

ANALYSIS OF INVERTEBRATES, MEASURES OF DISPERSION AND CLASSIFYING VARIABLES BI 311

Laboratory 03 and Homework 03

Introduction:

Next to microbes and plants, the most dominant life forms on the planet are arthropods. There are more species of arthropods than any other zoological group. For ants alone, if the mass of all the ants on the planet were measured they would outweigh the mass of all the humans. Beetle species predominate in the terrestrial environment. More than 160 beetle species were found in a single tree in Panama. Insects make some of our worst pests and transmit disease but also are critical for ecological processes like pollination. For many ecosystems, we can monitor the health of the system by keeping tabs on the richness, abundance and diversity of the arthropods. Arthropods are also useful for field studies as they are easily encountered compared with most species of vertebrates.

Skills:

1. Learn invertebrate sampling techniques using the Hess sampler, plankton net, Wilding sampler and Berlese-Tullgren funnel.
2. Learn how to use EXCEL to estimate the range, variance and standard deviation to assess dispersion (variation) in data.
3. Learn how to use EXCEL to calculate the standard error and confidence intervals to assess confidence in the estimated mean.
4. Learn how to distinguish different types of variables in preparation for inferential statistical analyses.

Materials:

- Hess samplers
 - Macroinvertebrate key
 - Plankton nets
 - Wilding sampler
 - EXCEL file titled "SVLizards"
 - PDF of Hokit and Branch 2003
-

EXERCISE 1-USING MACROINVERTEBRATES TO ASSESS BIOTIC HEALTH

Background:

Streams and rivers are highly prized in Montana for their recreational and agricultural value. Rafting, kayaking and fishing all require clean flowing water. However, historic mining activity (mining before the 1960's) has challenged Montanans to reclaim rivers and streams. **Acid mine drainage** has polluted many of Montana's favorite trout fisheries including the famous Blackfoot River. Mining activity exposes mineralized bedrock material and can lead to a decrease in pH with an increase in heavy-metal concentrations. Arsenic, cadmium, lead and zinc are artificially high in many Montana rivers and have led to destruction of viable fisheries.

Living in the 21st century, we may think hi-tech solutions are needed to solve the water problems in Montana but, ironically, several low-tech solutions are proving to be very effective. Many of Montana's rivers and streams are being reclaimed using settling ponds and artificial wetlands. We have also found that **biological monitoring** or **bioassessment** (using biological organisms to assess stream health) is often better at identifying stream problems than alternative hi-tech gadgets. Hi-tech probes are usually restricted to measuring one factor (e.g. pH or conductivity) over a short period of time. Biological organisms must live in the stream environment their entire life and deal with all of the interacting chemical components. For example, **macroinvertebrates** (stream invertebrates larger than 1000 microns) have been used to assess the recovery efforts of many rivers (McGuire 1999). Together, biological monitoring and **bioremediation** are cleaning up most of Montana's polluted streams and rivers.

Field Procedure:

1. Each team will use a Hess sampler to obtain macroinvertebrate samples on Ten Mile Creek. Operation of the sampler will be demonstrated but generally consists of randomly selecting a sampling point, placing the net at a 45° angle in the stream channel, and then stirring up about 0.5 m² of the substrate directly upstream from the net to allow the invertebrates to drift into the sampler.
2. Either spread the net out on the shore or pour the sample into an insect tray or white enamel pan and separate the invertebrates from other debris. Count the total number of macroinvertebrates in each sample and record the results in your field notebook. Using the key provided, separate the invertebrates into 3 groups: 1) stonefly nymphs (Plecoptera), mayfly nymphs (Ephemeroptera), and caddisfly larvae (Trichoptera) (collectively called PET's), 2) fly larvae (Diptera), and 3) all other invertebrates. Count and record the number of invertebrates in each category.

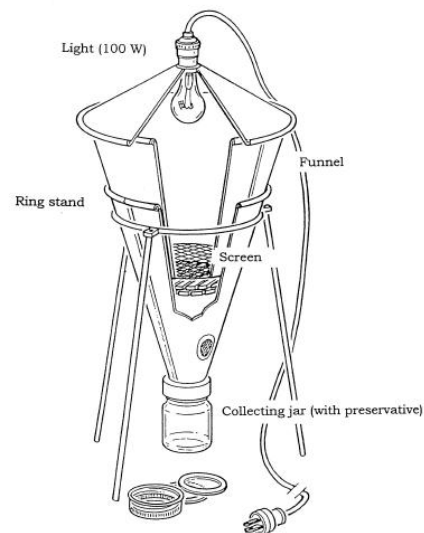
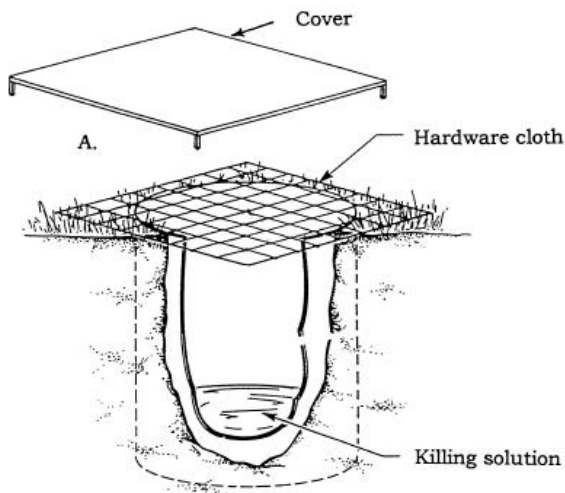
| | |
|---------------------|--|
| PET's | |
| Diptera | |
| Other Invertebrates | |
| Total Invertebrates | |
| PET/PETD ratio | |

- Repeat the sampling protocol for as many samples as you can before the end of the lab period. Sum the total number of macroinvertebrates and then calculate the ratio: PET/PET+Diptera (i.e. sum all the PET and Diptera together and divide into the number of PET's). Because PET's are more sensitive to chemicals than diptera and other invertebrates, the PET/PETD ratio will tell you what the chemical load is for a stream. You may use these techniques to compare two or more streams. Those with larger ratios are considered less polluted.
- You can use the "total invertebrates" for an assessment of sedimentation. Streams with more sediments have fewer macroinvertebrates and, consequently, fewer fish.

Follow Up: Other Invertebrate Analysis Tools

There are many types of invertebrates other than aquatic invertebrates that can be used to assess biodiversity patterns.

- Demonstration of the use of Wilding sampler (stovepipe)
- Demonstration of the use of plankton net and Sedgewick plate counter.
- Demonstration of the use of pitfall and sticky traps.
- Demonstration of a Tulgren funnel.



EXERCISE 2- MEASURES OF VARIATION AND CONFIDENCE

Background:

In the last homework you estimated averages. Averages only give you a middle value for a sample population. They do not indicate how large the spread of values may be around the average estimate. Chapter 5 of Ambrose et al. (2007) provides an introduction to measures of dispersion (also known as variation).

The **range** (the difference between the smallest and largest value) is the simplest measure of variability. However, variance and standard deviation are used more often because they have unique relationships with inferential statistics (to be practiced later in the course). The **variance** (traditionally represented by the symbol S^2) is calculated as follows:

$$S^2 = \frac{\sum x^2 - ((\sum x)^2/n)}{n-1},$$

where x is a sample value, and n is the sample size. The **standard deviation** is calculated by taking the square root of the variance. Although the variance is more important for calculating inferential statistics (as we will see later in the course), the standard deviation

is most often reported with other descriptive statistics because it has the same measurement units as the sample mean and thus, can be plotted in the same figure as the mean. Fortunately, EXCEL can easily calculate both measure of variation and measure of confidence, sparing you the tedious task of using a hand calculator.

In addition to measures of dispersion (variation), measures of confidence are useful descriptive statistics for assessing data. Measures of confidence (e.g. standard error and confidence intervals) are used to assess the precision of the estimated mean. **Precision** refers to the repeatability of the mean estimate (i.e. how close are the estimates to each other) and not necessarily to **accuracy** (how close an estimated mean is to the true mean). The **standard error** is a useful measure of the precision of the mean and is calculated by simply dividing the standard deviation by the square root of the sample size. Thus, **sample size** directly affects the precision of the mean estimate: the larger the sample size, the greater the precision. The **confidence interval** is an interval on each side of the estimated mean that should contain the true mean with a specified level of confidence (e.g. 95%). It is calculated by multiplying the standard error by a *t*-value obtained from a statistical distribution called Student's *t*. The *t*-value will depend on the samples size and on the selected significance level.

Laboratory Procedure:

Open the EXCEL file titled *SVLLizard* available on moodle. The snout-vent length (SVL) is a measure of body size that measures the length of the lizard from the tip of the snout to the cloaca (or vent). This practice data set will allow you to estimate the range, variance and standard deviation for male versus female individuals of the Florida Scrub Lizard.

You may need to activate the *Data Analysis ToolPak* for EXCEL to calculate measures of dispersion. Select the *DATA* tab and look for an option for *Data Analysis* located on the right side of the toolbar. If there is one, skip to section B below. If not, you need to activate the tool pack by following the directions in Section A.

Part A: Activating Data Analysis Toolpak

1. Select the *File* tab in the upper left corner of the screen.
2. Select *Options*, and then *Add Ins*.
3. Select *Go* and check the box for the *Analysis ToolPak* and *OK*.
4. EXCEL will ask you if you would like to install the tool pack: select *Yes* and wait.
5. Now when you select the *DATA* tab, *Data Analysis* should appear on your tool bar at the top.

Part B: Estimating Measures of Variation and Confidence in EXCEL

Use EXCEL to calculate measures of dispersion for male scrub lizards. You can calculate these separately like you did in the previous homework for the mean, median and mode by pasting (or typing) in a formula. For example, if you type `'=STD(B1:B150)'`, EXCEL will calculate the standard deviation of the data in the cells B1 through B150. Alternatively, there is a shortcut for calculating all your descriptive statistics in one step. (I didn't tell you this last week because I wanted you to practice entering formulas.)

Here is the shortcut: You can calculate mean, median, mode, variance, standard deviation, standard error and the confidence level all in one step by using the *Data Analysis* option in EXCEL.

1. Select the *Data* tab, then *Data Analysis*, then *Descriptive Statistics* and then *OK*.
2. Select the red arrow button next to the *Input Range*.
3. Highlight the input data you want to analyze (e.g. all the male lizard data in column B) and select *Enter*.
4. Select *Summary Statistics* and *95% confidence*, and then *OK*.

The calculations will appear on a new worksheet. The new table in the worksheet will have the following calculations: Mean, Standard Error, Median, Mode, Standard Deviation, Sample Variance, Kurtosis, Skewness, Range, Minimum, Maximum, Sum, Count, Confidence Level (95.0%). You know all about the Mean, Median and Mode from last week. Don't worry about the Kurtosis or Skewness for now. We will return to these when we deal with tests of inference. The Sum is simply the sum of all the data that you highlighted (not useful in this case of lizard SVL but could be useful for other data) and the Count is the number of cells you highlighted which is equivalent to the sample size *n*.

The variance, range and standard deviation give you an indication of the variability of the data. The variance is useful for calculating inferential statistics (later in the course) but is usually not reported with descriptive statistics. Instead, the standard deviation (which is the square root of the variance) is usually reported because it is in the same measurement units as the estimated mean (e.g. male lizard SVL = 50.4 ± 8 mm). Sometimes (not always) authors will report the range because it is so easy to calculate. The range can be reported as simply the difference between the minimum and maximum samples or, more common, the range is reported as the minimum and the maximum with a dash between them (e.g. male lizard SVL ranged between 40 – 59 mm). I prefer you do the latter because a range without a minimum or maximum doesn't tell you much about the extremes.

Measures of Confidence: The standard error is in the same units as the mean and gives you a measure of the precision of the mean (e.g. male lizard SVL = 50.4 ± 2 mm). With normally distributed data, we can assume the true mean is within 2 mm of 50.4 mm. Confidence intervals give us a further measure of confidence about the mean. To get the confidence interval, you need to take the confidence level reported by EXCEL and subtract the level from the mean to get the lower confidence limit and then add it to the mean to get the upper confidence limit. For example, with a mean male SVL of 50.4 mm and a confidence level of 0.6, the confidence interval would be 49.8 – 50.9. Because the level of confidence we selected was 95% we can say that we are 95% confident that the true mean SVL for male lizards is between 49.9 and 51.0.

Repeat the analysis for the female SVL data.

EXERCISE 3- DESCRIBING TYPES OF VARIABLES

Background:

In designing a study, you should select variables that are relevant to the pattern being studied, are clearly defined, and can be measured accurately and precisely. The definition of a variable must be clear enough that other researchers can duplicate the measurement. For example, a variable such as “habitat patch isolation” is meaningless unless you specify how isolation was defined (e.g. isolation = average distance to neighboring habitat patches).

The two most fundamental types of variables are continuous versus discrete. **Continuous variables** include measurements of subjects that can be recorded as integers with decimal points that reflect the precision of the measurement. Length, mass and temperature are examples of continuous variables. **Discrete variables** involve counts of subjects and can only be recorded as whole numbers. For example, the abundance of a population of scrub lizards could be 200 or 201 but not 200.5. Discrete variables may also come in the form of a name or nominal variable such as male versus female.

Once we have identified the variables for a study we can measure the variables using one of four possible scales of measurement: **Nominal**, **ordinal**, **interval** and **ratio** scales. Knowing which scale is used will allow you to better select appropriate statistical analyses later.

Variables may also be categorized for inferential studies into dependent versus independent variables. Typically when we use inferential statistics we have measured a variable we call the **dependent variable** (or **response variable**) that may or may not be influenced by **factors** (or **independent variables**). We often make comparisons between populations that differ with respect to their independent variables and we wish to know if this difference influences the dependent variable. For example, we could test whether upstream versus downstream from a confluence on a river (independent factor = location) influences one or more measurement variables (e.g. dependent variable = stream flow rate). Note that, in this case, the independent variable is a nominal variable (upstream versus downstream) while the dependent variable is a continuous variable (e.g. 100.5 m³ per second). This is not always the case. Some independent variables may be continuous, some dependent variables discrete, and any combination of independent and dependent variable is possible. Also, a nominal (sometimes called **categorical**) variable can have 2 or more levels (or groups). For example, the location variable has 2 levels, upstream and downstream. This is important because the number and type of independent and dependent variable will partially determine which statistical analysis is appropriate.

Laboratory Procedure:

You should first read chapter 3 of Ambrose et al. (2007) for a complete description of scales of measurement. Second, I've attached one my articles to the MOODLE site [Hokit, D.G. and L.C. Branch. 2003. Habitat patch size influences the demographics of the Florida scrub lizard (*Sceloporus woodi*). JOURNAL OF HERPETOLOGY 37:257-265]. Read the abstract (**note: you do not have to read the entire paper. The abstract, figure 1 and figure 2 should provide enough information**) and list all of the variables. For each variable describe which measurement scale was used, whether it is discrete or continuous, and whether it is a dependent or independent variable.

HOMWORK SUBMISSION

Each person should create a WORD file named *Homework 3*, answer the following questions, and submit their file to me via email.

1. Using your EXCEL output for measuring dispersion and confidence in female lizards, copy and complete the following table in your homework file.

| Parameter | Female SVL |
|-------------------------------|------------|
| Range | |
| Mean \pm standard deviation | |
| Mean \pm standard error | |
| Confidence interval | |

2. Complete the following table regarding the variables in the Hokit and Branch 2003 article:

| Variable | Continuous or Discrete? | Measurement Scale? | Dependent or Independent |
|--------------------------|-------------------------|--------------------|--------------------------|
| Patch Size in Fig 1 | | | |
| Patch Size in Fig 2 | | | |
| Survivorship in Fig 1 | | | |
| Abundance in Fig 2 | | | |
| Recruitment in Fig 2 | | | |
| Female Growth Rate Fig 2 | | | |