

## Very Brief Guide to Minitab (Version 14)

Gustavo Cepparo, last updated January 22, 2007

Send comments and corrections to [gcepparo@austincc.edu](mailto:gcepparo@austincc.edu)

To use the textbook datasets, I recommend that you locate the folder which has the data sets as Text (ASCII) files. Find the dataset you want, and then download and open with Notepad or any word-processing program. Next, left-click with the mouse, and highlight the data. Click **Edit>Copy**, open Minitab, put your cursor at the top of the worksheet, and click **Edit>Paste**. (This same method works on data from other sources as well as our textbook datasets, and it also is the first step in putting data into the worksheet for the statistics program CrunchIt.)

Practice with the data below (if you are reading this document on a computer) by left-clicking from Chol/Urban to 205, and pasting to Minitab under C1 (Column one). Be careful: the name Chol/Urban must be on the gray box (under C1) and NOT in Row 1! If you make a mistake about where to put the labels, don't try to erase it. Simply insert a new worksheet and start over.

### Chapter 1 (Picturing Distributions with Graphs)

#### Histogram

Cholesterol Level among urban Guatemalans

Chol/Urban

155

134

227

228

236

241

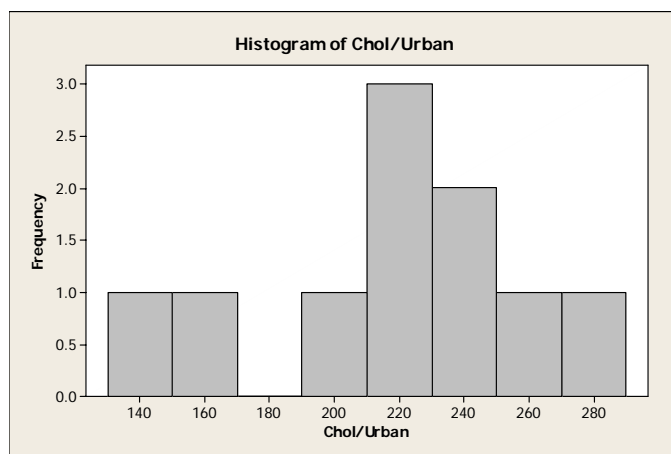
222

252

273

205

Click **Graph > Histogram > Simple**, then click **OK**. Type C1 (under Graph Variables), then click **OK**.



## Stemplot

Cholesterol Level among urban Guatemalans

Chol/Urban

155

134

227

228

236

241

222

252

273

205

Click **Graph > Stem-and-Leaf**. Type **C1** (under Graph Variables), then click **OK**.

**Stem-and-leaf of Chol/Urban    N    = 10**

**Leaf Unit = 10**

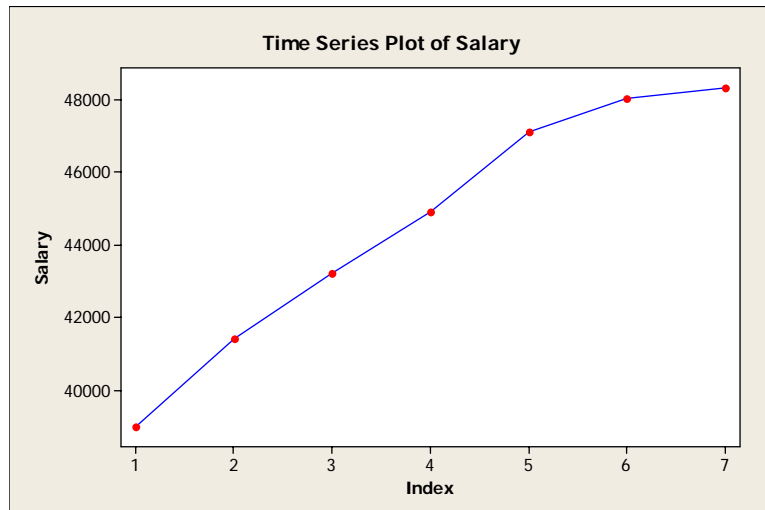
1	1	3
2	1	5
2	1	
2	1	
3	2	0
(4)	2	2223
3	2	45
1	2	7

## Time-plot

Median Salary of Full-Time Professors at Fort Riley Community College since 2000

Salary  
39000  
41400  
43200  
44900  
47100  
48000  
48300

Click **Graph>Time Series Plot>Simple**, then click **OK**. Type **C1** (under Graph Variables), then click **OK**.



Note: For other graphs and more options for changing the look of these graphs, refer to *Minitab Manual*, by Betsy Greenberg.

## Chapter 2 (Describing Distributions with Numbers)

### Numerical Summaries

Cholesterol levels among Urban and Rural Guatemalans.

Urban	Rural
155	108
134	114
227	155
228	131
236	220
241	156
222	154
252	171
273	
205	

Click **Stat > Basic Statistics > Display Descriptive Statistics**. Next, under Variables, type **C1 C2**, and click **OK**.

### Descriptive Statistics: Urban, Rural

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Urban	10	0	217.3	13.5	42.7	134.0	192.5	227.5	243.8	273.0
Rural	8	0	151.1	12.6	35.5	108.0	118.3	154.5	167.3	220.0

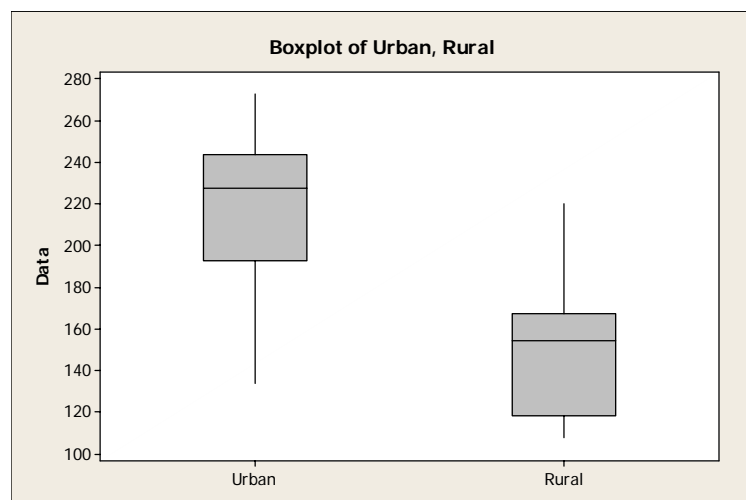
Note: For working with stacked data, meaning data where there are two different sets of observations in the same column, see the *Minitab Manual*.

## Boxplot

Cholesterol levels among Urban and Rural Guatemalans.

Urban	Rural
155	108
134	114
227	155
228	131
236	220
241	156
222	154
252	171
273	
205	

Click **Graph > Boxplot**. Next, you want to choose the bottom left choice “Multiple Y’s Simple”, then click **OK**. To conclude, under Graph Variables, type **C1 C2**. Click **OK**.



Note: For working with stacked data, meaning data where there are two different sets of observations in the same column, see the *Minitab Manual*.

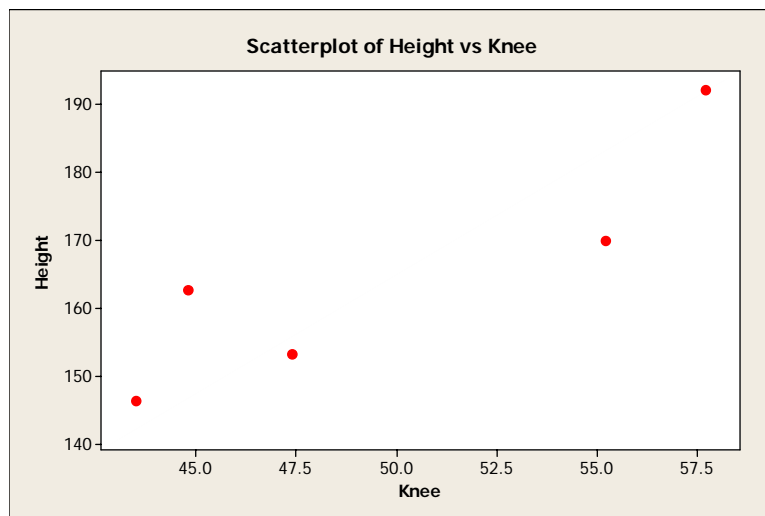
**Chapter 3.** Students are not required to learn to use software for the problems in Chapter 3. The *Minitab Manual* does explain how to use **Calc > Probability Distributions > Normal** to find values that, by hand, can be found in the Normal Table. Use “cumulative probability” and “inverse cumulative probability.”

## Chapter 4

### Scatterplot

Knee	Height
57.7	192.1
47.4	153.3
43.5	146.4
44.8	162.6
55.2	169.9

Click **Graph> Scatterplot>Simple**, then click **OK**. Under Y variables, type **C2**. Under X variables, type **C1**.



### Correlation

Knee	Height
57.7	192.1
47.4	153.3
43.5	146.4
44.8	162.6
55.2	169.9

Click **Stat>Basic Statistics>Correlation**. Under Variables, type **C1 C2**, and click **OK**.

Pearson correlation of Knee and Height = 0.885  
P-Value = 0.046

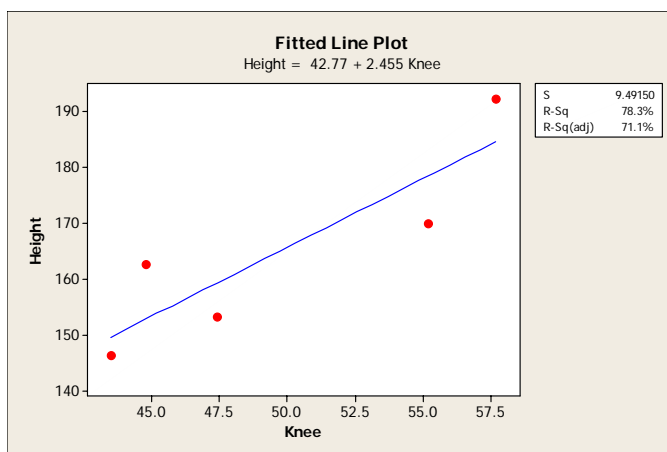
Note: For discussion of various options for scatterplots, see the *Minitab Manual*.

## Chapter 5 (Regression)

Knee	Height
57.7	192.1
47.4	153.3
43.5	146.4
44.8	162.6
55.2	169.9

### Short Regression Output

Click **Stat>Regression>Fitted Line Plot**. Next to **Response**, type **C2**, and next to **Predictor**, type **C1**. Then click **OK**.

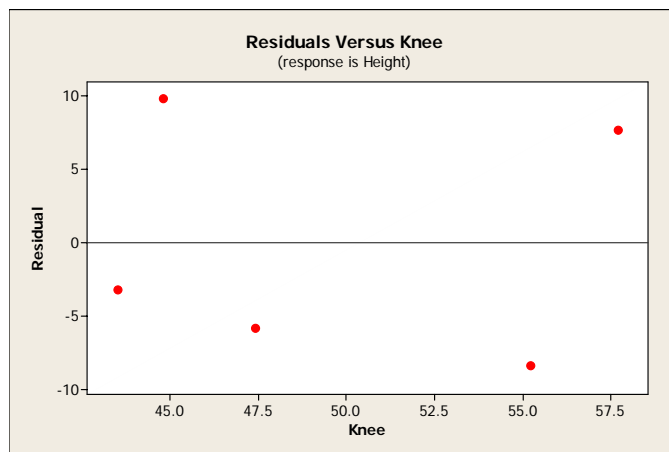


### Residual Plot

Knee	Height
57.7	192.1
47.4	153.3
43.5	146.4
44.8	162.6
55.2	169.9

Click **Stat>Regression>Fitted Line Plot**. Next to **Response**, type **C2**. Next to **Predictor**, type **C1**. Next, click on **Graphs ...**, under **Residuals versus the variables**, type **C1**. Then click **OK**, then **OK**.

See next page for residual plot. →



Note: For additional options, see the *Minitab Manual*.

## Long Regression Output

```
Knee  Height
57.7  192.1
47.4  153.3
43.5  146.4
44.8  162.6
55.2  169.9
```

Click **Stat>Regression>Regression**. Next to **Response**, type **Height** and next to **Predictors**, type **Knee**.

Next, let's compute a point prediction by going to **Options**. Under **Prediction intervals from new observations**, type 55. Then click **OK**, then **OK**.

The regression equation is  
 Height = 42.8 + 2.46 Knee

Predictor	Coef	SE Coef	T	P
Constant	42.77	37.32	1.15	0.335
Knee	2.4555	0.7458	3.29	0.046

S = 9.49150    R-Sq = 78.3%    R-Sq(adj) = 71.1%

### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	976.67	976.67	10.84	0.046
Residual Error	3	270.27	90.09		
Total	4	1246.93			

### Predicted Values for New Observations

Obs	Fit	SE Fit	95% CI	95% PI
1	177.82	5.79	(159.40, 196.25)	(142.44, 213.21)

### Values of Predictors for New Observations

Obs	Knee
1	55.0



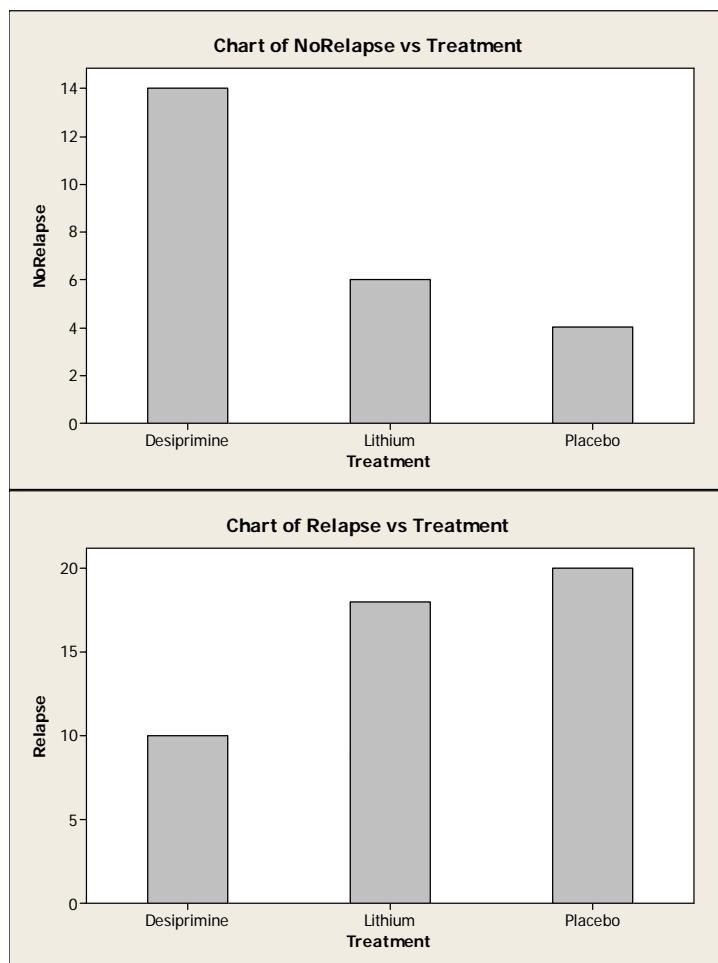
## Chapter 6 (Two-Way Tables)

### Bar Graphs

Treatment	Relapse	NoRelapse
Desiprimine	10	14
Lithium	18	6
Placebo	20	4

When you enter the data on Minitab, you must type **Treatment**, **Relapse**, and **NoRelapse** in the gray cells. If you put them in the white cells, nothing will work correctly. If you make a mistake and start to put them in the white cells, you must go to an entirely new worksheet to fix that – not just retype them in the right place.

Go to **Graphs>Bar Chart**, go down on the pull down menu “Bar represent” to **Values from a Table**, then stay on **Simple** and click **OK**. Next, under “Graph variables,” type **Relapse**, **NoRelapse**, and under Categorical variable, type **Treatment**. To finish, click **OK**.



Note: If you want percents instead of counts after you typed **Treatment** under **Categorical variable**, click on **Bar Chart Options...** and check **Show Y as a percent**.

Note: For table summaries (cross tabulation), see the *Minitab Manual*.

## Chapter 8 (Sampling)

### Simple Random Sampling

Situation:

Suppose you work in the Experimental Psychology Lab at UT and your boss asks for a SRS of 15 rats from the 357 rats in room F, (every rat has a collar with a unique number form 1 to 357).

Go to **Calc>Make Patterned Data>Simple Set of Numbers**. Now, next to “**Store patterned data in**”, type C1. Next to “**From first value**”, type 1. Next to “**To last value**”, type 357, then click **OK**.

Now, in column 1, you have every rat, from number 1 to 357.

Next, we have to take 15 rats from C1. Go to **Calc>Random Data>Sample From Columns...**. Next to “**Number of rows to sample:**”, type 15. Under “**From columns:**”, type C1, and under “**Store samples in:**”, type C2, then click **OK**.

C2  
128  
242  
279  
246  
198  
170  
67  
304  
115  
53  
73  
326  
247  
31  
188

Now, you must go to room F to find the rats that match the above numbers. These rats will participate in a diet experiment.

Note: For more information, see the *Minitab Manual*.

## Chapter 9 (Experimental Design)

### Randomize

Situation:

Suppose you work as a nutritionist and you have 16 subjects and 4 diets (treatments). You can use Minitab to randomize. Type from 1 to 16 in C1.

Go to **Calc>Random Data>Sample From Columns...**, next to **“Number of rows to sample:”**, type 16. Under **“From columns:”**, type C1, and under **“Store samples in:”**, type C2, then click **OK**.

C1	C2
1	7
2	11
3	3
4	8
5	16
6	13
7	10
8	15
9	14
10	9
11	12
12	6
13	1
14	5
15	4
16	2

Next, just read C2 and assign each subject a diet (from 1 to 4) in order: for example, subject 7 takes diet 1, subject 11 takes diet 2, subject 3 takes diet 3, subject 8 diet 4, subject 16 takes diet 1, subject 13 takes diet 2 and so forth ... all the way to the end where subject 2 takes diet 4.

Note: For more information, see the *Minitab Manual*.

## Chapter 10 (Probability)

In this chapter, we mainly use software to simulate random data. Below is a description of how to simulate random data from a discrete distribution. We can also use Minitab to simulate random data from continuous distributions. See the *Minitab Manual* for more examples and details of how to do simulations from continuous distributions and how to summarize the results from simulations both graphically and numerically.

### Discrete

HouseholdSize	1	2	3	4	5	6	7
Probability	.251	.321	.171	.154	.067	.022	.014

In Minitab, under C1 (Number of bedrooms for rental properties) and C2 (Probability), type the data you see below:

C1	C2
HouseSize	Prob
1	0.251
2	0.321
3	0.171
4	0.154
5	0.067
6	0.022
7	0.014

Next, I want the computer to make a simulation by sampling many rental properties with the assigned probabilities from the above table.

Go to **Calc> Random Data > Discrete**. Next to number of rows, type 1000. Under **Store in columns:**, type C4. Next to **Values in:**, type C1. Next to **Probabilities in:**, type C2, then click **OK**.

Now, C4 will have 1000 random rental properties, and we want to summarize these data, so go to **Stat>Tables>Tally Individual Variables**, then check **percents**, and click **OK**.

**Tally for Discrete Variables: C4**

C4	Percent
1	25.50
2	32.00
3	16.50
4	15.60
5	7.10
6	2.50
7	0.80

Suppose I want to find  $P(X \leq 2) = ?$  for my simulated data.  
From my simulation of 1000 randomly selected households,  
I have  $P(X \leq 2) = 0.255 + 0.32 = 0.575$ .

Notice that, in the original population distribution, the probability of getting a house size less than or equal to 2 was  $0.251 + 0.321 = 0.572$ .  
Since a sample size of 1000 is quite large, we would expect that the observed proportion from our simulation would be close to the value of the probability from the population, and it is, since 0.575 is very close to 0.572.

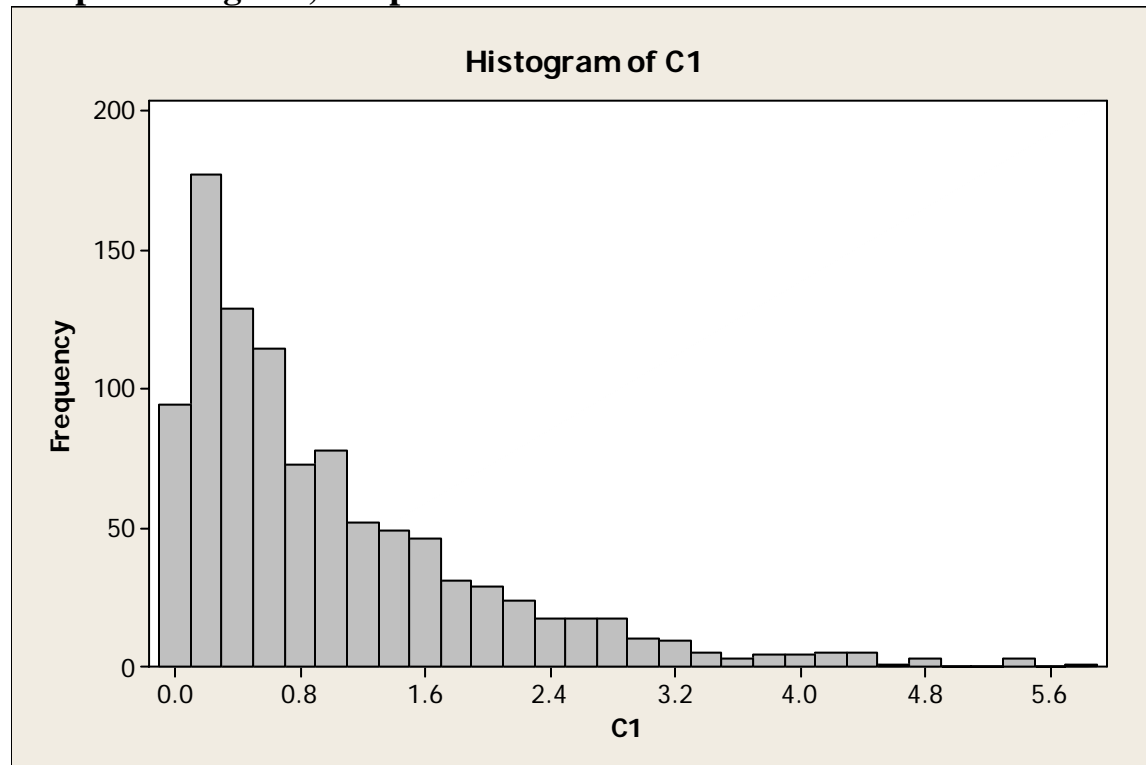
## Chapter 11 (Sampling Distributions)

**Note:** For additional illustrations and discussion of the material below, see the *Minitab Manual*.

### Central Limit Theorem

Example 11.8 uses the central limit theorem. The original distribution (sampling one observation at a time) is exponential with mean 1 hour and standard deviation of 1 hour. Let's simulate this distribution in minitab.

Go to **Calc>Random Data>Exponential**. Next to **Number of rows of data to generate:** enter 1000, under **Store in column(s):** type **C1**, then click **ok**. Now in column 1 you must have 1000 observations from an exponential distribution with mean 1 and standard deviation 1. To conclude lets make a histogram by going to **Graph>Histogram, Simple ok**.



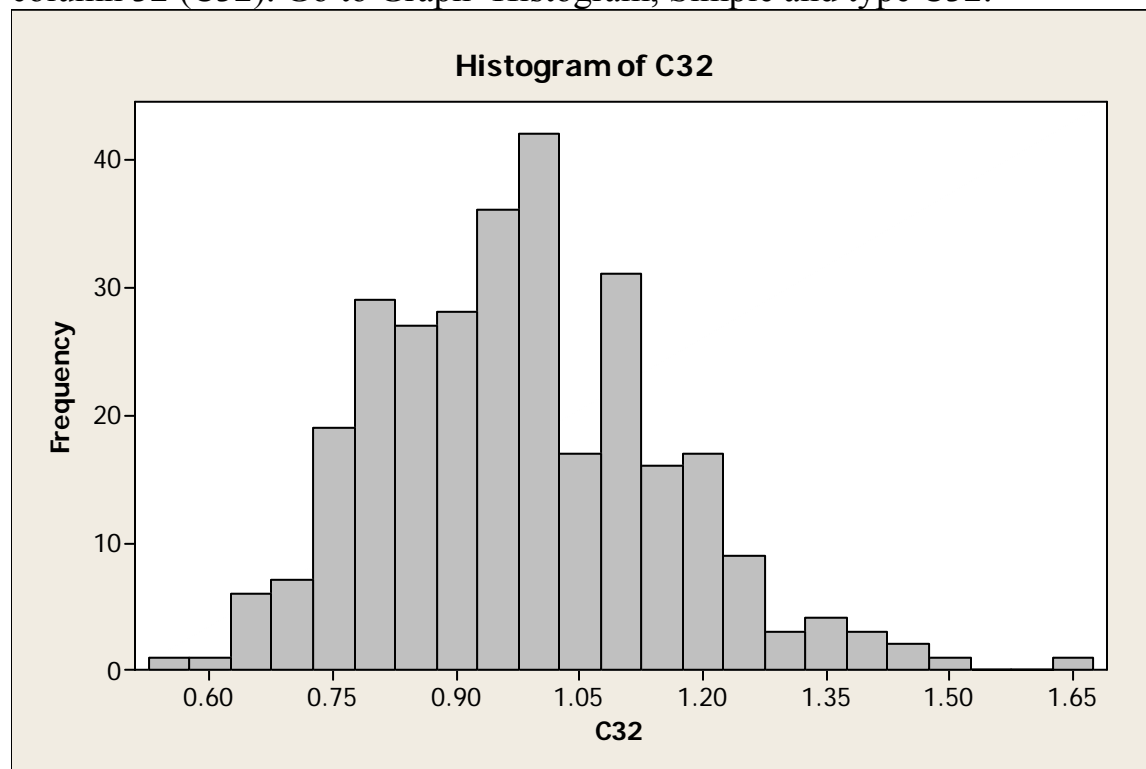
Note: The above distribution is strongly right skewed.

**Now we want to use Minitab to illustrate the central limit theorem.**

Lets start by generating 300 rows in 30 columns, go to **Calc>Random Data>Exponential**. Next to **Number of rows of data to generate:** type **300**, under **Store in column(s):** enter **C1-C30**.

To calculate the sampling distribution for the mean of 30 observations, go to **Calc>Row Statistics**, fill the bubble **Mean**, then under **Input variables:** type **C1-C30**, next to **Store result in:** type **C32**, to conclude click **ok**.

Now we want to graph the simulation of the sampling distribution located in column 32 (C32). Go to **Graph>Histogram, Simple** and type C32.



Next go to **Stat>Basic Statistics>Display Descriptive Statistics**, under variables type C32, then click **ok**.

#### **Descriptive Statistics: C32**

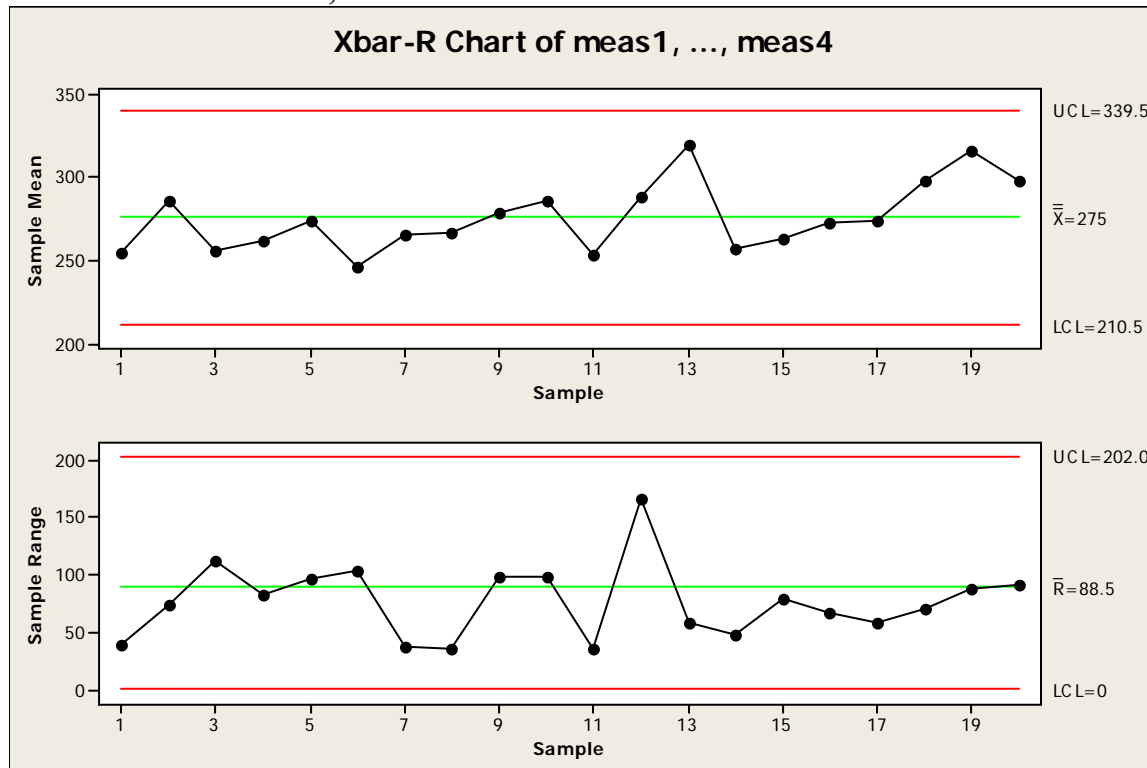
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
C32	300	0	0.9781	0.0101	0.1745	0.5741	0.8555	0.9666	1.0893

Variable	Maximum
C32	1.6359

To conclude, go to the book to compare our simulation with the central limit theorem results.

## Control Charts

Example 11.9 of BPS discusses a manufacturer of computer monitors. The data is on Table 11.1. We will be working with the raw sample data in columns C2-C5. Go to **Stat>Control Charts>Variable Chart for Subgroups>X-bar**. Next, change from **All observations for a chart are in one column:** to **Observations for a subgroup are in one row of columns:**. Continue by typing **C2-C5** and going to **Xbar-R Options**. There, enter **275** next to **Mean:** and **43** next to **Standard deviation:**, to conclude click **ok ok**.





## Chapter 14 (Confidence intervals for the mean)

Note: For illustrations of the dialog boxes, discussion of various options, and interpretation of the output, see the *Minitab Manual*.

In order to compute confidence intervals, you must click on **Stat>Basic Statistics>1-Sample Z...** . If your data is in column 1, you must type **C1** under **Samples in columns**, and next to **Standard deviation**, type the sigma given in the problem. It is NOT CORRECT to find the standard deviation of the sample data and use that standard deviation for problems in this chapter, because in this chapter you are given the population standard deviation to use. And that means you will use a z-statistic.

If you want to change the confidence level, go to **Options**.

Notice that you have a choice of whether to use the full data set or to enter just the summary statistics. And you have a choice of whether to do a confidence interval or a hypothesis test.

## Chapter 15 (Test of Significance)

Note: For illustrations of the dialog boxes, discussion of various options, and interpretation of the output, see the *Minitab Manual*.

In order to test hypotheses, follow the above commands and check the box next to **Perform hypothesis test**. Also, you must check the box next to **Hypothesized mean:..** If you want to change the alpha level or want a one-sided test, go to **Options**. Notice that, if you choose a one-sided test, the confidence interval accompanying it will be a one-sided interval, which is not covered in our course.

## Chapter 18 (Inference about a Population Mean)

Note: For illustrations of the dialog boxes, discussion of various options, and interpretation of the output, see the *Minitab Manual*.

### One sample t test and confidence intervals

In order to compute confidence intervals or hypothesis tests, click on **Stat>Basic Statistics>1-Sample t...**, (similar to Chapter 14 commands). It is different from the Chapter 14-15 commands because it uses the t-statistics rather than the z-statistics and you DO NOT give a value for sigma, the population standard deviation. Instead the software automatically uses the sample standard deviation to estimate sigma. But, for putting in the data and setting up the options, it is similar to the commands described for Chapters 14 and 15.

### Matched pairs t procedures

Go to **Stat>Basic Statistics>Paired t...**, (similar to Chapter 14 commands)

It is sometimes a bit tricky to figure out which direction to use when doing a one-sided hypothesis test here. You may not be able to clearly and correctly identify which direction it should go except by trying one and looking at the result. Try one direction and see if the printed alternative hypothesis is what you intended. If not, then go back and try the other direction in the dialog box.

Alternatively, you can calculate the differences first (**Calc** menu and use one column minus another column to compute a column of differences, which you put in C4 or somewhere like that.) Then do a 1-sample t test on the column of differences. **Stat>Basic Statistics>1-Sample t . . .**

## Chapter 19 (Two-Sample Problems)

Note: For illustrations of the dialog boxes, discussion of various options, and interpretation of the output, see the *Minitab Manual*.

Go to **Stat>Basic Statistics>2-Sample t...**, (similar to Chapter 14-18 commands)

There are more choices here.

The sample data may be in two columns, in one column, or already summarized. Choose the appropriate one in the dialog box.

For this course, we NEVER assume that the variances are equal, so don't check that box.

It is sometimes a bit tricky to figure out which direction to use when doing a one-sided hypothesis test here. You may not be able to clearly and correctly identify which direction it should go except by trying one and looking at the result. Try one direction and see if the result makes sense. If the result is a P-value greater than 50%, you should look very carefully at the output to see whether the hypothesis you tested was really the one you intended. If it wasn't, try the other direction.

Notice that the degrees of freedom MINITAB uses here is from the exact formula given in the book, not from the simpler method you use when doing the problems by hand. Since these may be quite different, it can lead to a fairly different P-value when you do it by hand and when you use software to get a more accurate answer.

## Chapter 20 (Inference about a Population Proportion)

Note: For illustrations of the dialog boxes, discussion of various options, and interpretation of the output, see the *Minitab Manual*.

Go to **Stat>Basic Statistics>1 Proportion...** Most of the time, the data is summarized, so check **Summarized data**. Next to **Number of events**:, write the successes, and next to **Number of trials**:, write the total number of successes plus failures.

See Example 20.4 on page 497.

**Number of trials:** 2673

**Number of events:** 170

Go to **Options** to change to a 99% CI, next click **OK**.

Sample	X	N	Sample p	99% CI
1	170	2673	0.063599	(0.052041, 0.076741)

To use the plus-four procedure, just change the number of events and number or trials appropriately.

## Chapter 21 (Comparing Two Proportions)

Note: For illustrations of the dialog boxes, discussion of various options, and interpretation of the output, see the *Minitab Manual*.

Go to **Stat>Basic Statistics>2 Proportions...** Next follow similar commands to Chapter 20.

It is somewhat confusing that the dialog boxes are arranged so that you always do both a confidence interval and a hypothesis test at the same time, but, of course, you need a different standard deviation formula for the two of those, so the numerical answers you get in the same output will not be correct. Be sure that you look in the Options box and check the box for the pooled p if you want to compute a hypothesis test, and make sure it is not checked to compute a confidence interval.

## Chapter 23 (Two Categorical Variables: The Chi-Square Test)

Note: For illustrations of the dialog boxes, discussion of various options, and interpretation of the output, see the *Minitab Manual*.

Do not look for data files for this chapter. You will enter the table in the Minitab worksheet, and DO NOT INCLUDE THE TOTALS. Then go to **Stat>Tables>Chi-Square Test**.

## Chapter 24 (Inference for regression)

This chapter in the textbook is very computer-intensive. It relies heavily on the same MINITAB commands used in Chapter 5 (long regression output.) The textbook explains how to read this output to find information about confidence intervals and hypothesis testing. Analysis of the residuals is also covered.

See the *Minitab Manual* for more details about producing and reading the output for intervals and tests and about how to produce the graphs for analysis of residuals.