# Statistical Techniques for Biological and Environmental Sciences

Brad Duthie

2022-08-29

# Contents

*Contents*

Contents

# IX. Counts and Correlation 135

# Week 9 Overview 137

# 40. Frequency and count data 139

# 41. Chi-squared goodness of fit 141

# 42. Chi-squared test of association 143

# 43. Correlation key concepts 145

# 44. Correlation mathematics 147

# 45. Correlation hypothesis testing 149

# 46. *Practical*. Analysis of count data, correlation, and regression 151

# X. Linear Regression 153

# Week 10 Overview 155

# 47. Regression key concepts 157

# 48. Regression validity 159

# 49. Introduction to multiple regression 161

# 50. Model selection (maybe remove this?) 163

# 51. *Practical*. Using regression 165

# Preface

Welcome to the module. This workbook will be used throughout the semester and contain all of the information that you need for the statistical techniques (SCIU4T4) module.

## 0.1. What is statistics?

An explanation of the material, and what will be taught.

## 0.2. Why this module is important

Some discussion of module importance

## 0.3. Teaching overview

Here is how you will be taught, with online lectures, reading assignments, and face-to-face practicals.

## 0.4. Assessment overview

You will have one formative test and two summatitive tests. You will also have one mock exam and one exam exam.

### 0.4.1. Test 1F

Information about Test 1F

### 0.4.2. Test 1S

Information about Test 1S

### 0.4.3. Test 2S

Information about Test 1S

### 0.4.4. Mock Exam

Information about the mock exam

### 0.4.5. Exam

Information about the exam

## 0.5. Jamovi statistical software

Introduction to Jamovi, and why we are using it instead of other software.

## 0.6. Textbooks

Introduction to the primary textbook Learning statistics with jamovi, and a mention of other sources.

## 0.7. Canvas

How we will use Canvas, and how this book relates to it (Learning and Teaching content, where lectures, assessments, and discussions can be found).

## 0.8. Timetable

# Part I.

# Background mathematics and data organisation

# Week 1 Overview

In week 1, we will focus on a refresher of some necessary background mathematics for this module. We will then turn to the topic of how to organise data sets. We will then practice organising datasets and saving them in a usable format.

Week: 1 Dates: Suggested Readings: Textbook intro, Hadley's paper Assessments: Practice quiz Practical: Quick summary of topics covered

# 1. Background mathematics

Some of this will be review, but it is important. Suggested reading for this (some mathematics text).

## 1.1. Numbers and operations

A very broad reminder of mathematics, which you will need for this module

## 1.2. Order of operations

This is easy to forget

# 2. Data organisation

It is important to organise data properly so that statistical analysis can be done. Here I explain the tidy approach to data. Suggested reading Hadley Wickam's paper.

## 2.1. Tidy data

## 2.2. Data files

# 3. *Practical*: Preparing real datasets

In this practical, we will use a spreadsheet to organise datastes.

## 3.1. LibreOffice Calc

LibreOffice Calc is a free and open source spreadsheet program. The instructions for this section will be identical to the more popular commercial Microsoft Excel.

## 3.2. Exercise Organising data 1

Walks through Exercise 1.3.2

## 3.3. Exercise Organising data 2

Walks through Exercise 1.3.3

## 3.4. Exercise Organising data 3

Walks through Exercise 1.3.4, saving all of these as CSV files

## 3.5. Summary of exercises and why they are useful.

# Part II.

# Statistical concepts

# Week 2 Overview

| | |
|---|---|
| **Dates** | 30 January 2022 - 3 February 2022 |
| **Reading** | **Required:** SCIU4T4 Workbook chapters 4-8 |
| | **Recommended:** Navarro and Foxcroft (2022) Section 3.3-3.9 |
| | **Optional:** Rowntree (2018) Chapter 2 |
| | **Advanced:** Spiegelhalter (2019) Chapters 1-3 |
| **Lectures** | 2.0: Introduction to Week 1 (2 min.) |
| | 2.1: Why study statistics? (18 min.) |
| | 2.2: Populations and samples (7 min.) |
| | 2.3: Types of variables (11 min.) |
| | 2.4: Units, precision, and accuracy (9 min.) |
| | 2.5. Uncertainty propagation (11 min.) |
| **Practical** | Introduction to Jamovi |
| **Assessments** | Week 2 Practice quiz on Canvas |

In Chapter 4, we focus on key concepts that will be used throughout this module. In particular, it is important to understand the difference between a **population** and a **sample**, and to recognise that there are many types of variables in statistics.

In Chapter 5, we look at different variable types. Different types of variables have different characteristics, which will affect how these variables are best visualised in figures and analysed with statistical hypothesis tests introduced later in the semester. Identifying variable types will therefore be important, both for module assessments and when working with new datasets. A variable's type will rarely be stated explicitly when doing scientific research, and will not always be provided in assessments for this module. Being able to infer variable type is therefore an importat skill.

Chapter 6 focus on units of measurement, and how these units are communicated in text. Units are essential in scientific measurement, and we will use them throughout the module to indicate the type and scale of data measurement. We are not expecting you to memorise all scientific units, so a table on units is provided at the end of this workbook.

Chapter 7 will introduce the propogation of measurement errors. This is important to understand because no measurement is perfectly accurate, and predicting how errors in measurement combine is fundamental to understanding measurement accuracy.

Finally, Chapter 8 guides you through the Week 2 practical, which is an introduction to Jamovi. This aim of this practical is to become familiar with the Jamovi interface and comfortable importing data into Jamovi to collect some descriptive statistics.

# 4. Populations and samples

When we collect data, we are recording some kind of observation or measurement. If we are working in a forest, for example, we might want to measure the heights of different trees, or measure the concentration of carbon in the soil. The idea might be to use these measurements to make some kind of inference about the forest. But as scientists, we are almost always limited in the amount of data that we can collect. We cannot measure everything, so we need to collect a *sample* of data and use it to make inferences about the *population* of interest. For example, while we probably cannot measure the height of every tree in a forest, nor can we measure the concentration of carbon at every possible location in the forest's soil, we can collect a smaller number of measurements and still make useful conclusions about overall forest trees and carbon concentration.

Statistics thereby allows us to approximate properties of entire populations from a limited number of samples. This needs to be done with caution, but before getting into the details of how, it is important to understand the difference between a **population** and a **sample** to avoid confusing these two concepts. A **population** is the entire set of possible observations that could be collected. Some examples will make it easier to understand:

- All of the genes in the human genome
- All individuals of voting age in Scotland
- All common pipistrelle bats in the United Kingdom

These populations might be important for a particular research question. For example, we might want to know something about the feeding behaviours of pipistrelle bats in the UK. But there is no way that we can find and observe the behaviour of every bat, so we need to take a subset of the population (a sample) instead. Examples of samples include the following:

- A selection of 20 human genes
- A pub full of Scottish voters
- 40 caught common pipistrelle bats

It is important to recognise that the word "population" means something slightly different in statistics than it does in biology. A biological population, for example, could be defined as all of the individuals of the same species in the same general location. A statistical population, in contrast, refers to a set of observations (i.e., things that we can measure). Sokal and Rohlf (1995) provide a more technical definition for "population",

> In statistics, population always means the *totality of individual observations*
> *about which inferences are to be made, existing anywhere in the world or*
> *at least within a definitely specified sampling area limited in space and time*
> [p. 9, emphasis theirs].

They define a sample to be "a collection of individual observations selected by a specified procedure" (Sokal and Rohlf, 1995). For our purposes, it is not necessary to be able to recite the technical definitions, but it is important to understand the relationship between a population and a sample. When we collect data, we are almost always taking a small sample of observations from a much larger number of possible observations in a population.
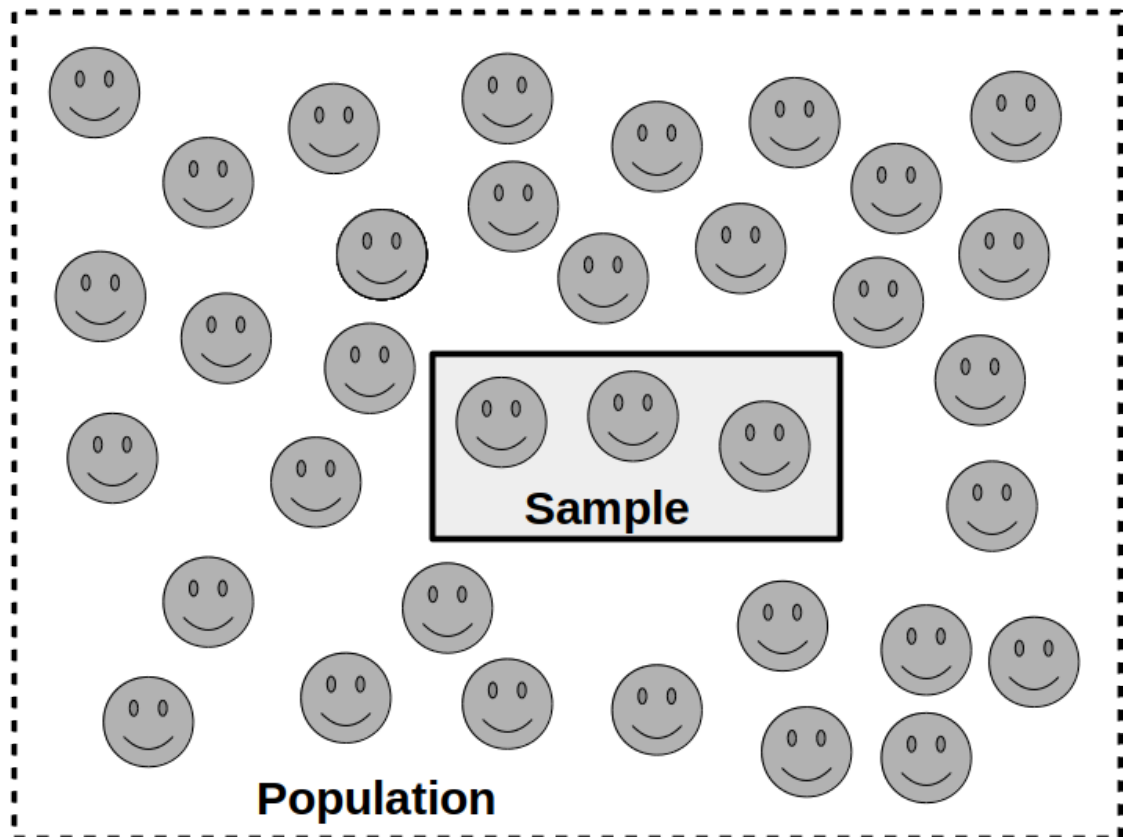


Figure 4.1.: A conceptual figure illustrating how a statistical population relates to a statistical sample. The population is represented by 35 smiling faces enclosed within a dashed box. The sample is represented by a solid box within the dashed box, within which there are 3 smiling faces. Hence, we have a sample of 3 measurements from the total population.

# 5. Types of variables

A variable is any property that is measured in an observation (Sokal and Rohlf, 1995); i.e., anything that varies among things that we can measure (Dytham, 2011). We can summarise how these measurements vary with summary statistics, or visually with figures. Often, we will want to predict one variable from a second variable. In this case, the variable that we want to predict is called the **response variable**, also known as the **dependent variable** or **Y** variable ('dependent' because it *depends* on other variables, and 'Y' because this is the letter we often use to represent it). The variable that we use to predict our response variable is the **explanatory variable**, also known as the **independent variable** or **X** variable ('independent' because it does not depend on other variables, and 'X' because this is the letter most often used to represent it). There are several different types of variables:

- **Categorical** variables take on a fixed number of discrete values (Spiegelhalter, 2019). In other words, the measurement that we record will assign our data to a specific category. Examples of categorical variables include species (e.g., "Robin", "Nightingale", "Wren") or life history stage (e.g., "egg", "juvenile", "adult"). Categorical variables can be either nominal or ordinal.

  - **Nominal** variables do not have any inherent order (e.g., classifying land as "forest", "grassland", or "urban").
  - **Ordinal** variables do have an inherent order (e.g., "low", "medium", and "high" elevation).

- **Quantitative** variables are variables represented by numbers that reflect a magnitude. That is, unlike categorical variables, we are collecting numbers that really mean something tangible (in contrast, while we might represent low, medium, and high elevations with the numbers 1, 2, and 3, respectively, this is just for convenience; a value of '2' does not always mean 'medium'). Categorical variables can be either discrete or continuous.

  - **Discrete** variables can take only certain values. For example, if we want to record the number of species in a forest, then our variable can only take discrete counts (i.e., integer values). There could conceivably be any whole number of species (1, 2, 3, etc.), but there could not be 2.51 different species in a forest; that does not make sense.

 – **Continuous** variables can take any real value within some range of values (i.e., any number that can be represented by a decimal). For example, we could measure height to as many decimals as our measuring device will allow, with a range of values from zero to the maximum possible height of whatever it is we are measuring. Similarly, we could measure temperature to any number of decimals, at least in theory, so temperature is a continuous variable.

The reason for organising variables into all of these different types is that different types of variables need to be handled in different ways. For example, it would not make sense to visualise a nominal variable in the same way as a continuous variable. Similarly, the choice of statistical test to apply to answer a statistical question will almost always depend on the types of variables involved. If presented with a new data set, it is therefore very important to be able to interpret the different variables and apply the correct statistical techniques (this will be part of the assessment for this module).

# 6. Accuracy, precision, and units

The science of measurements is called "metrology", which, among other topics, focuses on measurement accuracy, precision, and units (Rabinovich, 2013). We will not discuss these topics in depth, but they are important for statistical techniques because measurement, in the broadest sense of the word, is the foundation of data collection. When collecting data, we want measurements to be accurate, precise, and clearly defined.

## 6.1. Accuracy

When we collect data, we are trying to obtain information about the world. We might, for example, want to know the number of seedlings in an area of forest, them temperature of the soil at a location, or the mass of a particular animal in the field. To get this information, we need to make measurements. Some measurements can be collected by simple observation (e.g., counting seedlings), while others will require measuring devices such as a thermometer (for measuring temperature) or scale (for measuring mass). All of these measurements are subject to error. The *true* value of whatever it is that we are trying to measure (called the "measurand") can differ from what we record when collecting data. This is true even for simple observations (e.g., we might miscount seedlings), so it is important to recognise that the data we collect comes with some uncertainty. The **accuracy** of a measurement is defined by how close the measurement is to the *true* value of what we are trying to measure (Rabinovich, 2013).

## 6.2. Precision

The **precision** of a measurement is how consistent it will be if measurement is replicated multiple times. In other words, precision describes how similar measurements are expected to be. If, for example, a scale measures an object to be the exact same mass ever time it is weighed (regardless of whether the mass is accurate), then the measurement is highly precise. If, however, the scale measures a different mass each time the object is weighed (for this hypothetical, assume that the true mass of the object does not change), then the measurement is not as precise.

One way to visualise the difference between accuracy and precision is to imagine a set of targets, with the centre of the target representing the true value of what we are trying to measure (Figure 6.1)[1].
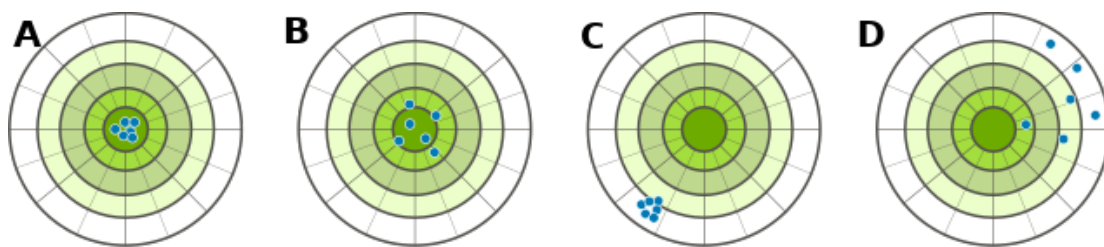


Figure 6.1.: A conceptual figure illustrating the difference between accuracy and precision. Points in (A) are both accurate and precise, points in (B) are accurate, but not precise, points in (C) are precise but not accurate, and points in (D) are neither accurate nor precise.

Note again that accuracy and precision are not necessarily the same. Measurement can be accurate but not precise (Figure 6.1B) or precise but not accurate (Figure 6.1C).

## 6.3. Systems of units

Scientific units are standardised with the Système International D'Unités (SI). Having standardised units of measurement is highly important to ensure measurement accuracy (Quinn, 1995). Originally, these units were often defined in terms of physical artefacts. For example, the kilogram (kg) was once defined by physical cylinder metal housed in the Buereau International des Poids et Mesures (BIPM). In other words, the mass of a metal sitting at the BIPM *defined* what a kg was, with the mass of every other measurement being based on this physical object (Quinn, 1995). This can potentially present a problem if the mass of that one object changes over time, thereby causing a change in how a kg is defined. Where possible, it is therefore preferable to define units in terms of fundamental constants of nature. In 2019, for example, the kg was redefined in terms of the Planck constant, a specific atomic transition frequency, and the speed of light (Stock et al., 2019). This ensures that measurements of mass remain accurate over time because what a kg represents in terms of mass cannot change.

We can separate units into base units and derived units. Table 6.1 below lists some common base units for convenience (Quinn, 1995). You do not need to memorise these units, but it is good to be familiar with them. We will use these units throughout the module.

---

[1]This figure was released into the public domain by Egon Willighagen on 8 March 2014.

Table 6.1.: Base units of SI measurements. For details see Quinn (1995).

| Measured Quantity | Name of SI unit | Symbol |
|---|---|---|
| Mass | kilogram | $kg$ |
| Length | metre | $m$ |
| Time | second | $s$ |
| Electric current | ampere | $A$ |
| Temperature | kelvin | $K$ |
| Amount of a substance | mole | $mol$ |
| Luminous intensity | candela | $cd$ |

We can also define derived SI units from the base units of Table 6.1; examples of these derived SI units are provided in Table 6.2 below. Again, you do not need to memorise these units, but it is good to be aware of them.

Table 6.2.: Examples of derived SI units.

| Measured Quantity | Name of unit | Symbol | Definition in SI units | Alternative in derived units |
|---|---|---|---|---|
| Energy | Joule | $J$ | $m^2\ kg\ s^{-2}$ | $N\ m$ |
| Force | Newton | $N$ | $m\ kg\ s^{-2}$ | $J\ m^{-1}$ |
| Pressure | Pascal | $Pa$ | $kg\ m^{-1}\ s^{-2}$ | $N\ m^{-2}$ |
| Power | Watt | $W$ | $m^{-2}\ kg\ s^{-3}$ | $J\ s^{-1}$ |
| Frequency | Hertz | $Hz$ | $s^{-1}$ | |
| Radioactivity | Becquerel | $Bq$ | $s^{-1}$ | |

When numbers are associated with units, it is important to recognise that the units must be carried through and combined when calculating equation. As a very simple example, if want to know the speed at which an object is moving, and we find that it has moved 10 metres in 20 seconds, then we calculate the speed and report the correct units as below,

$$speed = \frac{10\ m}{20\ s} = 0.5\ m/s = 0.5\ m\ s^{-1}.$$

Notice that the final units are in metres per second, which can be written as $m/s$ or $m\ s^{-1}$ (remember that raising $s$ to the $-1$ power is the same as $1/s$; see below for a quick reminder about superscripts). It is a common error to calculate just the numeric components of a calculation and ignore the associated units. Often on assessments, we will ask you not to include units in your answer (this is just for convenience on the tests and exam), but recognising that units are also part of calculations is important.

## 6.4. Other examples of units

Remember that an exponent (indicated by a superscript, e.g., the 3 in $m^3$) indicates the number of times to multiply a base by itself, so $m^3 = m \times m \times m$.

### 6.4.1. Units of density

Density ($\rho$) is calculated by,

$$\rho = \frac{mass}{volume} = \frac{kg}{m^3}.$$

The units of density are therefore mass per unit volume, $kg\,m^{-3}$.

### 6.4.2. Mass of metal discharged from a catchment

The mass of metal carried by a stream per unit time ($M$) is given by multiplying the concentration of metal per unit volume ($C$) of water by the volume of water discharged per unit time ($V$),

$$M = C \times V.$$

This equation is useful in showing how units can cancel each other out. If we calculate the above with just the units (ignoring numbers for $C$ and $V$),

$$M = \frac{mg}{l} \times \frac{l}{s} = \frac{mg}{s}.$$

Notice above how the $l$ units on the top and bottom of the equation cancel each other out, so we are left with just $mg/s$.

### 6.4.3. Soil carbon inventories

For one final example, the inventory of carbon $I$ within a soil is given by the specific carbon concentration $C$ ($g$ of carbon per $kg$ of soil), multiplied by the depth of soil analysed ($D$, $m$), and by the density ($\rho$, $kg\,m^{-3}$),

$$I = C \times D \times \rho = \frac{g \times m \times kg}{kg \times m^3} = \frac{g}{m^2} = g\,m^{-2}.$$

Notice above how the $kg$ on the top and bottom of the fraction cancel each other out, and how one $m$ on the top cancels out one $m$ on the bottom, so that what we are left with is grams per metre squared ($g\,m^{-2}$).

# 7. Uncertainty propogation

Nothing can be measured with perfect accuracy, meaning that every measurement has some associated error. The measurement error might be caused by random noise in the measuring environment, or by mistakes made by the person doing the measuring. The measurement error might also be caused by limitations or imperfections associated with a measuring device. The device might be limited in its measurement precision, or perhaps it is biased in its measurements due to improper calibration, manufacture, or damage from previous use. All of this generates uncertainty with respect to individual measurements.

Recall from Chapter 6 the difference between precision and accuracy. We can evaluate the precision and accuracy of measurements in different ways. Measurement precision can be estimated by replicating a measurement (i.e., taking the same measurement over and over again). The more replicate measurements made, the more precisely a value can be estimated. For example, if we wanted to evaluate the precision with which the mass of an object is measured, then we might repeat the measurement with the same scale multiple times and see how much mass changes across different measurements. To evaluate measurement accuracy, we might need to measure a value in multiple different ways (e.g., with different measuring devices). For example, we might repeat the measurement of an object's mass with a different scale.

Sometimes it is necessary to combine different values. For example, we might measure the mass of 2 different bird eggs in a nest, then calculate the total mass of all the 2 eggs combined. The measurement of each egg will have its own error, and these errors will propagate to determine the error of the total egg mass for the nest. How this error propagates differs depending on if they are being added or subtracted, or if they are being multiplied or divided.

## 7.1. Adding or subtracting errors

In the case of our egg masses, we can assign the mass of the first egg to the variable $X$ and the mass of the second egg to the variable $Y$. We can assign the total mass to the variable $Z$, where $Z = X + Y$. The errors associated with the variables $X$, $Y$, and $Z$ can be indicated by $E_X$, $E_Y$, and $E_Z$, respectively. In general, if the variable $Z$ is calculated by adding (or subtracting) 2 or more values together, then this is the formula for calculating $E_Z$,

$$E_Z = \sqrt{E_X^2 + E_Y^2}.$$

Hence, for the egg masses, the error of the combined masses () equals the square root of the error associated with the mass of egg 1 squared ($E_X^2$) plus the error associated with the mass of egg 2 squared ($E_Y^2$). It often helps to provide a concrete example. If the error associated with the measurement of egg 1 is $E_X^2 = 2$, and the error associated with the measurement of egg 2 is $E_X^2 = 3$, then we can calculate,

$$E_Z = \sqrt{2^2 + 3^2} \approx 3.61.$$

## 7.2. Multiplying or dividing errors

Multiplying or dividing errors works a bit differently. As an example, suppose that we need to measure the total area of a rectangular field. If we measure the length ($L$) and width ($W$) of the field, then the total area is the product of these measurements, $A = L \times W$. Again, there is going to be error associated with the measurement of both length ($E_L$) and width ($E_W$). How the error of the total area ($E_A$) is propagated by $E_L$ and $E_W$ is determined by the formula,

$$E_A = A\sqrt{\left(\frac{E_L}{L}\right)^2 + \left(\frac{E_W}{W}\right)^2}.$$

Notice that just knowing the error of each measurement ($E_L$ and $E_W$) is no longer sufficient to calculate the error associated with the measurement of the total area. We also need to know $L$, $W$, and $A$. If our field has a length of $L = 20$ m and width of $W = 10$, then $A = 20 \times 10 = 200$. If length and width measurements have associated errors of $E_L = 2$ m $E_W = 1$ m, then,

$$E_A = 200\sqrt{\left(\frac{2}{20}\right)^2 + \left(\frac{1}{10}\right)^2} \approx 28.3.$$

Of course, not every set of measurements with errors to be multiplied with be lengths and widths. To avoid confusion, the general formula for multiplying or dividing errors is below, with the variables $L$, $W$, and $A$ replaced with $X$, $Y$, and $Z$, respectively, to match the case of addition and subtraction explained above,

$$E_Z = Z\sqrt{\left(\frac{E_X}{X}\right)^2 + \left(\frac{E_Y}{Y}\right)^2}.$$

Note that the structure of the equation is the exact same, just with different letters used as variables. **It is necessary to be able to apply these equations correctly to estimate combined error.**

## 7.3. Applying formulas for combining errors

It is not necessary to understand why the equations for propagating different types of errors are different, but a derivation is provided in Appendix B for the curious.

# 8. *Practical.* Introduction to Jamovi

This practical focuses on learning how to work with datasets in Jamovi. Jamovi is available in the university laboratory computers through AppsAnywhere. You can also download it on your own computer for free or run it directly from a browser. For an introduction to what Jamovi is and why we are using it in this module, see Section 0.5 of this workbook and Sections 3.3-3.9 of Navarro and Foxcroft (2022). In this module, we will work with two datasets, both of which are based on real biological and environmental studies conducted by researchers at the University of Stirling.

The first dataset includes measurements of soil organic carbon (grams of Carbon per kg of soil) from the topsoil and subsoil collected in a national park in Gabon. These data were collected by Dr Carmen Rosa Medina-Carmona in an effort to understand how pyrogenic carbon (i.e., carbon produced by the charring of biomass during a fire) is stored in different landscape areas (Santín et al., 2016; Preston and Schmidt, 2006). **Download these data here: soil_organic_carbon.csv**

The second dataset includes measurements of figs from trees of the Sonoran Desert Rock Fig (*Ficus petiolaris*) in Baja, Mexico. These data were collected by Dr Brad Duthie in an effort to understand coexistence in a fig wasp community (Duthie et al., 2015; Duthie and Nason, 2016). Measurements include fig lengths, widths, and heights in centimeters from 4 different fig trees, and the number of seeds in each fruit. **Download these data here: fig_fruits.csv**

This lab will use the soil organic carbon dataset in Exercise 8.1 for summary statistics. The fig fruits will be used for Exercise 8.2 on transforming variables and Exercise 8.3 on computing a variable.

## 8.1. Exercise for summary statistics

Download the soil organic carbon dataset if you have not already done so, and save it in a location where you know you will be able to find it again, then open Jamovi. Once Jamovi is open, you can import the dataset by clicking on the three lines in the upper left corner of the tool bar, then selecting 'Open' (Figure 8.2).

You might need to click 'Browse' in the upper right of Jamovi to find the file. Figure 8.3 below shows how this will look when you browse for a data file.

*8. Practical. Introduction to Jamovi*
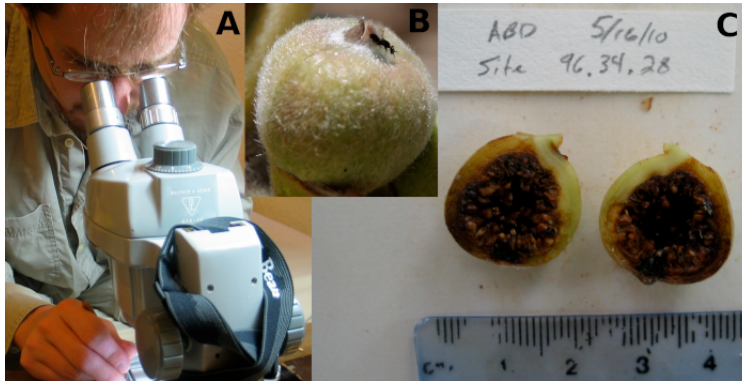


Figure 8.1.: Three images showing the process of collecting data for the dimensions of figs from trees of the Sonoran Desert Rock Fig in Baja, Mexico. (A) Processing fig fruits, which included measuring the diameter of figs along three diferent axes of length, width, and height, (B) a fig still attached to a tree with a fig wasp on top of it, and (C) a sliced open fig with seeds along the inside of it.
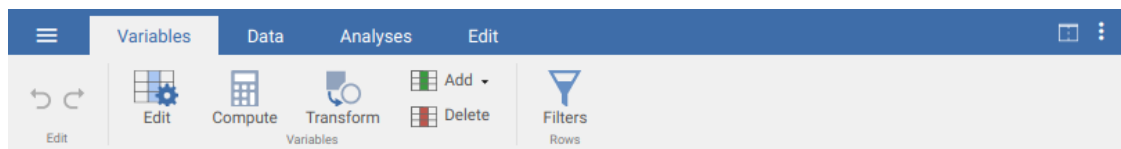


Figure 8.2.: The Jamovi toolbar including tabs for opening files, Variables, Data, Analyses, and Edit. To open a file, select the three horizontal lines in the upper right
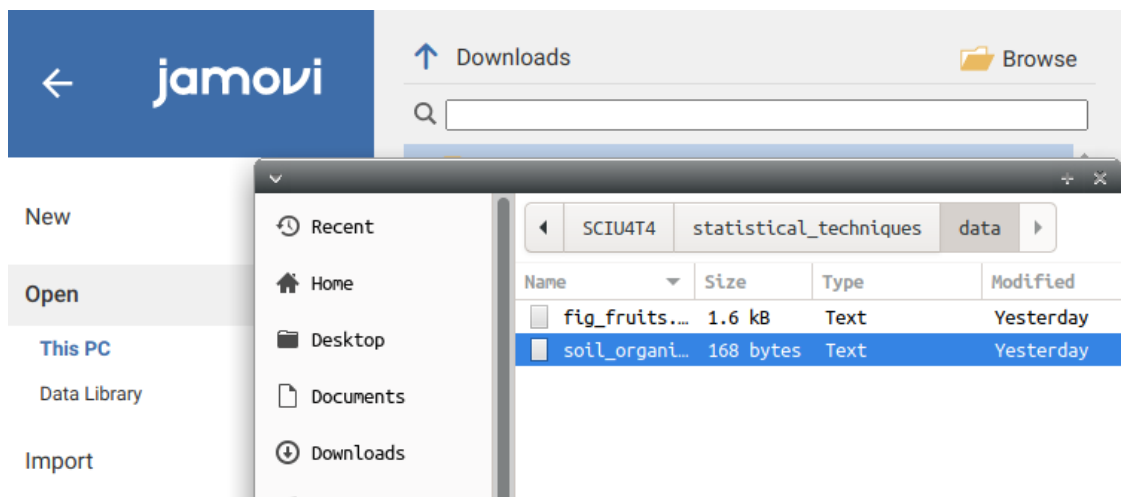


Figure 8.3.: The Jamovi interface for opening a file with the 'Import' tab selected. Options for browsing to a file on the computer are available in the upper right, which opens the window in the foreground.

38

Once the data are imported, you should see two separate columns. The first column will show soil organic carbon values for topsoil samples, and the second column will show soil organic carbon values for subsoil samples. Notice right away that these data are not formatted in a tidy way. We need to fix this so that each row is a unique sample, and each column represents some property of the sample (see the previous chapter section on data organisation for a reminder of why). It might be easiest to reorganise the data in a spreadsheet such as LibreOffice Calc or Microsoft Excel. But you can also edit the directly in Jamovi by clicking on the 'Data' tab in the toolbar (see Figure 8.2). The best way to reorganise the data in Jamovi is to double-click on the third column of data next to 'subsoil' (see Figure 8.4).
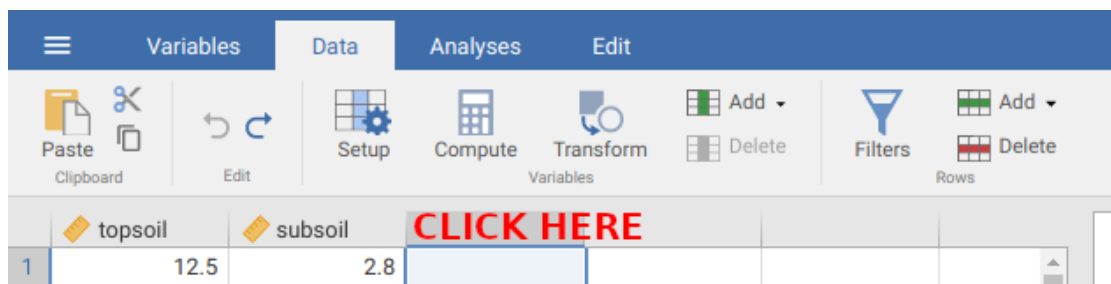


Figure 8.4.: The Jamovi toolbar is shown with the soil organic carbon dataset. In Jamovi, double-clicking above column three where it says 'CLICK HERE' will allow you to input a new variable.

After double-clicking on the location shown in Figure 8.4, there will be three buttons visible. You can click the 'New Data Variable' to insert a new variable named 'soil_type' in place of the default name 'C'. Keep the 'Measure type' as 'Nominal', but change the 'Data type' to 'text'. When you are done, click the **>** character to the right so that the variable is fixed (Figure 8.5).
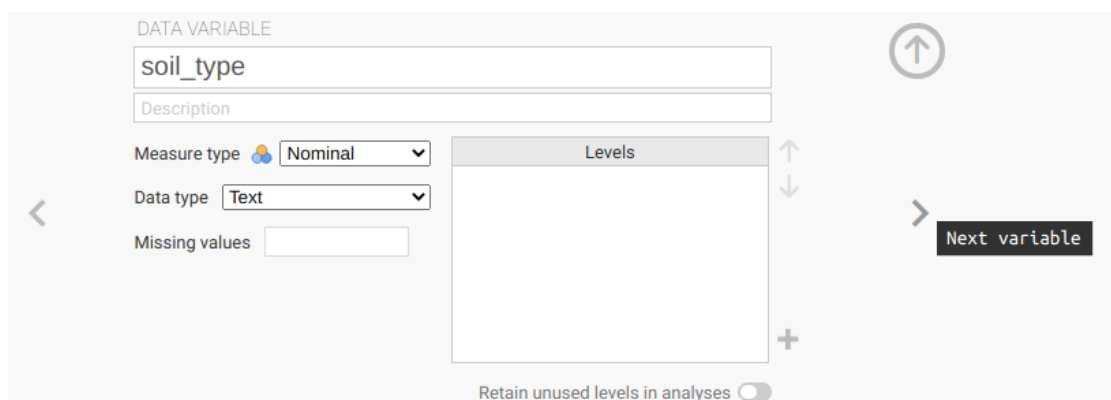


Figure 8.5.: The Jamovi toolbar is shown with the input for creating a new data variable. The new variable added is to indicate the soil type (topsoil or subsoil), so it needs to be a nominal variable with a data type of text.

After typing in the new variable 'soil_type', add another variable called 'organic_carbon'. The organic_carbon variable should have a measure type of 'Continuous' and a data type of 'Decimal'. After both soil_type and organic_carbon variables have been set, you can click the up arrow with the upper right circle (Figure 8.5) to get the new variable window out of the way.

With the two new variables created, we can now rearrange the data in a tidy way. The first 19 rows of soil_type should be 'topsoil', and the remaining 15 rows should be 'subsoil'. To do this quickly, you can write 'topsoil' in the first row of soil_type and copy-paste into the remaining rows. You can do the same to write 'subsoil' in the remaining rows 20-34. Next, copy all of the topsoil values in column 1 into the first 19 rows of column 4, and copy all of the subsoil values in column 2 into the next 15 rows. After doing all of this, you column 3 (soil_type) should have the word 'topsoil' in rows 1-19 and 'subsoil' in rows 20-34. The values from columns 1 and 2 should now fill rows 1-34 of column 4. You can now delete the first column of data by right clicking on the column name 'topsoil' and selecting 'Delete Variable'. Do the same for the second column 'subsoil'. Now you should have a tidy data set with two columns of data, one called 'soil_type' and one called 'organic_carbon'. You are now ready to calculate some descriptive statistics from the data.

First, we can calculate the minimum, maximum, and mean of all of the organic carbon values. To do this, select the 'Analyses' tab, then click on the left-most button called 'Exploration' in the toolbar (Figure 8.6).
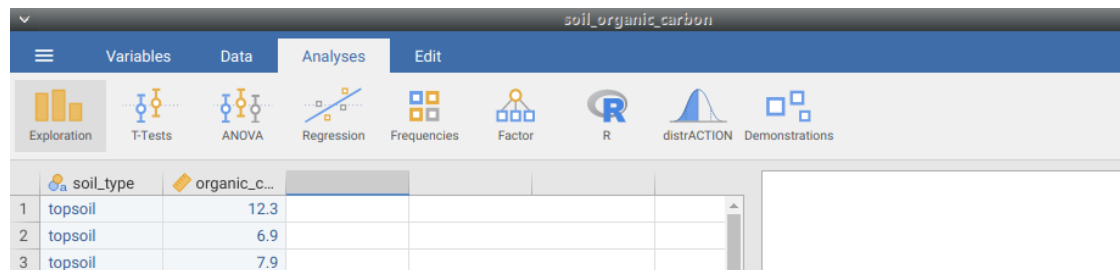


Figure 8.6.: The Jamovi toolbar where the tab 'Analyses' can be selected at the very top. Below this tab, the button 'Exploration' can be clicked to calculate descriptive statistics.

After clicking on 'Exploration', a pull-down box will appear with an option for 'Descriptives'. Select this option, and you will see a new window with our two columns of data in the left-most box. Click once on the 'organic_carbon' variable and use the right arrow to move it into the 'Variables' box. In the right-most panel of Jamovi, a table called 'Descriptives' will appear, which will include values for the organic carbon mean, minimum, and maximum. Write these values on the lines below, and remember to include units.

Mean: _____

Minimum: _____

Maximum: _____

These values might be useful, but recall that there are two different soil types that need to be considered, topsoil and subsoil. The mean, minimum, and maximum above pools both of these soil types together, but we might instead want to know the mean, minimum, and maximum values for topsoil and subsoil separately. Splitting organic carbon by soil types is straightforward in Jamovi. To do it, go back to the Exploration → Descriptives option and again put 'organic_carbon' in the Variables box. This time, however, notice the 'Split by' box below the Variables box. Now, click on 'soil_type' in the descriptives and click on the lower right arrow to move soil type into the 'Split by' box. The table of descriptives in the right window will now break down all of the summary statistics by soil type. First, write the mean, minimum, and maximum topsoil values below.

Mean: _____

Minimum: _____

Maximum: _____

Next, do the same for the mean, minimum, and maximum subsoil values.

Mean: _____

Minimum: _____

Maximum: _____

From the values above, the mean of organic carbon sampled from the topsoil appears to be greater than the mean of organic carbon sampled from the subsoil. Assuming that Jamovi has calculated the means correctly, we can be confident that the topsoil sample mean is higher. But what about the *population* means? Think back to concepts of populations versus samples from Chapter 4. Based on these samples in the dataset, can we really say for certain that the population mean of topsoil is higher than the population mean of subsoil? Think about this, then write a sentence below about how confident we can be about concluding that topsoil organic carbon is greater than subsoil organic carbon.

What would make you more (or less) confident that topsoil and subsoil population means are different? Think about this, then write another sentence below that answers the question.

Note that there is no right or wrong answer for the above two questions. The entire point of the questions is to help you reflect on your own learning and better link the concepts of populations and samples to the real dataset in this practical. Doing this will make the statistical hypothesis testing that comes later in the module more clear.

## 8.2. Exercise on transforming variables

In this next exercise, we will work with the fig fruits dataset. Open this dataset into Jamovi. Note that there are five columns of data, and all of the data appear to be in a tidy format. Each row represents a separate fig fruit, while each column represents a measured variable associated with the fruit. The first several rows should look like the below.

```
##   Tree Length_cm Width_cm Height_cm Seeds
## 1    A      1.5      1.8       1.4   238
## 2    A      1.7      1.9       1.5   198
## 3    A      2.1      2.1       1.6   220
## 4    A      1.5      1.6       1.4   188
## 5    A      1.6      1.6       1.5   139
## 6    A      1.5      1.4       1.5   173
```

The dataset includes the tree from which the fig was sampled in column 1 (A, B, C, and D), then the length, width, and heights of the fig in cm. Finally, the last column shows how many seeds were counted within the fig. Use the Descriptives option in Jamovi to find the overall (i.e., not split by Tree) mean length, width, and height of figs in the dataset. Write these means down below (remember the units).

Mean length: _____

Mean height: _____

Mean width: _____

Now look at the different rows in the Descriptives table of Jamovi. Note that there is a row for 'Missing', and there appears to be one missing value for fig width and fig height. This is very common in real datasets. Sometimes practical limitations in the field prevent data from being collected, or something happens that causes data to be lost. We therefore need to be able to work with datasets that have missing data. For

now, we will just note the missing data and find them in the actual data set. Go back to the 'Data' tab in Jamovi and find the figs with a missing width and height value. Report the rows of these missing values below.

Missing width row: _____

Missing height row: _____

Next, we will go back to working with the actual data. Note that the length, width, and height variables are all recorded in cm to a single decimal place. Suppose we want to transform these variables so that they are represented in mm instead of cm. We will start by creating a new column 'Length_mm' by transforming the existing 'Length_cm' column. To do this, click on the 'Data' tab at the top of the toolbar again, then click on the 'Length_cm' column name to highlight the entire column. Your screen should look like the image in Figure 8.7.
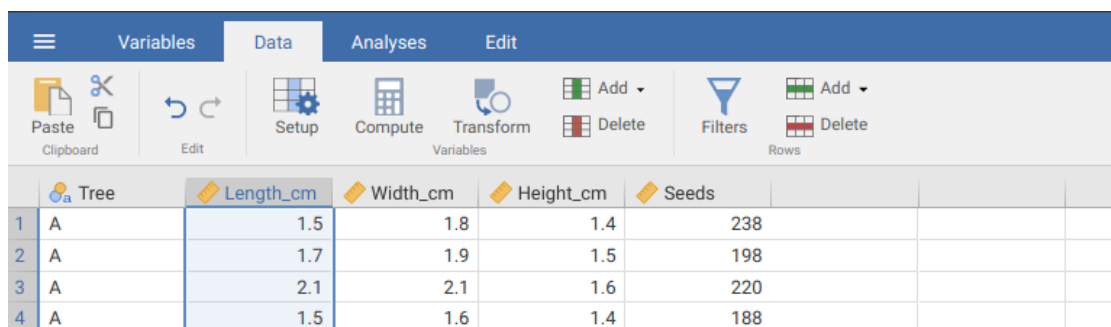


Figure 8.7.: The Jamovi toolbar where the tab 'Data' is selected. The length (cm) column is highlighted and will be transformed by clicking on the Transform button in the toolbar above

With the 'Length_cm' column highlighted, click on the 'Transform' button in the tool-bar. Two things happen next. First, a new column appears in the dataset that looks identical to 'Length_cm'; ignore this for now. Second, a box appears below the toolbar allowing us to type in a new name for the transformed variable. We can call this variable 'Length_mm'. Below, note the first pulldown menu 'Source variable'. The source value should be 'Length_cm', so we can leave this alone. The second pulldown menu 'using transform' will need to change. To change the transform from 'None', click the arrow and select 'Create New Transform' from the pulldown. A new box will pop up allowing us to name the transformation. It does not matter what we call it (e.g., 'cm_to_mm' is fine). Note that there are 10 mm in 1 cm, so to convert from cm to mm, we need to multiply the values of 'Length_cm' by 10. We can do this by appending a `* 10` to the lower box of the transform window, so that it reads `= $source * 10` (Figure 8.8).

When we are finished, we can click the down arrow inside the circle in the upper right to get rid of the transform window, then the up arrow inside the circle in the upper right to get rid of the transformed variable window. Now we have a new column
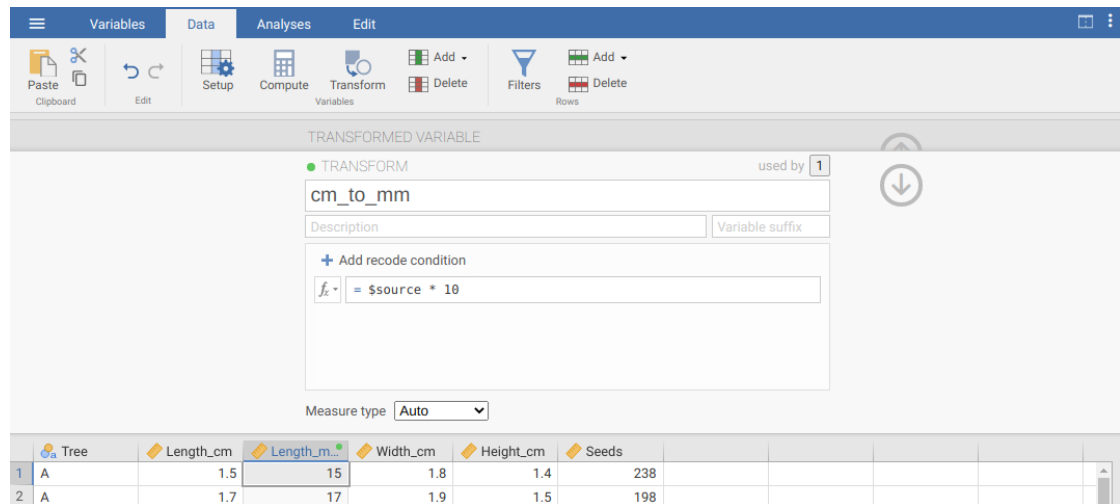
Figure 8.8.: The Jamovi toolbar where the tab 'Data' is selected. The box below shows the transform, which has been named 'cm to mm'. The transformation occurs by multiplying the source (Length mm) by 10. The dataset underneath shows the first few rows with the transformed column highlighted (note that the new 'Length mm' column is 10 times the length column.

called 'Length_mm', in which values are 10 times greater than they are in the adjacent 'Length_cm' column, and therefore represent fig length in mm. If we want to, we can always change the transformation by double-clicking the 'Length_mm' column. For now, apply the same transformation to fig width and height, so we have three new columns of length, width, and height all measured in mm (note, if you want to, you can use the saved transformation 'cm_to_mm' that you used to transform length, saving some time). At the end of this, you should have eight columns of data, including three new columns that you just created by transforming the existing columns of Length_cm, Width_cm, and Height_cm into the new columns Length_mm, Width_mm, and Height_mm. Find the means of these three new columns and write them below.

Mean length: _____

Mean height: _____

Mean width: _____

Compare these means to the means calculated above in cm. Do the differences between means in cm and the means in mm make sense?

44

## 8.3. Exercise on computing variables

In this last exercise, we will compute a new variable 'fig_volume'. Because of the way that the dimensions of the fig were measured in the field, we need to make some simplifying assumptions when calculating volume. We will assume that fig fruits are perfect spheres, and that the radius of each fig is half of its measured width (i.e., 'Width_mm / 2'). This is obviously not ideal, but sometimes practical limitations in the field make it necessary to make these kinds of simplifying assumptions. In this case, how might assuming that figs are perfectly spherical affect the accuracy of our estimated fig volume? Write a sentence of reflection on this question below, drawing from what you have learned this week about accuracy and precision of measurements.

Now we are ready to make our calculation of fig volume. The formula for the volume of a sphere $(V)$ given its radius $r$ is,

$$V = \frac{4}{3}\pi r^3.$$

In words, sphere volume equals four thirds times $\pi$, times $r$ cubed (i.e., to the third power). If this equation is confusing, remember that $\pi$ is approximately 3.14, and taking $r$ to the third power means that we are multiplying $r$ by itself 3 times. We could therefore rewrite the equation above,

$$V = \frac{4}{3} \times 3.14 \times r \times r \times r.$$

This is the formula that we can use to create our new column of data for fig volume. To do this, click on the first empty column of the dataset, just to the right of the 'Seeds' column header. You will see a pull down option in Jamovi with 3 options, one of which is 'NEW COMPUTED VARIABLE'. This is the option that we want. We need to name this new variable, so we can call it 'fig_volume'. Next, we need to type in the formula for calculating volume. First, in the small box next to the $f_x$, type in the (4/3) multiplied by 3.14 as below.

```
= (4/3) * 3.14 *
```

Next, we need to multiply by the variable 'Width_mm' divided by 2 (to get the radius), three times. We can do this by clicking on the $f_x$ box to the left. Two new boxes will appear; the first is named 'Functions', and the second is named 'Variables'. Ignore the functions box for now, and find 'Width_mm' in the list of variables. Double click on this to put it into the formula, then divide it by 2. You can repeat this two more times to complete the computed variable as shown in Figure 8.9.
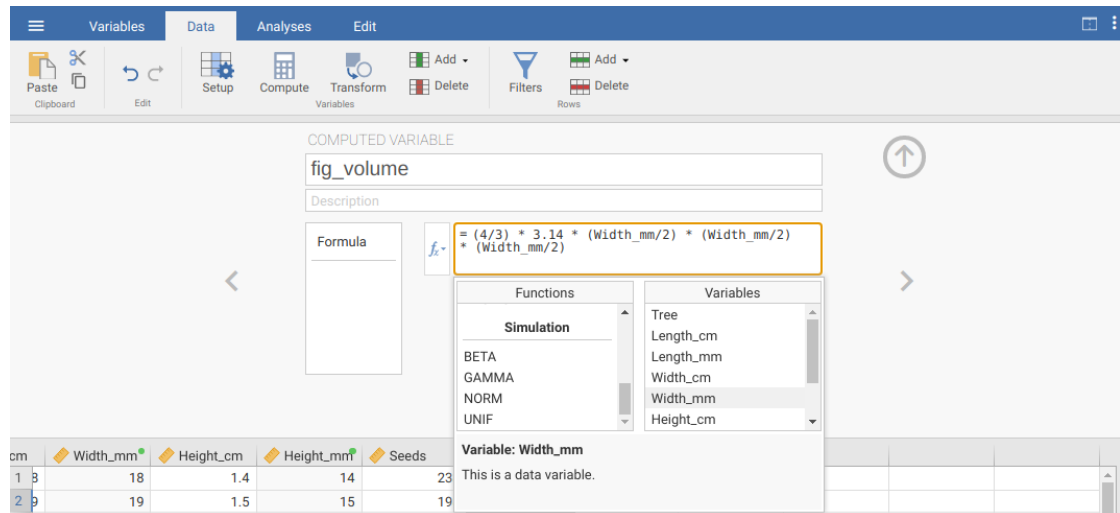


Figure 8.9.: The Jamovi toolbar where the tab 'Data' is selected. The box below shows the new computed variable 'fig volume', which has been created by calculating the product of 4/3, 3.14, and Width mm three times.

Note that we can get the cube of 'Width_mm' more concisely by using the carrot character (^). That is, we would get the same answer shown in Figure 8.9 if we instead typed the below in the function box.

```
= (4/3) * 3.14 * (Width_mm/2)^3
```

Note that the order of operations is important here, which is why there are parentheses around `Width_mm/2`. This calculation needs to be done before taking the value to the power of 3. If we instead had written, `Width_mm/2^3`, the Jamovi would first take the cube of 2 ($2 \times 2 \times 2 = 8$), then divided `Width_mm` by this value giving a different and incorrect answer. When in doubt, it is always useful to use parentheses to specify what calculations should be done first.

You now have the new column of data 'fig_volume'. Remember that the calculations underlying this new variable need to be done for the units. The width of the fig was calculated in mm, but we have taken width to the power of 3 when calculating the volume. In the spaces below, find the overall mean, minimum, and maximum volumes of figs and report them in the correct units.

Mean: _____

Minimum: _____

Maximum: _____

Finally, it would be good to plot these newly calculated fig volume data. These data are continuous, so we can use a histogram to visualise the fig volume distribution. To make a histogram, go to the Exploration → Descriptives window in Jamovi (the same place where you found the mean, minimum, and maximum). Now, look on the lower left-hand side of the window and find the pulldown menu for 'Plots'. Click 'Plots', and you should see several different plotting options. Check the option for 'Histogram' and see the new histogram plotted in the window to the right. Draw a rough sketch of the histogram in the area below.

# Part III.

# Summary statistics

# Week 3 Overview

Overview of what this week will include.

Week: 3 Dates: Suggested Readings: Textbook intro to Jamovi Assessments: Practice quiz

# 9. Decimal places and significant figures

It is important to know how to write a number you have calculated to the appropriate number of digits, typically either defined as a number of decimal places or as a number of 'significant figures'. This is especially important in any module (for example, this one) where you are required to put numeric answers into a test or exam, as the computer will only recognise the answer as being correct if it is expressed in the way stated in the question. If you are unfamiliar with how to work out the right number of decimal places or significant figures, then these guides should be useful, although there are many other web-sources that would explain the concepts too.

# 10. The mean

Explanation of the mean

# 11. The mode

Explanation of the mode

# 12. The median and quantiles

Notes on the median and quantiles

# 13. Mean, mode, median, and resistance

Resistance in statistics

# 14. Plots

Graphics are critical for visualising data, which is always important

## 14.1. General principles

Some points about plots

## 14.2. Histograms

Histograms are special, and introduce the concept of a distribution.

## 14.3. Box-whisker plots

# 15. *Practical.* Real data with Jamovi

Using some real datasets in Jamovi

## 15.1. Some biological example

Remember to first put it in a tidy format. Get summary statistics here too.

## 15.2. Some environmental example

Transform and compute a new variable, plotting in Jamovi.

## 15.3. Summary of exercises and why they are useful.

Useful for scientific publications, reading, and dissertation work.

# Part IV.

# Probability models and the Central Limit Theorem

# Week 4 Overview

General overview of what will be the focus of this week.

Week: 4 Dates: Suggested Readings: Textbook intro to probability Assessments: Practice quiz Practical:

# 16. Introduction to probability models

Some background

## 16.1. A practical example

How to think about probability

## 16.2. Probability distributions

Some more useful examples

### 16.2.1. Binomial distribution

Explanation, fairly straightforward

### 16.2.2. Poisson distribution

Another example

### 16.2.3. Normal distribution

Why this is so important

# 17. The Central Limit Theorem (CLT)

General overview

## 17.1. Examples of the CLT in action

## 17.2. The standard normal distribution

## 17.3. What are z-scores?

# 18. *Practical.* Probability and simulation

Some of these examples will be similar to what will be on the assessment

## 18.1. Calculating probability exercise 1

Example exercise 1 with some simple probability calculations

## 18.2. Calculating probability exercise 2

Example exercise 2 with some simple probability calculations

## 18.3. Calculating probability from normal distribution

Example exercise for getting a the probability of some value sampled above, below, or between some threshold in Jamovi.

## 18.4. Normal distribution and sample size

Showing how we get closer to the normal distribution as sample size increases in Jamovi.

## 18.5. Simulating the central limit theorem

Doing the example from a uniform distribution in Jamovi.

# Part V.

# Statistical inference

# Week 5 Overview

General overview of what will be the focus of this week.

Week: 5 Dates: Suggested Readings: Textbook intro to probability Assessments: Practice quiz Practical:

# 19. Sample statistics and population parameters

An explanation of this

# 20. Standard Normal Distribution

What this means, and why it is important.

# 21. Confidence intervals

How these are calculated, and how to interpret them

# 22. The t-interval

What this is and how it relates to the normal distribution, and why it is important.

# 23. *Practical.* z- and t- intervals

**23.1. Example constructing confidence intervals**

**23.2. Confidence interval for different levels (t- and z-)**

**23.3. Proportion confidence intervals**

**23.4. Another confidence interval example?**

# Part VI.

# Review of parts I-V

# Week 6 Overview (Reading week)

This is a special chapter for week 6, which is a reading week, and it will function as a very brief pause for review. It will also ensure that the numbers of chapters will correspond to weeks.

# Part VII.

# Hypothesis testing

# Week 7 Overview

General overview of what will be the focus of this week.

Week: 7 Dates: Suggested Readings: Textbook Assessments: Practice quiz Practical:

# 24. What is hypothesis testing?

An explanation of this, and that we are starting to get into some of the more interesting bits of inferential statistics.

# 25. Making and using hypotheses and types of tests

What this means, and why it is important.

# 26. An example of hypothesis testing

Errors

# 27. Hypothesis testing and confidence intervals

Relationship between these two.

# 28. Student t-distribution and one sample t-test

What this is and how to do it in Jamovi.

# 29. Another example of a one sample t-test

From the lectures

# 30. Independent t-test

What this is and how to use it in Jamovi.

# 31. Paired sample t-test

Another explanation, example, and how to do it in Jamovi.

# 32. Violations of assumptions

What to do in this case

# 33. Non-parametric tests, and what these are.

Explanation of how to do them in Jamovi.

# 34. *Practical.* Hypothesis testing and t-tests

## 34.1. Exercise on a simple one sample t-test

## 34.2. Exercise on an independent sample t-test

## 34.3. Exercise involving multiple comparisons

## 34.4. Exercise with non-parametric

## 34.5. Another exercise with non-parametric

# Part VIII.

# Analysis of Variance (ANOVA)

# Week 8 Overview

General overview of what will be the focus of this week.

Week: 8 Dates: Suggested Readings: Textbook Assessments: Practice quiz Practical:

# 35. What is ANOVA?

General explanation

# 36. One-way ANOVA

Explain what this is.

# 37. Two-way ANOVA

More explanation

# 38. Kruskall-Wallis H test

Non-parametric explanation

# 39. *Practical.* ANOVA and associated tests

## 39.1. ANOVA Exercise 1

## 39.2. ANOVA Exercise 2

## 39.3. ANOVA Exercise 3

## 39.4. ANOVA Exercise 4

# Part IX.

# Counts and Correlation

# Week 9 Overview

General overview of what will be the focus of this week.

Week: 9 Dates: Suggested Readings: Textbook Assessments: Practice quiz Practical:

# 40. Frequency and count data

General explanation

# 41. Chi-squared goodness of fit

Explain what this is.

# 42. Chi-squared test of association

More explanation

# 43. Correlation key concepts

# 44. Correlation mathematics

# 45. Correlation hypothesis testing

# 46. *Practical.* Analysis of count data, correlation, and regression

**46.1. Chi-Square Exercise 1**

**46.2. Chi-Square association Exercise 2**

**46.3. Correlation Exercise 3**

**46.4. Correlation Exercise 4**

# Part X.

# Linear Regression

# Week 10 Overview

Week: 10 Dates: Suggested Readings: Textbook Assessments: Practice quiz Practical:

# 47. Regression key concepts

# 48. Regression validity

# 49. Introduction to multiple regression

General explanation

# 50. Model selection (maybe remove this?)

Seriously consider moving the regression into this week. and ease the amount of material in previous weeks.

# 51. *Practical.* Using regression

## 51.1. Regression Exercise 1

## 51.2. Regression Exercise 2

## 51.3. Regression Exercise 3

## 51.4. Regression Exercise 4

# Part XI.

# Randomisation approaches

# Week 11 Overview

The aim of this lecture is to introduce the randomisation approach to statistical hypothesis testing. We will first introduce the general idea of what randomisation is and how it relates to the hypothesis testing that we have been doing since week five. We will then consider an instructive example in which a randomisation approach is used in place of a traditional t-test to test whether or not the mean values of two different groups are identical. We will then compare the assumptions underlying randomisation and how they differ slightly from the assumptions of traditional hypothesis testing. We will then look at how randomisation can be used to build confidence intervals and test hypotheses that would difficult to test with other approaches. In learning about randomisation approaches, we will also review some key concepts from earlier in the module. The aim is not to understand all of the nuances of randomisation, but to understand, conceptually, what is going on in the methods described below.

Week: 11 Dates: Suggested Readings: Textbook Assessments: Practice quiz Practical: R starts creeping in now?

# 52. Introduction to randomisation

General explanation

# 53. Assumptions of randomisation

How these differ

# 54. Bootstrapping

What this is and why we use it.

# 55. Monte Carlo

# 56. *Practical.* Using R

## 56.1. R Exercise 1

## 56.2. R Exercise 2

## 56.3. R Exercise 3

# Part XII.

# Statistical Reporting

# Week 12 Overview

Week: 12 Dates: Suggested Readings: Textbook Assessments: Practice quiz Practical: R starts creeping in now?

# 57. Reporting statistics

General explanation

# 58. More introduction to R

How these differ

# 59. More getting started with R

Just more to do.

# 60. *Practical.* Using R

## 60.1. R Exercise 1

## 60.2. R Exercise 2

## 60.3. R Exercise 3

# Part XIII.

# Review of parts (VII-XII)

# Module summary

This chapter will be specifically to prepare for exam.

# A. Statistical units

# B. Uncertainty derivation

It is not necessary to be able to derive the equations for propagating error from week 2, but working through the below might be interesting, and provide a better appreciation for why these formulas make sense. Another derivation is available in Box et al. (1978) (page 563), but this derivation is expressed in terms of variances and covariances, which is likely to be less helpful for this module.

**Propagation of error for addition and subtraction**.

For adding and subtracting error, we know that we get our variable $Z$ by adding $X$ and $Y$. This is just how $Z$ is defined. We also know that $Z$ is going to have some error $E_Z$, and we know that $Z$ plus or minus its error will equal $X$ plus or minus its error plus $Y$ plus or minus its error,

$$(Z \pm E_Z) = (X \pm E_X) + (Y \pm E_Y).$$

Again, this is just our starting definition, but double-check to make sure it makes sense. We can now note that we know,

$$Z = X + Y.$$

If it is not intuitive as to why, just imagine that there is no error associated with the measurement of $X$ an $Y$ (i.e., $E_X = 0$ and $E_Y = 0$). In this case, there cannot be any error in $Z$. So, if we substitute $X + Y$ for $Z$, we have the below,

$$((X + Y) \pm E_Z) = (X \pm E_X) + (Y \pm E_Y).$$

By the associative property, we can get rid of the parenthesis for addition and subtraction, giving us the below,

$$X + Y \pm E_Z = X \pm E_X + Y \pm E_Y.$$

Now we can subtract $X$ and $Y$ from both sides and see that we just have the errors of $X$, $Y$, and $Z$,

$$\pm E_Z = \pm E_X \pm E_Y.$$

*B. Uncertainty derivation*

The plus/minus is a bother. Note, however, that for any real number $m$, $m^2 = (-m)^2$. For example, if $m = 4$, then $(4)2 = 16$ and $(-4)2 = 16$, so we can square both sides to get positive numbers and make things easier,

$$E_Z^2 = (\pm E_X \pm E_Y)^2.$$

We can expand the above,

$$E_Z^2 = E_X^2 + E_Y^2 \pm 2E_X E_Y.$$

Now here is an assumption that we have not told you about elsewhere in the module. With the formulas that we have given you, we are assuming that the errors of $X$ and $Y$ are independent. To put it in more statistical terms, the covariance between the errors of $X$ and $Y$ is assumed to be zero. Without going into the details (covariance will be introduced later in the module), if we assume that the covariance between these errors is zero, then we can also assume the last term of the above is zero, so we can get rid of it (i.e., $2E_X E_Y = 0$),

$$E_Z^2 = E_X^2 + E_Y^2.$$

If we take the square root of both sides, then we have the equation from Chapter 7,

$$E_Z = \sqrt{E_X^2 + E_Y^2}.$$

**Propagation of error for multiplication and division**.

Now that we have seen the logic for propagating errors in addition and subtraction, we can do the same for multiplication and division. We can start with the same point that we are getting our new variable $Z$ by multiplying $X$ and $Y$ together, $Z = XY$. So, if both $X$ and $Y$ have errors, the errors will be multiplicative as below,

$$Z \pm E_Z = (X \pm E_X)(Y \pm E_Y).$$

Again, all we are doing here is substituting $Z$, $X$, and $Y$, for an expression in parentheses that includes the variable plus or minus its associated error. Now we can expand the right hand side of the equation,

$$Z \pm E_Z = XY + Y E_X + X E_Y + E_X E_Y.$$

As with our propagation of error in addition, here we are also going to assume that the sources of error for $X$ and $Y$ are independent (i.e., their covariance is zero). This allows us to set $E_X E_Y = 0$, which leaves us with the below,

$$Z \pm E_Z = XY + YE_X + XE_Y.$$

Now, because $Z = XY$, we can substitute on the left hand side of the equation,

$$XY \pm E_Z = XY + YE_X + XE_Y.$$

Now we can subtract the $XY$ from both sides of the equation,

$$\pm E_Z = YE_X + XE_Y.$$

Next, let us divide both sides by $XY$,

$$\frac{\pm E_Z}{XY} = \frac{YE_X + XE_Y}{XY}.$$

We can expand the right hand side,

$$\frac{\pm E_Z}{XY} = \frac{YE_X}{XY} + \frac{XE_Y}{XY}.$$

This allows us to cancel out the $Y$ variables in the first term of the right hand side, and the $X$ variables in second term of the right hand side,

$$\frac{\pm E_Z}{XY} = \frac{E_X}{X} + \frac{E_Y}{Y}.$$

Again, we have the plus/minus on the left, so let us square both sides,

$$\left(\frac{\pm E_Z}{XY}\right)^2 = \left(\frac{E_X}{X} + \frac{E_Y}{Y}\right)^2.$$

We can expand the right hand side,

$$\left(\frac{\pm E_Z}{XY}\right)^2 = \left(\frac{E_X}{X}\right)^2 + \left(\frac{E_Y}{Y}\right)^2 + 2\left(\frac{E_X}{X}\right)\left(\frac{E_Y}{Y}\right).$$

Again, because we are assuming that the errors of $X$ and $Y$ are independent, we can set the third term on the right hand side of the equation to zero. This leaves,

$$\left(\frac{\pm E_Z}{XY}\right)^2 = \left(\frac{E_X}{X}\right)^2 + \left(\frac{E_Y}{Y}\right)^2.$$

*B. Uncertainty derivation*

Note that $XY = Z$, so we can substitute in the left hand side,

$$\left(\frac{\pm E_Z}{Z}\right)^2 = \left(\frac{E_X}{X}\right)^2 + \left(\frac{E_Y}{Y}\right)^2.$$

Now we can apply the square on the left hand side to the top and bottom, which gets rid of the plus/minus,

$$\frac{E_Z^2}{Z^2} = \left(\frac{E_X}{X}\right)^2 + \left(\frac{E_Y}{Y}\right)^2.$$

We can now multiply both sides of the equation by $Z^2$,

$$E_Z^2 = Z^2\left(\left(\frac{E_X}{X}\right)^2 + \left(\frac{E_Y}{Y}\right)^2\right).$$

We can now take the square root of both sides,

$$E_Z = \sqrt{Z^2\left(\left(\frac{E_X}{X}\right)^2 + \left(\frac{E_Y}{Y}\right)^2\right)}.$$

We can pull the $Z^2$ out of the square root,

$$E_Z = Z\sqrt{\left(\frac{E_X}{X}\right)^2 + \left(\frac{E_Y}{Y}\right)^2}.$$

That leaves us with the equation that was given in Chapter 7.

# C. Statistical tables

# Bibliography

Box, G. E. P., Hunter, W. G., and S, H. J. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building.* John Wiley & Sons, New York.

Duthie, A. B., Abbott, K. C., and Nason, J. D. (2015). Trade-offs and coexistence in fluctuating environments: evidence for a key dispersal-fecundity trade-off in five nonpollinating fig wasps. *American Naturalist*, 186(1):151–158.

Duthie, A. B. and Nason, J. D. (2016). Plant connectivity underlies plant-pollinator-exploiter distributions in Ficus petiolaris and associated pollinating and non-pollinating fig wasps. *Oikos*.

Dytham, C. (2011). *Choosing and Using Statistics: A Biologist's Guide.* John Wiley & Sons.

Navarro, D. J. and Foxcroft, D. R. (2022). *Learning Statistics with Jamovi.* (Version 0.75).

Preston, C. M. and Schmidt, M. W. (2006). Black (pyrogenic) carbon: A synthesis of current knowledge and uncertainties with special consideration of boreal regions. *Biogeosciences*, 3(4):397–420.

Quinn, T. J. (1995). Base units of the Système International d'Unités, their accuracy, dissemination and international traceability. *Metrologia*, 31(6):515–527.

Rabinovich, S. G. (2013). *Evaluating Measurement Accuracy.*

Rowntree, D. (2018). *Statistics Without Tears.* Penguin.

Santín, C., Doerr, S. H., Kane, E. S., Masiello, C. A., Ohlson, M., de la Rosa, J. M., Preston, C. M., and Dittmar, T. (2016). Towards a global assessment of pyrogenic carbon from vegetation fires. *Global Change Biology*, 22(1):76–91.

Sokal, R. R. and Rohlf, F. J. (1995). *Biometry.* W. H. Freeman and Company, New York, 3rd edition.

Spiegelhalter, D. (2019). *The Art of Statistics Learning from Data.* Penguin UK.

Stock, M., Davis, R., De Mirandés, E., and Milton, M. J. (2019). Corrigendum: The revision of the SI - The result of three decades of progress in metrology (Metrologia (2019) 56 (022001) DOI: 10.1088/1681-7575/ab0013). *Metrologia*, 56(4).