36-220 Lab #8
Confidence Intervals

Please write your name below, tear off this front page and give it to a teaching assistant as you leave the lab. It will be a record of your participation in the lab. Please remember to include whether you are in Section A or B. Keep the rest of your lab write-up as a reference for doing homework and studying for exams.

Name:

Section:

- The symbol ♣ at the beginning of a question means that, after you answer that question, you should raise your hand and have either the TA or lab assistant review your answer. Once they have reviewed your work they will place a check in the appropriate space in the table below. The purpose of this check is to be sure you have answered the question correctly.
- **You should try to complete as much of the lab exercise as possible. We understand that students work at different paces and have tried to structure the exercise so that it can be completed in the allotted time. If you work systematically through the handout and still don't complete every question don't worry. The important thing is that you understand what you are doing. Nonetheless, you are encouraged to complete the lab on your own.**

Check-Problem ♣	Instructor's Initials
Question 4	
Question 7	
Question 12	

36-220 Lab #8
Confidence Intervals

1 Getting the necessary data

First let's download the two worksheets we will need for the lab. Go to the class website and dowload the "speed.MTW" and "clouds.MTW" data-sets to your desktop.

2 Calculating Confidence Intervals

1. Select **File** > **Open Worksheet**. In the box that says "Look in:", choose the **Desktop**. Highlight "speed.mtw". Click **Open**.

2. Early scientists (including Kepler and Descartes) thought light was transmitted instantaneously. While other scientists were able to verify that light has finite velocity, the first reasonably accurate measures of the speed of light were made in 1882 by A. A. Michelson and Simon Newcomb. In one experiment, they measured the amount of time it took for light to travel from Newcomb's laboratory on the Potomac River to the base of the Washington Monument and back. The total distance was about 7442 meters (a little over 4 1/2 miles).

3. The data in column C1 are the 66 measurements made in this experiment. The number that appears in the worksheet was calculated by taking the actual measurement (in millionths of a second), subtracting off 24.8 and dividing by 10^{-3}. (Notice that this standardization can create negative values of your "speed" variable).

4. Make a histogram of the speed data.

   Question #1: Is there anything you would consider changing about the data set before calculating a confidence interval for the speed of light?

5. To delete the worst outlier, select **Data** > **Delete Rows**. Indicate that you want to delete row 31 from column C1. Click **OK**.

6. In order to calculate the confidence interval, select **Stat** > **Basic Statistics** > **1-sample t**. Select *SPEED*. Click on the **Options** button and choose a confidence level of 95. Leave the "Test Mean" field blank. Click **OK**.

   Question #2: What is the 95% confidence interval for the speed of light?

7. Now get rid of the other outlier (-2 can be found in row 54).
8. Recalculate the 95% confidence interval.

   Question #3: What is the new confidence interval?

   ♣Question #4: How does this confidence interval compare to the previous one (before you omitted another outlier), in terms of center and spread? Do the changes seem reasonable given the value that was omitted?

   Question #5: The "true value" for this experiment was later determined to be 33.02. Is this value contained in either of the confidence intervals?

It was later determined that there was something about Newcomb's method of measurement that created a **systematic** downward bias in his measurements. The moral of the story is that confidence intervals (and most other statistical methods) aren't very useful if your data are not measuring what you intended.

3 Simulating Confidence Intervals

In this section, we will use simulations to drive home the proper interpretation of confidence intervals.

1. Select **File** > **New** and select **Minitab Worksheet**.
2. First, we will simulate data from a Normal distribution with known mean and variance. Select **Calc** > **Random Data** > **Normal**. Generate 100 rows of data in each of 36 columns. Store in columns C1-C36. Select a mean of 20 and a standard deviation of 5. Click **OK**.
3. Each row is a sample of size 36 from this normal distribution. In all, you have 100 samples of size 36. We will now calculate a 95% confidence interval for each sample.
4. We need to calculate the mean of each row. Select **Calc** → **Row Statistics** from the pulldown menus. Under "Statistic" select "mean." Under "Input Variables:" enter "C1-C36". Enter "C37" in the "Store result in:" field.
5. Now, we need to calculate the standard deviation of each row. Select **Calc** → **Row Statistics** from the pulldown menus. Under "Statistic" select "Standard deviation". Under "Input Variables:" enter "C1-C36". Enter "C38" in the "Store result in:" field.
6. $t_{35,.025} = 2.03$. Therefore, we can calculate a 95% confidence interval using the following formula:
$$\overline{X} \pm 2.03 \left(\frac{S}{\sqrt{n}} \right)$$
7. Let's store the lower bound of the 95% confidence interval in C39. To do this, select **Calc** → **Calculator** from the pull-down menus. Then, type "C39" in the "Store result in" field. Enter "C37 - 2.03*C38/6" in the "Expression" field. Click **OK**.
8. Now let's create the upper bound of the 95% confidence interval in C40. To do this, select **Calc** → **Calculator** from the pull-down menus. Then, type "C40" in the "Store result in" field. Enter "C37 + 2.03*C38/6" in the "Expression" field. Click **OK**.
9. Column C39 contains the lower bound of the 95% confidence interval. Column C40 contains the upper bound of the 95% confidence interval.

Looking at columns C39 and C40, count up how many of the 100 confidence intervals do not contain the true population mean of 20.

Question #6: What fraction of the confidence intervals you calculated contain the population mean?

♣**Question #7: If you were to repeat this exercise over and over, on average, what fraction of the confidence intervals you calculate would you expect to contain the population mean?**

4 2-Sample Confidence Intervals

In this section, we will see how to use confidence intervals to compare two samples.

1. Select **File** > **Open Worksheet**. In the "Look in:" field, make sure "Desktop" is selected. Open "clouds.mtw".

2. This data gives rainfall in acre-feet from 52 clouds, 26 of which were chosen at random to be seeded with a compound containing silver nitrate. *SEEDED* equals 1 if the cloud was seeded and 0 if it was not.

3. The team who developed this seeding compound believe that it is possible that it will affect rainfall, but they do not know whether it will increase or decrease rainfall.

   To get a preliminary idea of this, it is useful to first look at the data. Make a boxplot of *RAINFALL* by *SEEDED*. Select **Graph** > **Boxplot**. Choose type "One Y, With Groups." The "Graph Variable" is *RAINFALL*, and the "Categorical variable for grouping" is *SEEDED*.

   Question #8: Compare the two distributions. Which one has the larger spread?

4. Select **Stat** > **Basic Statistics** > **Display Descriptive Statistics**. In the "Variables" box put *RAINFALL*, in the "By Variables" box put *SEEDED*.

Question #9: Compare the mean and median rainfall for the seeded and unseeded clouds. Does the seeding seem to have an impact on rainfall? If so, does it increase it or decrease it?

5. To calculate the confidence interval for the difference in the mean rainfall for unseeded and seeded clouds, select **Stat** > **Basic Statistics** > **2-sample t**. Highlight the circle next to "Samples in 1 Column". Indicate that the Samples are in C2 and the Subscripts are in C1. Notice that there is a box we can mark to tell Minitab to assume equal population variances. The rule of thumb is that, if the ratio of the larger sample variance to the smaller one is greater than four, you should not assume equal population variances. An alternative to following this rule of thumb is to conduct a hypothesis test to see if the variances are equal. For now, assume that the variances are equal.

6. Both the confidence interval and hypothesis test results will appear in the Session Window. For now, ignore the hypothesis test results.

   Question #10: What is the 95% confidence interval for the difference of the means?

7. Repeat the process to compute the 90% confidence interval. (You'll need to click on the **Options** button to change the confidence level.)

   Question #11: What is the 90% confidence interval?

♣Question #12: Suppose you had been asked to analyze this data for a client. Consider your answers to Questions #10 and #11 above. What would you tell your client about the relationship between cloud seeding and rainfall?