Lab 1: Hypothesis Testing: Resource Allocation and Seed Size

What you will be provided:

Seed pods from Palo Verdes
Dial calipers
Analytical balance

Objectives:

- Generate specific, testable hypotheses to explain why different pods have different numbers of seeds.
- Collect data that allows us to test these hypotheses.
- Perform statistical tests that allow us to reject or accept the **null hypothesis**.
- Learn important aspects of **summary** or **descriptive statistics**: mean, median, mode, range, standard deviation, variance.
- Learn how to use these descriptive statistics in order to test differences between experimental treatments or populations.

Introduction to Resource Allocation

Optimal investment in reproduction requires an allocation of resources to reproductive tissues that thereby become unavailable for vegetative growth, maintenance, etc. Palo verde (*Parkinsonia* sp.) trees produce pods that have from 1-5 seeds per pod. Each of these seeds requires an energy investment. In plants, each individual flower and fruit can be considered as arising from a separate system, meaning that each fruit is invested with fixed amount of resources. Thus, we might expect that flowers that invest in a greater number of seeds may end up with smaller individual seeds.

Introduction to Beetle Parasitism

The seeds of palo verdes are also a valuable resource for many animals, as they are rich in energy and protein. In particular, a number of species of seed beetles in the family Bruchidae (this family is sometimes merged into the family Chrysomelidae) make use of palo verde seeds for raising larvae. In some species, such as *Mimosestes* spp. (right), the eggs are laid on the fruit (pod) and the larva burrows into a seed. When the larva has finished its development, it chews its way out of the seed and the fruit, leaving an obvious exit hole. Other species lay their eggs directly on seeds and must therefore use an exit hole or other fruit damage to get in. Like plants, beetles will attempt to optimize their reproductive success, and one potential factor is choice of fruits upon which they lay their eggs.

Procedure

1. You will work in pairs for this lab. If time permits, you can use the available laptops to begin your data analysis. Your instructor will inform you if you are to look for beetle damage or to only examine the relationship between seed size and number (this will depend on the age and condition of the available pods).
2. Generate hypotheses for explaining how and why seeds might differ (or not) between pods with more than one seed and pods with a single seed.
3. If requested to do so, generate hypotheses for explaining how and why beetle parasitism rates might differ (or not) between pods with more than one seed and pods with a single seed.
4. If examining beetle parasitism, start by carefully examining each of your pods for beetle exit holes, which will be circular and about 2 mm in diameter. Separate the parasitized and unparasitized pods, and count the number of single-seeded and multiple-seeded pods in each group.
5. Remove the seeds from each pod individually and weigh the seed using the balance. Make sure you keep track of whether each seed came from a pod with a single seed or from a pod with more than one seed.
6. Record your data in the data sheet provided below.

DESCRIPTIVE STATISTICS

Measurements of Central Tendency

Central tendency refers to what a typical or most common value in a population. Several different measures are commonly used, depending on the data and question:

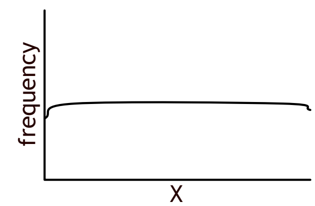
- **Mean:** the numerical average. Calculate by summing the measurements of your samples, then dividing by the number of samples. This is often written as $\overline{X}$. This is the most commonly used measure of central tendency.
  - In *Excel*: =AVERAGE(*array*)
- **Median:** the middlemost value. Calculate by organizing your readings in a progressive sequence, then finding the middle value. If you have an even number of samples, the median falls between the central values. The median is often used with nonparametric statistics, which are used when the data do not fit the assumptions of parametric statistics.
  - In *Excel*: =MEDIAN(*array*)
- **Mode:** the most common value. Easily seen in a progressive sequence, as used to determine the median. Note that while there can be only a single mean or median, there can be multiple modes.
  - In *Excel*: =MODE(*array*)

Distribution of the Data

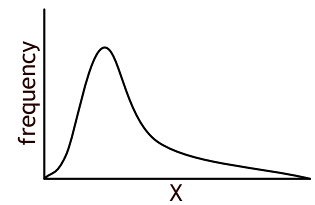
Nearly as important in describing data as measurements of central tendency are measurements of how variable the data are, how much deviation there is from the mean or median and the pattern of that distribution. The first step is to produce a **histogram**. Histograms plot some measurement along the horizontal (X) axis and the frequency of those observations on the vertical (Y) axis. When producing histograms, the measurements should be assigned into categories of equal width and there should be fewer categories than there are observations (around 10 categories is probably suitable for many of the data sets in this class).

The type of distribution can be determined by examining the shape of the histogram, as follows:

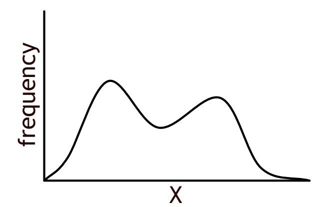
- **Uniform distribution:** the histogram has a flat shape; all values within the range are equally likely. Rare for continuous biological data.



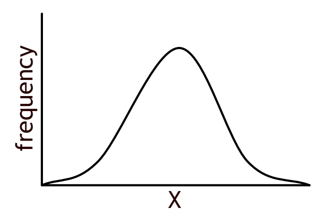
- **Skewed distribution:** the distribution has an asymmetric shape; one side rises steeply while the other has a longer "tail."



- **Bimodal** (or multimodal) **distribution**: the histogram has two (or more) separate humps.



- **Normal distribution**: the distribution has a symmetrical bell-like shape (called a "bell curve"). This is the most common type of distribution in nature. Most parametric statistics assume the data are normally distributed.



Dispersion around the Central Tendency

In addition to the shape of the distribution, we also want to know the width of the distribution, that is, how variable the data are. As with measurements of central tendency, there are several different measurements of dispersion, as follows:

- **Range:** "distance" between the lowest and highest measurements, often represented as lowest value–highest value. The range is sensitive to sample size (larger samples increase the chance of having a distant outlier.
  - In *Excel*: =MIN(*array*) and =MAX(*array*)

- **Standard deviation (*s*):** A measure of the dispersion around the mean within the normal distribution. This is a standard measurement of variation and takes into account different possible shapes of the standard distribution. 95% of the measured values lie within 1.96 standard deviations of the mean and 99% within 2.58 standard deviations of the mean. Its equation takes the form:

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

  Where:

  s = the standard deviation
  $\sum$ = summation across indicated values
  x = any one of the data points
  $\sum x^2$ = the square of each data point, and then sum all of these squares
  $(\sum x)^2$ = the sum of all data points, and then square that sum
  n = the number of data points (sample size).

  - In *Excel*: =STDEV(*array*)

- **Standard error (*SE*):** The standard error is the standard deviation divided by the square root of the sample size, as follows:

$$SE = {}^{s}\!/\!_{\sqrt{n}}$$

  Standard error can be useful for a quick and dirty test of statistical significance: if the standard errors of two means don't overlap, then it is likely that those means are statistically significantly different. Mostly used on graphs.

- **Variance (s^2):** the square of the standard deviation.
  - In *Excel*: =VAR(*array*)

STATISTICAL HYPOTHESIS TESTING

Beyond simply describing data, we will also want to test specific hypotheses. For example, we may wish to determine if two treatments or groups of subjects differ in some variable. If we measure two groups of subjects and discover that their means are not exactly identical, this by itself does not allow us to determine if a difference actually exists; the observed difference may simply be due to chance given the variability in the data. Statistical tests use the dispersion of data (such as the variance) to determine the probability that any difference or relationship observed in the data is due simply to chance.

To do this, we need two *statistical hypotheses*, the **null hypothesis** and **alternative hypothesis**. (Statistical hypotheses are different from *scientific hypotheses*, which are statements of possible general explanations for observed phenomena, such as your hypothesis from Procedure 2, above. The null and alternative hypotheses are more akin to formal statements of predictions.)

- **Null hypothesis (H_0)**: There is no real difference between the treatments (or no association between the variables).
- **Alternative hypothesis (H_A)**: There is a difference between the treatments (or an association between the variables).

Statistical analysis involves determining the validity of the null hypothesis. If we reject the null hypothesis, then we are left with the conclusion that the alternative hypothesis is correct. In deciding if the null hypothesis is true or not, we can either make a correct decision or an error, as follows:

		Null Hypothesis was:	
		Accepted	Rejected
Null Hypothesis is:	True	*Correct decision*	*Type I error*
	False	*Type II error*	*Correct decision*

Scientific hypothesis testing generally takes the conservative approach: it is better to make a type II error (conclude no effect when in acuality there is one) than a type I error (conclude there is an effect when in actuality there is none). Thus, statistical tests measure the probablity (***P*-value** or simply ***P***) of making a type I error. Unless that probability is very low, we accept the null hypothesis. How low a probability is low enough? The (arbitrary) standard cutoff point, or critical level (designated α) is 0.05 or less. In other words, we accept the null hypothesis unless there is less than a 5% chance of making a type I error by rejecting it.

The majority of statistical tests serve to determine the *p*-value or probability of making a type I error. If the value of ***P*** is less than α, the critical value (usually 0.05), we can reject the null hypothesis and thus accept the alternative hypothesis. Which test one uses depends on the type and characteristics of the data and the hypothesis being tested.

Several statistical tests are set up on an *Excel* worksheet that can be downloaded from the course web site. This worksheet has a few commonly used statistical tests and covers most of the tests that will be needed for the labs and your independent project. The following key will help you choose the test (it applies only to those given on the Statistical Analysis *Excel* file), but don’t hesitate to also ask your instructor for assistance.

A) Type of Data:
- Measurements[1] — **go to B**
- Counts — **go to G**

B) Type of Hypothesis:
- Comparison of means among two or more samples/treatments — **Go to C**
- Determine if there is an association between two variables (i.e. does one variable change in a predictable way as the other variable changes?) — **Go to F**

C) Number of samples or treatments:
- 2 — **Go to D**
- 3 or more — **Go to E**

D) Are samples paired (e.g. multiple testing of the same subject) or independent?
- Independent — Use *Student's T-test*
- Paired — Use *Paired T-test*

E) Are samples matched (e.g. repeated on same subject) or independent?
- Independent — Use *One-way Analysis of Variance (ANOVA)*
- Matched — Use *Repeated Measures ANOVA*

F) Are both variables independent?
- Yes — Use *Pearson's Correlation*
- No — Use *Linear Regression*

G) Are the expectations known?
- Yes — Use *Chi-square Goodness of Fit Test*
- No — Use *Chi-square Contingency Test*

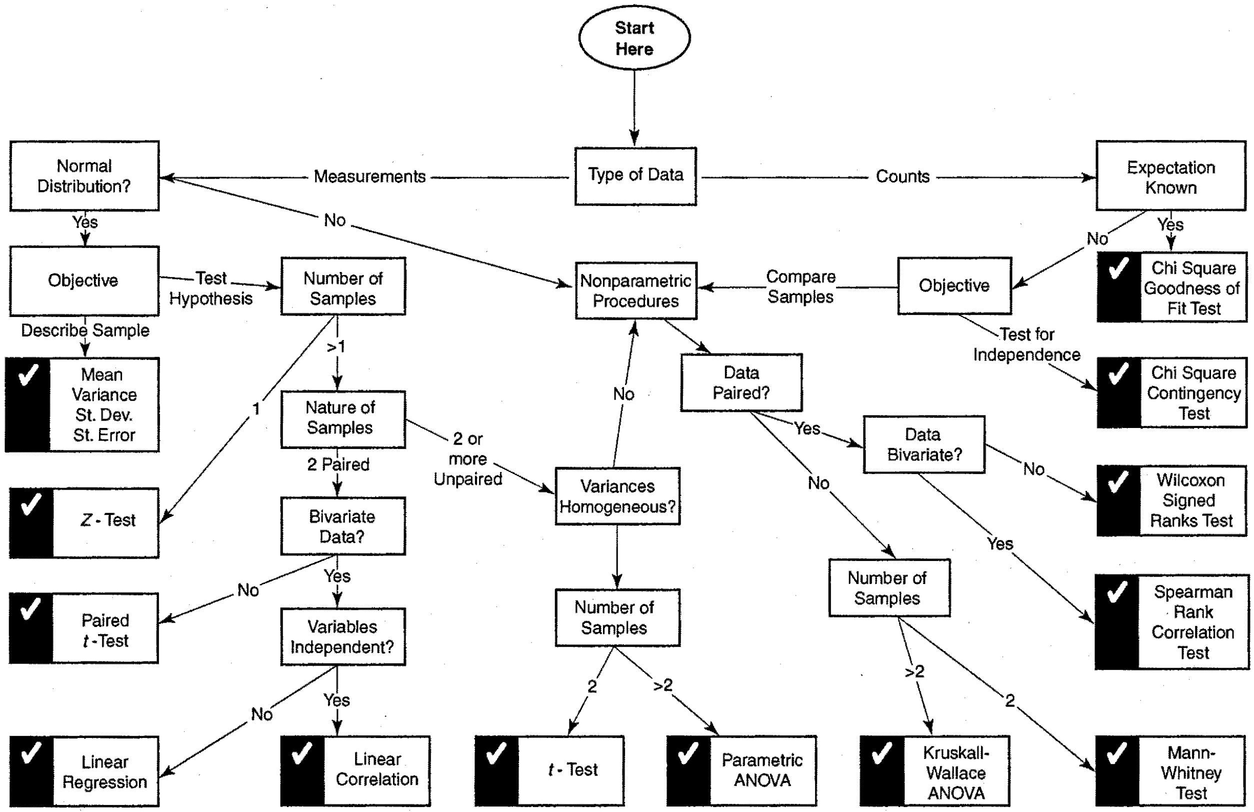
The diagram on the following page outlines even more choices, including non-parametric tests for if your data are not normally distributed.

For the comparison of seed size from single-seeded and multi-seeded pods, the most appropriate test would be the Student's t-test, since the data involve measurements, we want to compare the means of two samples, and the samples are not paired. While the t-test statistic can be calculated by hand, it is rather tedious to do so, and much easier to do in *Excel* or statistical software. Therefore, the equations will not be presented here.

For determining if beetles are more likely to parasitize single-seeded or multiple-seeded pods, the most appropriate test would be the Chi-square contingency test, since these data involve simple counts of fruits in two types of categories. The number of parasitisms expected under a null hypothesis are not known and would be based on the relative number of each category of fruit type.

---

[1] Measurements can include counts if the counts are data collected from multiple subjects. For example, if the number of plants are counted in multiple plots, each plot (not each plant) is a subject, in which case it is possible to average the number of plants per plot.



Additional Considerations.

When performing statistical tests, one usually first calculates a test statistic (such as a t-value, F-value, or χ^2 value) which is compared to a probability distribution to determine the actual p-value. This last step can be done by looking up the statistic in an appropriate statistical table or it can often calculated by statistical software. Two additonal pieces of information are often necessary when looking up (and subsequently reporting) statistical results:

- **Degrees of freedom (DF or df):** Generally, you will have n–1 degrees of freedom, where n is the sample size. This is because that is the number of sample values that can vary independently. Think about it, once the value of the mean has been fixed, if you have n–1 values, then you know what the last value is! If you have two samples, such as when comparing two means, the total degrees of freedom are sum of the degrees of freedom for each sample (e.g. $n_1 + n_2 - 2$).
- **Number of tails**. A *two-tailed test* is non-directional: deviation can occur in either direction. This is the more common situation and all of the tests in the statistical analysis worksheet are two-tailed. In contrast, in a *one-tailed test* the deviation can only theoretically occur in one direction. the null hypothesis would only be rejected if the results deviated in one direction but not in the other.

Reporting Statistical Results.

If we are able to reject the null hypothesis of no difference between the means, then we say that the means are *statistically significantly different*. On the other hand, if the null hypothesis was not rejected, then we say that there was *no statistically significant difference* between the means. (Although it is common to leave out the word statistical, it should be emphasized that statistical significance, or lack thereof, does not necessarily mean the difference is biologically significant.) In most journals, statistical results are presented in a specific shorthand, as in the following examples:

> "Peanut M&Ms are significantly larger than plain M&Ms (Student's t-test, t = 3.071, d.f. = 40, P = 0.0038)."

> "There was no statistically significant difference in the relative number of each color in bags of plain and peanut M&Ms (Chi-square contingency-test, χ^2 = 3.51, d.f. = 4, P = 0.48)."

Note that both examples begin with a *verbal description of the key results* (whether or not there was a difference/relationship/association and in what direction). The statistics provide the evidence to support the validity of the observed difference or pattern, and are meaningless without reference to what pattern or relationship that is being tested. The statistical information typically includes the name of the test used, the test statistic (in these examples, t and χ^2), the degrees of freedom (or sample size), and the p-value. If p-values are very small or not calculated exactly (for example if they were looked up on a table), then they can be represented as an inequality; for example, $P < 0.05$ or $P < 0.001$. Also note that there is no mention of the null or alternate "hypotheses" as these are just part of the logic of statistical testing and are understood when reporting if a test is statistically significant.

Assignment:

Record your data and answer the questions on the accompanying worksheet. This assignment will be due in your next lab.