**Graphing and Statistic Guide**

The following guide provides step-by-step instructions with picture guides for foundational graphing and statistics in Excel. While the specific interface, version of Excel, and operating system you use may be somewhat different from those used to produce this guide, the core functions are likely similar. If there are differences between the interface you interact with and the guide instructions, remember that Google has all the answers to based data sheet management questions no matter what version of Excel you’re using! A key skill for all scientists is to be able to solve problems in procedures using outside information often gathered from the internet so we can consider this a realistic experience with science!

**Line Graphs**

Line graphs are used to track changes in a continuous variable over time. For example, consider the data showing changes in seagull population size over time.

Graphical user interface, application, table, Excel

Description automatically generated

To begin, highlight *only* the data for the dependent variable (in this case, population size) without highlighting the column title. Once you’ve highlighted the appropriate data, click on the Insert tab and select the line graph option.

Graphical user interface, application, table, Excel

Description automatically generated

Excel will automatically generate a graph that looks like this:

Chart, line chart

Description automatically generated

Our first edit is to change our x-axis labels to reflect the year in which the data was collected. Right click anywhere on the figure to pull up an editing menu and click the ‘Select data’ option.

Graphical user interface, application, table, Excel

Description automatically generated

This will call a Select Data Source pop-up. Click the Edit button under horizontal axis labels (see next page for pictures). Another pop-up called Axis Labels will appear. Highlight *only* the data for the time series excluding the column header and then click OK (see pictures on next page). Select OK on the Select Data Source pop-up and you’ve added years to your x-axis!

Graphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

Now we need to add a title and axis labels. Click on Add Chart Element and add a chart title and x- and y-axis labels.

Graphical user interface, application, table, Excel

Description automatically generated

Your completed figure will look like this:

A picture containing text, screenshot, font, line

Description automatically generated

**Trendlines**

Finally, scientists often use lines of best fit, or trendlines, to determine the relationship between time and a dependent variable. In our example, we want to understand how seagull population size has changed over time. You can add a line of best fit to the line graph by right-clicking on the line, going to “Add trendline,” and making sure that the linear option is selected (see below). Remember to check the “Display equation on chart box” to show the equation of the best fit line you’ve added.

Graphical user interface, application, table, Excel

Description automatically generated

Once you’ve added your line of best fit and added the chart and axis titles, your line graph should look something like the one below. The equation of your trendline provides you with important information about the relationship between your two variables. Of particular importance is the slope of your line. The slope tells you how much your dependent variable changes per unit time increase of your independent variable. In our example, we can see seagull population declined by roughly 58 birds per year (derived from the slope of the line).

Chart, line chart

Description automatically generated

**An important note about organizing temporal data for line graphs and trend lines:**

Climate scientists often use long-term data sets to determine how variables like temperature, precipitation, and associated biological variables have changed over time. The most common way to plot temporal data like these is to use a line graph. Scientists can them use a trend line to determine the rate of change in the variable over time. There are several important considerations in organizing temporal data to ensure that the rate of change is calculated accurately. First, the time series must be organized from the earliest to latest time points to avoid misordered data effects. Second, if any time points are missing, they must be included in the data as a year with a missing data point rate than simply deleted. This treatment of the data as mising ensures that the rate of change in your variable is calculated over the correct period of time rather than artificially shortening the time period by deleting the time point with the missing data.

**Bar Graphs**

Bar graphs are useful for categorical data that can be organized into two or more groups. The first step in any data analysis is organizing your data in a logical way. The example data we’ll use compares the total number of bird species on islands classified as either large or small. Eight total islands are shown in the table with four large and four small islands.

Graphical user interface, application, table, Excel

Description automatically generated

While the individual data points for each island are interesting, we’d like to depict the average number of birds based on island size in our graph. For that, we first need to do some basic statistics: means and standard deviations.

A mean is simply the sum of all numbers in a group divided by the total number of samples and can be easily calculated in excel. Simply type the formula as below (=AVERAGE(cell:cell)) and highlight the cells with values that you would like to average. Hit enter once your formula is complete and the mean will be displayed. (See images on next page)

Graphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

Next, we need to calculate some estimate of how representative the mean is of the individual data points in each group. For example, the average of 1, 1, 1, 100, 100, and 100 is 50.5. However, 50.5 isn’t a particularly accurate reflection of any of those individual data points. To estimate the spread of individual data points around the mean, we calculate standard deviation as shown below. As with the mean, simply type the formula (=stdev(cell:cell)), highlight the relevant cells, and hit enter.

Graphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

To show how much variation there was about the mean, scientists usually plot standard error bars. Standard errors, much like standard deviation are just another way of quantifying spread in the individual data points. A handy rule of thumb is that non-overlapping standard error bars usually mean two groups are significantly different from each other. Next, we’ll calculate the standard errors that will go on our bar graph!

While there’s no automatic formula for standard errors in excel, we can calculate them really easily. The formula for standard error is:

=STDEV(cell:cell)/(SQRT(COUNT(cell:cell))

The image below show the calculation for the standard error of the total number of bird species on large islands.

A screenshot of a computer

Description automatically generated

Once we calculate both standard errors, we’re ready to move onto the next step!

Check with the picture below to make sure you’ve done the calculations correctly.

Graphical user interface, application, table, Excel

Description automatically generated

You can learn more about the differences between standard error and standard deviation at this link: <https://statisticsbyjim.com/basics/difference-standard-deviation-vs-standard-error/>.

Now that we’ve calculated some basic statistics for each of the test groups, we can plot our data in a graph. Excel has lots of ready-made graph options all of which can be found under the “Insert” tab under “Charts” including bar graphs. (called column charts in Excel when the bars are oriented vertically) as shown below.

First, select the data that you’d like to graph. I’ve highlighted both the group means and the group categories and excel will automatically label the columns on my graph accordingly.

Graphical user interface, application, table, Excel

Description automatically generated

Once I click the chart option I want, excel automatically produces a basic, unfinished graph with no error bars and lacking important axis labels.

Next, we need to add error bars to show our measure of spread, standard deviation, on the graph. Click anywhere on your chart to pull up the “Chart tools” tab. Under “Design,” you’ll see “Add chart element.” Clicking on it brings up a list of options. Click “Error bars” and then “More Error Bar Options.”

A side pane will pop-up: select “Custom” and then “Specify Value.” (see below)

Graphical user interface, application, table, Excel

Description automatically generated Graphical user interface, application, table, Excel

Description automatically generated

Another pop-up will appear that allows you to select the appropriate values for your error bars. Click the funny looking button to the right of the “Positive Error Value” cell, highlight both of your standard deviation values, and then click to funny looking button again. Repeat this exact same process for the Negative Error Values and then hit “OK.”

Graphical user interface, application, table, Excel

Description automatically generated A screenshot of a computer

Description automatically generated

Voila! You now have error bars on your graph!

Graphical user interface, application, table, Excel

Description automatically generated

Alas, our work isn’t quite finished yet. We still have no title and lack axis labels. Thankfully, that’s easily remedied!

Click on Add Chart Element and add a chart title, x- and y-axis labels, and a legend if needed.

Graphical user interface, application, table, Excel

Description automatically generated Chart, box and whisker chart

Description automatically generated

**T-tests**

In our example of categorical data above, we looked at the relationship between island size and the total number of bird species found on the island. Now we ask, are the total number of bird species found on large vs. small islands significantly different from each other or are they relatively similar? The bar for the large islands *looks* bigger than for the small islands, but the standard deviations are also pretty big… How can we objectively determine whether the two groups differ from each other?

Meet the **t-test.** T-tests are used to compare two means while integrating information about sample sizes and spread of the data about the mean of each group.

To perform a t-test, Excel will calculate a value called a t-statistic. Take a look at the formula for the t-statistic below. Essentially, the t-statistic answers the question, “How certain are we that these groups are truly different from each other based on how different the means are, how much the individual data points vary in each group, and how big the sample size is.”

Or, put in words:

We’ve already calculated the means and we know the sample size for each group is four. The last thing Excel will calculate for the t-test is the variance. Variance is simply another measure of the spread of the individual data points about them mean. While excel will do this automatically, you need to calculate it yourself to select the appropriate t-test type. Use the equation (=VAR(cell:cell)) to calculate the variance for each group. (You can learn more about the similarities and differences between standard deviation and variance at this link: <https://www.mathsisfun.com/data/standard-deviation.html>).

Graphical user interface, application, table, Excel

Description automatically generated

To perform a test in excel, we’ll use the T.TEST function. In a cell near your data, type =T.TEST(

This will call the t-test function which has four elements: array 1, array 2, tails, and type. Each element is defined below and will be describe in greater detail later.

Array 1 = the data for your first group

Array 2 = the data for your second group

Tails = whether you predicted a directional difference or any difference between groups

Type = whether your groups had equal or unequal variance.

Graphical user interface, application

Description automatically generated

After entering the formula, select the data for your first group (shown below). Once range appears, enter a comma.

Graphical user interface, application, table, Excel

Description automatically generated

Next, select the data for your second group (show below). Once the range appears, enter a comma.

Graphical user interface, application, table, Excel

Description automatically generated

Next you need to determine the ‘tails’ of your test. If you hypothesized that there would be a directional difference between your group (i.e., large islands will have more bird species than small islands), your test is 1- tailed. If you hypothesized that your two groups would be different, but you weren’t sure in which direction, your test is 2-tailed. Since I hypothesize that there will be more bird species on larger islands than smaller islands, I’m doing a 1-tailed t-test.

Enter a 1 and then a comma (see below).

Graphical user interface, application, table, Excel

Description automatically generated

Finally, we determine the type of t-test based on variances in your group. If your variances are similar, select ‘Two-sample equal variance’ and if your variances are quite different, select ‘Two-sample unequal variance.’ A general rule of thumb is that if your larger variance is within 4 times the value of your smaller variance, they are similar enough to use an equal variance t-test. For example, if your smaller variance is 42 and your larger variance is 136, since 42\*4 = 168 and 136 < 168, you could use an equal variance t-test. (You can learn more about equal variance assumptions in various statistical tests here: https://www.statology.org/equal-variance-assumption/.)

Graphical user interface, application, table, Excel

Description automatically generated

Once you’ve entered your type, add end parentheses and you’re done! The t-test will produce your **p-value.**

P-values, or probability-values, provide an estimate of how confident we are that the difference observed between groups is real and not simply due to random chance. A very low p-value, for example 0.01, means that there is only a 1% chance that your results are due to random chance and that we are 99% confident that the difference between groups is due to some true difference between treatments. In contrast, a very high p-value like 0.5 means there’s a 50% chance that our observed difference between groups is due to random chance. In this case, we say that there is no significant difference between groups.

Scientists tend to be rather cautious about saying that groups are significantly different from each other. Typically, scientists require that a p-value be 0.05 or *less* for two groups to be significantly different from each other. Everything above p = 0.05 typically is classified as *not* significantly different.

Our p-value for this test is 0.004. This p-value indicates a very low probability that the difference in bird species number between large and small islands is due to random change. Thus, we conclude that large islands have significantly more bird species than small islands on average.

Graphical user interface, application, Word

Description automatically generated