Week 1: Computers

Computers play an ever increasing role in all areas of life, including fisheries science. No matter what job you may find, computers will be there. Filling out computer "bubble forms" for creel surveys, using hand-held or remote data-loggers, managing large databases, modeling fish populations, and writing papers are all computer-based.

We will be working with past years’ Rock Creek data, which can be accessed from the course website. The data include size habitat variables (length, width, depth, stream velocity, substrate size, and distance to bank) as well as counts of fish species collected from a three stream reaches in two habitat types.

By the end of the lab you should be familiar with the following Excel techniques:
- Formulas
- Sort
- Filter
- Pivot table
- Graphs: scatter plots and bar charts
- Histograms

Techniques demonstrated in class:

1) Formulas. These allow you make calculations based on your data. You can do these either manually or using predefined excel functions such as average, standard deviation, etc.

2) Sort. This allows you to arrange data in a particular order. For example you may want to sort all fish by length from smallest to largest. To sort data, select the data you wish to sort then use then Data  Sort.

3) Filter. This allows you to select specific data to analyze. For example, you may wish to just select the fish data for cutthroat trout. To use a filter, select your data then Data  Filter  Autofilter.

4) Pivot tables. This allows you to calculate summary statistics (i.e., mean, standard deviation, sample size, maximum, minimum, etc) grouped by certain attributes of the data (i.e., “pool” or “riffle”). Highlight the data that you are interested in, then go to Data  Pivot Table Report. You will see a menu that will allow you to arrange the data (based on column headings) into a table, and calculate the summary statistics. For example, using the “count” field will let you count the number of fish caught by species or habitat type.

5) Graphs are a critical component of any scientific paper. There are many options for the layout and formatting of graphs in Excel. We will provide some demonstrations in lab but it is best that you do some experimenting and practice on your own.

6) Histograms. These allow you to count the number of records that fall into bins within a continuous data series. For example, you may want to group sculpins into 5 mm length bins to create a length-frequency histogram. To create a histogram, you first need to determine the bins you wish to use and enter these somewhere on your spreadsheet. Then choose
Tools ➔ Data Analysis ➔ Histogram. This will generate the binned count data, and the next step is to graph the results to complete the length-frequency histogram.

**Exercises:**

1) **Habitat data** - Compute the average, standard deviation (SD), and the coefficient of variation (CV) for each habitat variable from one reach (area) of the stream. Use the spreadsheet functions to calculate the mean and SD. To calculate the CV, use the following equation:

\[
CV = \left( \frac{SD}{\text{average}} \right) \times 100.
\]

The Excel commands for average and SD are “average” and “stdev”, respectively.

Why do we want you to do this? There are two reasons. First, we want you to learn how to take advantage of the commands and functions that spreadsheets offer. Second, the coefficient of variation can be an ecologically meaningful measure. The CV gives you a sense of the variability within a group of measurements. The larger the CV, the larger the variability. For example, let’s say we have a population of data with a mean of 5 and a SD of 5. The CV in this case would be \((5/5\times100) = 100\%\). Suppose you have another data set with a mean of 50 and an SD of 5. Now the CV is only \((5/50\times100) = 10\%\). Although the SDs are equal, the first data set is more variable than the second.

2) **Species data** - Calculate the Shannon Index \((H')\) for the fish assemblage data in one reach of Rock Creek. The Shannon index is a metric that fish biologists often use to measure the diversity of an ecological community. The Shannon Index is calculated as follows:

\[
H' = \sum [-Pi \times (\ln Pi)]
\]

Pi is the proportion of species i in the assemblage
\(\ln Pi\) is the natural log of that proportion.
\(\sum\) is the symbol for sum

You are given raw catch data, and you will need to convert this to counts by species (use a pivot table). You will need to compute the proportion for each species, the natural log of that proportion, the product of the two, and the sum of the products for the entire assemblage. It is the sum that is important, not the individual values for each species. The number should be positive.

So, what is the Shannon Index? In a nutshell, the SI allows you to quantitatively describe species diversity by incorporating two important elements of diversity: species richness and relative abundance. The higher the Shannon Index, the higher the diversity. The best way to prove this to yourself is to play around with the data on the spreadsheet. Add a species to your assemblage and see how SI changes. Rearrange the relative abundance of the species in the assemblage. If you make one species much more dominant, the SI will decrease. If you make all species about equal in representation, the SI will increase. You can generate identical SI values from two communities that differ greatly in species richness (number of
species) and evenness (relative abundance of species). Please remember that this one number cannot capture all the elements of a community of animals or plants, and should not be relied upon without interpretation.

It is also useful to calculate the Evenness index, which is a measure of relative diversity. Thus how evenly distributed are the fish among the species observed? If there are 5 species, does each contribute 20% of the total or is one 90% of the total and the others 2.5% each? This is calculated from $H'$, the Shannon Index calculated from the actual data, and the theoretical maximum value for $H'$ for a community with the given number of species.

Evenness Index: $J' = \frac{H'}{H'_{\text{max}}}$

where $H'_{\text{max}}$ (the maximum diversity) = $\ln r$

$r$ is the total number of species

The evenness index ranges from 0 to 1 and has positive values.

3) **Create graphs and tables** - Throughout this class, and throughout your fisheries careers, you will need to graphically present data in an effective manner. There are four things I would like you to produce:

A) Make a scatter plot showing the relationship between two of the habitat variables.

B) Make a bar graph of the densities of five species of fish in the three reaches or in the two habitat types. You will likely need to use a pivot table to generate fish counts by species, and formulas to find habitat area. Density is calculated by dividing fish abundance in a certain section (reach or habitat) by the total area surveyed in that region. Bar graphs are useful for presenting categorical or discrete data, such as numbers of fish in several specific locations. Be sure to put labels on the axes of both graphs.

C) Create a table that shows your calculated means, SDs, and CVs for all the habitat variables. Tables are used when you present specific data that the reader might want to know precisely; figures are used to present relationships where the pattern or trend is being communicated rather than the precise data points. Tables should never have vertical lines.

D) Create a length-frequency histogram for torrent sculpins. A length frequency histogram has length bins on the x-axis and counts of fish within each bin on the y-axis. To do this you will need to use the histogram technique described above.

The figures and the table must have complete, self-explanatory captions and labels. Show your TA your table and graphs before you leave class.
Figure X: this caption should describe the variables shown and the relationship between them. Always include UNITS (e.g. m/s) with your variable labels.

Figure X: insert description of variables shown and pattern observed....
Figure X. Sample length-frequency histogram of torrent sculpin