Exercise 2. Analyzing and Presenting Scientific Data
Effectiveness of Antibacterial Substances

During the last lab, you developed a series of hypotheses about the effectiveness of different types of household products, and initiated an experiment to test those hypotheses. Now you will analyze the data to determine whether your hypotheses are supported or refuted.

Goals
- Create scientific graphs in MSExcel.
- Interpret data using statistical tools.
- Understand the concept of statistical significance.
- Use graphs and statistics to support conclusions.
- Write figure captions and conclusions using scientific language.

Data Collection
Collect your group’s plates from last week. Draw pictures of your plates and their zones of inhibition:

A. Soap solutions:  
B. Household cleaners:  
C. Medical disinfectants:

Using the rulers provided, measure the diameter of the zone of inhibition (in mm) for each of your substances and controls. Enter the data in the table below:

<table>
<thead>
<tr>
<th>Disk</th>
<th>Zone of inhibition diameter (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soap solutions</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Household cleaners</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>Medical disinfectants</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
</tr>
</tbody>
</table>
Statistics and Sample Size
Individuals in a study represent a single sample of the population that they come from. In our section-wide experiment, each group is one sample. Your experimental treatments could differ from each other for two basic reasons:

1) There is an actual difference in the effects of different treatments.
2) The difference is due to chance.

Statistical tests give us the probability that the difference is due to chance. If this probability is low (< 0.05), then we can say that the difference is statistically significant.

By collecting many samples, we minimize the probability that our results are due to chance alone.

Entering Data into Spreadsheets
Input your data into the spreadsheet provided by your instructor, or on the whiteboard. Then, enter the section’s soap solution data into your own spreadsheet. Spreadsheets are a valuable tool scientists use to compile, analyze, visualize, and summarize data. They contain a series of cells with row (1, 2, 3) and column (A, B, C) identification. Both text and numbers can be entered into cells. Enter the column headings “Sample,” “Antibacterial soap,” “General soap,” and “Control.” Your spreadsheet should look like this:

![Spreadsheet Example](image)

Enter the sizes of the zones of inhibition for the entire section in their designated columns. Calculate the section mean and standard deviation for each treatment and the control. The mean is just the average. The standard deviation is a measure of the data’s variability. If the samples for a particular treatment are very different from each other, the standard deviation will be high. If the standard deviation is high, then there is a higher probability that treatment differences are due to chance.

Calculate the Mean
1. Type “Mean” in the sample column below your last sample.
2. Type “=AVERAGE(highlight data)” into the next cell.
3. Move the cursor to the lower right corner until it forms a small box.
4. Click and drag the small box over the next two cells to copy the formula for the other data.

Calculate the Standard Deviation
1. Type “StDev” in the sample column below mean.
2. Type “=STDEV(highlight data)” into the next cell.
3. Repeat the click and drag maneuver to copy this formula for the other treatments.

Look at the standard deviations. Which samples are the most variable compared to their means, those for the control, antibacterial, or general soap?
Calculate the 95% Confidence Interval
The true population mean has a 95% chance of falling within the 95% confidence interval. If your samples are highly variable, the confidence interval will be large. If your samples are all the same, the confidence interval will be small. The mean minus the confidence statistic is the low end of the interval, and the mean plus the confidence statistic is the high end of the interval. You will use the standard deviation to calculate the confidence statistic.

1. Type “Confidence statistic” into the sample column below the standard deviation.
2. Type “=confidence(0.05, highlight stdev, type in sample size)”
3. Repeat the click and drag maneuver to copy this formula for the other treatments.

Data analysis and Visualization
Now that you’ve calculated some statistics, it’s time to evaluate your results.

Making a Graph in Excel
To make a graph, you first need to enter the data on an Excel spreadsheet in the format shown below. From this, you will generate your graph. (Note: you should not include the table you used to generate your graph in your lab report).

<table>
<thead>
<tr>
<th>Antimicrobial agents</th>
<th>#1</th>
<th>#2</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Averages</td>
<td>11.00</td>
<td>5.83</td>
<td>0.3333</td>
</tr>
<tr>
<td>Confidence statistic</td>
<td>1.43135</td>
<td>1.178</td>
<td>0.4132</td>
</tr>
</tbody>
</table>

Putting 95% confidence intervals onto your graph
1. Right click on one of the graph columns and choose “Format data series.”
2. Choose the “y error bars” tab.
3. Choose “custom.”
4. Click in the “+” box and then highlight the row of data labeled “confidence statistic.”
5. Then click in the “-“ box and highlight the same row of data labeled “confidence statistic.”
6. Hit enter.

You should now have a graph that shows the means of each of the kinds of products tested in a category. The bar graph should have error bars so that you can determine whether the products differ significantly in their effectiveness.

The following is an example of how your graph might look:
Evaluating your Results

We will evaluate results using the confidence intervals. There are many formulas that would allow us to numerically evaluate the statistical significance of the difference between our different treatments, but we will use visual analysis for now.

If the confidence interval is small, there is not a lot of variability in your samples, and your confidence in the result is strong. For example, you would have a small confidence interval if all of the plates you measured for antibacterial soap had the same size zone of inhibition. Columns 2 and 3 in the example graph have relatively small confidence intervals. Column 1 in the example graph shows data with a large confidence interval.

You can use the confidence intervals to determine whether there is a statistically significant difference between your products. If the confidence intervals overlap (as in Column 1 and 2), there is no statistically significant difference even though the means may appear to be different. There is a > 5% chance that the difference is due to chance.

You would conclude that those two products did not differ in their antimicrobial properties.

However, if the confidence intervals do not overlap (as in Columns 2 and 3, and 1 and 3), there is a statistically significant difference between the treatments. There is a <5% chance that the difference is due to chance, and that is acceptable to scientists.

From this graph, you would conclude that both Products 1 and 2 are significantly more effective than Product 3 but that there is no difference between 1 and 2.
Zone of Inhibition Lab Report (10 points)

Scientists communicate their results to other scientists by writing and publishing scientific papers. Papers include figures and tables along with a written description of the data and their interpretation.

**Format**
Your initial lab report will be in an abbreviated format, in order to develop your writing and data analysis skills over the course of the semester. The report must be typed, spell-checked, and double-spaced. Each comparison type should be on its own page (3 pages total). Late lab reports will not be accepted.

The report must include 4 components:
1. Clearly stated hypothesis for each type of product (you will have 3).
2. Figure pasted from Excel.
3. A well-written, clear figure caption (see below for guidelines).
4. Logical conclusions including whether AND how your data support or refute your hypothesis. In the conclusion, you should refer to your figures by number. For example:

   Leaves of plants grown with the inoculum were 46% larger than those that were not (Figure 1).

**Writing figure captions**
Each figure should have a caption that does three things: 1) states what the data is, 2) tells where the data came from (Bacteria name, number of samples, etc…), and 3) describes the important features of the data in the figure. Note that the figure caption goes below the figure. The first sentence in a figure caption is a fragment. All subsequent sentences are full. DO NOT simply write “Figure 1. Graph of zone of inhibition versus type of product.” Read the captions on the figures below as examples – it is not important that you understand the graph.

---

**Hypothesis:**

**Figure 1.** Antimicrobial activity…

**Conclusions:** Bah blah…

---

**Fig. 5.** The relation between the average temperature and bio-volume of zooplankton in the upper 10 m.

**Fig. 2.** Graphical model of *Bythotrephes* growth. Development time on the horizontal axis represents one complete life cycle of parthenogenetic females. Mean P contents at instar midpoint ages are indicated.
**Writing scientifically**

Whole books have been written about the subject of scientific writing. You will get a brief intro here. Here are a few rules we expect you to follow:

**Do not use casual language or slang.**

**Bad example:** “Barr’s (1980) statement is merely a “cop out” because he refuses to acknowledge that there are major differences between the two groups.”

**Revised:** “Barr (1980) fails to address this issue, because he does not acknowledge…”

**Be concise.**

- Don't use more words than are necessary.
- Don't use complicated jargon when simple terms will do.
- Read over your work to make sure sentences are not redundant.

**Bad example:** The procedure was that approximately one hour prior to the initiation of the experiment, each avian subject was transported by the experimenter to the observation cage. That specific individual subject was presented with various edible materials and ingestion preferences were investigated utilizing the method developed by Wilbur (1965). When indices and measurements of the data collection were finalized and the experiment terminated, the subject was transferred back to the holding cage.

**Revised:** One hour before the experiment, I put each bird in the observation cage, where I studied feeding preferences using Wilbur's (1965) method. The bird was then returned to the holding cage.

**Use active voice, not passive voice.**

- It has been reported by Smith = Smith reported
- There is reason to hypothesize = My hypothesis is

**Change from the following sentences from the passive voice to the active voice.**

- Lactate was produced by *S. aureus*.

- The breakdown of hydrogen peroxide is catalyzed by peroxidase.

- It has been shown by evidence from genetic studies that genes are arranged in a linear order.

- Nearly half the seedlings were eaten by woodchucks.

- It was found that the population decreased with time.

- Absorbance measurements were taken by the author every 15 minutes.
Writing for conciseness: Reduce these jargon terms to a single word.

1. red in color ______________
2. a majority of ______________
3. a great number of ______________
4. due to the fact that ______________
5. during the course of ______________
6. small in size ______________
7. fewer in number ______________
8. for the purpose of ______________
9. in all cases ______________
10. was of the opinion that ______________
11. with the possible exception of ______________
12. take into consideration ______________
13. at the present time ______________
14. it is often the case that ______________

Revising Sentences: These are actual sentences from past student papers. Rewrite them.

1. After analyzing the information attained through our calculations and illustrated in the graphs and charts that all of my hypotheses were unsupported by the results.

2. Granitic rocks are harder than basic rocks, so when there is an interaction between the two, the basic rocks are the victim and, so to speak, lose ground to the granitic rock.

3. I am a little puzzled as to why there would be higher error and affect for regularly dispersed populations because I would think that because it is regularly dispersed it would be easier to get an accurate representation than with the clumped group.

4. There are many environmental/chemical factors that may be looked at.

5. The relationship between zooplankton abundance and the different chemical/environmental factors has become clearer thanks to figures 2-A, 2-B, and Table 2.

6. This refers to the diamonds ability to scathe virtually any surface, while at the same time it is virtually unscatheable.

7. Corundum is a mineral that is created naturally in nature.
8. We got the ideas for what variables to test by previous research that has been done.

9. In our experiment we predicted that the higher the discharge the lower the abundance of most zooplankton taxa. We did show this to be true in our case because the discharge was higher in Amity and it had a lower density than Chester.

10. Many factors are involved when looking at a stream’s features.

11. This study has shown us that there was either too much error in our data collection or there were other factors that we did not hypothesize and test that were a bigger factor than the ones that we tested.

12. The diversity of the creeks were also determined and it was found that Chester Creek had a higher taxon richness, with 28 species present, than in Amity Creek, which only had 22 species present.

**Figure Captions: These are poor figure captions. Discuss how to improve them.**

![Figure 1](image1.png)

**Figure 1.** This figure shows Chlorophyll A and Secchi depth for each of the four sites. There is also a trend line to show the overall shape of the data.

![Figure 2](image2.png)

**Figure 2.** Temperature (blue) Dissolved Oxygen (pink). The cooler the water, the easier oxygen is absorbed, hence DO increases.