Introduction to Minitab

Throughout the semester, you will find it necessary to use spreadsheets and statistical packages to help you analyze data obtained in the lab. Most commonly, you will likely use Microsoft Excel to perform necessary calculations. If you need a refresher, directions for using Excel to help analyze data obtained in the analytical lab can be found at http://chemlab.truman.edu/CHEM222manual/pdf/spreadsheet.pdf. As a spreadsheet, Excel is very useful for data organization and manipulation. When you need to do a statistical analysis, you will find that it is often easier and more appropriate to use software that is designed for that purpose. The program we will use in this course is Minitab, a statistical software package available on any campus computer or through Truman's virtual network. We will use Minitab to help obtain descriptive statistics and confidence intervals, perform regression analysis to create calibration curves and to help display our data.

Accessing Minitab:
Minitab is available on any computer running the campus image. To access Minitab on the Dell Duos in the lab (or your personal computer connected into the Truman network), you must first log into Truman's virtual network. (If you have not set your computer up to access Truman's virtual network, see https://secure.truman.edu/its-s/viewclient/ to learn how to connect both on and off campus.)
On the Dell Duos:

1. Open the VMWare View client ( ) to launch the network.
2. If it isn't already entered, type view.truman.edu in the box labeled "Connection server" then click Connect.
3. Enter your username and password and choose Login.
4. Choose to Display Full Screen or in a window as you prefer from the drop down menu and click Connect.
5. After logging in, open Minitab16 from the Program Menu.

Alternatively, you can access the virtual network through any browser by going to https://virtualdesktop.truman.edu/, choosing the "VMware Horizon View HTML Access" link on the right and logging in using your Truman username and password.

Brief Overview of Minitab
Minitab is a powerful statistical software package. We will be using just a small portion of the features of the program in this class. If you are interested in exploring further, detailed tutorials are available at http://case.truman.edu/resources/tutorials.asp

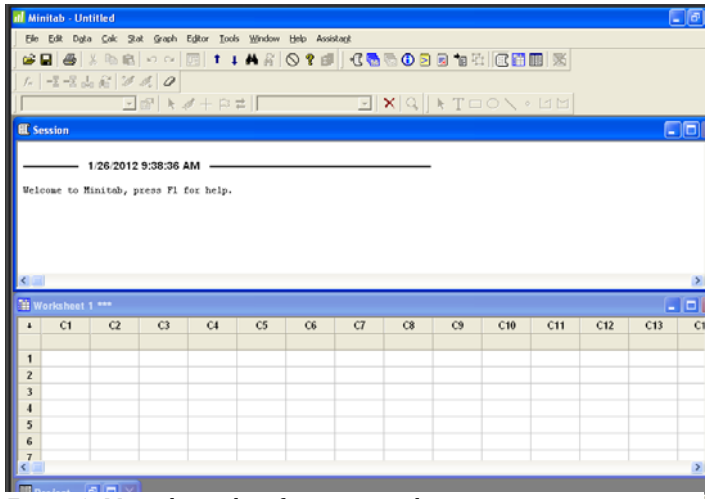

Figure 1. Minitab window for a new analysis.

The main window for Minitab (Figure 1) consists of two important windows – Session and Worksheet. You will enter data into the Worksheet window and, after

performing statistical analysis, results will be given in the Session window (as well as in pop-up windows).

Obtaining descriptive statistics and presenting data

In the following example, you will use Minitab to find the descriptive statistics of the data obtained in an experiment to calibrate a volumetric pipet and a micropipet (Table 1).

Table 1. Calibration data obtained for volumetric pipet and micropipet

5-mL volumetric pipet		100-μL micropipet	
Trial	True Vol (mL)	Trial	True Vol (mL)
1	5.2348	1	0.0983
2	5.1303	2	0.0986
3	4.9237	3	0.0991
4	5.0754	4	0.0995
5	5.1877	5	0.0992
6	5.0226	6	0.0994
7	5.0652	7	0.0992
8	5.0671	8	0.0993
9	5.0799	9	0.0993

1. Type a name for each column in the gray boxes directly under C1 and C2. (For example, one could read "5 mL" and one "micro".)
2. Now enter the true volumes (determined after buoyancy and density correction in Excel). These data may be entered by typing directly or by copying the corresponding information from Excel and pasting it directly into Minitab.
3. Under the "Stat" menu, choose "Basic Statistics" and "Display Descriptive Statistics."
4. Choose the variable(s) you want to analyze by first clicking in "Variables" and then double clicking on the name(s) of the data set(s).
5. Click on "Statistics" to see the types of results that will be displayed. You may explore these different results, but the most useful for our purposes will be the mean, standard deviation, minimum, maximum, median and first and third quartiles. Click "OK" when the results you wish to see are chosen.
6. If all you want to see are the descriptive statistics, then click "OK" again. If you wish to display the data graphically, click on "Graphs." In the "Graph" window, you may choose to see a histogram of your data (with or without a normal curve), an individual value plot and/or a boxplot. Any of these will give you a visual idea of the variation in your results. For now, choose all four options so you can see what type of results each shows. Click "OK."
7. Now, click "OK" in the Display Descriptive Statistics window to see your results.
8. A series of pop-up windows will open displaying each graph. You can click on each component of a graph to change the display. You can print each graph or copy and paste it into a document.
9. In the Session window, scroll up to the top to see a list of the descriptive statistics you selected for each of your variables.
10. You can also use the procedure above to simultaneously observe the results of similar measurements for two or more different categories of data on one graph. See http://case.truman.edu/Documents/Minitab%20Introduction%20to%20Minitab.pdf for more information.

Determining a confidence interval

In order to determine the range of results in which we have a specific probability of finding the true value given our sample data, we will use Minitab to determine a confidence interval for the 5 mL pipet and the micropipet.

1. Enter your data into the Worksheet as before.
2. From the "Stat" menu, select "Basic Statistics" and then "1 sample t."
3. In the pop-up menu, click in the box below "Samples in columns." The data set(s) you have entered should then appear in the window on the left.
4. Double click on the data set on which you wish to perform the analysis.
5. Select "Options." In the pop-up window, enter a confidence interval. Most commonly, you will enter 95.0. Select "OK."
6. After clicking "OK" in the 1-sample t window, scroll back up to find your results in the Session window. You should see the number of samples (N), their mean, the standard deviation, the SE (standard error) of the mean (calculated as the standard deviation divided by the square root of N) and the interval indicating the range of results within which the true value is likely to lie (at a given confidence) if all of the error in the data is random.

Hypothesis testing

Often, you will want to compare the results you have obtained in lab to a "true" value or to compare the results obtained by two different methods of analysis. To do this, we will use Minitab to perform hypothesis testing. (See your analytical or introductory statistics notes for more information.) For a detailed introduction to using Minitab for this process, see http://case.truman.edu/Documents/Minitab%20Hypothesis%20Testing.pdf

Comparing your results to a "true" value: Are they equal? (H_0: $\bar{x} = \mu$; H_A: $\bar{x} \neq \mu$)

1. Enter your data into the Worksheet as before.
2. From the "Stat" menu, select "Basic Statistics" and then "1 sample t."
3. In the pop-up menu, click in the box below "Samples in columns." The data set(s) you have entered should then appear in the window on the left.
4. Double click on the data set on which you wish to perform the analysis.
5. Make sure the box "Perform hypothesis test" is checked. Enter the "hypothesized mean" – in this case, it will be 5.00 mL for the volumetric pipet.
6. Select "Options." In the pop-up window, enter a confidence interval. Most commonly, you will enter 95.0. For the "Alternative," choose "not equal" in this case. Select "OK."
7. After clicking "OK" in the 1-sample t window, scroll back up to find your results in the Session window. In addition to the results included in the confidence interval directions above, you will also see a value of "T" and "P." You can use these to determine whether to reject or fail to reject the null hypothesis.

For these results, "T" is the calculated value of Student's t for the given degrees of freedom. "P" is the *p*-value that helps us determine if the null hypothesis should be rejected at the given confidence level. The *p*-value represents the probability that, under the given conditions of degrees of freedom and standard deviation, a value of the sample mean ($\bar{x}$) this extreme or more would be found if the null hypothesis were true. The smaller the value of *p*, the less likely that this will be possible. For our purposes, if we want to make the comparison at a 95 % confidence level, we will "reject the null" if the value of *p* is less than 0.05 (this value is found as 100 – confidence level).

For the volumetric pipet data, the *p*-value was determined to be 0.020 (see Figure 2). In this case, p<0.05. So, we reject the null hypothesis (that the mean is equal to the "true" value of 5.00 mL). This indicates that, at the 95 % confidence level, there is evidence that there is a statistically significant difference between the mean volume dispensed by this pipet and the manufacturer's claimed volume of 5.00 mL.

One-Sample T: 5 mL

```
Test of mu = 5 vs > 5


                                          95% Lower
Variable  N    Mean   StDev  SE Mean      Bound     T      P
5 mL      9  5.0874  0.0906   0.0302     5.0312  2.89  0.010
```

Figure 2. Results obtained for 1 sample, 2-tailed t-test for 5 mL volumetric pipet in Minitab.

The example above only looked at whether the sample average was equal to the true value, not whether it was greater or less than the true value (a 2-tailed t-test). In order to make this measurement, we must perform a 1-tailed t-test as shown below.

Comparing your results to a "true" value: Is the sample bigger (or smaller)? (H_0: $\bar{x} = \mu$; H_A: $\bar{x} > \mu$)

1. Repeat Steps 1-5 from the previous directions.
2. Select "Options." In the pop-up window, enter a confidence interval. Most commonly, you will enter 95.0. For the "Alternative," choose "greater than" in this case (i.e., for our purposes, we are trying to see if the average of 5.0874 mL is greater than 5.00 mL). Select "OK."
3. After clicking "OK" in the 1-sample t window, scroll back up to find your results in the Session window. In addition to the results included in the confidence interval directions above, you will also see a value of "T" and "P." You can use these to determine whether to reject or fail to reject the null hypothesis.

For the volumetric pipet data, the *p*-value in this case was determined to be 0.010. Again, here p<0.05. So, we reject the null hypothesis (that the mean is equal to the "true" value of 5.00 mL). This indicates that, at the 95 % confidence level, there is evidence that the mean value is greater than manufacturer's claimed volume of 5.00 mL.

For either case, if p>0.05, we would fail to reject the null hypothesis and have no evidence that there is a statistically significant difference between our measured value and the manufacturer's claimed volume of 5.00 mL.

See http://case.truman.edu/Documents/Minitab%20Hypothesis%20Testing.pdf for directions for using Minitab to compare the averages of **two sets of data** or the results of **paired analyses**.

Creating a scatterplot

To analyze your weigh bottle and buret data, you will create a scatterplot (x-y graph) in Minitab.

1. Enter your data into the Worksheet creating a column for each variable.
2. From the "Graph" menu, select "Scatterplot."
3. In the pop-up window, choose the type of scatterplot you would like to use. If you just wish to observe the data, choose "Simple." In the case of the weigh bottle data, you may choose "With Connect Line" as each measurement was made one after the next and is time-based. For your buret data, choose "With Regression" in order to also determine a line of best fit for your data. If you wish to plot more than one line on the same graph, choose one of the "Groups" options. After choosing the graph style, click "OK."
4. In the next window, select a data set for the x and y variables by first clicking in the column and then double clicking the variable (or highlighting it and choosing "Select").
5. You may use "Scale" and "Labels" to change the appearance of your graph. These may also be changed after the graph is prepared.
6. If performing a regression analysis (for your buret data, for example), choose "Data View." You can add a connecting line in the "Data Display" window if it is appropriate (it is NOT appropriate for your buret data). Click on the "Regression" tab and confirm that "Linear" is chosen and that "Fit Intercept" is selected. Choose "OK" in each window to plot the data.
7. In the pop-up window, click on the title and axes labels to change them to something reasonable. By placing your mouse over the regression line, you will see the equation for the line. Once again, you may copy and paste the plot or print it to a PDF to later include it in your ELN.

Performing simple linear regression

To analyze your buret data, you will perform a regression analysis of your data (similar to finding a line of best fit) in Minitab.

1. First create a scatterplot as in the previous section.
2. In the "Stat" menu, choose "Regression" and then "Regression."
3. For "Response" choose your y variable (your dependent variable). For your buret data, that should be the mass. For "Predictors" choose your x variable (your independent variable). For your buret data, that should be the volume. Hit "OK."
4. Scroll back up in the Session window to see your results. The regression equation will be given at the top in the format "y = b + mx" with b as the y-intercept and m as the slope. Rather than "x" and "y," the actual name of the variables will be given.
5. For more information about the remaining information presented in the Session window, see: http://case.truman.edu/Documents/Minitab%20Simple%20Regression%20Analysis.pdf
6. More information will be given in how to use these results in class.