Biology 116 Lab Manual

Dr. Tristyn Hay and Dr. Robin Young

2021-08-30

Contents

Welcome

Welcome to the BIOL116 lab manual!

The material in this manual is designed to be used as part of the BIOL116 labs. It integrates heavily with additional material that can be accessed only through the BIOL116 Canvas page, including assignments, quizzes, and other material that is specific to this year's offering of the course. As such you will be expected to access BOTH this website and Canvas regularly.

Your first lab, Lab 1, is asynchronous (meaning that you complete it online, in your own time during that week) and the material is posted exclusively on Canvas. **Your second lab**, Lab 2, will be your first chance to physically enter the lab, and meet your TA and your fellow lab mates. The material for Lab 2 can be found in this text, under the Lab 2A and Lab2B headings (with links to this content also on Canvas).

The rest of the labs will continue to alternate between asynchronous, online labs, and synchronous, in-person labs. **For in-person labs it is vital that you attend the lab in which you are registered** (Check your registration online to be sure of where you are supposed to be!).

There's much more information about what to expect from labs, the grading policies, and the weekly schedule in the Canvas Module for Lab 1. For now, we will say that we're looking forward to the start of term, and to meeting you (both virtually and in-person). Science is amazing and exciting and cool, and we can't wait to share our love of science and biology with you!

Copyright

Please use the following for citing this document

Hay, T., Young, R. (2021). *Biology 116 Lab Manual*. https://ubco-biology.github.io/BIOL-116-Lab-Manual/

All source files are available from https://github.com/ubco-biology/BIOL-116-Lab-Manual.

Lab 1

Introduction

Last updated 2021-08-30

Content for the first lab will all be on Canvas, the learning management system used at UBC. Enrolled students can access this content at https://canvas.ubc.ca/courses/90147

Lab 2A

The Process of Science

Last updated 2021-08-30

Note on Copyright

Objectives

At the end of this lab students should be able to:

- List the main components involved in the "process of science"
- Provide examples of scientific activities and / or list where these activities would generally fit in the "process of science" chart.
- Explain - in your own words or through examples - what is meant by each of the following statements regarding the process of science:
  - The process of science involves testing ideas with data
  - Though they can be supported by evidence, it is often difficult for scientific hypotheses to be proven or directly observed
  - Scientific knowledge is subject to change as new evidence and perspectives emerge
  - The process of science is often non-linear
  - The process of science involves observation, exploration, discovery, testing, communication, and application
  - Scientists can use multiple methods - experiments, observations, models - to test their ideas
  - Science involves curiosity and creativity
  - Scientists work together to share their ideas

  - Scientists aim for their studies to be replicable
  - Over time, scientific ideas or hypotheses that are supported by multiple methods and replicated in various studies become established theories that are widely accepted by the scientific community

Discussion Questions

1. What is science?
2. What do scientists do and how - in general - do they do it? Discuss this with a partner.

Mystery Cylinders

Mystery Cylinders An exercise in the scientific process

Imagine a container, a cylinder that you are unable to open, and you cannot see inside. How would you manage to figure out what is inside the cylinder? What would your thinking process be?

This activity is broken down into 3 steps:

1. Observe
2. Share and Collaborate
3. Submit

Figure 1: Figure 1. Mystery cylinders.Image by UBCO, licensed under CC BY 4.0

Phase 1 – "Observe" your cylinder

Outside of opening the container, you have the option to hold it, shake it etc. What types of observations would you like to make? During this activity, you and your group will be able to interact with the cylinder. Think about the question you are addressing and how you would go about answering it.

Work in groups to come up with three suggestions of ways you can manipulate your cylinders to try and determine what is inside. Remember that not all of your initial ideas will be plausible - you don't always have all the ideal tools available to you. You do not have an X-ray machine in the lab . Offer three possible tests that could likely be done in lab.

Designate one person in your group to write down all observations and information you attain during your experiment. Following this, you and your group now must try to reach a consensus about what is inside the interior of each cylinder.

Note

If you do not take extensive notes throughout the experimentation process - electronically or on paper - you will not have the information that you need to complete this assignment.

Phase 2 - Share & Collaborate

Your instructor will then help your group to discuss findings with other groups.

- Discuss whether you are using similar techniques to analyze your cylinders.
- Collaborate with others to see if you can arrive at a consensus guess for the interior of your cylinder(s).

Phase 3 – Submit for marks

Once you feel confident with your predictions, you will need to submit them to Canvas for marks. There is a separate file posted with the assignment instructions.

Discussion Question

1. With your original group, make a list of the **specific things** that you did during this activity and how / why they are a part of doing science - e.g. twisting the cylinders = ??

NOTE

Wait for your instructor before proceeding to the next section *Science Flowchart*.

Science Flowchart

A Simple Flowchart

Study the simple science flowchart with your instructor:

A Complex Flowchart

Now let's view the complex science flowchart:

Class Discussion

Follow your instructor's directions as a class to place examples of the types of things you did in the mystery box activity into each circle. See the list generated for the previous discussion question.

This discussion marks the end of the first part of this lab (Lab 2). If you have not already done so, make sure you have the information you need to complete your submission of the Mystery cylinder assignment before the end of the day. Once you leave your lab you will not have the opportunity to come back if there is something you forgot or missed.

When you're ready, you can proceed to Part 2 of Lab 2, where the research project is discussed.

Mystery Cylinder Assignment

Purpose To help you learn how to keep good lab notes and make predictions.

Rationale A key piece of science is to record the work you do so that you can learn from what worked and what didn't. Keeping accurate records of the work that you do can make all of the difference. Patents and other scientific

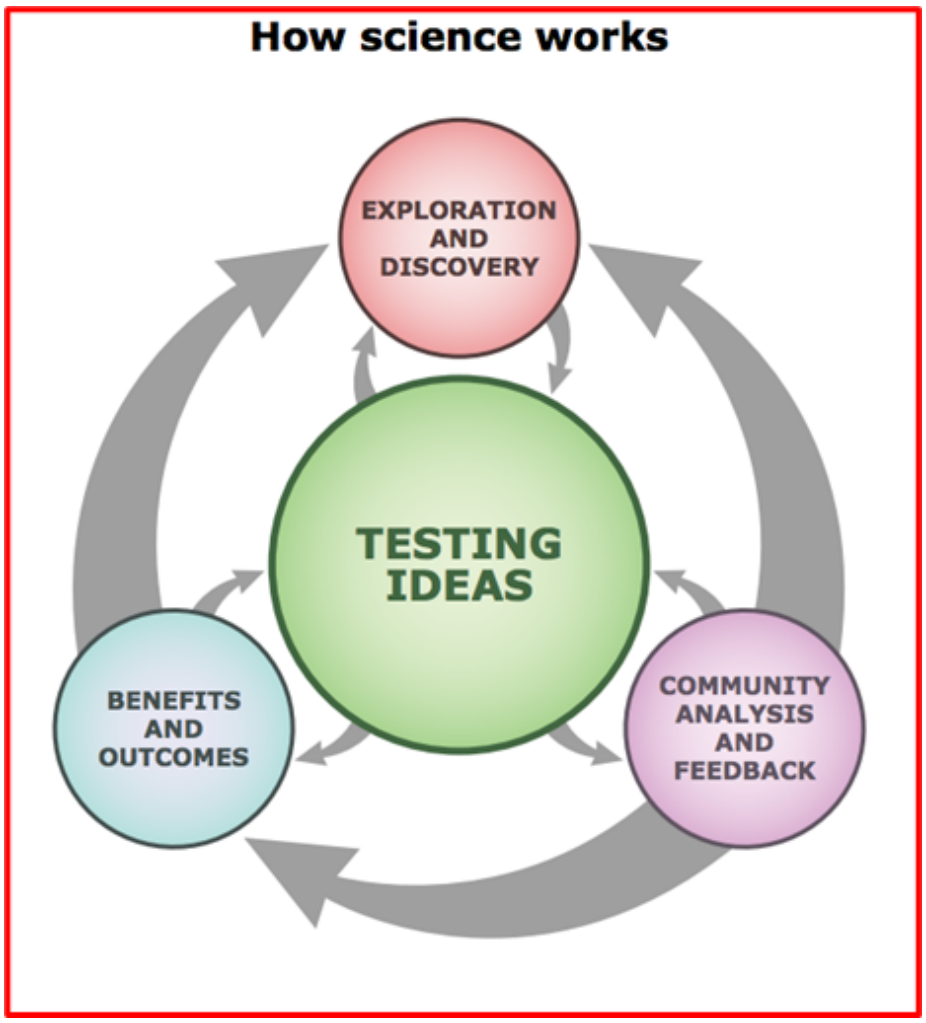

Figure 2: Figure 2. Simple science flowchart.Image used with permission from The University of California Museum of Paleontology, Berkeley, and the Regents of the University of California.

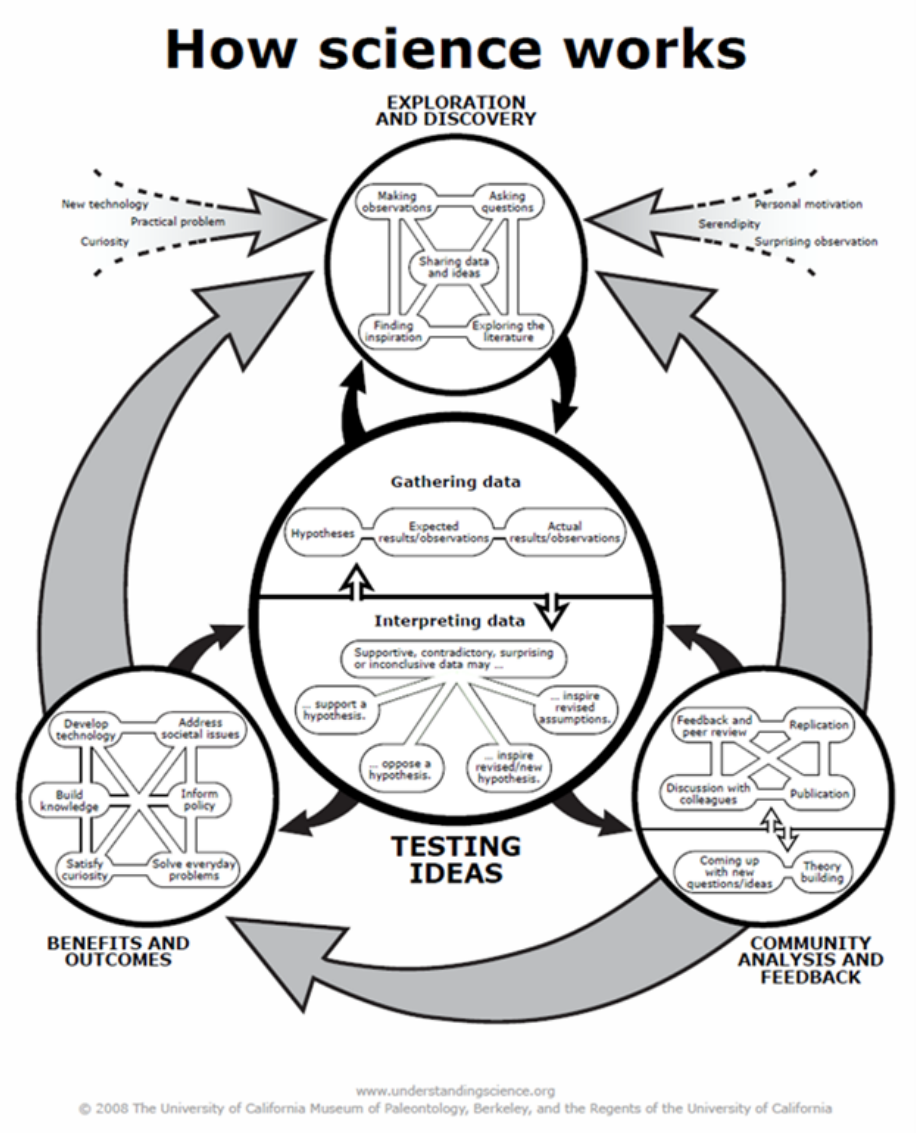

Figure 3: Figure 3. Complex science flowchart.Image used with permission from The University of California Museum of Paleontology, Berkeley, and the Regents of the University of California.

discoveries can be worth a lot of money to those that can prove that the work they did was theirs. This requires accurate and honest note keeping.

Instructions For this assignment, you will be working with your teaching assistants to determine the contents of a "mystery cylinder". You will have the opportunity to see, touch, and manipulate each cylinder in the lab and you are expected to work within your lab group to come up with 3 experiments that your group will conduct. Once you have completed the experiments, you will be expected to make a prediction regarding the contents of your "mystery cylinder".

If you do not take extensive notes in your lab notebook throughout the experimentation process, you will not have the information that you need to complete this assignment.

Submission For this assignment, you will submit your predictions for the contents of your mystery cylinder, along with a rationale that explains how the experiments you ran, which were documented in your lab notes, support your prediction.

You have two options for submission of this part of your work:

1. You can submit a 1-page written prediction, with rationale. It should be no more than one page (double-spaced, Arial 12pt font, 1cm margins)
2. You can submit a video (no more than 2 min)

You will upload your mystery cylinder assignment on Canvas via the dedicated assignment entitled "Mystery Cylinder Submission". This assignment is **due before 11:59 pm the day of your lab**.

While it is encouraged that you work together prior to submission, your lab submission must be entirely your own work.

This means:

- No copying and pasting from other sources - even if you plan to 'tweak'. This is not your own words.
- Do not work it out together, and then alter the final draft to make it look less similar. Figure it out together, and then go away and write your final submission yourself.

Mystery Cylinder Assignment Grading Rubric

Criteria

Ratings

Pts

Predictions - All prediction(s) are present - Prediction(s) are clear and consise - Prediction(s) are reasonable

5pts – Full Marks: All criteria are met

3pts – Proficient: Two of the three criteria are met

2pts – Unsatisfactory: One fo the criteria are met

0pts – No Marks: No criteria are met

5 pts

Rationale - A rationale is provided for the prediction(s) - Each rationale explains how the experiments conducted supported each prediction - The rationale(s) provided are clear and easily understood

5pts – Full Marks: All criteria are met

3pts – Proficient: Two of the three criteria are met

2pts – Unsatisfactory: One fo the criteria are met

0pts – No Marks: No criteria are met

5 pts

Total points: 10

Lab 2B

Research Project

Last updated 2021-08-30

Over the term, you and your partner will be designing and conducting an experiment, analyzing the data you generate and then presenting your results. In order to prepare you for this task, you will need to have read through the material below in advance of lab.

Learning Outcomes

Students will participate in the process of science and demonstrate scientific thinking. Students will be introduced to the characteristics of the experimental method of research and learn:

- How to form a hypothesis.
- How to design an experiment to test your hypothesis.
- How to analyze and display your results.
- How to interpret your results so as to support or reject your hypothesis.
- How to present your results as an oral presentation.

In addition, you will begin to explore the principles and practices of Open Science, as a way to help ensure reproducibility and transparency in the work that you do as scientists.

Science & The Scientific Method

Science is just one of many different lenses through which we, as humans, attempt to better understand the world around us. Science as a discipline is built on the foundation of observable, measurable evidence that allows us to make testable predictions about the world. There are other ways of learning about our world (faith, intuition, memory, imagination, etc.). Each of these "ways of knowing" uses different kinds of evidence to help us make sense of our world.

Science is also a highly collaborative endeavor, where each scientist builds on the work of their colleagues and mentors, to collectively construct scientific knowledge. This collaborative aspect is central to the ideas of Open Science, which is a key part of what you'll be learning about throughout your UBCO degree.

The Scientific Method

The way that we *do* science is that we follow a set of guidelines called "The Scientific Method". This method helps us make sure that we test our ideas in such a way that we've done our best to remove our own opinions and personal biases from the process. Removing personal opinions and biases, and focusing on what the evidence is telling us is absolutely vital to science. It is the heart of everything that we, as scientists, try to do every day.

Deductive Logic

Some scientists are interested in measuring and recording observations in nature. This is sometimes called "descriptive science". Scientific sampling and collection methods allow researchers to describe nature accurately. We can then use these observations to make generalizations about the world using "deductive logic".

Inductive Logic

Other scientists wonder why things are the way they are. They form ideas, or hypotheses using "inductive logic" about how things must work and then test these ideas in experiments that they design.

Whether a scientist is conducting descriptive (observational) science or experimental science, they will be using evidence to support the claims that they make, and they will be using the framework of the scientific method to help them do this in a way that avoids bias and focuses on the evidence they have.

In this term's research project we will be focusing on experimental research, rather than descriptive science.

Reproducibility

One of the most important concepts in science is the idea of reproducibility. This means that if our methods are carefully recorded, and shared with others, they will be able to reproduce the work that we did, and, ideally, get similar results that yield identical conclusions about the hypothesis. It also means that scientists never base their conclusions on only one experiment. We always attempt to **replicate** our results, by running the same experiment, following our proposed method, at least 2-3 more times (and maybe more?) to ensure that the results we are getting are consistent, and not some kind of random fluke. Shortly, we will give you information about how to build a workflow, based on Open Science practices and principles, that sets you up to best be able to replicate your results, troubleshoot what went wrong if necessary, and also to help others understand and reproduce your work.

Understand that scientists do these things because they do not accept anything on faith. Because science is not a belief system; scientists are convinced only by evidence and data. Any idea is up for debate and everything can be criticized. It takes many years and many experiments to convince scientists that something is true or not true. For example, the idea that the continental plates are moving slowly over the surface of the Earth took decades to take root. It was debated and tested and tested again by many people before it was fully accepted as fact. This is as it should be… scientists don't much like to be wrong about the big stuff. We don't accept any explanation as "true" before every other possible explanation has been tested and rejected. It is only when we have done our best to disprove an idea, and it has stood up to everything that we've thrown at it, that we then begin to accept that maybe… just maybe… it might be true… at least for now, until new evidence is found that throws everything into doubt again.

Open Science

Open Science is a movement to make scientific research transparent and accessible to everyone. This not only gives the best opportunity for research to be critically examined, to ensure reproducibility, but it also makes it easier for scientists to share their work with others, and build on the work that has been done before. You will be learning more about the principles and practices of Open Science in Lab 3. For now, know that as you work on your research project, you will follow the stages of a typical registered report and implement Open Science practices including:

- Throughout the experiment, using appropriate version control on electronic documents and proper file and data management practices (see below).
- Performing a literature review on your research topic and documenting a list of consulted studies, how they were found, and the strengths, limitations, and weaknesses of each.
- Submitting a written proposal with an established *a priori* hypothesis, experimental design, and plan for presenting and analyzing your data. This will be marked before the experiment implementation phase and TA feedback incorporated into the project as needed. Creating a detailed, thorough plan for your research often takes as much time as running the experiment and collecting and analyzing your data. The more you plan, including anticipating potential problems, the easier the implementation!
- Implementing the study according to your plan, and noting any deviations from that plan (Note: deviations often happen, and that's OK! The key is to document them). These reflections will be submitted for marks.
- Submitting and presenting a poster that details your experiences implementing the research plan (including any changes recorded, justification for changes, analysis of the data, and your interpretation and conclusion).
- Conducting a peer review of other students' poster presentations using the poster presentation rubric as a guideline.

The Research Workflow

Now that we've briefly introduced you to Open Science, and what it means in the context of this project, let's talk more about what you're going to be doing. In a nutshell, you are going to use the scientific method to learn something about an organism, in the controlled setting of a lab experiment.

We expect that you have been exposed to the principles of the Scientific Method before this, while you were still in high school. As such, we are assuming that you have a little bit of prior knowledge to draw from. However, we need to update traditional representations of the scientific method by including the steps of a registered report. The diagram below shows the steps in a nutshell.

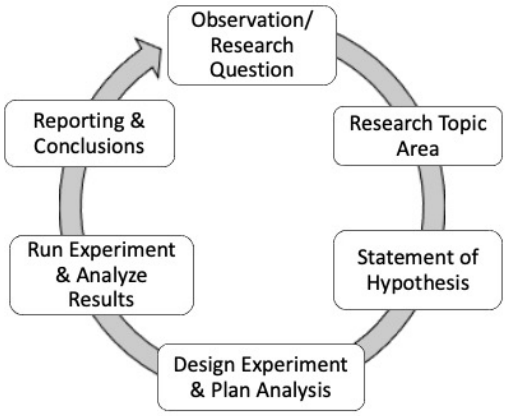

Figure 4: Figure 4. Visual representation of the steps of the Scientific Method.Image by Efbrazil, licensed under CC BY 4.0

In the following sections, we will discuss each step of the scientific method, and how it applies in the context of this research project.

Research Question

The first step in any research is to decide on a *problem* or focus for the investigation. This often happens naturally when you observe something that interests you, but that you can't entirely explain. In your head, a question forms, such as "why did it do that?"

Deciding on a problem to explore and formulating a question is your very first task.

As all good scientists, the question you ask should be informed by existing knowledge and relevant research. A review of the published scientific literature is a good way to do this.

For this research project you will be using mealworms to conduct your study. Your research question must include the effect of "x" on some physiological parameter such as survival, some aspect of behaviour, some aspect of vision or colour vision, or some aspect of hearing. Your TA will be available to help you consider your options as you work on this.

Statement of Hypothesis

A **hypothesis** is an unproven explanation for the observed phenomena. In its simplest form, a hypothesis is an "educated guess" or intuitive hunch that is proposed as a possible answer to the question you're interested in answering. There's a couple of things to know about hypothesis building before you get started:

A hypothesis is not a question, it is a statement

For example, "over a given time period, plants will grow taller at higher temperatures" is a hypothesis, whereas "over a given time period, will plants grow taller at a higher temperature?" is a question. They're generally related, but they're not the same.

A hypothesis must be testable

The hypothesis does not need to be "correct" (after all, there's really no way to know that at this point) but you do have to be able to test whether it is correct or not. In our example from above, we can test the hypothesis by growing the plants at different temperatures, and measuring their heights after a set amount of time. Thus, we have a way to measure the effect of interest and test our hypothesis.

A hypothesis comes before the experiment, not the other way around

We call this an *a priori* hypothesis, meaning that we made the hypothesis before we ran the experiment and learned the answer. Sometimes, because we want to show that we knew what we were doing, we feel the need to change the hypothesis we started with, so that it better reflects the results we got. This is known as HARKing (Hypothesizing After Results are Known), and is not considered to be good science. It's important to present your hypothesis as you originally developed it, and then discuss what you have learned about the topic based on your testing of that hypothesis.

Design Experiment & Plan Analysis

Note

BIOL202 is the course designed to teach you the foundations of biological statistics, including experimental design, how to decide what statistical test to use when, and how to visualize data. Here we are giving you only enough information so that you will be able to complete your BIOL116 research project. We recommend that you take BIOL202 as early in your degree as possible, as it will make all of your lab work easier.

Designing Your Experiment

Once you have a research question and a clear and testable *a priori* hypothesis, you can design an experiment to test it.

Variables

Ideally, experiments should be conducted in such a way that the experimenter has control over every **variable** that might have an influence on your results (in reality this is much harder than it sounds!). A variable is any factor that might affect the outcome of the experiment. The experimenter therefore manipulates the **independent variable** and observes the effects of this manipulation on the **dependent** (or **response**) **variable**.

For example, if the goal is to determine the effects of temperature on plant height after 14 days, then height is the dependent variable because it "depends" on the temperature to which the plant is exposed. All other variables must be controlled or held constant, to the extent possible. Temperature is the independent variable because that is the variable which is manipulated by the experimenter.

The ideal way to perform such an experiment is to arrange a set of tests that are identical in all ways (light, soil moisture, etc.) except for the one specific factor that is being tested (in this case, temperature). Thus, for our plant experiment, a greenhouse with multiple temperature-control chambers would be ideal; each chamber would host a different temperature "treatment group". It's worth noting here that we don't always have the option to do what's ideal, but that doesn't mean we shouldn't work to control as much as possible. We should also consider carefully how the things we didn't or couldn't control might be affecting the results of our experiment.

Control

Crucially, one of the treatment groups must serve as a **control** group against which all other treatment groups are compared. **The importance of the control group cannot be over emphasized**. It is essential to know how the system you are investigating works under normal circumstances (i.e., before you started messing with it), before you can be sure the results obtained from the experimentation are actually due to the manipulation of the independent variable(s).

To continue our example above, if you wanted to investigate the effects of temperature (the independent variable) on the height of plants after 14 days (the dependent variable), you would measure the heights of plants grown at their normal, expected temperature (most likely room temperature) as the control group, and then compare the data collected from this group to the heights of plants exposed to higher and / or lower temperatures, depending on what you're hoping to learn. The control group would provide the "normal standard" against which the other treatment groups would be compared.

Sample Size

Another important rule governing experimentation is that each treatment group (which includes the control group) should include a decent number of individual test subjects or "replicates" (and ideally an equal number of individuals in each group). The more replicates you include in your experiment, or in other words the larger your "sample size" (indicated by the letter n) per treatment group, the more confident you would be in your results and the more power your study has. However, in most situations, increasing the number of replicates increases the cost and / or logistical difficulty of the experiment.

The key is to have a sufficient number of replicates per group to ensure your experiment has the **power** to detect meaningful treatment effects (if they exist). Determining what the minimum sample size per group should be is beyond our scope here, but for our purposes you can assume that three is the bare minimum, and ten or more is desirable.

Variation & Random Assignment

Biological variation is the inherent differences among organisms in a study that arise due to differences in genetic makeup, age, sex, health, etc. This natural variation has the potential to obscure or confuse experimental treatment effects. Thus, it is important to attempt to minimize this variation when designing your experiment (e.g., by using organisms of the same age, sex, etc.). Even when potential sources of variation are accounted for, it is crucial that subjects be **randomly assigned** to the treatment groups, so that any inherent variation among individuals will be distributed at random among all treatments, including the control group.

Planning Ahead

Raw Data

The data you collect from the experiment are generally called the **raw data**. You should always make sure that you have a copy of these raw data stored somewhere safe, so that you or someone else could start the data analysis from scratch should the need arise (like, say, if your files got corrupted or deleted by accident). Saving a safe copy could be as simple as taking a picture of your recorded data in your lab notes with your phone, or keeping a copy of a file on both your computer and a usb stick.

How do you plan to save a copy of your raw data?

Checking for Errors

The next step is to plan how to check for any mistakes that might have been made when recording your observations, and how to deal with them. For instance, if one of your data points is an order of magnitude larger than all others (e.g. a ”100” instead of a ”10”), this is likely a typo. If so, then simply state this and make the correction. But sometimes you'll see an observation that appears unusual compared to the other data points, and it's not a mistake. These are sometimes considered ”outliers”, and you need a plan in place to deal with these outliers. For instance, an honest and transparent approach is to conduct any analyses you do both with and without the outliers included, presenting both sets of results. If the exclusion of the outlier(s) doesn't change your conclusions, then great! But if it does, then you'll need to discuss this. The key is to have a clear plan in place, and to document what you did, and be honest and transparent about it!

Typos and outliers are often best revealed using graphs; they'll show up as observations that are far from the other observations in your graph. Thus,

visualizing your raw data with effective graphs should be the next step in your plan.

Visualizing & Describing

Your experimental design determines what type of data you will collect, which then determines the appropriate method for describing, visualizing and analyzing the data.

For your experiment, the data you collect will depend on the question you're exploring, and the hypothesis that you're testing. As such, the way that you will need to describe, present, and analyze the data will likely be different than your classmates', since their hypotheses will be different from yours.

Types of Data

In our example experiment, our response (dependent) variable "plant height after 14 days" is a continuous numeric variable. Additional examples of continuous numeric variables are temperature, weight, time, or distance. If instead (or in addition) we had decided to measure the number of leaves on each plant after 14 days, then this would be a discrete numeric variable, as it can only take on discrete values. Another example of a discrete numeric variable would be "number of hairs on the thorax" of a fly, or number of petals on a flower.

Perhaps the species of plant we opted to use in our experiment can produce different colours of flower on different plants. If this is something we planned to measure, then the "flower colour" (red, pink, white) produced by each plant would be an example of a nominal categorical variable. Another example of a nominal categorical variable would be "birth country", or "hair colour". Lastly, if we had planned to judge the "odour strength" of the flowers, we might have scored odours as "weak", "moderate", and "strong", which constitutes an ordinal categorical variable, because the categories have a logical order to them.

But what about our independent variable? In our example experiment, temperature is the independent variable, and let's say we subjected the plants to three different temperatures: 10, 20 (control), and 30 degrees Celsius. Strictly speaking, temperature is another example of a continuous numeric variable. However, in our experiment we are manipulating the temperature to be exactly 10, 20, or 30 degrees Celsius. Thus, our independent variable can be considered a discrete numeric variable, or in practice, it could also be treated as an ordinal categorical variable (either approach would be ok). In human health research, experimental studies often test the effects of different drugs on some health outcome, in which case "drug type" would be an example of a nominal categorical independent variable.

Types of Experiments

The example experiment we've been describing is a type of measured-response experiment, in which a numeric response variable (plant height) is measured in relation to a manipulated, independent treatment variable (temperature) that, in our case, is handled as a discrete numeric variable or ordinal categorical variable.

An example of an experiment in which the response variable is a categorical variable is a choice experiment. Here, organisms such as insects or mice are presented with two or more categories of, say, food, to choose from. With categorical variables, what is measured and analyzed is the *frequency* of the different categories. For example, consider the *a priori* hypothesis that mice prefer high-protein food over low protein and high fibre foods. One could implement a single choice experiment in which 20 individual female mice (of similar age and health) were each independently provided 2 minutes within an experimental "arena" (the apparatus) to make a choice between the 3 food types. Across these 20 independent "trials", the researcher tallies the *frequency* with which each category of food is chosen. In this example, the dependent variable is "food type", and there is no independent variable.

Deeper Dive: Optional

If our *a priori* hypothesis had been that female mice show a preference for high protein foods whereas males do not, then we could have randomly selected 20 male mice and 20 female mice to undergo the same type of choice experiment. In this case, the "sex" of the mouse would be the independent, nominal categorical variable, and "food type" would be the nominal categorical dependent variable.

Planning the Analysis

Although it is best practice to make all your data available (e.g. as a table in an appendix), it is also important to summarize and describe your data for the reader, and to present your summaries in a table. The focus here is describing and summarizing the data from the dependent variable. As with graphs, the best approach for describing data depends on the data types.

Note that if you have an independent variable, and if it varied at all during the experiment despite being under your control (most instruments will vary a bit), it is important to plan to describe this also, though often this information is placed in an appendix. Since you'll be presenting your work as a poster, you should be prepared to answer questions about this, in case someone asks you.

Descriptive Stats

In these section we will cover the following desriptive stats:

- mean
- median
- mode
- variance
- standard deviation
- inter-quartile range
- proportion

Scenario 1 Dependent Variable is Continuous Numeric

In this scenario you describe your dependent variable with a pair of descriptors: a measure of centre and a measure of spread.

Measures of centre include the arithmetic mean (a.k.a. the average), the median, and the mode. Most of the time we use the average, and it is meant to give an idea of a "typical" value in your dataset.

On its own, a measure of centre is not particularly useful, because one doesn't know if all data points are close to the "typical" (average) value, or if most of them are quite different from the average (in which case your average value isn't particularly "typical"!). This is why we need to describe the spread of your data too. Measures of spread include the variance ($s2$), the standard deviation ("s"), and the inter-quartile range (IQR). The IQR is the difference between the values in the dataset that lie at the 25th and 75th percentiles (or first and third quartiles). We do not discuss the IQR further here.

You'll learn in BIOL202 that characteristics of your data dictate which pair of descriptors are best suited for describing your data. For now, you should know that the average should be paired with the standard deviation, and this pair is typically the preferred pair to use. The median should be paired with the IQR. The mode is less commonly used in biology.

Variance s^2 is a measure of how far, on average, data values deviate from the mean. A small variance indicates the data are tightly clustered around the mean. The larger the variance, the more spread out the data. Variance is calculated by summing all the squared deviations from the mean (a deviation is the difference between an individual measurement and the mean) and dividing this sum by the number of data entries minus one.

$$s^2 = \frac{\sum (x_i - x)^2}{n - 1}$$

Standard deviation s is simply the square root of the variance.

$$s = \sqrt{(\frac{1}{n-1}) \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

Calculating and presenting your descriptive statistics

PRO TIP: Most scientific calculators have a "DATA" mode that includes a set of functions for calculating descriptive statistics such as the average and standard deviation. But it is still important that you understand how to do calculations by hand, and what the descriptors represent.

For detailed instructions on how to calculate the mean and standard deviation by hand, consult this web resource. If you're keen, you can consult these instructions on how to use the "R" software to do the calculations.

Based on the UBCO Biology Guidelines for data presentation, measures of spread should be reported to one more decimal place than the number of decimal places that the data entries contain.

If your experiment included a treatment variable that is categorical, like in our plant experiment where the three temperature treatments are handled as ordered categories (10, 20, and 30 degrees), then you should plan to calculate

your descriptive statistics on the dependent variable (plant height) using the data from each of the treatment groups separately. So in our example you'd calculate the average and standard deviation of plant height for each of the three groups of height measurements, and report these in a table as described in the Procedures and Guidelines document.

Scenario 2: Dependent Variable is Categorical

Recall that for categorical dependent variables, like "food type" in our example choice experiment, we tally the frequency with which each category occurs. Imagine that the low protein food was chosen by 6 of the mice, the high fibre food was chosen 4 times, and the high protein food was chosen 10 times. These data are very straightforward to present "as is" in a table, but we should also calculate the main descriptive statistic for categorical data, called the "proportion" p. That is simply the frequency of the particular category (say, high fibre food) divided by the total number of trials (or total sample size), here 20. Thus, for the high fibre food, the corresponding proportion $p = 4/20 = 0.2$. The proportion value should always fall between 0 and 1. You should plan to report both the raw frequencies of each category alongside their corresponding proportions.

Be sure to report your sample size n for each treatment group, regardless of what type of variable your dependent variable is. Also make sure to tally any missing values in any of the groups (e.g. that may have arisen due to problems during the experiment)

Deeper Dive: Optional

If your choice experiment includes an independent categorical variable, such as sex (male / female), then you should calculate and report the raw frequencies and corresponding proportions for each category of the dependent variable (food type) for each category of the independent variable (here, male and female).

Visualizing Data

For your project you will use a Shiny app to construct the appropriate graph(s).

For common measured-response experimental designs, in which a continuous numeric dependent variable is analysed in relation to a categorical independent variable, the appropriate way to visualize the data is with a boxplot or stripchart.

For study designs in which the dependent variable is categorical, as in the food choice experiment, the appropriate way to visualize the frequency data (the tallies of the different food types chosen) is a bar graph.

Deeper Dive: Optional

When there is both a dependent and independent categorical variable in the study design, then one can use a grouped bar graph or a mosaic plot to visualize the data.

Creating Effective Graphs

Regardless of the type of graph used, the independent variable (i.e., temperature in our example) is always placed on the horizontal (or X) axis and the dependent variable (i.e., height after 14 days) appears on the vertical (or Y) axis. For bar graphs, the y-axis will have a label "frequency" (or "count"), as that is what was measured for the categorical dependent variable. All figures must have a number (and be numbered sequentially) and a detailed title (called a "figure heading") that are placed below the figure. For more details consult the Biology Procedures and Guidelines document.

Statistical Tests

Now that you have decided what graph will be most appropriate for visualizing and presenting your data once they're collected, and have planned how to describe and summarize your data, it is time to make the key decision on how to best analyse your data to test your hypothesis. This decision is based on the experimental design and the type of data involved. You'll learn in BIOL202 the formal statistical tests that are most appropriate for your study. For now, we'll use less formal but generally effective approaches.

First, we need to cover some foundational statistical concepts. We'll do our best to keep things simple, but the reality is that there's a lot to know to make these decisions appropriately, so at the end of the day there's a limit to how simple this can be, and still be correct. Try not to worry if it doesn't all make sense at first. Take it slow, read it more than once, work with your group, and ask lots of questions of your classmates and your TA… after all, collaboration and teamwork is what science is all about!

Inferential Statistics

In the previous section about planning to describe our data, we learned that the average (mean) was a useful descriptor of a "typical value" for a continuous, numeric variable, and that a <a href="https://ubco-biology.github.io/Procedures-and-Guidelines/glossary#Proportion>proportion was the best descriptor for categorical variables. We were planning our "descriptive statistics". Now our goal is to plan for drawing *inferences* from our data about the world at large. We are embarking on *inferential statistics*.

We now need to cast the descriptors we calculate, like the mean, into a different light: no longer is it a simple, calculated, fixed value; now we consider it an *estimate* of some true mean in a population at large, and we need to recognize that this calculated value has uncertainty associated with it.

For instance, for our plant experiment, we planned to describe our data, including calculating the mean plant height for each of the three treatment groups. But when it comes time to analyse our experiment data, we switch modes and consider each of those means as *estimates* of their respective, true population means, specific to each treatment group. But wait, we don't have populations in our experiment! We certainly don't, and experiments never do. However, so long as we randomly assigned our plants to our treatment groups, then we can safely assume that the average response we observe among the individuals in a given treatment group is *representative* of what we would observe if we were to subject a different random sample of plants from the same population to the exact same experimental conditions.

We need to remember, however, that individuals within every population vary in many ways, and therefore if someone else were to conduct the experiment using the exact same conditions, but a new set of randomly chosen subjects (plants), randomly assigned to each treatment group, they will most certainly not get the exact same calculated values for the treatment means. This reflects something called "sampling error", and sampling error is entirely expected! This sampling error is what introduces uncertainty to our estimates. We must recognize that, even though we might have imposed strongly different treatments in an experiment, we can't simply interpret any resulting differences between treatment means at face value; we need to take account of the possibility that those differences could have arisen solely due to sampling error.

Null and Alternative Hypotheses

As scientists, we should approach new ideas and hypotheses with skepticism (even if we're actually excited about them!). We therefore undertake studies like our plant experiment with the working assumption that our research hypothesis is wrong, and we need to be convinced otherwise with good evidence. We recognize that sampling error is ever present, and that it could easily cause us to draw incorrect conclusions about our data.

To formally account for the potential influence of sampling error, and help guard against drawing incorrect conclusions about our study results, we need to formulate two statistical versions of our *a priori* research hypothesis: a null **hypothesis (H0)** and an **alternative hypothesis (HA)**. These provide clear and *testable* statements about what study outcomes would look like if the hypothesis was NOT supported (H0), and what would be observed if the hypothesis was supported (HA). In other words, the null hypothesis is what we expect to observe if only sampling error were at play. The alternative hypothesis is what

we expect to see if there was a biological effect at play that was strong enough to overcome the influences of sampling error.

Continuing with our plant experiment example, and recognizing that for this study design the suitable analysis is to compare average values of the dependent variable across treatment groups (see next section), we could write our null and alternative hypotheses as follows:

H0: The average height of plants grown under different temperatures does not differ after 14 days

HA: The average height of plants grown under different temperatures does differ after 14 days

Correspondingly, the null hypothesis represents the status quo, i.e. nothing going on, and it is this hypothesis that we objectively and directly test through experimentation and statistical analyses; only if the evidence is strong enough to reject the null hypothesis would we conclude that the data are consistent with the alternative hypothesis, which itself reflects what we expect to see if our research hypothesis were correct.

It is important to emphasize that a failure to reject the null hypothesis **does not mean that the null hypothesis is true; it simply means that there is insufficient evidence to reject it at this stage**. Likewise, evidence consistent with the alternative hypothesis is just that: evidence that is *consistent* with there being an effect of temperature on plant height. Only after sufficient and independent replication of the experiment should we conclude that our research hypothesis is true. After all, there are many factors that influence how robust your experiment is, and how likely it is to reflect the "truth". We call this the **power** of your study. Your methods, sample size, and the true effect size all influence the power of your study.

Significance & Confidence

The null and alternative hypotheses are typically formulated at the same time you decide which test is best suited for your study (see the next section). Also, at the same time you need to decide a clear criterion upon which to base your decision about whether the evidence is strong enough to reject the null hypothesis in favour of the alternative. This decision should be guided by a number of factors, but for the present purposes we'll go with the traditional approach, which is to guard strongly against making a mistake of a "false positive", i.e. rejecting the null hypothesis when in fact nothing was going on (it should not be rejected). Specifically, we set what's called a "significance level" (denoted with the Greek letter alpha) at 0.05, or 5%. This means that we're willing to make that "false positive" mistake at most 5% of the time, on average. In practice, this means that the evidence needs to be pretty strong to reject the null hypothesis. The corollary of the 5% significance level is something called the "level of

confidence". That is, using a 5% significance level is the same as having a 95% level of confidence in something. We'll see how the significance level and level of confidence are used below.

You might reasonably ask: "Hold on a minute: how come we use such an arbitrary and hard criterion for deciding when something becomes "statistically significant" (i.e. evidence strong enough to reject the null hypothesis)?" That's a fantastic question! It turns out that the practice of statistics is slowly moving away from this overly rigid approach. That, however, is an idea to be explored later in your degree in BIOL202. For now, however, we'll stick with tradition.

Test Statistics

Lastly, with your testable hypotheses and significance level stated, you'll need to figure out which statistical test is best suited to your study design (next section). All statistical tests calculate something called a "test statistic", and it is ultimately this quantity that is used to evaluate whether your criterion for significance is met. Specifically, based on the significance level you chose, and your sample size n, you are provided - in "statistical tables" - a "critical value" of the test statistic that, if met or exceeded, leads to the rejection of the null hypothesis in favour of the alternative.

The critical value demarcates an outcome that is sufficiently convincing - based on your significance level chosen - that it is unlikely to have arisen solely from sampling error. After stating your significance level, and deciding which test is appropriate (see below), it is good practice to write down what the critical value of the corresponding test statistic is. Using your experiment data, you will then calculate (or the computer will provide) an observed value of the test statistic, and assess this against the critical value.

Note

In BIOL116 you'll be using a Shiny app to conduct your analyses, and it will tell you whether you've exceeded the critical value or not. Thus, there is no need to write down the critical value in the planning stage.

P-value

When you conduct a statistical test using the computer or app, it will provide you with something called a "P-value". The magnitude of this P-value depends on how close or far your calculated test statistic value is to your critical value. They key thing to remember is that if you chose a significance level of 0.05 (the standard), then a P-value of 0.05 or lower indicates a "statistically significant" outcome, i.e. evidence consistent with your alternative hypothesis. Thus, this is complementary to evaluating your calculated test statistic value against the critical value of the test statistic.

Which Statistical Test to Use

In this section we will cover the following inferential statistical tests

- Student's *t*-test
- Analysis of Variance
- Chi squared (χ^2) goodness-of-fit
- Chi squared (χ^2) contingency test

Comparing Means Among Treatment Groups

Scenario 1: Continuous numeric Dependent Variable and Categorical Independent Variable

This scenario is typical of **measured response experiments** like our plant experiment.

In this scenario the typical approach is to compare the average value (the mean) of your dependent variable among the categories of the categorical independent variable (that is, the treatment groups).

Our approach aims to determine whether any of the treatment groups' means differ "significantly" from any of the other group means. At first glance, that seems straightforward: If the mean height of plants subjected to 30 degrees was 22.2cm, and the mean height of plants subjected to 20 degrees was 18.2cm, then clearly the warmer temperature treatment yielded a significantly greater average height, right?

We must remember: we're now conducting inferential statistics, so each of the treatment group means has *uncertainty* associated with it owing to sampling error, and that uncertainty should cause us some pause. We need to be convinced, with some level of confidence, that the difference we observe between the group means could not simply have arisen due to sampling error, but rather was most likely due to our experimental treatments.

We start by writing down our statistical null and alternative hypotheses:

H0: The average height of plants grown under different temperatures does not differ after 14 days

HA: The average height of plants grown under different temperatures does differ after 14 days

Next, as instructed in the previous section, we need to decide what "significance level" we wish to use. You'll learn in BIOL202 that there are many factors that influence this decision. For now, we'll follow the standard and use a 5% significance level. In short, this means that we're willing to make a false positive mistake in our conclusion at most about 5% of the time, on average.

Next we need to ask: how many categories (treatment groups) do we have in our categorical independent variable?

If there are only 2 categories (one of which should be a control group), then we perform something called a **"Student's t-test"**.

If there are more than 2 categories then we conduct an **"Analysis of Variance"** (ANOVA). Yes – a strange name for a test that analyzes means!

These tests, like all statistical tests, have some assumptions associated with them, and you'll learn more about those in BIOL202. For now, you will practice implementing these tests using a Shiny App.

As you'll see when you practice in the Shiny app, the test statistic provided by the Student's t-test is, you guessed it, a "t" statistic. If the value of "t" that is calculated using your experiment data exceeds or is equal to the "critical value" of "t", then the P-value provided by the computer will be less than or equal to 0.05 (assuming this was the significance level you chose), and you would reject the null hypothesis in favour of the alternative, and conclude that the average height of plants grown under different temperatures does differ after 14 days. If, on the other hand, your calculated value of "t" is less than the critical value, and correspondingly the P-value is greater than 0.05, then you fail to reject the null hypothesis and conclude that, at present, the data are consistent with the conclusion of no difference among average heights of plants grown under different temperatures. **Remember**, this doesn't necessarily mean the null hypothesis is true! We simply don't have any evidence to suggest it's false. The same procedure would be used for the ANOVA test, but in this case you evaluate the "F" statistic.

Comparing Frequencies to a Baseline Expectation

Scenario 2: Categorical Dependent Variable Without an Independent Variable

This scenario is typical of **choice experiments,** like the one described above with mice choosing among food types. The single categorical response variable is "food type". You can imagine that if there was a strong preference for one of the three food types, then that food type would be chosen more frequently by the mice than the others. Alternatively, if there was no preference, then one would expect all three food types to be selected with similar frequency.

As skeptical scientists, this latter "status quo" scenario should be our working hypothesis, and evidence would need to be strong and clear to convince us otherwise. We need a way to quantitatively test this.

For this scenario in which we have a single categorical dependent variable, we use something called a χ^2 goodness-of-fit test (when we say it, we often call it the chi-squared test, and pronounce the Greek letter, χ, as Kai), which quantifies

the "fit" of observed frequencies to those expected if nothing were going on. This test uses the χ^2 test statistic.

We now formulate suitably worded null and alternative hypotheses:

H0: The three food types are chosen with equal frequency by the mice (or something similarly clear).

HA: The three food types are NOT chosen with equal frequency by the mice (or something similarly clear).

Although the test is relatively straightforward to undertake, you can make use of shiny app to do the test. In brief, as the overall difference between your observed frequencies (from the experiment) and those expected by the null expectation increase, the value of your calculated χ^2 will increase. If it increases in magnitude to the point that it equals or exceeds the "critical value" of χ^2 that you would have established before, then the P-value will be less than or equal to 0.05, and you would reject the null hypothesis in favour of the alternative.

Pro Tip

If there are only 2 categories in the dependent variable, then the most powerful statistical test to use is a **binomial test**, but a χ^2 goodness-of-fit test will still work.

Scenario 3: Categorical Dependent Variable and one Categorical Independent Variable

This scenario is also typical of "choice experiments", and above we provided one example in which we hypothesized that female mice showed a food preference whereas males do not. In this case, we plan to conduct something called a χ^2 contingency test, also called a χ^2 test of association. For example, if indeed we were correct with our research hypothesis, then the evidence would show that a preference for food type is *contingent* on the sex of the mouse.

The appropriate null and alternative hypotheses are:

H0: The three food types are chosen with equal frequency by male and female mice.

HA: The three food types are not chosen with equal frequency by male and female mice.

An alternative but less effective wording that is common to see is:

H0: There is no association between food preference and sex.

HA: There is an association between food preference and sex.

The latter statements are more ambiguous with respect to quantitative predictions. Nevertheless, they are acceptable.

Again, we plan what significance level to use: 5% or 0.05. Based on this significance level, and on the number of categories in our dependent and independent

variables, we would figure out what the critical value of χ^2 is for our test. But in our case, we'll again use an online app for the test.

If our calculated value of χ^2 is greater than the critical value of χ^2, then we reject the null hypothesis in favour of the alternative, and conclude that "Food preference is contingent on the sex of the mouse, because males and females chose the food types with different frequencies."

For more details on how to report the results of statistical tests, refer to the UBCO Biology Guidelines for data presentation.

Reporting & Conclusions

This final step in the scientific method is to provide a straightforward description of your findings, and to interpret the results of your statistical analyses. The first step is to include clear and concise statements that clearly indicate the outcomes of your statistical analyses.

Specifically, the key statistical interpretation is whether your statistical analyses yielded a "P-value" that was less than your stated level of significance (0.05). This dictates the wording of your main statement of results, though in all cases your statement should reflect the wording of your null and alternative hypotheses. For our plant experiment, for example

> We found that temperature significantly affected plant height.

Or

> We found that temperature had no significant effect on plant height.

Pro Tip

The statistical results are typically summarized in parentheses at the end of your main statement of results, as follows

> We found that temperature significantly affected plant height (ANOVA: F = 4.42; n = 20 per group; P = 0.023).

The parentheses should include the type of test conducted, the value of the test statistic (provided by the Shiny app), the sample sizes in each treatment group, and the P-value (also provided by the Shiny app).

You should then add another statement that provides more information about your data, referencing your graph and any patterns it shows. For example

> On average, plant height was greatest within the highest temperature treatment, and smallest within the lowest temperature treatment (Figure 1).

Or

> Plant height varied substantially within each treatment group, and showed no consistent pattern among groups (Figure 1).

The key biological interpretation is whether or not the experiment yielded outcomes that were consistent with your research hypothesis and associated predictions, and what biological processes underlie your findings. Were all your expectations met? None? If you conducted multiple runs of your experiment, were their results all consistent, or did some runs yield different outcomes? If they differed, suggest why this might have been.

It is important to remember that scientific investigations often don't yield the anticipated results. If there are discrepancies between your results and those of others, or what you expected to find based on your reading of the scientific literature, this is the place to try and explain those discrepancies. As a general rule, this means looking at the published results of other scientists, and critically comparing the work they did to yours, to see if similarities make sense, and discrepancies can be explained. Doing scientific experiments well is extremely challenging, so don't be discouraged if mistakes are made along the way! The key thing is to document those mistakes and discuss how / if they may have influenced the outcomes.

When possible, you should not only repeat your experiment as many times as is reasonable (only if planned in advance), but also compare your results to those of other investigators working on the same problem. This helps you to determine how reproducible you would expect your result to be, which also helps provide evidence for how likely your results reflect the "truth". All information obtained from other sources, or any ideas that are not your own, must be properly cited in the body of your poster presentation and included in a "Literature Cited" or "References" list at the end. This is an essential part of science, and academic endeavours. No scientist ever works in a vacuum, and comparing to others is expected, so it's perfectly normal and expected that you will learn from the work of others, and cite them.

Maximizing Reproducibility

Ensuring your experimental design and your plans for analysis are sound *before* you undertake the research is crucial. Mistakes in experimental design, or in your plans for statistical analyses, can render results entirely meaningless. To avoid such costly mistakes, and also to promote transparency, scientists often submit their experimental design and analysis plan for independent review by peers, and once the design is finalized (often after several rounds of revision), they "register" it so that there exists a formal record of the original design for others to view.

For this project you will submit your initial experimental design and plans for analysis for feedback.

Workflow & Information Management

As mentioned earlier, when we introduced the concept of Open Science, your group is expected to follow proper file and data management practices. Having your files and data organized appropriately will save you time. The best rule of thumb for lab experiments is to NEVER assume that you will remember exactly what you did days, weeks or months prior… you won't, and your science (as well as your lab grade) will suffer. So being organized and meticulous is absolutely vital to the success of your project.

Considering how you are going to keep track of everything is one of the **first** decisions that you need to make as a group. If you don't write things down right from the moment you start developing your experimental plan, how will you be able to run the experiment the exact same way twice? For example,

- Are you going to use paper or electronic files?
- How will you share methods and data with each other?
- How are you going to keep track of the changes you make over time, in case you change your plan and need to back up to an earlier version of your experimental plan?
- Once you know the data you want to collect, how are you going to record it?
- How should that file be organized to make it as easy as possible to accurately record the data your experiment generates?
- How will you make sure you know which test subject/ trial each data point belongs to at all times?
- Once your data are recorded, how will you explore them to check for mistakes, such as typos? And how exactly will you deal with such mistakes to make the data "clean"?
- With the "clean" data in hand, were you able to implement the statistical analysis as originally planned? Or did you need to modify your analysis in any way?
- In **Lab 4** (online, asynchronous lab) you will be exploring best practices for file management and naming conventions. For now, here is an exam-

ple of naming a data file for an experiment investigating how mealworm movement is affected by the presence of light.

- 20200626_MealwormProject_Light-movement-data.csv
- Here is an example of naming a written proposal for the same experiment:
- 20200724_Mealworm-project_Proposal_V01.docx

As you design your experiment, and develop (and troubleshoot!) the method you expect to use, it can be very helpful to draw out your proposed methodology into a flow chart. This can help you better visualize what you want to do, and may help you realize where you need to think a little more critically about your proposed plan and how it will work.

PRO-TIP: visual representations of your methods work really well on posters! So, if you start working on it right from the start, it should be awesome by the time you're ready to build your poster!

As stated earlier, all your decisions about experimental design, analyses, and data management and cleaning should be made and documented (see below) *before* you collect any data, otherwise you may subconsciously (or consciously) let the data influence your actions, which will *bias* your conclusions.

Following these steps should ensure the experiment is **repeatable** by anyone that wishes to do so (including you!). Others will not get exactly the same results (because of sampling error), but they should have no trouble replicating the experiment - your documentation should be clear and thorough.

Conducting the Research

At long last, you've submitted your experimental design and analysis plans for feedback, made some adjustments that were "ok'd", and now you can finally undertake your research!

Keep lots of notes, especially with respect to any deviations from your original plans. We all know that stuff happens, so even the best laid plans can unravel. The key point is to document what you do.

If your analyses were meticulously planned, including the null and alternative hypotheses, statistical tests, significance levels, critical values, etc… they should be relatively straightforward to implement once you have your data collected. Indeed, many researchers have `R` "scripts" written in advance that they tested on made-up data, such that when the real data are available, the analyses practically do themselves!

Citing

Citing the work of others

Science is, at its heart, a collaborative endeavour. No scientist ever works alone, especially in the modern times we live in. As such, all scientists share ideas and information in one way or another. Information is our currency: we trade it, share it, and make it grow. This also means that giving credit to others for their ideas and information is a vitally important part of science, not to mention academic life and the Open Science movement. Again, you will learn more about Open Science and how it works next week, in the asynchronous material of Lab 3.

For your written work, every time you mention an item of information or any idea that is not your own, the source must be credited in the text. This refers not only to published material but can also include personal communications from colleagues and professors. In scientific writing, we avoid using direct quotations and footnotes wherever possible. We do not copy *verbatim* from our sources - i.e. copy & paste. Instead, put the source away when you write, so that you naturally rephrase the material into your own words. Finally, acknowledge the source using the appropriate style and format for the work you're trying to cite.

While many different formats exist for citing your sources, for the purposes of this project the APA 7th edition reference and citation style is to be followed. The Procedures and Guidelines: APA Citations has a quick reference guide and other resources that you can follow up with.

Academic Integrity & Copyright

Citing falls under two broader categories, academic integrity and copyright. Academic integrity is about honest, responsible conduct in academia. Copyright is the legal framework that governs how the things we make - the things we write and draw - can be copied and distributed by others. The Procedures and Guidelines: Academic Integrity and Copyright sections provide further guidance on each of these.

Closing Remarks...

Time to start doing science!

You will conduct your research project over the course of the term, alternating between in person sessions, when you will work with your group to plan and / or implement your research plan, and asynchronous online sessions, when you will work through Canvas modules to help you gain the skills and information you need to be successful in this project.

The weekly plan is written in your course syllabus, along with which weeks are in person - called synchronous in the syllabus - and which ones are online - called asynchronous in the syllabus. Most weeks you will have something to submit that is related to this project in one way or another (except for the weeks you're actually running your experiments and collecting the data!. You will get more information about this in you first in-person lab this week.

One final thought… while science is serious and noble and logical and all of the rest… it's is also exciting and fun! We can be rigorous and careful in the work we do, and also be really excited by the cool questions we're exploring. So even while you are doing your best to learn the principles of science, it is our sincere hope that you also find the joy in research and discovery!

Happy Sciencing!

Assignment

Please use the following template for this assignment:

20210813_Lab2b_Experimental-Design-PreLab-Assignment.docx (18 KB)

NOTE This assignment is due at the start of your next in-person lab (i.e. Lab 4).

Lab 3

Open Science

Last updated 2021-08-30

This week's lab content can be found here. You are asked to cover Part 1, Principles of Open Science.

The accompanying quiz can be found in Canvas.

Lab 4

Experimental Research Pilot

Last updated 2021-08-30

This week you will be trying out your experimental design.

This week is all about learning! Be prepared that things may not go as expected and that is okay. This week is about taking the time to figure out what works, what doesn't work and what we need to change.

This is all part of research but will help better prepare you for your next few weeks of data collection.

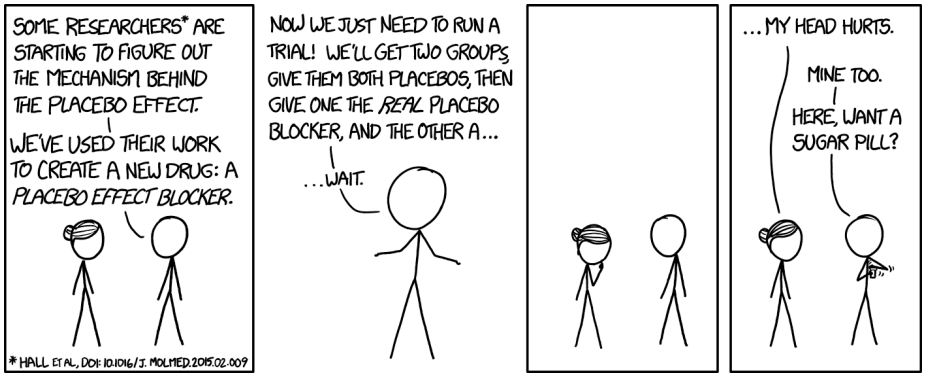

Figure 5: Figure 1. Placebo comic.Image by xkcd. Licensed under CC BY-NC 2.5

Questions

Remember, there are 20 of you and only one teaching assistant in your lab. You may be feeling anxious and frustrated at times as you may not be able to get the help you need immediately but please be patient.

Your teaching assistant will come around to provide whatever support they can, however there may be a bit of a wait. While you are waiting on your teaching assistant see if you and your partner aren't able to figure things out together. You may find you do not need your teaching assistant when you put both heads together.

Materials & Safety

All the materials needed will be along the far bench with some items being held in the fume hood. For those items in the fume hood please ensure you follow the guidelines indicated by your teaching assistant. There are safety issues that need to be considered and so we must ensure we are following all laboratory safety rules. If you are in doubt of how to dispose of something or where it should be placed please ask rather than guess.

Tidy Up

There are other students coming in immediately after you so please be considerate.Your station should look the same as it did when you arrived.Take the time to clean up your area and put everything away.

Have Fun

Yes, this may be a challenging week for you but we expect that it will also be a very enjoyable experience… you get to test your own theories! Have fun with it!

5

Naming Conventions

Last updated 2021-08-30

This week's lab content can be found here. You are asked to read

- Chapter 1: File and Data Management and
- Chapter 2: File Naming.

The accompanying quiz can be found in Canvas.

Lab 6

Data Collection

Last updated 2021-08-30

Now that you have conducted your pilot experiment and saw how it went, you are ready to collect some data.

Make sure you have worked through any major changes to your experimental methodology with your teaching assistant and you have carefully documented everything that happened during your pilot week (Lab 4). This information will come in handy when it comes time to explaining your results, but for this week you can just focus on collecting data for your research!

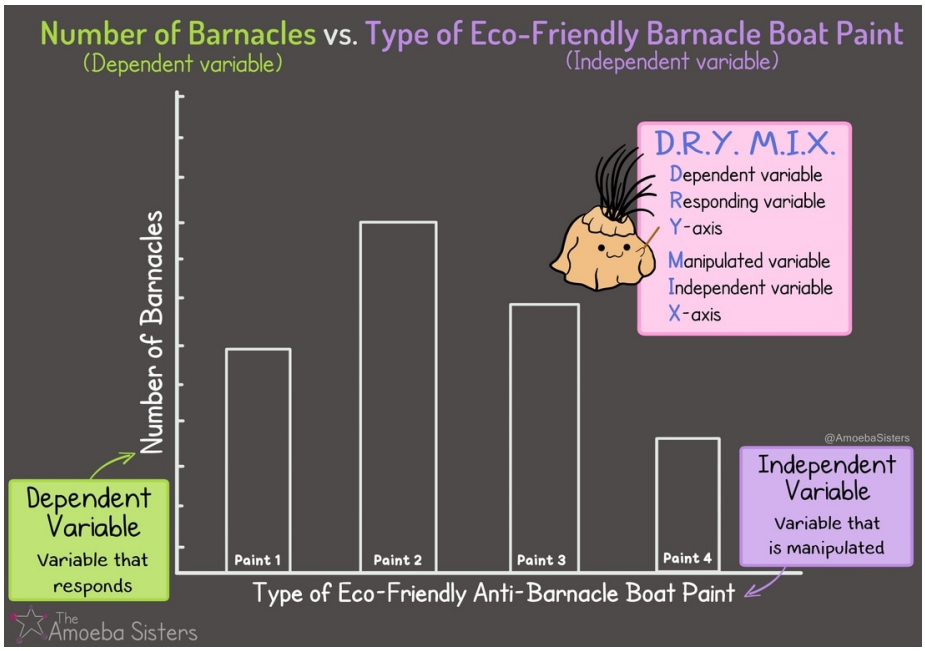

Figure 6: Figure 1. Types of variables comic.Image by Amoeba Sisters. Used in accordance with the creators' terms of use

Lab 7

Lab 7: Intro to R & Shiny Apps

Last updated 2021-08-30

This lab will have you putting into practice through visualizations and data tables much of what was covered in Lab 2B.

Data Tables & Figures

Before we tackle data visualization and analysis, we should talk about data presentation conventions.

The UBCO Biology faculty has a Procedures and Guidelines document outlining conventions for data presentation based on current "best practices" in Biology. Keep in mind that you may encounter small differences when working with data or reading the results of research from other disciplines. That being said, the aim of these conventions is to achieve consistency among faculty, instructors, and students in how data are summarized and presented within lab reports and research papers.

Before continuing through this lab, please take some time individually or with your groups to read sections 5.1, 5.2, and 5.4 of the UBCO Biology Procedures and Guidelines document. For this lab's assignment and all future assignments in this course, you wil be expected to follow these guidelines!

Intro to R

Visualizing data is a key part of science communication. Until now, it is likely that you've used or interacted with visualizations built using tools like Excel. In Biology, it is increasingly common to not use Excel, but instead to use an application, or programming language, called R.

R offers several advantages when it comes to building visualizations. R is what we call a scripted progamming language. When we use scripts to manage our data and create visualizations, we're engaging in reproducible workflows, specifically computationally reproducible workflows.

Computational reproducibility

What is computational reproducibility?

When you load data into an application like Excel, you click through a series of options to clean up or organize and then visualize your data. If you wanted to repeat this process, you would have to manually go through the same series of steps over and over again, which is a lot of mouse clicking. It's also very difficult to communicate your workflow from raw data to visualization as you would have to write down every step and do so in a way that was clear enough for someone else to reproduce exactly.

Working in a scripted programming language like R, instead of clicking buttons, we write code that tells the program how to clean, organize, and then visualize our data. When we want to repeat the process, we just run our script - or mini program - and it exactly, computationally, reproduces our workflow. It's less work for us and it's reproducible. And, since anyone we share our script with can read our script to see the steps taken to go from raw data to visualization, it's transparent. We just don't get this level of reproducibility and transparency with Excel.

Shiny Apps

In BIOL 116 we're not going to learn how to write R, we'll save that for later. What we are going to do is start working with learning to visualize data in an application built in R called Shiny Apps. Shiny Apps allow a user to interactively visualize a data set. It will also show you the R code that is running in the background to build the visualization. This will be your first window into R.

In this lab, we'll work with a couple of data sets that are preloaded into the Shiny App, so you can explore mapping different variables to a visual space.

The first is a data set about penguins, called `palmerpenguins`. From the authors: "The `palmerpenguins` data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica." Read more about the `palmerpenguins` data set here if you'd like.

The second is about cars, called `mtcars`. From the authors: "The data was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models)." Read more about the `mtcars` data set here if you'd like.

Getting Started

The Shiny App that we'll be using can be found at this link.

The following is a descriptive overview of what you'll find in the two panes of the application.

Left Pane

On the left navigation you have four menu options: Welcome; Upload Data; Plot; Descriptive Stats; Analysis.

Welcome

This is an overview of the app itself.

Upload Data

For now, ignore the option to upload your own data. Instead, use either the `palmerpenguins` data or the `mtcars` data, selected from the drop down menu in the Plot menu. In **Lab 9**, you'll upload your own data set from your experimental research project to build visualizations for your report.

Plot

Here we select the variables we'd like to visualize.

You'll want to first select a dataset, `penguins` or `mtcars`, and then select an `x` and a `y` variable from the drop down menus. Lastly, tell the Shiny App if your variable is `quantitative` or `categorical`. In Lab 2 we learned about each of these variable types, but here is a brief refresher:

Quantitative, or numeric, data may be either discrete or continuous.

- Discrete quantitative data are whole integers - population numbers are a good example of this, as we can't have half a person!
- Continuous quantitative data on the other hand are data that lies on a continuum - an example is temperature, where there are infinite potential temperatures between 25 and 26 degrees Celsius, but our data collection tool or convention determines to what decimal point we'll record a given temperature.

Categorical data, per its name, deals with categories of things and may be either nominal or ordinal.

- Ordinal categorical data has an intrinsic order where one thing is more or less than another - storm severity is often classified by stages - Stage 1, Stage 2 … Stage 5 - where Stage 1 is less severe than stage 5. We don't know how much more severe one stage is than the next - that is we can't quantify the difference - but we know there is an intrinsic order.
- Nominal categorical data has no natural order - the ordering of the colours blue, pink, or white makes no difference to the data or analysis.

Once you've selected your data sources and identified the data types you're working with, you'll be presented with a series of plotting options. You may also be presented with the option to group one of your variables. You will also be presented with the option to save a copy of your plot.

Descriptive Stats

Here you can run basic descriptive stats on the data that you've plotted.

Similar to the instructions under the Plot tab, here, you must first select a dataset, `penguins` or `mtcars`, and then select two variables from the drop down menus, `Variable 1` and `Variable 2`. Again, you'll need to tell the Shiny App whether each of these variables are `quantitative` or `categorical`.

Once you've made your selections, the Shiny App will automatically calculated the preferred descriptive statistics for the type of variables you have selected. For example, if you choose

- **two quantitative variables**, the Shiny App will provide you with the sample size `n`, mean, standard deviation `sd`, median, and inter-quartile range `iqr`.
- **one categorical and one quantitative variable**, you will be provided with the sample size `n`, mean, standard deviation `sd`, median, and inter-quartile range `iqr` for each group of your categorical variable.
- **two categorical variables**, then a frequency table will be displayed showing you how many subjects fall into each group.

Analysis

Here is where you will perform statistical tests using your selected data. The type of test performed by the Shiny App depends on the type of variables you select.

- **T-test**: This analysis is used when examining a single quantitative (numeric) response variable in relation to a single categorical variable that has only 2 groups.
  - When performing a T-test in this app, you will be asked for a few additional parameters.
  - Type in the confidence level you would like to use for the T-test. For example, if you'd like a 95% confidence interval, type in 0.95.
  - One assumption of the T-test is that the variance for each sample is approximately equal. However, the t-test used by this app (Welch's t-test) is somewhat robust to deviations in this assumption. For now, we will assume that both of your samples have equal variance. As such, please select 'Yes' when prompted for this option.
- **ANOVA**: This analysis is used when examining a single quantitative (numeric) response variable in relation to a single categorical variable that has more than 2 groups.
- **Fisher's Exact Test**: This analysis is used when testing for an association between two categorical variables. It is only used when both categorical variables have exactly 2 levels/groups. For example, if one variable is sex (male/female) and the other is survival (yes/no).
- **Chi-Squared Contingency Analysis**: This analysis is used when testing for an association between two categorical variables. It is only used when at least one of the categorical variables has more than 2 levels/groups. For example, if one variable is flower colour (pink/red/white) and the other is season (spring/summer/winter/fall).

Once you've conducted a statistical test, the Shiny App will provide you with instructions for 'Interpreting the Output'. For this course, the main output of interest is the p-value. Recall from Lab 2 that if the p-value is less than our pre-determined significance level, there is a significant relationship between the variables. However, if the p-value is less than the significance level, the relationship between the variables is insignificant.

Right Pane

Under each menu option (i.e. Plot), the right pane displays three sections: Instructions; Output; Source Code

Instructions

This green box shows detailed instructions for how to use that specific page. For example, under the `Plot` menu option, there are detailed instructions for how to create, visualize, and save your plot.

Output

This blue box shows the output that the Shiny App creates based on your selections. For example, under the `Plot` menu option, this box will show the plot you created.

Source Code

This orange box shows the script used by R to generate the output. If there are any packages that need to be loaded in order to run the script, they are shown within the `library()` function at the beginning of the script. You don't need to worry about doing anything with this part for now. Rather, the script is included to enhance the transparency and computational reproducibility of your analysis. For example, since you know the script used to create your plots, calculate descriptive statistics, and perform statistical analysis; if you or someone else wanted to reproduce your methods, they could load your data, copy this script into R, and re-run everything exactly as shown in the Shiny App.

Screencast Demo

The following is a brief screencast introduction to using the Shiny App with the included datasets.

Activity

This is your opportunity to play around with several different data types, visualizations and descriptive statistical summaries. The assignment will ask that you submit a couple of examples of both data visualizations and tables, all properly formatted according to Section 5 of the UBCO Biology Procedures and Guidelines document.

Assignment

Please use the following template for this assignment:

20210810_Lab7_Shiny-App-Visualization_Assignment_V1.docx (22 KB)

Grading Rubric

Question 1

1 mark for correct figure based on types of variable inputs.

- Two categorical variables = mosaic or barplot
- Two numeric variables = scatterplot
- One numeric and one categorical variable = stripchart (if < 20 observations in each group), boxplot (if > 20 observations in each group)

1 mark for figure caption that follows guidelines. At a minimum a caption:

- is located below the figure (0.25 marks)
- includes a figure number (0.25 marks)
- describes both variables (0.25 marks)
- includes sample sizes (0.25 marks)

Question 2

4 marks for a correctly formatted table.

- Correct values in the table (1 mark)
- Correct descriptive statistics were used for each variable type (1 mark)
- The heading is informative (1 mark)
- Sample sizes and units are always included (1 mark)

Question 3

For each scenario (4 marks total):

- 1 mark for correct p-value
- 1 mark for correctly determining significance (or lack thereof) based on p-value

Lab 8

Data Collection Continued

Last updated 2021-08-30

By now you should have already successfully run your entire experiment once, and collected some data.This is great!Take your time in your lab to attain as many trials as you can.The more trials you conduct the more assured you can be in the trends you are seeing.This is your last opportunity to collect your data so make the most of it.

Remember that you should be working to minimize the differences between today's trials and the trial(s) you ran in the last data collection period. Your methods should be as similar as possible - assuming, of course, your experiment worked last time.

Note

Making changes adds a new variable to your experiment, which will have an unknown impact on your results. If you must make changes for some reason, make sure that you carefully document everything once again.

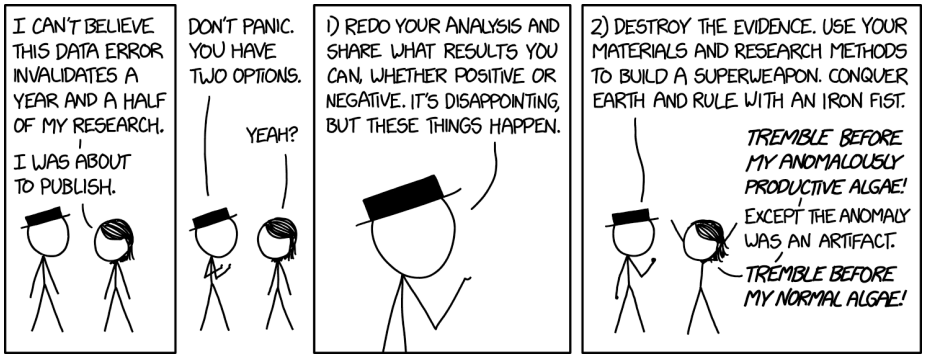

Figure 7: Figure 1. Data error.Image by xkcd. Licensed under CC BY-NC 2.5

Collaboration is a key part of science. During this time, you may want to start informally discussing what you are seeing with your partner, peers and / or your teaching assistant to see if you can explain what you are seeing.

- If you are seeing what you hypothesized, ask yourself, is it for the reason you thought? Why or why not?
- If you are seeing something different than what you had originally hypothesized why is that?

Scientists often use each other to discuss their findings and / or issues in order to help better understand what is going on. Your peers and your teaching assistants are resources for you so take advantage. Collaboration provides the opportunity to learn much much more about science and the scientific method so use it wisely.

Lab 9

Data Analysis & Shiny Apps

Last updated 2021-08-30

CONTENT PENDING

Lab 10

Poster Presentation

Last updated 2021-08-30

CONTENT PENDING