Appendix E: Understanding and Representing Data

Introduction

Scientists commonly use graphs and tables to represent and communicate numerical data. In order to represent data from an experiment accurately, a scientist needs to know what type of data he/she is collecting and what the data mean.

To make sense of data, you can start by answering the following questions:

1. What are the units of measurement in your data set?
2. Is your data qualitative or quantitative?
3. What are the dependent and independent variables?
4. Is your data best represented as a table or a graph?

Basic Concepts Relating to Data

Unit of Measurement: A standard of basic quantity or increment by which something is divided, counted, or described, such as ml, kg, mm, m/s, °F, etc.

Qualitative Data: Data for which the scale of measurement is a set of unordered categories called a nominal scale. For example, types of trees, types of compounds, etc. Qualitative variables are considered discrete variables, because they vary in some quality but not magnitude—one category is not greater than the other.

Quantitative Data: Data for which the scale of measurement has magnitude. Most quantitative data are either interval or ordinal. Quantitative values are continuous, so each possible value may be greater or less than any other value. Ordinal data has an order but the distance between values does not have precise numerical meaning. For example, rank in a graduating class or ranking of runners in a race. Interval data uses a scale that has a specific numeric distance between values—it has a unit of measurement. For example, the time in which the runners ran the race, is interval data.

Identifying Variables

Tables and graphs serve to organize findings and represent relationships among variables. A variable is what is measured or manipulated in an experiment. Variables provide the means by which scientists structure their observations. Identifying the variables in an experiment provides a solid understanding of the experiment and what the key findings in the experiment are going to be. Variables are usually classified as independent and dependent.

An **independent variable** is the variable you have control over; the one you can manipulate. It is usually what you think will affect the dependent variable—what you are measuring in the experiment. In some cases, you may not be able to manipulate the independent variable. It may be something that is already there and is fixed, something you would like to evaluate with respect to how it affects something else—like color, kind, and time. Independent variables are usually charted on the x-axis of a graph.

A **dependent variable** is what you measure in the experiment and what is affected during the experiment. The dependent variable responds to the independent variable. It is called dependent because it "depends" on the independent variable. In a scientific experiment, you cannot have a dependent variable without an independent variable. Dependent variables are usually charted on the y-axis of a graph.

It is possible to have experiments in which you have **multiple variables**. There may be more than one dependent variable and/or independent variable. This is especially true if you are conducting an experiment with multiple stages or sets of procedures. In these experiments, there may be more than one set of measurements with different variables.

When conducting an experiment, scientists can manipulate variables and record all relevant data. Usually, standard scientific experiments have control groups and treatment groups. A **control group** is used as a baseline measure. The control group is identical to all other items or subjects (organisms, specimens, etc.) that you are examining with the exception that it does not receive the treatment or the experimental manipulation that the treatment group receives. For example, when determining the effects of various types of fertilizers on tall fescue grasses, the control would be a plot of grass that is identical to all other plots of grass in the experiment with the exception of lacking fertilizer. The **treatment group** is the item or subject that is manipulated. In our example, all other plots of grass that receive the different types of fertilizers would be the treatment groups.

Tables vs. Graphs

An initial decision that has to be made about your data is whether it should be displayed in a table or a graph. Though there are no hard rules, there are general guidelines you can use to make this determination. Questions to ask yourself:

- Are the independent and dependent variables qualitative or quantitative?

When both your independent and dependent variables are qualitative, there will not be a suitable graph format to plot your data.

- What is the total number of data points to be shown?

If it's just a few, you don't really need a graph. For example, if you only have four data points, you would use a table, unless you wanted to show a rate of change between each pair of points. When you have over 6-10 data points, you almost always represent the data in a graph.

- Is there more than one independent variable?

When a second independent variable is introduced, a graph is usually called for in order to understand the relationship between the two sets of data.

- How important is it to be able to see individual values?

If showing individual values is important, you may need to include a table or choose a graph that allows for distinction of exact data values.

- How important is it to understand the overall trend?

Graphing your data will allow you to reveal an overall trend. The type of graph will depend on the type of data you have (see below). You may also need to insert a trend line (regression line) on your graph (see below).

Types of Graphs

Qualitative data is most often graphed using bar graphs, whereas quantitative data may be represented with multiple bar graphs, histograms, line graphs, and scatter plots.

Bar Graphs:
Bar graphs are used when you have different categories that you're comparing to each other. Usually, you have a result for each category, and graphing these results lets you see the differences between the categories. For example, you would use a bar graph if you wanted to show the number of wins that different football teams have had this season. Bar graphs are used when the independent variable is qualitative.

Histograms:
Histograms are similar to bar graphs, but each bar represents a range of data in a larger distribution of data. Histograms allow you to summarize the independent variable data. For example, if you wanted to know how many individuals in a population were between the ages of 0-4, 5-9, 10-13, etc., you would use a histogram.

Line Graphs:
Line graphs are used when you're looking at how something changes over time. A line graph allows you to look at the change in one specific variable instead of comparing different categories. For example, you would use a line graph to show how the usual temperature changes in different seasons, or to show how your height has changed as you have gotten older.

Scatter Plot Graphs:
Scatter plot graphs are used when you need to show individual data points on a graph in order to visualize a certain distribution. Unlike line graphs, data points are not connected, but a trend line (regression line) can be inserted onto a scatter plot to express a trend in the data. Trend lines represent the correlation between the independent and dependent variable. For example, as the x-value increases, the y-value increases. How well the line "fits" the data is called "goodness of fit" in statistical terms. This can be described by a constant known as the correlation coefficient, R^2 (R-squared). R-squared refers to how well the line describes the trend of the data. The closer R-squared is to 1.0, the better the fit of the data to the line.

Some Things to Consider:

Assuming that you have chosen the proper graph type to display your data, most of the final adjustments concern the labeling and display of your variable axes and your title.

- Have you labeled each axis with a variable? Spell out the name or use standard abbreviations.

- Have you shown a legend for your variables? A legend is needed if you have coded a second independent variable with color, pattern, or shape. You do not need a legend if you have a single independent variable. (Note: this legend is different from the "figure legend" described below).

- Have you put the units of measurement in parentheses after the variable(s)? Again, only use standard abbreviations.

- Are scale values labeled on the axis? Make sure the labels are adequately spaced. Label every other tick mark if they are too crowded.

- Do the axes and graph regions have an appropriate density of tick marks and grid lines? Generally speaking, less is better. Readers are not usually expected to be able to read individual values from a graph.

- Is the text in the graph an appropriate size for the final presentation of the graph? The graph may need to be placed in the document and printed out to test this.

- Does the graph have an appropriately descriptive title? Simply listing the variables is usually not adequate. Conversely, the title does not have to be a full sentence. The graph

needs to be a stand-alone representation of the data, so it must be titled to convey enough information.

- o In some cases, a figure legend (or caption) is used instead of a title to describe the graph. The figure legend would come after the figure number (see next point below).

- When placed in the document, graphs simply have the figure number put below the graphic (e.g., Figure 1). No figure legend is needed if the graph has a title. If you are using a photo or other graphic, a figure legend may be needed (a figure legend may also be used as described in the point above).

- Tables are labeled on top. For example, Table 1. The title of the table, if the title is not embedded in the first table row, follows the table number. For example, Table 1. Number of children by age groups in North Carolina counties.

Resources:

For more help with determining the type of data, go to:
http://labwrite.ncsu.edu/res/gh/gh-datatype.html

For more help with identifying the right type of graph for the data, go to:
http://labwrite.ncsu.edu/res/gh/gh-graphtype.html

For help on the basics of using an Excel® spreadsheet, or creating various types of graphs with Excel®, visit the website page listed below. It has a collection of tutorials for basic data input and different types of graphs, in addition to more advanced techniques:
http://labwrite.ncsu.edu/res/gt/gt-menu.html

For help with tables, go to:
http://labwrite.ncsu.edu/res/gh/gh-tables.html

NOTE: These guidelines are based on materials from the LabWrite Project©, produced at NCSU, funded by the National Science Foundation, 2000-2007.