APPENDIX

INSTRUCTIONS FOR USING MICROSOFT EXCEL – PERFORMING DESCRIPTIVE AND INFERENTIAL STATISTICS AND GRAPHING

(Developed by Dr. Dale Vogelien, Kennesaw State University)

** For a good review of basic statistics, visit the following web sites (developed by Dr. Matt Laposata for KSU’s Science 1101 course:

Statistical Analyses Using Microsoft Excel

- Double click on Excel icon OR click on (start), (Programs), (Microsoft Excel)
- At the top of the worksheet, type in text information that will help you or others understand later what study the data reflect, computations performed, statistical test(s) performed, etc.
- Enter data/observations for a sample in columns or rows; label column/row for later reference. Note: the accompanying instructions are for data observations entered in columns. Each data set is entered in a separate column.
- To calculate the MEAN for a set of observations:
  1. position cursor in the cell where you would like the computed mean value to appear
  2. click on (fx) in the toolbar, click on (statistical), click on (average); note: once a function has been used it can be quickly found under the “Most Recently Used” category.
  3. a box will appear asking you for the observations you want to calculate the average of; position cursor in first cell of the column containing observations for a sample and, holding the left click down, select all values of a sample. Alternatively, you can hold the shift key down while pressing the down arrow key or the page down key.
  4. click (OK) cell you first set your cursor in.
  5. Steps 1 – 4 can be repeated for other data sets OR you can copy the formula and paste it to another cell. The number of observations must be the same! Not only the formula, but all specifications associated with it will be copied. You can check this by clicking on the cell with the formula; in the formula bar at the top of the spreadsheet, the formula and the specifications are given (e.g. AVERAGE (C3:C17) indicates that the calculation is the average of cells C3 through C17).
- To determine the MODE AND MEDIAN for a data set, use (fx) to select mode and median (again, under the category Statistical)
- To calculate the VARIANCE for a set of observations:
  1. position cursor where you would like the computed variance value to appear
  2. click on (fx) in the toolbar, click on (statistical), scroll down the list in the box at the right and click on (VAR)
  3. a box will appear asking you for the observations you want to calculate the variance of; position cursor in first cell (of row or column) containing observations for a sample and, holding the left click down, select all values of a sample
  4. click (OK)
  5. the variance for the values selected will appear in the box you first set your cursor in. Steps 1 – 4 can be repeated for other data sets OR you can copy the formula and paste it to another cell. The number of observations must be the same! Not only the formula, but all specifications associated with it will be copied. You can check this by clicking on the cell with the formula; in the formula bar at the top of the spreadsheet, the formula and the specifications are given (e.g. AVERAGE (C3:C17) indicates that the calculation is the average of cells C3 through C17).
• To calculate the **STANDARD DEVIATION** for a set of observations:
  1. position cursor where you would like the computed standard deviation value to appear
  2. Click on (fx) in the toolbar, click on (statistical), scroll down the list in the box at the right and click on (STDEV)
  3. a box will appear asking you for the observations you want to calculate the standard deviation of; position cursor in first cell (of row or column) containing observations for a sample and, holding the left click down, select all values of a sample
  4. click (OK)
  5. the standard deviation for the values selected will appear in the box you first set your cursor in.
Steps 1 – 4 can be repeated for other data sets OR you can copy the formula and paste it to another cell. The number of observations must be the same! Not only the formula, but all specifications associated with it will be copied. You can check this by clicking on the cell with the formula; in the formula bar at the top of the spreadsheet, the formula and the specifications are given (e.g. AVERAGE (C3:C17) indicates that the calculation is the average of cells C3 through C17).

• To perform a **t-TEST**:
  1. Enter the sample data sets to be compared in adjacent columns. It is recommended that you label the columns.
  2. Position cursor where you want the computed p value for the t-test to appear
  3. Click on (fx) in the toolbar, click on (statistical), scroll down the list of calculations in the box on the right and click on (TTEST)
  4. A box will appear asking you to define/select the following: array 1 (e.g. those observations in a given sample); array 2 (e.g. those observations in the other sample), tails (enter 2), type (depends on the relationship of data; if the data is “before and after” measurements or data for the same individuals under different treatments, then select “paired” – or 1; if the analysis is for unpaired data sets and the variances for the two data sets are similar, select 2; if the analysis is for unpaired data sets and the variances are clearly very different, select 3).
  5. When all boxes are filled in, click on (OK). If you want to change information in a box, simply click the mouse in the box, delete the information, and enter new information.
  6. A calculated p value will appear in the box you first set your cursor in.

**Generating a Bar Graph**

• If you enter the information to be graphed carefully, this can be a breeze! I recommend typing in your data as follows (label the rows and columns also, specifically noting the category of the I.V.!).

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.V. case 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I.V. case 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I.V. case 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(etc.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• Select any cell outside of the data range. Click on the Chart Wizard button on toolbar at the top of the page. Select: Column chart type and Clustered column chart sub-type, then click the Next button.
• A box will appear that will allow you to enter specifics about the chart: highlight only those cells that contain the I.V. names and the values for the means only (left click and hold to highlight a group of cells; do not highlight column head). A preview of the graph will appear within a window of the box. Click the Next button.
• Another box will appear that will allow you to provide a title and legend and axes labels and units. Click the Next button.
• The next box to appear asks you where you want to place the new graph – on the same page as the one you are currently working in or on a new sheet. The choice is yours. To add standard error bars, position the cursor within a column and right click. Select Format Data Series. Select Y Error Bars. Click on the first Display (e.g. Both) and click the button next to Custom. Click in the + box and then go to the spreadsheet, find the corresponding SDs and highlight them with the mouse. Click in the – box and the go to the spreadsheet and highlight the corresponding SDs again (remember that we want to display the
SD above, or +, and below, or -, a column). Click OK. The standard error bars should now appear on the graph.

- Any portion of the graph can be modified by placing mouse over the area that you want to change and right clicking. The options for change will occur in a pop-up box.

**Generating a Scatter Plot and Adding a Trendline**

- Enter the two data sets to be graphed in adjacent columns such that the variable you want to plot on the x axis is in the first (e.g. left most) column. It's a good idea to label the columns for your reference.
- Click on the Chart Wizard and select XY (scatter) as the chart type and Scatter as the chart sub-type.
- A box will appear that will allow you to enter specifics about the chart. Click the mouse in the data range line, and then use mouse to highlight all the data in both columns. This tells Excel what data you want to plot. Do not include the column labels. Click the Next button.
- Another box will appear that will allow you to provide a title and legend and axes labels and units. Click the Next button.
- The next box to appear asks you where you want to place the new graph – on the same page as the one you are currently working in or on a new sheet. The choice is yours. Regardless of location, you can change font size and type and any other feature of the graph.
- Once the data have been graphed, right click while positioning mouse over any one of the data points on the graph. A menu will appear - select “Add Trendline”. The box that appears will have two tabs – Type and Options. Under the Type tab, select linear (click the box). This will direct Excel to add a straight line that best reflects all data points graphed. Under the Options tab place a check in the box for the line equation. This will supply the equation for the generated trendline in “y = mx + b” form. Depending on your need, you may or may not want to check the “Zero intercept” box. Checking this box will direct Excel to originate the trend line through the origin.

**Generating a Histogram (Frequency Distribution)**

- Enter data into the worksheet such that all data from a single data set is in one column. Label the column.
- Go to Tools (on the toolbar), select Data Analysis. A box will appear. In this box select Histogram and then OK. Click mouse in the Input Range box, and then use the mouse to highlight all the data to be used.
- Click mouse in Output Range button, and then click inside the accompanying box. Move mouse to the cell in which you want the organized data to be placed. Then click OK. Excel generates the most logical “bins” (or categories) and the number of observations in each category. This is the information that is graphed. Alternatively, you can specify the “bins” you want Excel to use. In another column labeled bins, type in the bins you want to use (e.g. 1-10, 11-20, 21-30, and so on). In the adjacent column type the upper number of each bin to use as a reference in the analysis (Excel won't allow you to use 1-10, for example; so use 10 to represent this bin). Repeat the Histogram analysis, but this time select your bins to use.
- To graph the data, list bin labels in one column and the corresponding frequencies in the adjacent column. If two distributions are to be plotted on a single graph (e.g. height for males and height for females), place the list of frequencies for one data set (e.g. males) in the column adjacent to the bin labels, and the list of frequencies for the second data set (e.g. females) in the next adjacent column.
- Select any cell outside of the data range. Click on the Chart Wizard button on toolbar at the top of the page. Select: column chart type and clustered column chart sub-type, then click the Next button.
- A box will appear that will allow you to enter specifics about the chart: highlight the cells with the bin labels and number of observations. A preview of the graph will appear within a window of the box. Click the Next button.
- Another box will appear that will allow you to provide a title and legend and axes labels and units. Click the Next button.
- The next box to appear asks you where you want to place the new graph – on the same page as the one you are currently working in or on a new sheet. The choice is yours. Regardless of location, you can change font size and type and any other feature of the graph.
Generating a Line Graph

- Enter the two data sets to be graphed in adjacent columns such that the variable you want to plot on the x axis is in the first (e.g. left most) column. It’s a good idea to label the columns for your reference. If the y value is an average of three or more observations, place the SD in the next column (to the right of the corresponding mean).
- Select any cell outside of this area. Click on the Chart Wizard button on toolbar at the top of the page. Select: XY(scatter) chart type and Scatter with data points connected by lines as the chart sub-type, then click the Next button.
- A box will appear that will allow you to enter specifics about the chart. Click the mouse in the data range line, and then use mouse to highlight the data in all columns. This tells Excel what data you want to plot. Include the column labels, but DO NOT include the standard deviations. Click the Next button.
- Another box will appear that will allow you to provide a title and legend and axes labels and units. Click the Next button.
- The next box to appear asks you where you want to place the new graph – on the same page as the one you are currently working in or on a new sheet. The choice is yours. Regardless of location, you can change font size and type and any other feature of the graph.
- To add standard error bars, position the cursor over any data point on the graph and right click. Select Format Data Series. Select Y Error Bars. Click on the first Display (e.g. Both) and click the button next to Custom. Click in the + box and then go to the spreadsheet, find the corresponding SDs and highlight them with the mouse. Click in the – box and the go to the spreadsheet and highlight the corresponding SDs again (remember that we want to display the SD above, or +, and below, or -, a column). Click OK. The standard error bars should now appear on the graph.

Statistical Tests for BIOLOGY 2108 Lab. Exercises

We will be using the unpaired t-tests for lab. 3, “Physiology” and for lab. 4, “Ecology”.

We will use a “Chi-Square Goodness-of-Fit” test for lab. 1, the Drosophila Lab. Exercise.

To perform a Chi-Square Goodness-of-Fit Test:

1. Enter the sample data sets to be compared in adjacent columns, with the “observed” values in the first column, and the “expected” values in the second column. It is recommended that you label the columns.
2. Position cursor where you want the computed p value for the chi-square test to appear
3. Click on (fx) in the toolbar, click on (statistical), scroll down the list of calculations in the box on the right and click on (CHITEST)
4. A box will appear asking you to define/select the following: array 1 (= the observed values) & array 2 (= the expected values).
5. When all boxes are filled in, click on (OK). If you want to change information in a box, simply click the mouse in the box, delete the information, and enter new information.
6. A calculated p value will appear in the box you first set your cursor in.

For lab. 2, the Feral Cat exercise, and in lab. 4, “Ecology”, we will use a different chi-square test, an “R X C Contingency Table Test of Independence”.

Since this test is not readily available on Excel, we will use a program call “E-Z Stat”. Instructions for carrying out this test follow:
USING E-Z STAT

Chi-Square Test (Contingency Table)

1. Open with the EZ-Stat icon; click anywhere on the grid to remove the E-Z Stat logo.

2. Enter “Observed” values (your actual data counts) or your “group A” data beginning with row 1, column 1; press “Enter” after each entry.

3. Move cursor to row 1, column 2 & enter your “Expected” values (which you calculated) or your “group B” data. Be sure that the observed & expected values for each category appear on the same row.

4. Select “Analysis” on the bar at the table’s top; choose “Tests of Independence/Homogeneity”.

5. Select “OK” on the dialog boxes which appear asking you to “enter number of data groups” & those saying “select columns”.

6. Select “No” on the dialog box concerning the use of the Yates correction (if this box appears).

7. Copy the following data for your lab report:
   - “p for Chi-square” ("p-value").