinking about how to use models to assist in experimental design and analysis.

What is a model?

A model is a simplified representation of a complex system. In the context of data analysis, a model is a way to represent a dataset that allows us to shape testable hypotheses or make predictions about the data. Models can be simple or complex, depending on the data and the research question.

Eventually, you will be presenting what we call an **empirical model**, which looks like a mathematical equation, but actually does not really require *any* mathematical knowledge to understand.

An example of an empirical model is this:

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

Which roughly translates to "the value of y is equal to the value of x_1 times a constant, plus some error and an initial constant."

The above may look intimidating, but we can simplify it by thinking of it as a simple relationship, rather than an equation, therefore forming:

$$y \sim x_1$$

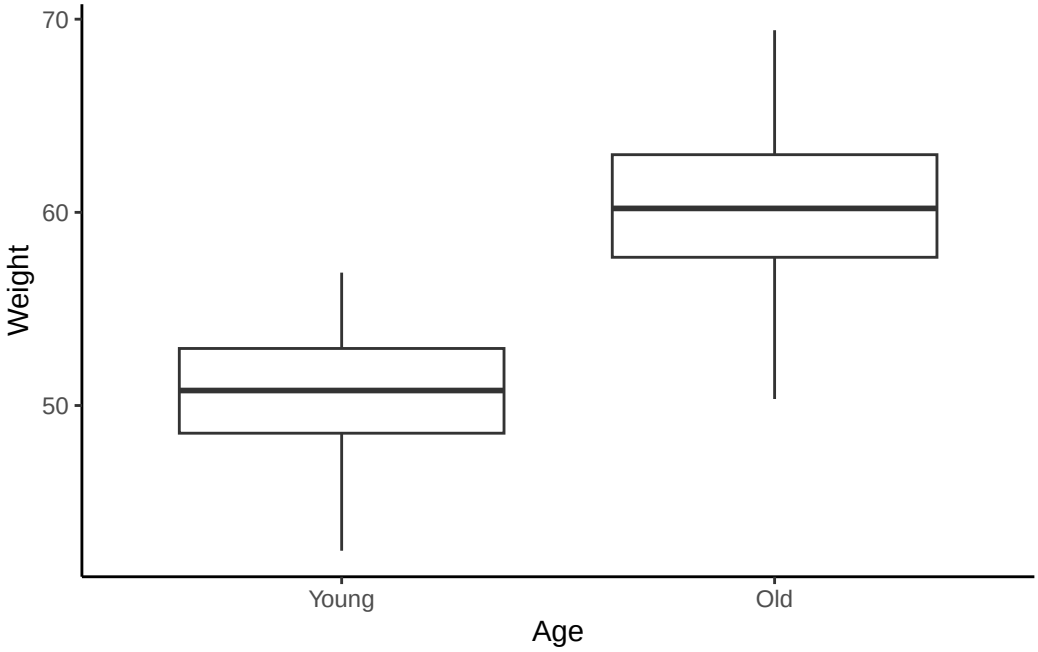
And so if y is weight and x_1 is age, then it is interpreted as:

$$\text{weight} \sim \text{age}$$

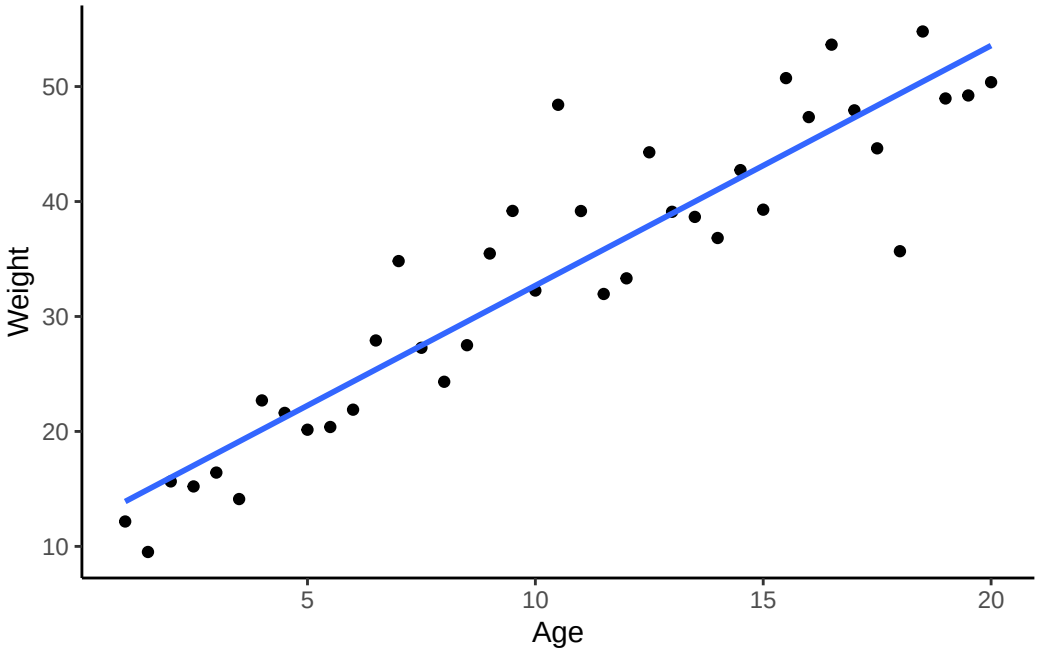
Which translates to "the weight of an object is influenced by its age".

Plots as models

For now, we will focus on **graphical models**. These are *plots*, but essentially they are models since they can represent a relationship between two or more variables. Using the same example as above, we can plot the weight of a species of animal against its age and height to see if there is a relationship between them:



Interestingly, how you consider your variables can drastically change the type of plot and model used for data analysis. For example, consider the same relationship as above, but plotted differently:



Notice how both plots explain the same kind of relationship, but their interpretation and the type of model used are different. This is the *essence* of study design and data analysis – the model you select can drastically change the data and results you get, **and planning for it is crucial to the success of your study**.

> Tip
>
> If you can **plot** it, you are already **modelling** it!

Tasks

Preparation

- MS Excel or similar: to read the files
- MS Word, Google Docs or similar: to write your answers
- Data: penguins.csv and possum.xlsx

Instructions

- **Step 1.** Download and explore the data: penguins.csv and possum.xlsx.
- **Step 2.** Identify variables from the data to formulate models. You do no need to analyse the data, neither do you need to explore it in detail. **The focus is on understanding the variables and how they can be modelled.** Use the variables to consider the following models:
  - Histogram to show the distribution of a variable
  - Scatterplot to show the relationship between two variables
  - Boxplot to compare two or more groups
  - Barplot to compare two or more groups, but with error bars. Consider what the error bars represent.
- **Step 3.** For plots that compare two variables or more, consider how you could alter the way you collect data to use a different model. Also consider this: is the alteration logical and feasible?
- **Step 4.** These plots are often associated with specific empirical models and statistical analyses, which we will cover in the upcoming weeks. Can you predict what sort of statistical model you would use to analyse the data in Step 1 and Step 2? **It is ok if you do not know most of the answers to this question yet.** Discuss this with your peers and demonstrators.

> **i** Example
>
> 1. Browse the **possum data** in MS Excel. *Let's consider a model of how the* ***weight*** *of a possum changes with* ***age****.*
> 2. Identify the variables: what type of variables are they? Are they continuous, categorical or something else? With that in mind, draw the plot(s) that you think will help you interpret the data. *Weight is continuous and age is categorical, so a boxplot would be a good start.*
> 3. Different model: *Age data to be collected in years, rather than categories. This would allow us to use a scatterplot to model the data. But it may not be possible to age possums as we do not know their birth dates, so this may not be feasible.*
> 4. *The first model is probably a two-sample t-test, while the second model could be a linear regression.*
>
> Notice that in the example above, you have not performed any data analysis. You are simply considering how to interpret the data, which is an important step in study design. This helps you determine the type of data you need to collect and how to analyse it.
> If you look at the data closely, **there is only one juvenile possum in the dataset**, so you would not have been able to analyse this data anyway!

End of practical

That's it for today! If you have any questions, please ask your demonstrators. They are here to help you. Remember to submit your feedback on the cheatsheets and to work on the exercises in your own time. We will see you next week for more study design and data analysis!

Week 2 - Study design

Welcome

Welcome to the Week 2 practical. Today we will be focusing on data collection and how to prepare your data for analysis.

Learning objectives

By the end of this practical, you should be able to:

1. Formulate a model and hypothesis based on observations.
2. Design a study to test your model.
3. Enter data into a spreadsheet.
4. Understand the importance of tidy data.

What to submit at the end of the practical

1. **A simple study design**, documented on Google Docs. In groups, you will describe your model and sampling schemes on a shared document.
2. **Data** – your collected data will need to be submitted to your demonstrators for analysis next week.

Workshop

The Week 2 workshop will go through **Tidy data principles**. Workshop slides are available on Canvas.

Exercise 1 – Modelling from observations

> ❗ Important
>
> Remember, the best models are often the simplest. Focus on logical and critical thinking when designing your study. **The sampling design is usually more important than the statistical model.** Keep this in mind as you work through the exercise!

Background

Images are commonly used to quantify patterns in biology, from molecules to ecosystems. In ecology, for instance, satellite images and aerial photographs are used in some ecological studies to quantify patterns of distribution and abundances of animals and plants across habitats.

Here you will explore experimental design, model formulation and data analysis by collecting data from available images and use both evidence and logic to work out how to explain the patterns you observe.

Work in groups – You will be working in groups of 3-4 students. Your task is to **observe** the images provided and formulate a **model** that explains the patterns you observe. You will then create a **study design** that will allow you to test your model.

> **!** Important
>
> Do not proceed to Exercise 2 until you have completed this exercise, and a Demonstrator has sighted your study design.

Getting started

You are provided with random, representative, replicate photographs of a common marine habitat that spans > 8,000 km of the Great Southern Reef, i.e. the coastline along the bottom half of Australia – intertidal rocky shores (Fig 1).

Please check the online Unit Manual for the link to the images. Your demonstrators will also be able to show you how to access the images.

Using these images and the species identification guide, formulate a model and a (testable) hypothesis. You will then think about what data you need to collect to test your hypothesis. See below for further details.

Study design

You should have a clear model in mind before collecting data. The model determines your sampling strategy and the type of data you need. Consider the following:

- **What is the purpose of your model?** Think about the biological question you are trying to answer, and the kinds of data you need to collect to answer it.

- Think of how you would collect data to test your model. **What are the variables you need to measure?** How would you measure them?
- Scrutinise your images with the model in mind. **Can the data be collected representatively from the images?** Are there issues of replication or confounding variables that you need to consider?
- Do you have the time and resources to complete the study in the time frame provided, i.e. by the end of next week?
- What is your **backup plan** if data assumptions are not met?

While you may not have all the answers right away, it's important to have a rough idea of what you want to do, such as comparing groups, measuring correlation, or creating a linear model.

Figure 1: The Great Southern Reef. © 2021 California Academy of Sciences.

Task

A Google Docs page will be shared with you to record your responses. Please make sure that all group members have access to this document. On the document, claim a page and write your group name and the title of your project.

Then, address your study design by answering the following questions:

1. **Model and hypothesis**: What is your model and hypothesis? What are you trying to test and how will you test it? What would the plot look like?
2. **Variables**: What are the variables you need to measure? What types of variables are they (e.g. continuous, categorical)?
3. **Data collection**: How would you collect data to test your model? What are the potential issues you need to consider?

> **i** Example
>
> Note that this example will probably not be relevant to your images, but it should give you an idea of what is expected.
>
> - **Project title:** Does rock size influence the number of species that can be found on a rocky shore?
> - **Group name**: Schist happens
>
> 1. **Model and hypothesis**:
>
>   - We hypothesise that the number of species on the rocky shore is positively correlated with the size of the rocks.
>   - We will model the relationship between the number of species and the size of the rocks using a linear model.
>   - The plot is expected to be a scatter plot with the number of species on the y-axis and the size of the rocks on the x-axis.
>
> 2. **Variables**:
>
>   - We need to measure the number of species and the size of the rocks.
>   - Number of species is an integer variable.
>   - Size of the rocks is continuous variable.
>
> 3. **Data collection**:
>
>   - For each image, we will divide it into a 10x10 grid.
>   - 3 random quadrats will be selected from each grid.
>   - The size of all visible rocks will be measured in each quadrat.
>   - The number of unique species in each quadrat will also be counted.
>   - We will then calculate the average size of rocks and the average number of species for each image to minimise pseudoreplication.
>   - Possible confounding could arise from the presence of seaweed, which could affect the number of species present.

- If the number of species is too high to count, we will group species into functional groups and count the number of functional groups instead.

4. **Backup plan**:
    - If the assumptions of the linear model are not met, we will transform the data, otherwise we will attempt a non-linear model.

Exercise 2 – data entry

It is time to sample data from the images. While doing so, consider how you might minimise bias in your sampling. Also, if have not already done so, plan for time:

- How long will it take to sample data from each image?
- How many images will you sample from?
- How many samples will you take from each image?

Enter your data into a spreadsheet as soon as you have them. You can use any spreadsheet software you like (e.g. Excel, Google Sheets, Numbers). Google Sheets is easiest if everyone in your group is sampling and entering data at the same time as it is easy to set up a shared spreadsheet, but you may also share the same computer and enter data one at a time.

! Discuss

If more than one person is entering data, will it influence the results? Why or why not? Think about how you can reduce the impact of this.

i Optional: tidy your data

To make your life easier, *recall* the tidy data principles outlined in the workshop:

1. Each variable must have its own column.
2. Each observation must have its own row.
3. Each value must have its own cell.

The table below is tidy:

person	treatment	count

John	A	5
Mary	A	3
Jane	A	2
John	B	6
Mary	B	4
Jane	B	5

The table below is not (although it may look "better" to you). Why?

person	treatment A	treatment B
John	5	6
Mary	3	4
Jane	2	5

Note: Tidy data will reduce the amount of time you spend cleaning your data during data analysis, but you can also choose not to follow these principles if you prefer.

Submit your data

Demonstrators will provide you with a USB drive to submit your data. Please make sure that your data are saved in a format that can be opened by Excel or Google Sheets. **Your group will be using this data next week so make sure that we have it!**

End of practical

Don't forget to submit your data! You should also start to look at the projects available for Module 2, as you will sign up for one of them next week to work on Report 1.

Week 3 – Data analysis

Welcome

In today's practical, we will focus on data analysis. You will work with your group to analyse the data you collected in the previous lab. Cheatsheets are available for you to use, and your demonstrators are ready to help.

You may need to clean your data before you can analyse it, so be prepared to spend some time on this.

Learning objectives

By the end of this practical, you should be able to:

1. Understand the importance of reproducible analyses.
2. Fit a model to your data.
3. Check if the data meets the assumptions of the model.
4. Interpret the model output.

What to submit at the end of the practical

Your plot and model output, on Google Docs (or equivalent). We will spend some time discussing the results at the end of the tutorial.

Workshop

In today's workshop we will go through **Reproducible analyses and why R is good at it**. We will also demonstrate some AI workflows that are implemented in RStudio. Workshop slides are available on Canvas.

Lab activity

When ready, work with your group to analyse the data you collected in the previous lab. You will need to:

- Fit your intended model to the data.
- Check if the data meets the assumptions of the model.

- Interpret the model output.

Of particular importance is the assessment of assumptions using residual plots, rather than formal tests of assumptions.

Your demonstrators are ready to help. We expect you to be able to select the appropriate model to your data, but you can consult our cheatsheets and your demonstrators for help.

Exercise 1 – Preparing your data for analysis

Your data is available to download on Canvas. Please check that you have all the data you need before proceeding. You also have access to the data collected by other groups should you wish to practice your data analysis skills.

Background

If you have followed study design principles, the data you have collected should be clean and ready for analysis. However, it is still a good idea to inspect your data further before proceeding with the analysis. This includes a check for:

- Missing data
- Outliers
- Data entry errors
- Assumptions of the statistical model, and whether the data meets them

Missing data, outliers and data entry errors

These are what we sometimes call systematic errors. They can be detected by looking at the data and checking for unusual values, or cross-checking methods within your group. Make sure to remove or correct these errors before proceeding with the analysis.

Exercise 2 – Data analysis

Fitting a model to data

Recall that you have an empirical model e.g. `y ~ x`. You will need to use an appropriate statistical model to fit this relationship to your data.

> **Note**
>
> You should already know what model you are fitting to your data as it it part of your study design!

Checking assumptions

Assumptions need to be checked as they basically tell you whether the model you have fitted is appropriate for your data. If the assumptions are not met, you may need to transform your data or use a different model.

Common assumptions include:

- Normality (of residuals)
- Homogeneity of variance
- Independence of observations

We recommend looking at residual plots, rather than formal tests of assumptions. Formal tests of assumptions can be overly sensitive to large sample sizes and can lead to you rejecting a model that is actually appropriate for your data.

What to do if assumptions are not met? This is a common problem in data analysis, and interestingly, it is not always a problem. Provided that you have ensured that your data is representative of the population you are interested in, violations of normality can actually be ignored in many cases, **although you should still report them**. A larger problem is homogeneity of variance.

Nevertheless, you get to decide on what to do here to avoid violation of assumptions. The first step would be to try and transform your data. If this does not work, you will need to use a different model or a non-parametric equivalent of the model you are using.

Statistical testing

Running the statistical test is probably the *least* eventful part of the analytical workflow and will take you a few seconds regardless of the software you are using. Make sure that you record the software used and the specific statistical technique selected such that the analysis is **reproducible**.

Interpretation

Interpreting the output of a statistical test is a skill that takes time to develop. You will need to be able to:

- Understand the important parts of the output e.g. F-statistic, p-value, degrees of freedom, etc.
- Explain what the output means in the context of your data.
- Explain what the output means in the context of your research question.

Your lectures and this week's workshop will help you develop these skills.

Submit your results

Once complete, please upload your results to Google Docs (or equivalent) and share the link with your demonstrator. We will discuss the results together.

End

That's it. Three weeks in and Module 1 is almost over We hope that the content has been useful and that you are starting to see how study design and data analysis are interlinked. In Module 2, you will need to use the skills you have developed to design and analyse an experiment for Report 1.

Module 2 (Weeks 4 – 8)

Preamble

I think one of the most exciting things about being a biologist is discovering how the world works. One skill that helps me do so, is knowing how to set up an experiment to test my ideas about the world. You may need to overcome that initial learning hump, but if you persist, you will have a skill at your fingertips that will stand you in really good stead for the future. It doesn't matter whether you intend to go on in research or not - knowing what makes good research and good evidence is important in so many walks of life.

My message is: whether you are doing this unit just because you have to, or because you are interested – take heart! I never did anything much more complicated than a t – test even during my PhD. But now I LOVE designing experiments and grappling with data. It can be frustrating at first, especially if you are time-poor and just want to get the job done. It can be infuriating. It can also be fun (yes – fun!). Most of all it's incredibly satisfying and extraordinarily useful. So, persist, try your best, and ask for help if you need it.

Please remember we want you to learn this stuff.

No question you ask is EVER silly! So ask for help any time. We *do* expect you to have a go at telling us what you *think* the answer is. This helps us understand what you get and what you don't get, and so helps us help you.

We don't judge you by the questions you ask, but by what you end up learning.

About this Module

In Module 1, you have learnt the principles for converting a biological observation or question of interest into practice. Here, you will build on this skill. You will sign up to a research project and go through the process of converting a biological idea into a research experiment, which you and your group will then do from start to finish, to answer it. **You must draw on all the lectures we present to you in this unit, so COME / LISTEN to them as soon as we give them.** You will need to use the information immediately in these practicals. You will use your knowledge and experience to design your project, collect the data, and analyse and interpret it appropriately.

What you need to submit for this practical component

This assessment requires two separate submissions in Canvas:

1. **Report 1 Dataset Submission**: The Excel file (spreadsheets) containing your Group's data. This needs to be submitted by one group member on behalf of your group.
2. **Report 1**:Your Individual Report on your experiment, in the format of a short scientific article (worth 25% of the total unit mark).

How you will get feedback

- You will receive informative, but not exhaustive, feedback on your report in relation to the marking matrix. This will show what you achieved against each criterion, and why.
- To help you understand how you could improve, it helps enormously if you make a conscious effort to compare your submitted work against others. To facilitate this, I will post the best report written for your particular project.
- Should you want feedback beyond this, please reflect on your report and write a summary of how your report compares to the best one, addressing each of the marking criteria. Once you've done that, please email me to request an appointment, attaching your summary to the email message. BUT please note that because of the large class size, I may not be able to meet with everyone who wants to. As I am sure you will appreciate, I will give priority to any students with low/fail marks.

Week	Todo	Summary	Details
2 – 3	Sign up to a project group for Module 2 in your practical session.	Everyone in your group must attend the same practical session.	
4	Attend your practical session.	Work with your group to discuss and **plan** your project. Pick up any gear that you need, to run your pilot study. **After the practical**, contribute to your group's pilot study practical work, read relevant literature, and start preparing for your written report.	**Form your hypothesis:** as a group, convert the general biological question into specific biological hypotheses and predictions. **Plan the design of your main experiment:** with your group, design a **simple** experiment to test the hypothesis. Ensure your design can be analysed using the stats you are familiar with (e.g. from Januar's and Clare's lectures, and Module 1 practicals). Draw a graph of how you expect your data to look (refer to Clare's lectures). Decide what stats analysis is appropriate and why (refer to Clare's lectures). If you can't work this out, you are not ready to run your experiment. **Draft your data sheets:** as a group, draft your data sheets for data collection (you may need to modify after your pilot study). This will ensure you collect all the data you should and collect it all in the same way. **Work out what equipment you'll need:** decide equipment you need to do the research based around what we can supply you. Consult Clare, technical staff and/or demonstrators then make a list. For some projects, you will be able to take gear home straight away to run a pilot study. For others, you need to plan now and pick up gear in consultation with the technical staff. When contacting the tech

Week	Todo	Summary	Details
			staff, ensure you state your **name** and **day**, **group number** and **project**, e.g. BIOL2022 WED 2 to 4pm, Group 7 birds, Joe Blogs. **Plan your pilot study in class and run it *before* week 5 practical session:** this will help you refine your main experiment. Everyone in your group must contribute. Clare, technical staff and demonstrators will be available in the Lab to discuss expectations for this practical module and plans for the coming weeks in session, and in your own time. *Your group should aim to be self-sufficient out of practical times. However, for technical help outside of the practical, or last-minute changes gear to be picked up, please contact heather.sowden@sydney.edu.au.*
5	Attend your practical session.	Finalise the design and plans for your main experiment based on your pilot study and your reading of the relevant scientific literature Pick up any gear that you need for your main experiment.	**Discuss** what you learnt from your pilot study with your group. **Tweak the design of your main experiment if you need to**: make any adjustments to your plans and finalise your main experiment and its experimental design. **Know what stat tests you'll be doing**: ensure you know how you expect to statistically analyse your results, e.g. what your response and explanatory variable(s) are, what form and hence what stats test you intend to do. *Clare, technical staff and demonstrators will be available in the Lab to discuss your plans and provide feedback and gear. For technical help outside practicals, or last-minute changes gear to be picked up, please contact heather.sowden@sydney.edu.au.*
5 – 7	Work with your group.	Set up your main experiment, run it,	Collate the data: collate your group's data into one Excel file – e.g. using Google Sheets. Include both a metadata worksheet

Week	Todo	Summary	Details
		collect and collate the data. Everyone in your group must contribute.	and a clean results array in its own worksheet for data analysis. You can also have other worksheets for anything you want to record, plot, etc. Finalise data collation before week 7 practical if you can, leaving you more time for data analysis in Week 7 practical. *For technical help outside practicals, or last-minute changes gear to be picked up, please contact heather.sowden@sydney.edu.au.*
7	Attend your practical session. Work with your group.	Finalise data collation of your main experiment (including metadata). Start data analysis of your main experiment (and if possible, complete it) Decide what plots and particular stats output you will need to report in the results section of your written report.	**If you have not done so already, finalise data collation**: spend no more than 30 mins of the practical finalising data collation into an Excel spreadsheet. Preferably, have this done before this practical session. **Metadata**: include a metadata worksheet that summarises important information referring to your data, such as: • names of group members • the dates the experiment was run • sites where you collected data • response (i.e. dependent) variables and their units of measurement, • explanatory (i.e. independent) variables, such as treatments (and the levels within each treatment) and any continuous variables • anything to help you understand your project and data if you were to publish it for the world to see and understand, e.g. what the columns in your data sheet mean **Plot, analyse and interpret**: in the remaining practical time (at least an hour) using your group's data set, each group member should individually learn to use Excel and R or SPSS to plot your

Week	Todo	Summary	Details
			data, run appropriate statistical analyses and interpret your results. You need to think about what stats results you need in the results section of your written report. Fantastic if you can complete this task by the end of week 7. **Note**: It is okay to work with other group members to plot and run the tests. However, each group member should do the analyses on their own so that they learn; watching someone else is not enough.
7 – 8	In your own time.	Continue your data analysis if you have yet to complete it. Start *drafting* your report.	**Data analysis**: as above **Drafting your report**: make a start on your report, e.g. framework for Introduction, Materials and Methods, and Results. You could leave the Discussion until you have finalized your Results. Refer to: • Clare's tips on writing your report • More tips on writing your report • Module 2: Marking Scheme for Project Report. *For technical help outside practicals, contact heather.sowden@sydney.edu.au.*
8	Attend your practical session. Work with your group.	Clare will go through writing your report. Finalise analyses, stats output and graphs.	As a group: prepare your Excel data file for submission. The filename should be of the form Day-Time-Project, e.g. "WED2-4pmProj4A". This file must include: • a metadata worksheet • a worksheet of the data array you will have imported to a stats package for analysis • any other worksheets with workings, plots, etc. • As individuals: each group member needs to work on writing their report by: • completing their own results and graphs

Week	Todo	Summary	Details
			• deciding what they will include in their individual report • deciding how they will present the data and the associated statistics
9	Submit your assignment.	Submit your group's data and your written report in separate assignments in Canvas.	There are two Canvas assignments for this assessment: • **Report 1 Dataset Submission**: one person from your group needs to submit the group's data for this Canvas assignment. • **Report 1 (individual)**: each person in your group needs to submit their own report for this Canvas assignment.

Report 1 – Projects

> ❗ Important
>
> **FOR ALL PROJECTS – KEEP YOUR EXPERIMENT SIMPLE.**
> You should be able to design an experiment that could be analysed using one of the following:
>
> - a paired t-test,
> - an independent t-test,
> - 1-way ANOVA,
> - 2-way ANOVA,
> - correlation,
> - linear regression,
> - Chi-Square analysis or
> - logistic regression.
>
> General linear models (which covers a lot of the analyses above anyway) and generalised linear mixed models are *fine* **PROVIDED you do them correctly**.
>
> - You will be rewarded for doing a simple appropriate analysis correctly.
> - **You will be penalised for doing a complicated analysis wrong.**

Project 1 – Plant-herbivore interactions: Does previous damage to tomato plants affect their vulnerability to subsequent herbivory? Does the amount of previous damage matter?

Background: Many plants have plant secondary metabolites (PSMs), which often play a role in defence against herbivores. In some plant species, PSMs can be induced (levels increase) after damage from herbivores chewing on leaves. This change can take place within a day, affecting subsequent leaf damage by herbivores, and result in less subsequent damage. Is the evidence consistent with this idea in tomatoes?

Variation: If you want to answer a more complicated question, you could ask whether the amount of initial damage (a little versus lots?) affects the outcome.

> **Note**
>
> You do not need to measure PSMs. Rather, they provide a potential mechanism to explain any difference in subsequent herbivory.

Project 2 – Plant-herbivore interactions: Does previous damage affect vulnerability of vegetables to subsequent herbivory? Does it depend on plant species?

Background: Many plants have plant secondary metabolites (PSMs), which often play a role in defence against herbivores. In some plant species, PSMs can be induced (levels increase) after damage from herbivores chewing on leaves. This change can take place within a day, affecting subsequent leaf damage by herbivores, and result in less subsequent damage. Is the evidence consistent with this idea? Does it depend on plant species?

> **Note**
>
> You do not need to measure PSMs. Rather, they provide a potential mechanism to explain any difference in subsequent herbivory.

Project 3 & 9 – Foraging strategies in relation to food type: Do ants prefer some food over others?

Background: Ants spend a lot of their time searching for food, but the food they find varies in its nutritional and energetic value. Do they just forage for any foods or do they show preferences? Honey and tuna, for example, are very different foods. Do ants care? Honey is a high-carbohydrate low-protein food, while tuna is high protein, low carbs; although there are lots of other differences between these food as well.

Scout ants go out and look for food, then return to the nest and recruit others to help them. We can understand something about the requirements ants have for different foods by looking at how quickly and/or how many ants are recruited to a food source once it has been located by the scout ant.

Project 4 & 10 – Foraging strategies in relation to costs and benefits: Do ants seek food that is easier to access?

Background: Ants spend a lot of time searching for food, but the effort needed to harvest a given food could affect how they value it. For example, food that is difficult or slow to access (e.g. surrounded by complicated 3-D local habitat or with narrow access points slowing ant traffic) may be less preferred than food that is easy to access.

Variation: IF you want to answer a more complicated question, you could ask whether ants balance the effort of getting food against the value of that food. Does food that is hard to access need to be higher in value (e.g. more concentrated) than food that is easy and less energetically costly to access?

Scout ants go out and look for food, then return to the nest and recruit others to help them. We can understand something about the requirements ants have for different foods by looking at how quickly and/or how many ants are recruited to a food source once it has been located by the scout ant.

Project 5 – Germination of native grass seeds in relation to water availability

Background: Germination of native plants is triggered by a range of factors in Australia. Some plants need heat initially, others simply need water. Microlaena stipoides (weeping grass) is a native grass that is also used for turf and feed for livestock. It occurs along much of eastern mainland Australia, Tasmania and even in SW Western Australia. It is eaten by our marsupial herbivores, e.g. wombats. Its seeds germinate in a range of conditions within 1 to 2 weeks (>10 oC, faster when warmer), do not need fire but do need moisture. How often it rains may affect whether seeds germinate, in turn affecting recruitment and regeneration. Understanding the relationship between rain (watering) frequency and probability of seeds germinating helps us understand how climate influences the dynamics of our grasslands and use of native plants in agriculture.

> **Note**
>
> In your main trial, if you get no germination after 10 days, your group can choose to switch to wheat (Project 6) to ask the same question, because wheat germinates quickly.

Project 6 – Germination and growth of wheat in relation to water availability

Background: Growing crops such as wheat is crucial for food security. In Australia, our rainfall is highly variable, and this can be daunting for farmers wondering whether to sow seed and whether it will germinate if they do. White wheat varieties in Australia can germinate under a range of temperatures (12 to 25oC), but whether the initial rainfall is light (e.g. shower) versus heavy (e.g. soaking) may affect whether seeds germinate and/or seedlings grow and/or survive. Understanding these relationships is one step in understanding the conditions required to help ensure our food security.

Project 7 & 11 – Are birds active at different times of day?

Background: Birds may be active at different times of day for a variety of reasons. As a first step in understanding bird abundance and diversity we need to be able to measure it and know whether when we measure it makes a difference. If you stand in the same spot and count birds, does the outcome depend on when you did it? Are there other factors you need to control or at least take into account?

> **Note**
>
> Animal Ethics conditions apply (see below).

Project 8 – Use of urban water bodies by birds: Do characteristics of urban water bodies influence which birds use them?

- **Background**: Bird diversity within cities is affected by many factors, including availability of food, sites for nesting and shelter from predators. Water bodies, such as ponds and lakes, in parks and gardens can attract native (and introduced) birds. The Sydney Council, for example, has upgraded and revegetated the ponds in Sydney Park, just south of Sydney Uni, to encourage water birds. To maximise bird diversity, as well as simply maximise bird numbers, it is important to understand the relationship between characteristics of different water bodies and the diversity of birds they support.

> **Note**
>
> Animal Ethics conditions apply (see below).

Animal Ethics requirements for Projects on birds

If you sign up to these projects, you must follow the animal ethics requirements as approved by the University of Sydney Animal Ethics Committee: Project "Ecology teaching – bird surveys". Project Number: 2022/2148.

You may collect data for up to 10 visits per site over a 2 to 4 week period. For the bird survey, you may observe birds, count them and identify each to species if you can, using binoculars where needed. You may observe birds at a site for up to ~ one hour per visit, but must not interfere with nor handle the birds.

At a given site, the surveyor (1 to 5 students per site) will find a suitable location to observe birds. They will stand or sit still for 15 to 30 minutes per location and count the birds they can see in their site. They may need to move to up to five other locations per site if they have difficulty identifying the birds from afar. The number of each species will be recorded with reference to bird guides, and a total of about one hour will be spent at each site per observation period.

Depending on your project, you may quantify other data, such as vegetation assessments and size of any habitat types including water bodies, at each site, enabling a test of the factors affecting bird abundance and species diversity.

Observers will remain far enough from birds that the birds are not visibly affected by them (i.e. do not alter their behaviour, either by moving away or by altering what they are doing). This avoids any potential impact of the observer on time or energy budgets of the birds (as described in Bateman et al. 2011).

Reference: Bateman, P. W., and P. A. Fleming. 2011. Who are you looking at? Hadeda ibises use direction of gaze, head orientation and approach speed in their risk assessment of a potential predator. Journal of Zoology 285:316-323.

Conditions of bird project approval:

1. You must use the approved monitoring sheet "Bird survey monitoring sheet.xls" which will be provided to you. You may add additional columns or have additional sheets, as you decide, appropriate to your study.

2. You must take photos of the fieldwork sites and the observers watching the birds (i.e. to see the relative locations of the observers to the birds).
3. By the end of Week 9, you must post on Canvas:

    a. A complete legible copy of the final monitoring sheets from (1)
    b. A summary of the total number of (i) native birds, (ii) invasive birds hence (iii) all birds you counted at each site and across all sites over the course of your project, and a list of the sites you monitored, with GPS coordinates.
    c. Clearly and informatively labelled photos from (2).

4. Injured birds: In the very unlikely event that birds are found injured on site: you must:

    a. seek advice from Clare McArthur (clare.mcarthur@sydney.edu.au, subject heading: "BIOL2022 URGENT - bird project") or other academic in the unit of study,
    b. while ensuring your own safety first, if necessary arrange to appropriately transport seriously injured animals to the nearest veterinarian for treatment, AND
    c. you must notify Clare McArthur by email within 24h, summarising what has occurred (this will be reported as an adverse event to the Animal Ethics Committee even if you played no part in the injury).

Report 1 – Instructions

About

- The written report is the individual component of this assessment It is submitted via the Canvas assignment called Report 1.
- The other component of this assignment is the data file, which should be submitted by one group member via the Canvas assignment called Report 1: Group data.
- Date due: Week 10, 1 pm on the day of your prac session.
- Late penalty: you will incur a penalty of 5% (5 points out of 100) for each 24 hours or part thereof after the deadline (weekends and public holidays included), unless you have Special Consideration.
- Length penalty: you will incur a penalty of 5% (5 points out of 100; up to 30 points) for each page or part thereof over the page limit.

A note about plagiarism

Do not plagiarise or you will face penalties. Check the academic integrity website for access to:

- the policies and codes covering academic honesty and conduct at the University
- a link to the AHEM (Academic Honesty Education Module); complete the module if you haven't already done so. You cannot submit your report if you have not completed this module
- If you are still unsure about how to avoid plagiarism, check with Clare McArthur well before submitting your report.
- Your report will be checked for plagiarism using software.

Submitting your individual report

- **Submitting your anonymous report for marking**:
  - ONE electronic document through Canvas. As this needs to be anonymous, do not include your name anywhere in the report.
  - Use a filename of the form "Day-Time-Project-Your SID", e.g. "WED2-4pmProj4A-Your SID"

- **Using your report to help other students**: so we can provide useful feedback using real examples, your report or parts of it may be used as (anonymous) examples for current and future students. If you wish to opt out of this, please make a statement to this effect on your report.
- **Page limit**: maximum 10 pages double-spaced, including everything (i.e. abstract and references and everything else)
- **References**: minimum of 10 key references from the primary scientific literature, probably a mix of key recent and older papers. They must be relevant, demonstrating that you have read about the topic comprehensively. You will use them to show how you understand what people have already done and how your work fits into and/or extends what is already known, and that you have used the references to help understand what you have found.

Requirements for your report

- Use minimum of 12 pt font and no less than 2 cm margins all around.
- You must write all your report yourself, in your own words, on the experimental research project you have done.
- Your report should follow the format and general content of a **real scientific article**. To get a feel for what we expect and how to present each part of your report, read the published literature in your subject area.
- Choose a relevant journal style and follow it (but don't use "newsprint" columns); e.g. note that figure headings are below and table headings are above the figures/tables.
- We expect information from the primary literature; i.e. academic peer-reviewed sources, particularly scientific journals. Avoid websites: they usually lack quality control, and they come and go. I recommend that you use: Web of Science (and Google Scholar) as your main portal for accessing journal articles, searched using key words and EndNote, Mendeley or Zotero for organising your references (but it is not essential). Consult the Library for short courses on how to use these if you don't know already.
- **If you are uncertain about anything seek help early from the teaching staff in class.**

Clare's tips on writing your report

A great reference on how to write well:

Mensh B, Kording K (2017) Ten simple rules for structuring papers. PLoS Comp. Biol. 13. doi: 10.1371/journal.pcbi.1005619

START EARLY, START EARLY, START EARLY: you can start to draft your introduction as early as Week 5, once you have read the literature on the topic and have a feel for what your experiment will cover. Don't try to make it perfect – just get some words down. You can also start to draft your Materials and Methods section once you have started your experiment and before Week 7. You should be able to draft your result section after Week 7 and then finalise your whole report, including your discussion, in Week 8 and before submitting in Week 9.

Title: keep it brief, relevant, informative and interesting

Abstract

- Brief abstract (max. 120 words), including (1) background (providing the context for your problem/project), (2) hypotheses or aims, (3) what you did, (4) what you found, (5) what it means, hence could be five sentences
- Do NOT include references or details of statistical results (e.g. no F or P values)

Introduction

- What do we know already? Make sure it leads logically to your aim(s). Provide the context - previous studies on the subject, using published (mainly) primary scientific literature, NOT websites. Integrate information clearly and concisely without loss of original meaning, to provide a logical, well-interpreted review of existing research and other information.
- Aim of the experiment and biological hypotheses (IF…) and predictions (THEN…)
- Why is the study important and/or interesting
- Study system – species, location of experiment (brief)
- Definitions of scientific terms (if needed).

Materials and Methods

- Pilot study (if needed, keep it brief)
- Set-up of experiment
- Sampling design
- Data / variables recorded and how

- Statistical analyses used, including which are dependent/independent variable(s) (if applicable), unit of replication, actual test(s) & why they are appropriate, statistical package(s)

Results

- Written statements describing results (data and stats). Refer to tables and figures as needed. Do NOT simply provide the tables and figures without supporting text.
- Summary presentation of data (no raw data) as tables, figures or in text as appropriate. Use tables for large amounts of data where detail is important, use figures to illustrate patterns. Indicate sample size and errors, and note whether the latter are standard errors or standard deviations... Format consistently. Make sure decimal places reflect your real capability to measure (e.g. can you really measure a 1 kg plant to 6 decimal places?)
- Results of statistical analyses presented either in text, on figures, or in tables.

Discussion

- Main conclusions or findings
- Interpret results – what do they mean? How do they relate to original aims?
- Compare with and discuss in context of previously published work
- Every single study has some limitations to its approach or methods. This does NOT mean it is necessarily flawed. Take a positive approach to your research, given the usual constraints of ANY study, and if you do want to mention/discuss these constraints, do not reduce your discussion to a lengthy apology.
- Future studies – what would you do next to extend our understanding of the subject and/or improve what you did? Keep this brief.

General

- Write in PLAIN English
  - Be succinct and avoid trying to sound "scientific"
  - read it out aloud, if it sounds pompous it is, and it should NOT be
- Correct presentation (format/length)
- References used and cited correctly. You are expected to use at least 10 relevant references from the primary literature

More tips on writing your report

The following is a summary written by Clare's Lab Group of PhD and Honours students after critically discussing two papers — not for content — but for how they were written.

Structuring Research Papers: by Clare's Lab Group

"Rules" based on Mensh B, Kording K (2017) Ten simple rules for structuring papers. PLoS Comput Biol 13(9): e1005619. https://doi.org/10.1371/journal.pcbi.1005619

Papers discussed:

> Cline, B. B. and Hunter, M. L. (2014), Different open-canopy vegetation types affect matrix permeability for a dispersing forest amphibian. J Appl Ecol, 51: 319-329. doi:10.1111/1365-2664.12197
>
> Palmer, M. S., Fieberg, J. , Swanson, A. , Kosmala, M. and Packer, C. (2017), A 'dynamic' landscape of fear: prey responses to spatiotemporal variations in predation risk across the lunar cycle. Ecol Lett, 20: 1364-1373. doi:10.1111/ele.12832

General notes

- Do write for your reader, do not write for yourself
- If you read the first sentence of every paragraph, it should tell a story
- The order of your ideas in each section should be the same; i.e. consistent order in intro, MM, R and probably D of your aims, methods for each aim (provided that works), results and discussion of your results associated with each aim
- Subheadings can make sections easier to read and understand the flow of ideas
- Avoid zigzagging ideas, i.e. you should never need to say "as mentioned above"
- Use consistent wording – DO NOT use a variety of words to mean the same thing because it is confusing for readers. They will not know if you are referring to one thing or several. DO: pick one name for a variable and stick to it.
- Take the reader by the hand. DO NOT assume they know what you know. DO lead them through your work and explain things as you go along. Justify your

decisions. Think like a reviewer: what will they want to know? Pre-empt it by being transparent, explaining what you decided to do and why.

Introduction

- There will generally be more references in the intro than the discussion. The focus of the intro is on: what do we know already?
- References are used to provide background of the research area – do not just list references on a topic. Instead, use them to summarise what we know and understand as a result of these studies. Make it clear what we know from (a) empirical evidence (with examples, i.e. do we know it for lots of individuals/species/contexts/ecosystems or whatever? which? Just one? which?) OR (b) what we think we know (conceptual hypotheses).

Suggested Paragraph 1:

- Introduce the broad topic
- Describe what research has been done on this topic

Suggested Paragraph 2:

- Start to narrow down what area in this topic your paper will be covering Introduce gaps in knowledge Possibly describe issues with previous studies (that your paper will address)
- This information can show the importance of your paper

Suggested Paragraph 3 to 5:

- Specifically describe what your paper will investigate
- State your aims and hypotheses/predictions. These should follow and funnel down from your background. If you put your hand over your aims, you should be able to guess them from the background you provide above
- Briefly describe how you investigated the topic
- State any predictions with information to backup these predictions

Discussion

- The focus is now on what you found, what it means and so how you have advanced our state of knowledge & understanding as a result of what you have done.
- Therefore: generally fewer references in the discussion than the introduction
- References should be used to support your idea or relate your conclusions/results to previous studies

- The following paragraph order makes sense but you can obviously increase or decrease the number of paragraphs depending on the context and size of your study.

Suggested Paragraph 1:

- State any broad results that you found
- Provide evidence from your results that enabled you to come to that conclusion
- Compare the result to any predictions you made or previous studies

Suggested Paragraph 2 to 4:

- State more detailed results (e.g. species specific results if dealing with multiple species)
- Provide evidence from your results that enabled you to come to that conclusion
- Compare the result to any predictions made or previous studies

Suggested Paragraph 5:

- Limitations of your study
  - No study can answer everything and every study is limited in some way in what is achieves! Don't worry! That's life!
  - Don't dwell on this or go too much in depth – especially do not be apologetic.
  - Be positive: just outline the boundaries of the conclusions that can be made with your results
- Suggested direction of future studies i.e. what could be done next to move forward in our understanding and/or resolving any unanswered questions given what you have now discovered – again keep this brief

Suggested Paragraph 6:

- What are the implications of your study for your field of research
- Describe what this information can affect or be used for (e.g. management decisions)

Suggested Paragraph 7:

- You do not need a concluding paragraph – but if you do, it is not just a summary.
- It should be more big picture stuff but take care: it should not state something that could have been said whether or not you had done your study. It must rest on your study.

Module 3 (Weeks 9 – 13)

Preamble

The aim of this component of the practical series is to introduce you to several common techniques for analysing multivariate data. At the end of them you will:

- Be familiar with the principles underpinning principal components analysis (PCA), and non-metric multidimensional scaling.
- Be familiar with the principles underpinning multivariate hypothesis testing using permutational techniques such as ANOSIM and PERMANOVA.
- Planning and conducting experiments to test multivariate hypotheses.
- Know how to perform these analyses using relevant software.
- Be able to interpret, present and report on these analyses.

Background

Multivariate analyses are becoming more prevalent in biology as tools performing the necessary complex computations become more accessible. As the lectures outline, this accessibility has given rise to the uncritical use of methods leading to inappropriate analyses being used to deal with multivariate data sets. At the very least, biologists need to have a working understanding of how to choose appropriate tests and how to identify when someone has chosen poorly.

Some of the more common uses of multivariate approaches permit us to describe complex patterns simply, identify patterns of groupings in complex data sets and test statistical hypotheses using multivariate datasets.

In this section of the unit, you will be able to practice performing these multivariate techniques using sample data sets. Then you need to demonstrate your ability to use these multivariate methods to analyse data that you will collect in groups to highlight the nuances of different techniques. Whilst best practice normally dictates that we would normally decide on appropriate tests before we collect our data, this task will enable us to explore options for analysis more effectively.

Analyses and programs we'll use

The main programs we will use are **SPSS** and **PRIMER**. However, R instructions for all methods, in the form of R scripts, will be supplied on Canvas for those who prefer to use R.

- **PCA** – Principal components analysis lets us extract a reduced dataset from a larger dataset, accounting for as much variance as possible in the original more complex dataset. This will be useful for your habitat data as we can determine a reduced number of newly derived variables to describe them. We will do this in SPSS (or R).
- **nMDS and multivariate hypothesis testing (using ANOSIM, PERMANOVA and SIMPER)** – Mainly used to produce an ordination in order to visualize patterns in the multivariate cloud of data (by reducing dimensions in some way) and also to rigorously test explicit hypotheses concerning those patterns by reference to a priori groups or relationships with predictors, such as environmental variables. We will use PRIMER for this (or R).
- **ANOSIM** – testing hypotheses comparing different groups based on their similarity (Among c.f. within) using distance measures. Most effective with simple designs. We will use PRIMER for this (or R).
- **PERMANOVA** – more powerful approach to multivariate hypothesis testing using permutational methods, capable of analysing more complex designs. We will use PRIMER for this (or R).
- **SIMPER** – used when a significant difference is found to identify which components of the data set are driving the differences (e.g. the contribution each species makes to differences among communities. We will use PRIMER for this (or R).

Timeline

- **Week 9**: introduction, PCA of lecture data and designing model systems to test multivariate hypotheses (in groups) for your reports.
- **Week 10**: group presentations of model systems and experimental design, and how to perform nMDS, PERMANOVA (and ANOSIM) and SIMPER using sample data.
- **Week 11**: submitting your complete datafiles onto Canvas, analysing your report data.
- **Week 12**: additional analyses and summary of findings and guidelines for describing multivariate methods and results for your report.
- **Week 13**: final report due.

Week 9 – Part 1

Objectives

The aim of this component of the practical series is to introduce you to several common techniques for analysing multivariate data. At the end of them you will:

- Be familiar with the principles underpinning principal components analysis (PCA), and non-metric multidimensional scaling.
- Be familiar with the principles underpinning multivariate hypothesis testing using permutational techniques such as ANOSIM and PERMANOVA.
- Planning and conducting experiments to test multivariate hypotheses.
- Know how to perform these analyses using relevant software.
- Be able to interpret, present and report on these analyses.

Outline

For this specific practical you will:

- Use PCA to analyse "lecturer perceptions" data gathered in previous years.
- Design an investigation using a model system to test multivariate hypotheses (in groups) for your reports.

What you need to submit in Module 3

The assessment for Module 3 requires two separate submissions in Canvas:

1. **The Excel file (spreadsheets) containing your group's data.** This needs to be submitted by one group member on behalf of your group and will cover both the species data and habitat data. This must be submitted by 8am on the day of your prac session in Week 11 (worth 5% of the total unit mark).
2. **Report 2: your individual report on your study.** The report must be submitted in the format of a short scientific article (worth 15% of the total unit mark) – see assessment information for when this is due.

Analyses covered in this practical

- **PCA** – Principal components analysis is essentially a way to extract a reduced set of variables from a larger dataset, accounting for as much variance as possible in the original more complex dataset.

Part 1: Principal components analysis – "perceptions of biology lecturers"

You will use PCA to reduce your set of measured variables into a coherent smaller dataset. In addition, you will be comparing your lecturer perceptions between preferred systems (marine, freshwater or terrestrial), preferred taxon (animals or plants) and gender (male or female) using conventional univariate approaches (t-tests/ANOVAs). To perform these univariate approaches, make sure that you not only create the principal components (PCs) but also save the scores (see below) for each component.

The process in SPSS is listed below. R instructions are given on Canvas in an R Studio file

1. Running the PCA

a. Import your lecturer data from excel into SPSS. On the command bar, click Analyze, Dimension Reduction, Factor. Move the variables of interest into the Variables Box.
b. Click Descriptives and then check Initial Solution, Coefficients, and KMO and Bartlett's Test of Sphericity. Click Continue.
c. Click Extraction and then select Correlation Matrix, Unrotated Factor Solution, Scree Plot, and Eigenvalues Over 1. Click Continue.
d. Click Rotation. Select Varimax and Rotated Solution. Click Continue.
e. Click Scores and select Save Scores.
f. Click Options. Select Exclude Cases Listwise and Sorted By Size. Click Continue.
g. Click OK, and SPSS completes the Principal Components Analysis.
h. Take a good look at the correlation matrix. The PCA captures the essence of the correlations in this matrix. If there are any variables that are not correlated with the other variables, you might as well delete them prior to the PCA. Since we are using PCA to reduce the set of variables to a smaller set of components to be used in additional analyses, you can always reintroduce the unique variables (i.e. those not correlated with other variables) at that time.

i. Bartlett's Test of Sphericity, can be used to test the null hypothesis that our sample was randomly drawn from a population in which the correlation matrix was an identity matrix (a matrix full of zeros, except, for ones on the main diagonal). An identity matrix would occur only if each variable only correlated with itself and not any other variable. However, it is easier and more interpretable if you look at your data instead.

2. Interpreting the PCA

You need to:

a. Look at the correlation matrix.
b. Examine the scree plot and determine how many PCs you can identify (eigenvalues >1, Kaiser's Criterion).
c. Establish a meaningful name for each component based on the loadings for individual variables in the rotated solutions.

3. After the PCA

Examine whether gender, system biases (marine v. terrestrial) and taxon (animal v. plant) biases may affect the perceptions of biology lecturers. You will need to code the columns for each factor (i.e. males = 0, females = 1) for SPSS to perform its standard one-way ANOVAs or t-tests.

Questions to consider:

- Do males and females perceive lecturer qualities the same way?
- Do botanists and zoologists perceive lecturer qualities the same way?
- Do marine and terrestrial folk perceive lecturer qualities the same way?

Part 2: designing your multivariate experiment

The data and your model system

Biologists are becoming more interested in multivariate approaches because the questions we ask and the data we collect tend to be inherently multivariate, relating to both biotic and abiotic variables when dealing with assemblages and habitat assessments respectively. In groups, you will generate your own multivariate datasets using non-biological model systems like cars and the car parks they inhabit, gargoyles and the buildings they live on, or beer types and the type of drinking establishment in which they congregate.

An example of a non-biological model system

Cars have proven a useful vehicle for understanding analytical techniques in ecology (Gaston et al. 1993) and they are a good model system for our needs. We can study the assemblages of different cars (as analogues of "species") that inhabit different suburbs and look at some characteristics of their car parks (as analogues of "habitat use"). See the sample proforma attached illustrating how we can generate a multivariate system with testable hypotheses using car assemblages and the car parks they inhabit as an example.

Groups

In groups, you will need to create your own non-biological multivariate system. Using this model system, come up with several (2 to 3) multivariate hypotheses relating to the assemblages of "species" in different "treatments" (2 factor design is encouraged) and the "habitat variables" that may be influencing "habitat use". Make sure that your treatments are adequately replicated (you will need at least 3 of each treatment). You will need to make a sampling technique in your groups that will allow you to collect the data necessary to test your hypotheses. You should also decide who will co-ordinate the data preparation.

Each group will present a short (<5 min) presentation during the practical session in Week 11 (week 2 of the pracs). For this, you will have a short powerpoint presentation containing the following information. This is compulsory for this unit: This enables staff to give feedback so you can collect the appropriate data for analyses in the next 2 weeks. Note this feedback is meant to be friendly, we are not upset or angry if we point out flaws in your design.

- an introduction to your system (model species and your treatments);
- your hypotheses (there should be 2 or 3);
- a description of your taxonomy;
- your sampling design (including information on sampling effort and sample sizes);
- the habitat variables you will be measuring at each site (and how you will measure them);
- a map of your sites

Assessment

To make sure that you get the most time to analyse the data and understand its interpretation, you will bring your data in week 11 to be checked by demonstra-

tors. This is a group assessment so only one member of your group needs submit the data file for the whole group. This is worth 5% of your assessment, with near perfect datasets getting the full 5%. It is a group assessment, so Special Considerations, Simple Extensions and Academic Plans do not apply.

For the final assessment, each student will write a short report. The emphasis on the report will be on the methods and results sections, reflected in the attached marking scheme. You do not have to use your imagination in regard to references/ reasons for the study, but are welcome to. You do need to include your hypotheses in the Introduction section.

The report will consider the work you did and the ways you analysed it. It must be **no longer than four single-spaced typed pages** (not including the cover sheet, a title page with the abstract or any pages of references, figures and tables) and will be in the form of a manuscript suitable for submission to Austral Ecology. This report will focus on the methods you used, **in particular how you describe your analyses**, the type(s) of data you collected and **how you present your results**.

Your report is due at **11:59pm on Sunday, before STUVAC. We hope to return reports with summative feedback a week later**. Your report should be submitted electronically, anonymously (use your SID for identification) through Canvas.

Do not plagiarise or copy from others:

- Check the website http://sydney.edu.au/students/academic-integrity.html for access to the policies and codes covering academic honesty and conduct at the University
- We will take marks off if you have obviously used ChatGPT. You are welcome to use it to check grammar, or even R codes, but we can tell if you have used it extensively to write your reports. The “Notice to Contributors” for Austral Ecology is available at http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1442-9993/homepage/ForAuthors.html

Template

At the end of this section is a simple template to organise your group’s study design. You can download it on the HTML version of the Unit Manual.

References

See annotated reading list with lectures posted on unit web site.

- Gaston, K. J., T. M. Blackburn, Lawton. J. (1993). Comparing Animals and Automobiles – a Vehicle for Understanding Body Size and Abundance Relationships in Species Assemblages. Oikos 66: 172-179.
- Quinn G.P. and M.J. Keough. (2002). Experimental design and data analysis for biologists. Cambridge: Cambridge University Press. (appropriate chapters)
- Quinn G.P. and M.J. Keough. (2023). Experimental design and data analysis for biologists. 2nd edn Cambridge: Cambridge University Press. (appropriate chapters)

Project title:

Name of group members:

1.	4.	7.	10.
2.	5.	8.	11.
3.	6.	9..	12.

Hypotheses:

1.

2.

3.

4.

Factor 1:	**Factor 2:**
a) b) c) d)	a) b) c) d)
Provide details of how you are DEFINING each factor.	Provide details of how you are DEFINING each factor.

Write down your list of sites. For each site write how it will be classified for each factor.

Sites	**Factor 1**	**Factor 2**
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		
11.		
12.		
13.		
14.		
15.		
16.		
17.		
18.		
19.		
20.		

Add more if necessary

Sampling method:

(describe your sampling method – describe how you will be surveying your sites)

Sampling effort:

(describe your sampling effort e.g. # transects, # quadrats, time based)

Species taxonomy:

(list or describe your species in enough detail that every member of your group will be able to clearly identify your taxa to species level).

List the habitat variables that will be collected at each site

Habitat variable	**How it will be measured.**
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	
11.	
12.	
13.	
14.	
15.	
16.	
17.	
18.	
19.	
20.	

Add more if necessary

Project Title: Travelling the road to success: Do shopping centres and sports fields support different car communities?

Name of group members:

1.	4.	7.	10.
2.	5.	8.	11.
3.	6.	9.	12.

Hypotheses:

1. The assemblages of cars in car parks surrounding sport fields will be different from the assemblages of cars in car parks near shopping centres.
2. The assemblages of cars found in local government areas (LGA) with high mean income will be different from assemblages of cars in LGA with low mean income. The assemblages of cars in LGA with medium mean income will not differ from the car assemblages in LGA with high or low mean income.
3. Mean scores of principal components derived from habitat variables will differ between car parks adjacent to sporting fields and car parks adjacent to shopping centres (More specific hypotheses pending PCA)
4. Principal components derived from habitat variables will predict car species richness and/or car abundance at each car park. (More specific hypotheses pending PCA)

Factor 1: Car park functionality	**Factor 2: LGA mean income**
a) Car parks near sporting fields b) Car parks adjoining shopping centres c) d)	a) Low b) Medium c) High d)
Car parks must have at least 25 marked carspaces and at least one side must be adjoining either a sporting field or a multilevel shopping centre.	LGA mean income will be defined using the ABS metric for average individual taxable income (taxable individuals only) for 2010. We define low mean income to be: <$60000 year; medium between $65000 and $90000; high >$95000

Write down your list of sites. For each site write how it will be classified for each factor.

Sites	**Factor 1**	**Factor 2**
Hornsby Westfield (Hornsby LGA)	Shop	Medium
Pennant Hills Park (Hornsby LGA)	Sport	Medium
Rogers Park, Woy Woy (Gosford LGA)	Sport	Low
Deepwater Plaza Shopping Centre, Woy Woy (Gosford LGA)	Shop	Low
Gordon Shopping Centre (Ku Ring Gai LGA)	Shop	High
Golden Jubilee Field, Wahroonga (Ku Ring Gai LGA)	Sport	High
Top Ryde Shopping Centre (Ryde LGA)	Shop	Medium
Meadowbank Park (Ryde LGA)	Sport	Medium
Parramatta Park (Parramatta LGA)	Sport	Low
Westfield Parramatta (Parramatta LGA)	Shop	Low
Boronia Park (Hunters Hill LGA)	Sport	High
Gladesville Shopping Centre (Hunters Hill LGA)	Shop	High
Woollahra Oval (Woollahra LGA)	Sport	High
Eastpoint food fair (Woollahra LGA)	Shop	High
Royal Randwick Shopping Centre (Randwick LGA)	Shop	Medium
Matraville Sports Centre (Randwick LGA)	Sport	Medium
Booralee Park (Botany Bay LGA)	Sport	Low
Westfield Eastgardens (Botany Bay LGA)	Shop	Low

Sampling method:

We will not survey cars in roadside parking, only in car parks with designated car spaces. If car parks are multi-leveled only the ground-level will be surveyed (for both species composition and habitat variables). Each site will be sampled between 10am and 12noon on a Saturday so that we can capture car assemblages associated with children's sport. Each group member will survey 2 car parks in one LGA (ie, one shopping centre and one sporting ground). We will only survey car species that are parked in designated car spaces within the car park.

Sampling effort:

A timed sampling effort will be used. We will survey each site (car park) for car species for 30mins. All habitat variables (except % free car spaces and # of shaded car spots) will be measured after the 30min sampling time.

Species taxonomy:

We will classify cars based on their make (i.e. Ford, Holden, Kia, Nissan, Ferrari, Toyota, Jeep, Subaru etc.) and type. Each car will be classified into 1 of 11 types:

1. 4WD
2. Van
3. sedan
4. station wagon
5. hatchback
6. Convertible
7. People mover (7+ seats)
8. Ute
9. Smartcar and/or mini
10. Sportscar
11. Other

Ford fourwheeldrive	Ford hatchback	Ford smartcar
Ford van	Ford convertible	Ford sportscar
Ford sedan	Ford peoplemover	Ford other
Ford stationwagon	Ford ute	

As an example, there will be 11 possible 'species' of car from the 'genus' Ford:

List the habitat variables that will be collected at each site

Habitat variable	**How it will be measured.**
total # car spaces	Total # of spaces in car park or on ground floor of car park (estimate for large car parks) (continuous)
% available car spaces	# of car spots that were empty during the 30 min survey period/(#free spaces + # individual cars counted in 30 min) x 100 (continuous)
# of lights	# of individual light sources (continuous)
# of trees/shrubs	# of individual trees or shrubs within the boundary of the car park (continuous)
# of bins	# of bins within the boundary of the car park (continuous)
# of exits	# of exits for the car park (whole car park) (continuous)
cost/hr	0 = free; 1 = x < $3; 2 = $3 < x < $5.99; 3 = $6 < x < $8.99; 4 = 9 < x where x is the average fee/hour (ignore timed free parking in shopping centres). (ordinal)
timed parking	1 = x < 15min; 2 = 15min < x <59 min; 3 = 1 h < x < 2 h; 4 = 2 h < x < 4 h; 5 = untimed where x is the maximum allowed parking time and/or time without payment (ordinal)
area of car park	Total area of the car park (m2) – measured from satellite images (GoogleEarth) (continuous)
% man made cover	Man –made cover only – includes roofs, shade covers, awnings etc. 0 = no cover; 1 = 1-25% cover; 2 = 26 – 50% cover; 3 = 51-75% cover; 4 = 76-99% cover; 5 = 100% cover. Estimate. (ordinal)
# of car spots in shade	# of car spots in the 30 min sampling time that have over 33% shade. (continuous)
Secure parking	Is the car park secure (i.e. boom gate/ticket gate/valet/security person present)? 0 = not secure' 1 = secure (presence/absence)
Glass on ground	Is there smashed glass on the ground? 0 = no, 1 = yes (presence/absence)
# of rubbish items	# of individual rubbish items found within the survey area of the car park. (Cigarette butts or small pieces of glass not included) (Continuous)

Add more if necessary

Week 10 – Part 2

Objectives

The aim of this component of the practical series is to introduce you to several common techniques for analysing multivariate data. At the end of them you will:

- Be familiar with the principles underpinning principal components analysis (PCA), and non-metric multidimensional scaling.
- Be familiar with the principles underpinning multivariate hypothesis testing using permutational techniques such as ANOSIM and PERMANOVA.
- Planning and conducting experiments to test multivariate hypotheses.
- Know how to perform these analyses using relevant software.
- Be able to interpret, present and report on these analyses.

Outline

For this specific practical you will:

- Present group presentations of model systems and experimental design.
- Analyse sample data in the software package PRIMER or R to:
  - visualize data using cluster analysis and nMDS.
  - test one and two factor multivariate hypotheses using ANOSIM and PERMANOVA with a focus on permutational techniques (PERMANOVA).
  - determine variables contributing to the differences between groups using SIMPER.

Analyses covered in practical 2

- **nMDS and multivariate hypothesis testing (ANOSIM, PERMANOVA and SIMPER)** – Mainly used to produce an ordination in order to visualize patterns in the multivariate cloud of data (by reducing dimensions in some way) and also to rigorously test explicit hypotheses concerning those patterns

by reference to a priori groups or relationships with predictors, such as environmental variables.
- **ANOSIM** — testing hypotheses comparing different groups based on their similarity (Among c.f. within) using distance measures. Most effective with simple designs.
- **PERMANOVA** — more powerful approach to multivariate hypothesis testing using permutational methods, capable of analysing more complex designs.
- **SIMPER** — used when a significant difference is found to identify which components of the data set are driving the differences (e.g. the contribution each species makes to differences among communities.

Community structure/Assemblage structure

The process for looking at community structure is straight forward. We will use both cluster analysis and nMDS/ANOSIM/SIMPER to do this. This involves;

- Transforming/standardising your data as appropriate.
- Calculating Bray Curtis dissimilarities (you could use something else from BC if the data are suitable) and generating dissimilarity matrix.
- Performing MDS (and follow on below).
- PERMANOVA (or ANOSIM) to test for difference between treatments.
- SIMPER for identifying which variables contribute to the differences.

Using PRIMER (R instructions are given on Canvas within an R Studio files)

PERMANOVA, SIMPER and MDS analysis with PRIMER

1. Open the Excel document you wish to analyse.
2. Open PRIMER
3. In PRIMER go to Open and select Excel file you wish to analyse.
4. Select worksheet to be analysed. Keep “Data Type” as Sample Data. Clear button “Title”.
5. Check to ensure you have the correct number of rows and columns and all looks well.
6. Go to Edit, go to Factors, go to Add and name your new factor(s) appropriately, code each site for that factor and press OK.

7. Go to Pre-treatment and then to Transform (overall), and select an appropriate transformation for your data. (4th root or presence/absence, etc[1].)
8. To make an nMDS plot: Go to Analyse then Resemblance and click 'analyse between samples' and 'Bray-Curtis similarity'. Go to Analyse then to Non-metric MDS, change number of restarts to 20 and select configuration plot. You can modify how the nMDS plot is displayed by right clicking on the figure.
9. To perform a PERMANOVA you need to firstly create a PERMANOVA design, and then run the PERMANOVA you have designed

    a. To create a PERMANOVA design: Return to the resemblance matrix. Go to PERMANOVA+ then Create PERMANOVA design. Create a title for the design and select appropriate number of factors. Enter in each factor you want to incorporate into the design, and code as either fixed or random factor. Complex designs can be created at this point.
    b. To run the PERMANOVA: Return to the resemblance matrix. Go to PERMANOVA+ then PERMANOVA. Select the correct design (usually the highest number). Select Mains test and your permutational method (reduced model is usually fine), the sum of squares type (type III) and the number of permutations (999 recommended) then OK.

10. Interpret the PERMANOVA output. If one or more of your factors are significant, then you run the PERMANOVA design again, selecting Pair-wise test and select the factor/interaction of interest. Press OK.
11. If appropriate, return to transformed data set and go to Analyse and then SIMPER. Select appropriate factor and click OK.
12. Save result(s) to a file.

ANOSIM

As an alternative to PERMANOVA, many studies will test differences between treatments using Analysis of Similarities (ANOSIM). Unlike PERMANOVA which can be used for multi-factorial designs, ANOSIM is best used for simple one-factor analyses. PERMANOVA and ANOSIM will generate the same (or very similar results) for one-factor analyses.

To perform an ANOSIM

Return to resemblance matrix, go to Analyse then to ANOSIM, select appropriate factor which you added before.

[1] When you are analyzing your group data (week 11), you will be able to apply multiple transformations to your data. I'd encourage you to use a range of them to see how it affects your data. Note that PRIMER will be reluctant to undertake any analyses when data are not transformed!

Significance level in % in ANOSIM becomes probability if divided by 100. e.g. 0.2 ÷ 100 = 0.002 = P (in this case P < 0.05). If 'Significance level of sample statistic' in 'Global Test' section of the output is more than 5 (which is > 0.05), than there is no significant difference among any of the treatments even if 'Pairwise Tests' section has significant differences between pairs of particular treatments. ANOSIM results are presented as: (P = 0.076, Global R = 0.148, 999 permutations) where P is a global significance level.

For next week

Our first steps involve organizing the data so that we can get it into our analytical software (SPSS, PRIMER, R etc,) and generating summary statistics and figures. Different analytical software requires slightly different formats but there are usually some generalities that apply across programs.

For the start of next week's practical each group will need to have collated their group data into three spreadsheets; These spreadsheets will be marked out of 5 by demonstrators.

1. Habitat data: an SPSS or R ready excel file with sites as rows and habitat variables as columns (see below)
2. Assemblage/Community data: a PRIMER or R ready excel file with species ID as rows and sites as columns (see below)
3. Summary statistics: additional excel files as above (1 and 2) but with summary statistics (numbers of spp., individuals, habitat variables etc.) calculated for your data. This may be done on a site specific or treatment specific basis depending on your question.

1. Your habitat data for PCA

Habitat data – You will need to enter your data on habitat characteristics into SPSS or R for analysis next week. Make sure you are familiar with how to organise it for SPSS. Every site needs its own row where every habitat variable measured is listed, i.e. every site assessment has an independent row. If you have treatments you need to include a column labeling those treatments too (in addition to "site name").

2. Organising and entering assemblage data for PRIMER

Assemblage/Community data – You will need to generate your spp*site matrix into a PRIMER ready file so that we are ready to analyse your data at the labs next

week. These will take the form of a spreadsheet with a single column identifying species (unique species identifiers are fine) with the abundances for each site in rows (page 22). This sheet must start in the top left cell of that worksheet with only a single row of column labels. Make sure that you keep a reference file where you have recorded the specific identity of the unique species identifiers (e.g. A= Toyota Corolla, B= Leyland P76 etc.).

3. Generating summary statistics for additional (univariate) analyses

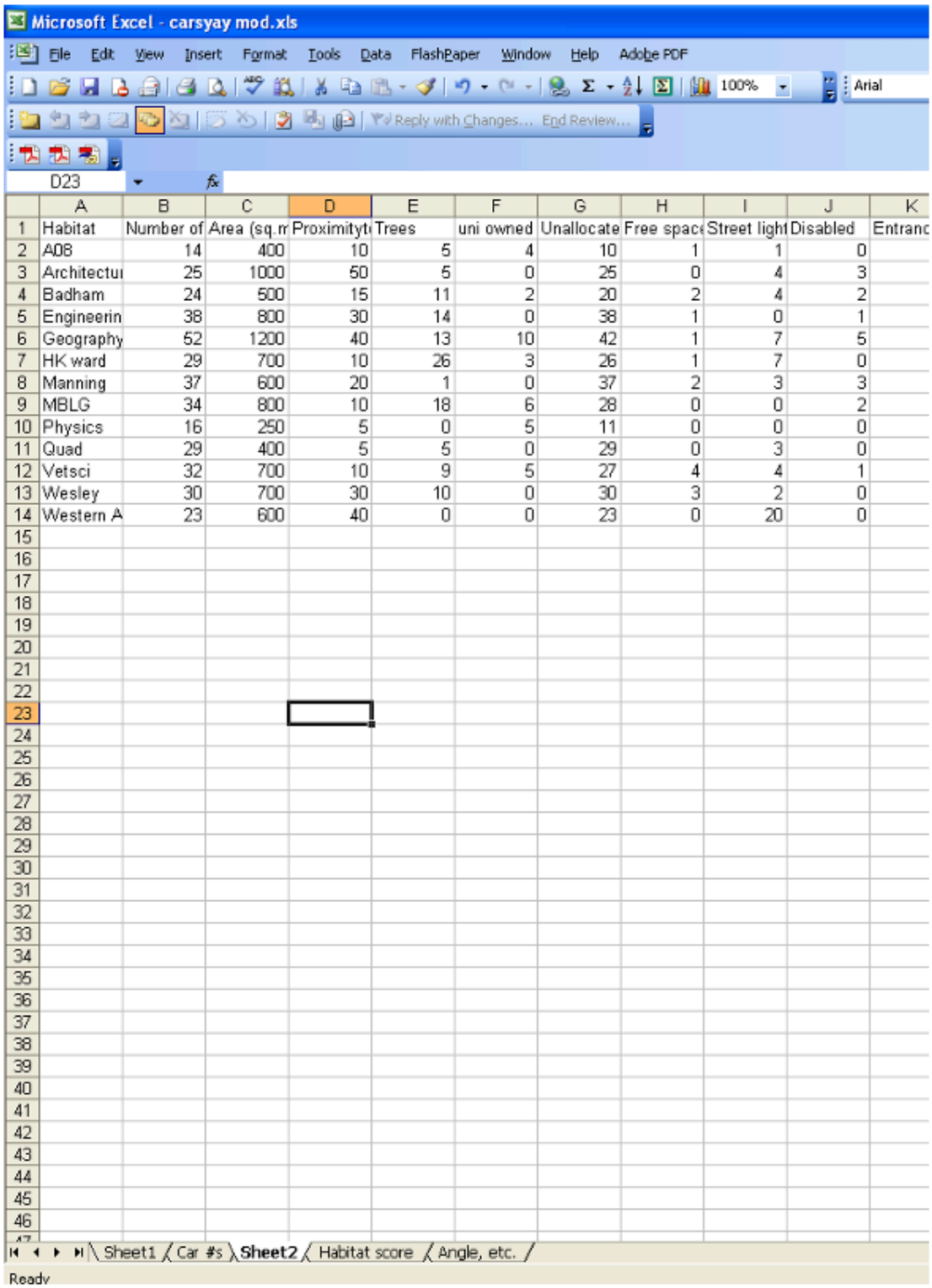
On an additional spreadsheet/workspace, you should calculate summary statistics to perform additional univariate analyses. Create a copy of your completed excel file and or worksheet. You should generate a column in the far right of your spreadsheet summing the number of things in each row. You can then sort the data (REMEMBER TO SELECT IT ALL!) by that column and delete all the zero rows (i.e. no evidence of that species in your system).

Owing to the high degree of variation and the number of singletons you will encounter it is possible (and likely) that some groups will pool species into more coarse groupings (i.e. “genera” or “family”).

The easiest way to calculate some of these summary statistics will be to use the pivot table command in Excel to sort your data for you. To do this:

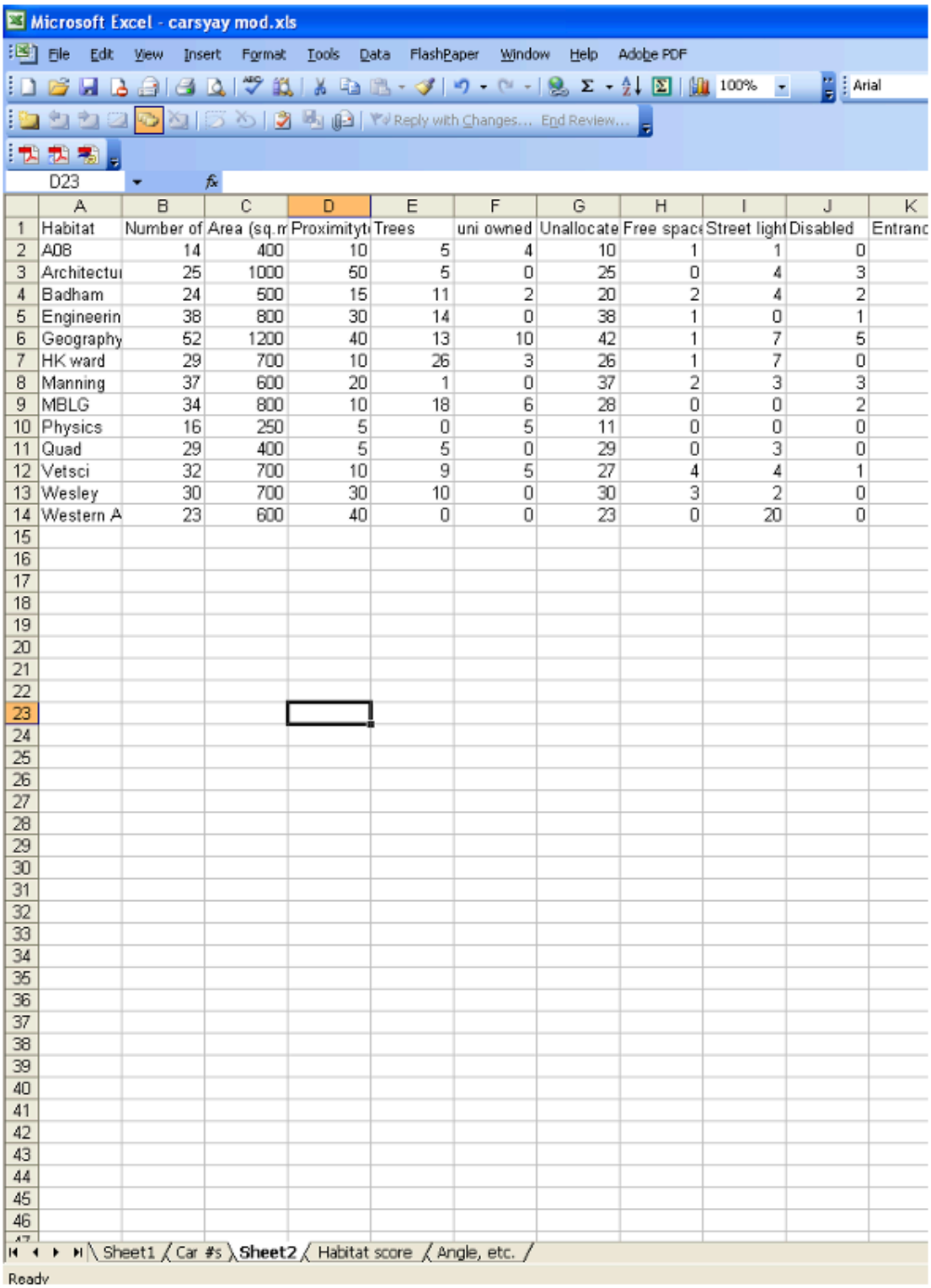
1. Highlight the dataset you want to use
2. Choose Pivot Table from the Data menu
3. Work through the wizard creating a pivot table in a new worksheet
4. Then change the drop down menu in the box to “to data area” and click on the different sites to add them.

You should have a functional summary pivot table of your data. You can change the pivot tables to look at “counts” (equivalent of presence absence data) or “sums” (abundance data) for your spreadsheet.



	A	B	C	D	E	F	G	H	I	J	K
1	Habitat	Number of	Area (sq.m	Proximityt	Trees	uni owned	Unallocate	Free space	Street light	Disabled	Entranc
2	A08	14	400	10	5	4	10	1	1	0	
3	Architectur	25	1000	50	5	0	25	0	4	3	
4	Badham	24	500	15	11	2	20	2	4	2	
5	Engineerin	38	800	30	14	0	38	1	0	1	
6	Geography	52	1200	40	13	10	42	1	7	5	
7	HK ward	29	700	10	26	3	26	1	7	0	
8	Manning	37	600	20	1	0	37	2	3	3	
9	MBLG	34	800	10	18	6	28	0	0	2	
10	Physics	16	250	5	0	5	11	0	0	0	
11	Quad	29	400	5	5	0	29	0	3	0	
12	Vetsci	32	700	10	9	5	27	4	4	1	
13	Wesley	30	700	30	10	0	30	3	2	0	
14	Western A	23	600	40	0	0	23	0	20	0	

Figure 1: Example of data sheet for importing data into SPSS for PCA



	A	B	C	D	E	F	G	H	I	J	K
1	Habitat	Number of	Area (sq.m	Proximityt	Trees	uni owned	Unallocate	Free space	Street light	Disabled	Entranc
2	A08	14	400	10	5	4	10	1	1	0	
3	Architectur	25	1000	50	5	0	25	0	4	3	
4	Badham	24	500	15	11	2	20	2	4	2	
5	Engineerin	38	800	30	14	0	38	1	0	1	
6	Geography	52	1200	40	13	10	42	1	7	5	
7	HK ward	29	700	10	26	3	26	1	7	0	
8	Manning	37	600	20	1	0	37	2	3	3	
9	MBLG	34	800	10	18	6	28	0	0	2	
10	Physics	16	250	5	0	5	11	0	0	0	
11	Quad	29	400	5	5	0	29	0	3	0	
12	Vetsci	32	700	10	9	5	27	4	4	1	
13	Wesley	30	700	30	10	0	30	3	2	0	
14	Western A	23	600	40	0	0	23	0	20	0	

Figure 2: Example of data sheet for importing data into PRIMER

Week 11 – Part 3

Objectives

The aim of this component of the practical series is to introduce you to several common techniques for analysing multivariate data. At the end of them you will:

- Be familiar with the principles underpinning principal components analysis (PCA) and non-metric multidimensional scaling.
- Be familiar with the principles underpinning multivariate hypothesis testing using permutational techniques
- Planning and conducting experiments to test multivariate hypotheses
- Know how to perform these analyses using relevant software
- Be able to interpret, present and report on these analyses. For this specific practical you will:
- Analyse and interpret your group data using multivariate techniques learnt in weeks 1 (PCA) and 2 (nMDS, PERMANOVA, SIMPER (if appropriate, etc).
- Perform analyses (ANOVA, regressions etc) to test additional univariate hypotheses.

Outline

At the end of the practical you should have:

1. Appropriate outputs for your community structure analyses:
    - Cluster analysis details and dendrogram (if appropriate)
    - nMDS (saved as suitable image file)
    - PERMANOVA results (in rtf/doc file)
    - SIMPER if justified
2. Appropriate outputs for your PCA:
    - RTF file with SPSS or R output
3. Univariate tests using your PCA results

- RTF file with SPSS or R output with Regression and/or ANOVA results (testing hypotheses using your newly derived PCs)

Your own data

At this stage it's very easy to fiddle with your data – is there any reason you might want to standardise or transform your data in any way? Some things to consider at this stage:

- Do the abundances matter?
- Are you intending to make every sample equal?
- Would it be worth analysing using your data at a higher "taxonomic" level?

Tips for PRIMER

The graphics are not fantastic – for some outputs you may need to save the coordinates and make a new graph. You can save the images as a windows metafile or bitmap if you wish, they are good enough for your report. Resolution is too low for most journals. You should save appropriate output files using standard formats you can incorporate into your report.

If you don't complete your analyses, I will make the program available to you outside of prac times. There will also be time in Week 12 to do any extras.

Week 12 – Support session

Objectives (you should know these by now!)

The aim of this component of the practical series is to introduce you to several common techniques for analysing multivariate data. At the end of them you will;

- Be familiar with the principles underpinning principal components analysis (PCA), and non-metric multidimensional scaling.
- Be familiar with the principles underpinning multivariate hypothesis testing using permutational techniques
- Planning and conducting experiments to test multivariate hypotheses
- Know how to perform these analyses using relevant software
- Be able to interpret, present and report on these analyses.

Outline

This week we will be in the lab to help with any analyses that groups wish to undertake. This may involve integrating the analyses (e.g. testing the relationships between PCs and univariate measures) or tweaking (not twerking) your analyses by examining subsets of your data using the options available in PRIMER.

Marking scheme

The marking scheme can be found on Canvas, in the Assignments section under **Report 2**.