

GUIDELINES: HOW TO MAKE FIGURES AND
DO STATISTICAL TESTS IN EXCEL MAC 2016

(These guidelines are modified from materials developed by C. Caceres for IB 449 Limnology).

The instructions below are for 2016 for Mac

T-TEST: 1 categorical independent variable with 2 treatments; 1 dependent variable

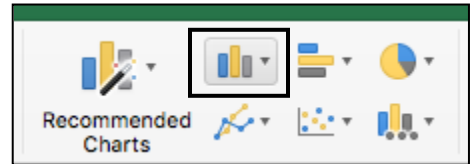
A. Use Excel to create a Figure

1. Calculate Average (mean), Standard Deviation, and Standard Error (SE) for each treatment.

Create this table:

Independent Variable	Average (mean)	Standard Deviation (stdev)	SE = SD/\sqrt{N} OR = $stdev/sqrt(N)$ (click on stdev to the R)
Treatment 1			
Treatment 2			

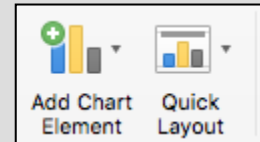
2. Select the **6 cells** in the **first 2 columns** of the table.
Click on INSERT located in the green toolbar.
Then select the “Column” icon in the middle-top of the ribbon.



Click on the type of chart you want to make (= 2-D Column: Clustered Column).

3. Add axes titles.

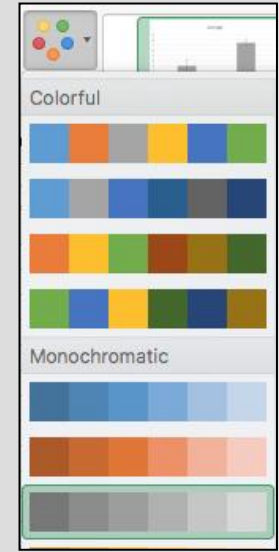
- a. Click on the graph. Within the ‘Chart Design’ tab in top ribbon, select ‘Add Chart Element’ at the far left → Axis Titles
- b. Add axis title below x-axis (horizontal with independent variable). Click in ‘Axis Title’ textbox on graph to edit
- c. Add axis title to y-axis (vertical (*rotated*) with dependent variable). Click in ‘Axis Title’ textbox on graph to edit



4. Add Standard Error bars to each average.

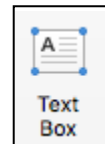
- a. Click on the graph. Within the ‘Chart Design’ tab in top ribbon, select ‘Add Chart Element’ at the far left → Error Bars → More Error Bars Options
- b. From the far right pop-up menu ‘Format Error Bars’ Select ‘both’ and ‘cap’. For Error Amount, select ‘Custom’, then ‘specify value’. Select red/blue box to Right in dialogue box.
- c. Highlight both values for SE from worksheet (for ‘+’ bar); click red/blue box to R. Repeat for ‘-’ SE bar, click red/blue box to R, then ‘OK’

5. Change color of bars to gray
 - a. Click on graph. Within 'Chart Design' tab in top ribbon select 'Change Colors' icon to the far left → select Monochromatic gray



6. Make vertical axis line visible
 - a. Double click on Y axis numbers. In the 'Format Axis' menu to the right select the paint can → line → solid line
 - b. Change color to black or gray by using paint can below the line options
7. Clean up graph. Click on or click/delete object directly.
 - a. No chart title
 - b. No legend
 - c. No (vertical value) gridlines (click on 2nd horizontal line on graph; delete).
 - d. To change the Y axis so that it starts at 0, click on the axis. In the Format Axis menu to the right select Bar Graph Icon → Axis Options → Axis Value → Type '0'.

8. Add figure legend. Within the Insert tab in the green toolbar select Text Box at the far right. Drag and create a box under your graph. The 'rule of thumb' is that the reader should completely understand the figure without reading any of the text of the paper. Include what the bar and vertical lines indicate (mean +/- SE). Keep it to one incomplete 'sentence'.



Example of figure legend:

Figure 1. Comparison of number of arthropods (mean +/-SE) in fields abandoned for one year vs. five years.

B. Use Excel to do t-test. Are means of two treatments significantly different?

1. Type '=T.TEST(' in a cell below data.
2. (Array 1 = select original data from treatment 1 = column A, Array 2 = select original data from treatment 2 = column B, Tails = 2 (2-tailed t-test), Type = 3 (2-sample test unequal variance)) OK.
e.g. in formula toolbar, type: =T.TEST(A2:A7,B2:B7,2,3), then (MAC) click on green check mark to the left of formula.
3. The p value for the t-test is printed on the selected cell on the worksheet. If it is $P < 0.05$, then there is less than a 1 in 20 probability that the difference between the means of the two populations is due to chance.

C. Interpretation of results.

1. Return to your 'Hypothesis/Prediction'.
2. Based on your analysis, what can you conclude? Do the data support your hypothesis?
3. To write up the "**Results**" section of a scientific manuscript, summarize the main pattern in your figure, state the statistical output, and refer to Figure 1 (in parentheses at end of sentence). Keep the results to 1 sentence.

RESULTS: The mean plant species richness was significantly (or not significantly, depending on what your result is) greater in the X-year than the X-year field (t-test $P < 0.05$) (Fig. 1).

ANOVA (Analysis of Variance): 1 categ. indep. var. with >2 treatments; 1 dep. var.

A. How to create figure in Excel:

1. Follow directions above for t-test. The only difference is to create a table with average and SE for each treatment, (not just 2).

B. How to do an ANOVA test in Excel: *Are the means of treatments different?*

1. Data must be organized with each treatment in a separate column and a label above each column.
2. Open StatPlus (in dock at bottom), while keeping open Excel with data.
3. Select 'statistics' on main tool bar, then 'Analysis of Variance (ANOVA)', then 'One-way ANOVA (simple)'.
4. For variables, 1) check the box 'labels in first row', 2) select icon to R, and 3) highlight all columns of data (and labels) in Excel sheet.
5. Click StatPlus, then 'ok' to get results page; copy/paste ANOVA table of results to Excel sheet near the data.

C. Interpretation of statistical results:

Whether the ANOVA is significant depends on the degrees of freedom. The critical value that the F-statistic (F crit) must be greater than to be significant ($P < 0.05$) is shown, as well as the F value and p-level.

Based on your analysis, what can you conclude? Are the means of your treatments significantly different? Do the data support your hypothesis?

If the ANOVA is not significant, the analysis is complete. If it is significant, continue.

Recall that the GUIDELINES ON BASIC STATISTICS states for ANOVA that if there is a significant difference among treatment means, ANOVA does not tell you anything about which treatments are significantly different from the others. Now you must do *t*-tests to determine which treatments are significantly different from the others (a separate *t*-test for each comparison of each pair of treatments).

To write the '**Results**', summarize the following, each in one sentence:

- 1) The conclusion based on ANOVA. Look at the figure and ANOVA. At the end of the sentence (put statistical text and P value in parentheses) (put Fig. 1 in parentheses).

RESULTS: The % cover of tree trunks by moss differed significantly among directions (ANOVA $P < 0.025$) (Fig. 1).

- 2) The conclusion based on multiple comparisons. Look at the t -tests. At the end of the sentence (put statistical test and P value in parentheses).

RESULTS: The north side had greater % cover than the south side (t -test, $P < 0.05$), but had similar % cover to the east and west sides (t -test, n.s. $P > 0.05$).

CHI-SQUARE TEST OF INDEPENDENCE: 2 categorical variables + count data

A. How to make a figure in Excel:

1. Create a table like this:

	edge	interior
snap-off	43	46
uproot	32	79

1. Select the 9 cells in bold in the 3 columns of the observed table.
MAC: Click on INSERT located in the green toolbar. Then select the “Column” icon in the middle top of the ribbon.
Click on the type of chart you want to make (= 2-D Column: Clustered Column).
2. Add axes titles.
 - a. Click on the graph. Within the ‘Chart Design’ tab in top ribbon, select add chart element at the far left → axis titles (PC = Green Chart Tools, then Layout)
 - b. Add axis title below x-axis (horizontal with independent variable). Click in ‘Axis Title’ textbox on graph to edit
 - c. Add axis title to y-axis (vertical (rotated) with dependent variable). Click in ‘Axis Title’ textbox on graph to edit
3. Change color of bars to gray
 - a. Click on graph. Within ‘Chart Design’ tab in top ribbon select ‘Change Colors’ icon to the far left → select Monochromatic gray
4. Make vertical axis visible by double clicking on Y axis numbers. In the ‘Format Axis’ menu to the right select the paint can → line → solid line
5. Clean up graph. Click on or click/delete object directly.
 - a. No chart title
 - b. No legend
 - c. No (vertical value) gridlines (click on 2nd horizontal line on graph; delete).
 - d. To change the Y axis so that it starts at 0, click on the axis. In the Format Axis menu to the right select Bar Graph Icon → Axis Options → Axis Value → Type ‘0’.

6. In a cell BELOW the figure, create a figure caption (we also call this a figure legend). Add figure legend. Within the Insert tab select Text Box at the far right. Drag and create a box under your graph. The ‘rule of thumb’ is that the reader should completely understand the figure without reading any of the text of the paper.

Example of figure legend:

Figure 2. Number of plots with high- vs. low- stem density in fields abandoned for one year vs. five years.

B. How to do a Chi-Square test with a calculator:

Are the distributions of the 2 categorical variables independent?

1. Create this table.

	OBSERVED			
	edge	interior	row total	
snap-off	43	46	89	
uproot	32	79	111	
column total	75	125	200	Grand total
	EXPECTED			
	33.375	55.625	89	
	41.625	69.375	111	
	75	125	200	

2. Calculate the marginal totals of observations (add up columns and rows). Grand total is the total of the two row totals (= total of the two column totals).
3. Complete the expected frequency table (E) above to one decimal place. This table will contain values expected if the type of tree death is **independent** of the location (This is the **null hypothesis**).
Multiply each row total X each column total and divide each product by the grand total. (E.g. Expected # Snap-offs in the Edge is (75 X 89)/200).
4. Calculate Chi-square by comparing the observed and expected frequencies for each of the four cells in the two tables. Subtract the difference, square it, and divide by the expected frequency for each cell. Add the results for each cell. Show your math.

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

5. Use the Chi-Square Statistics Table at the front of the manual after GUIDELINES TO BASIC STATISTICS to determine whether type of tree death is significantly independent of location. If the X^2 is greater than the value in a chi-square table with $P \leq 0.05$ and 1 degree of freedom (1DF) [$\# \text{ rows}-1)(\# \text{ columns}-1) = 1$], then the null hypothesis of independence is rejected. The two variables are associated. If $P > 0.05$, then you cannot reject the null hypothesis of independence.

C. Interpretation of results.

1. If $P < 0.05$, the variables are associated. E.g. the type of treefall is associated with the location in the forest. Look at the values of (o-e) for each cell and determine which type of treefall is associated with which location. E.g. snap-offs are more common at the edge and uproots in the interior than expected if the distributions of the two variables were independent.
2. If $P > 0.05$ the variables are independent. E.g. the type of treefall is not associated with the location in the forest. Look at the values of (o-e) for each cell. There is little difference between these values. E.g. a snap-off (or uproot) is as likely to be at the edge as in the interior.
3. **RESULTS: The distribution of types of Z was independent (or dependent) of location ($X^2 = 2.1$, $df=1$, $p>0.05$) (Fig.2). Overall, X were more than Y, but the ratio of X to Y was similar (or not) in both locations.**

CORRELATION: 2 CONTINUOUS VARIABLES

A. How to create a figure in Excel:

1. Have the labels and data for the 2 variables in the first 2 columns of the worksheet.
2. Highlight labels and data for making figure.
3. Select 'Charts' (PC = INSERT) in the green toolbar, then 'Scatter', then 'Marked Scatter'.
4. Add axes titles.
 - a. Click on purple Chart Layout in green toolbar.
 - b. Add axis title below x-axis (horizontal with independent variable).
 - c. Add axis title to y-axis (vertical (*rotated*) with dependent variable).
5. Clean up graph. Click on Chart Layout on toolbar (or click/delete object directly).
 - a. No chart title
 - b. No legend
 - c. No (vertical value) gridlines (click on 2nd horizontal line; delete).
 - d. To change the Y axis so that it starts at 0, click on the axis. Click Axes in toolbar, then vertical, then axis options, (MAC = then scale). Change Minimum = 0.0.
6. In a cell below the figure, create a figure caption. The 'rule of thumb' is that the reader should completely understand the figure without reading any of the text of the paper. Keep it to one 'sentence'.

Example of figure legend.

Figure 1. Number of fruits as a function of total leaf area (cm²) of a pawpaw tree.

B. How to calculate CORRELATION statistic in Excel:

Is there an association between X and Y? (no cause-effect of x and y)

1. Calculate the correlation coefficient = r , which measures the strength of association between the two variables.
 - a. Type 'correlation' in a cell below column A data. Select cell to its right.
 - b. Select 'Formulas' in green toolbar, then 'insert fx', (PC = select 'category'), then 'statistical', then 'correl'. (PC = OK).
 - c. (Array 1: Select data in column B,
Array 2: Select data in column A) (PC = then OK).
e.g. in formula toolbar, type: correl(B2:B7,A2:A7) then, in MAC, click green check mark to the left.
2. The correlation coefficient r is printed in the selected cell on the worksheet.

C. Interpretation of results

1. Whether $r = \text{significant}$ depends on the degrees of freedom. The critical value that r must be greater than to be significant ($P < 0.05$) can be looked up in Pearson's Correlation Coefficient in the GUIDELINES FOR BASIC STATISTICS.
2. If your r value is significant ($P < 0.05$), you have evidence that the two variables are associated. Whether the association is positive or negative depends on the sign (+ or -) of the value of r . A positive slope will have a positive r ; a negative slope will have a negative r . The closer the value of r is to 1 (or -1), the greater the strength of the association.

3. Based on your analysis, what can you conclude? Do the data support your hypothesis?
4. To write up the “**Results**” section of a manuscript, summarize the main pattern in your figure. State the statistical output and refer to Fig. 1 in parentheses at the end of the sentence. Keep the results to 1 sentence.

RESULTS: Number of fruits did not increase significantly as total leaf area of the tree increased ($r = 0.523$, $P > 0.05$) (Fig. 1).

REGRESSION (assuming cause-effect between two continuous variables):

A. Add a regression line to the figure made for the ‘correlation’.

1. Select chart area within the figure.
2. Select ‘Charts’ in the green toolbar, (PC = INSERT), then ‘Chart Layout’, then ‘trendline’, then ‘linear trendline’.

B. Calculate R^2 (values range from 0-1)

Does a change in X result in a change in Y?

1. Type ‘RSQ’ below ‘correlation’. Select cell to the right.
2. Select ‘Formulas’ in the green toolbar, then ‘insert fx’, (PC = select ‘category’), then ‘statistical’, then RSQ (PC = OK).
(Array 1 = select data in column with y values,
Array 2 = data in column with x values)
(e.g. =RSQ(B33:B39,A33:A39) then (MAC) click green check mark to the left.
3. The R^2 value is printed in the selected cell on the worksheet.

C. Interpret the R^2 value. Values range from 0-1. The closer the value to 1.0, the greater the strength of the relationship. How well does the line represent the data points?

D. Get an equation for the line ($y = mx + b$) where m = slope and b = Y intercept.

1. a. Type ‘slope’ below cell for ‘RSQ’.
b. Type ‘intercept’ below cell for ‘slope’
c. Type ‘line equation’ below ‘intercept’.
2. Obtain slope (and intercept) in same way as ‘RSQ’ above, selecting ‘slope’ (or ‘intercept’) from the list of statistical functions.
Be careful: use “y” values for Array 1, and “x” values for Array 2.
3. Type the equation to right of ‘line equation’.

E. To write up the “Results” section of a manuscript, summarize the main pattern in your figure. State the statistical output and refer to Fig. 1 in parentheses at the end of the sentence. Keep the results to 1 sentence.

RESULTS: An increase in light intensity resulted in an increase in algal biomass ($R^2 = 0.75$, $P < 0.01$) (Fig. 1).