

MINITAB TUTORIAL

by

Dr. Tim Zgonc

Thiel College

August 1996

Fourth Edition
Revised for Minitab Version 14
and Windows XP

by

Dr. Merv Newton

Thiel College

May 2005

Table of Contents

COPYRIGHT	iv
DEDICATION	v
PREFACE	vi
LESSON 1 - INTRODUCTION	1
ENTERING MINITAB	1
WINDOWS	1
ACTIVATING A WINDOW	1
MAXIMIZING AND MINIMIZING A WINDOW	2
THE DATA WINDOW	2
MANUAL DATA ENTRY	2
SAVING A WORKSHEET	3
RETRIEVING A WORKSHEET	4
THE SESSION WINDOW	5
DISPLAY DATA	5
PRINTING THE SESSION WINDOW	6
SAVING THE SESSION WINDOW	6
SAMPLING	6
SAVING AND OPENING PROJECTS	7
EXITING MINITAB	8
RESTARTING MINITAB	8
INSTRUCTIONS FOR ALL MINITAB ASSIGNMENTS	8
MINITAB ASSIGNMENT 1	9
LESSON 2 - FREQUENCY DISTRIBUTION	10
MINITAB ASSIGNMENT 2	11
LESSON 3 - INTRODUCTION TO GRAPHING	12
GRAPH WINDOWS	12
PRINTING GRAPHS	13
EDITING GRAPHS	13
SEPARATING BY VARIABLE	14
SAVING A GRAPH	14
MINITAB ASSIGNMENT 3	16
LESSON 4 - ADVANCED GRAPHING	17
BAR GRAPHS	17
UNTABULATED NOMINAL VARIABLE (RAW DATA)	17
TABULATED NOMINAL VARIABLE (FREQUENCY TABLE)	19

CONTROLLING HISTOGRAMS	20
POLYGONS	22
FINDING GROUPED FREQUENCIES	23
MINITAB ASSIGNMENT 4	24
LESSON 5 - MEASURES OF CENTRAL TENDENCY	25
SEPARATING STATISTICS BY VARIABLES	25
MINITAB ASSIGNMENT 5	27
LESSON 6 - MEASURES OF DISPERSION	28
MINITAB ASSIGNMENT 6	30
LESSON 7 - BOXPLOTS	31
MINITAB ASSIGNMENT 7	34
LESSON 8 - DISCRETE DISTRIBUTIONS	35
MEAN AND VARIANCE	36
MINITAB ASSIGNMENT 8	37
LESSON 9 - BINOMIAL DISTRIBUTION	38
CREATING A PROBABILITY TABLE FOR $X \sim B(n, p)$	38
SAMPLE PROBLEM	40
MINITAB ASSIGNMENT 9	43
LESSON 10 - CENTRAL LIMIT THEOREM	44
NORMALITY OF SAMPLE MEAN	45
MINITAB ASSIGNMENT 10	47
LESSON 11 - CONFIDENCE INTERVALS	48
MINITAB ASSIGNMENT 11	51
LESSON 12 - HYPOTHESIS TESTING; LARGE SAMPLE	52
MINITAB ASSIGNMENT 12	54
LESSON 13 -HYPOTHESIS TESTING; SMALL SAMPLE	55
ANDERSON-DARLING NORMALITY TEST	55
ONE SAMPLE t - TEST	58
OUTLIER EXAMPLE	58
SUMMARY DATA	60
MINITAB ASSIGNMENT 13	61
LESSON 14 - HYPOTHESIS TESTING; DEPENDENT SAMPLES	62
MINITAB ASSIGNMENT 14	64

LESSON 15 - HYPOTHESIS TESTING; INDEPENDENT SAMPLES	65
MINITAB ASSIGNMENT 15	67
LESSON 16 - ANOVA	68
MINITAB ASSIGNMENT 16	71
LESSON 17 - REGRESSION AND CORRELATION	72
REGRESSION ANALYSIS	72
GRAPH	72
CORRELATION	74
PREDICTION	74
SAMPLE WRITE-UP	75
MINITAB ASSIGNMENT 17	77
LESSON 18 - CONTINGENCY TABLES	78
MINITAB ASSIGNMENT 18	80
LESSON 19 - SIGN TEST	81
MINITAB ASSIGNMENT 19	82

© COPYRIGHT 2005

by

Mervin E. Newton

All rights reserved.

DEDICATION

This document is dedicated to the memories of

TIMOTHY ZGONC

1950 - 1998

and his wife

KATHLEEN ZGONC

1949 - 2004

Loyal servants and good friends of Thiel College
and of the Greenville, Pennsylvania community.

PREFACE

This material is used as part of the Elementary Statistics course at Thiel College. The text used in that course is *General Statistics* 4th Ed by Chase and Bown, published by John Wiley & Sons, Inc. All references in this document about page and problem numbers are based on that book. The order in which the chapters are covered in the course is 1, 2, 5, 6, 7, 8, 10, 3, 9, 11, and 12; but not every section of every chapter is included. The numbering of the lessons in this document is dictated by that chapter ordering. The table to the right indicates the section numbers from the text that provides the prerequisite statistical background for each lesson.

The current author would like to thank Professor Jennifer Curry, Professor Jerry Amon, Sara Franco Newton, Melanie Jovenall '00, Tammy Smith '04, Sharon Heilbrun '06, John Svirbly '06, Lesley Bobak '07, Bess Moran '07, and Lee Grable '08 for their help in proof reading this and/or earlier versions of this document.

Thanks also to Dr. Karl Oman who suggested many improvements for this and earlier revisions.

This document was printed from the WWW. The URL is:

www.thiel.edu/academics/math/UsflLnk/Mintb/Default.htm


Last revised : August 17, 2005

Lesson	Section
1	1.2
2	2.2
3	2.3
4	2.3
5	2.7
6	2.8 & 2.9
7	2.10
8	5.4
9	5.5
10	6.7
11	7.2
12	7.3 & 7.4
13	7.5
14	8.2
15	8.4
16	10.3
17	3.3, 3.4, & 9.4
18	11.3
19	12.2

LESSON 1 - INTRODUCTION

NOTE: It is recommended that you create a folder called something like "MintabWork" on your hard drive or J:\ drive before you start this tutorial. You will need to store many files as you work your way through this course, and this will give you a handy place to save them all.

ENTERING MINITAB

To enter Minitab double click on the Minitab logo . If the logo is not on your desktop, you can find it on the K drive: K:\MINITAB. Double click to start Minitab or right click to create a shortcut on your desktop. **NOTE:** This is only a shortcut to the program. Minitab is not transferred to your computer, so you must still be logged onto the Thiel network to use it, even if the shortcut is on your computer's desktop.

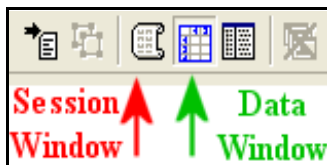
WINDOWS

You will encounter three types of windows in MINITAB. The **session window** contains comments, tables, descriptive summaries, and inferential statistics. The **data window** consists of all the data and variable names. **Graph windows** contain high resolution graphs. There is only one session window but there may be many data and graph windows. Data windows and the session window are discussed below, graph windows will be introduced in Lesson 3.

Upon entering the program, two windows appear. The session window occupies the upper portion of the screen and the data window is below it.


Although many windows may be shown at once, only one window is active at any time. You will know a window is active when its title bar at the top of the window is highlighted. When you first enter Minitab, the data window is active. Notice that the name of the data window is Worksheet 1.

ACTIVATING A WINDOW



Use the mouse to move the cursor anywhere in the window you wish to activate and click the mouse once. Try this with the session window. Click the mouse while the cursor is anywhere in the Session (top) window, and notice that "Session" is highlighted in blue on the top bar. Now click anywhere in the data window and notice that Worksheet 1 is highlighted. You may also activate a window by clicking the appropriate icon on the tool bar. A third method is by using the Window pull down menu from the very top of your screen. Click on Window. A submenu pops up. Click on 1 Session in the submenu. Note that the submenu disappears and that the session window is now active. (A sequence of menu choices such as this will be abbreviated: e.g. Window > 3 Worksheet 1)

MAXIMIZING AND MINIMIZING A WINDOW

The following buttons , which appear at the right end of the title bar of each window, control the size of the window. The button on the left minimizes the window so it is no longer visible. It can be restored by using the Window menu as discussed above. The middle button is a toggle that alternates between maximizing a window and restoring it to its



intermediate size. The right button closes a window. If the contents have not been saved they are lost forever. Notice that there is also a similar set of buttons at the very top of the screen that apply to the Minitab program. When you want to give a command to a particular window, make sure you are clicking on the window buttons and not on the Minitab program buttons. This is particularly important when you are working with a maximized window because the window buttons are much smaller than the program buttons as shown in the image to the left.

Maximize the data window then restore it by clicking on the middle button. Minimize the session window by clicking the left button then restore it using the Window menu.

THE DATA WINDOW

Each time you enter Minitab, the session window contains a date/time stamp and a welcome message, and the data window is empty and active. You will use one of two data entry methods: manually entering data or retrieving an existing file.

MANUAL DATA ENTRY

A new data set which has not been saved may be entered manually and then saved for future use. For example, the following are data collected on eight college freshmen:

STUDENT	SEX	HIGH SCHOOL GPA	H. S. CLASS RANK (% FROM TOP)	NUMBER OF SIBLINGS
1	M	2.65	46	2
2	M	3.72	18	0
3	F	2.95	37	1
4	F	3.59	20	1
5	F	3.05	41	4
6	M	3.10	32	2
7	F	2.85	51	6
8	M	3.86	9	1

Each column represents a variable. In Minitab, each variable name may be as much as 31 characters long, but we shall stick to about eight so that the fields don't get too wide. We will abbreviate the last three variable names as H.S.GPA, H.S.RANK, and SIBS.

Before entering a new data set by hand, be sure that the data window is empty. If you have just entered Minitab, this is the case. Otherwise click on the close button on the top of the data window and you will get a dialog box that will ask you if you want to save the current data window. Click "NO" and you will have a new data window with which to work. It is also a good idea to maximize the data window while you are working on it. Do that now.

Enter the variable names at the tops of the columns first. Click on the tan cell below C1 and type SEX. Use the <→> direction key to move to the top of succeeding columns and type the variable names. Your Data Window should look like the figure below.

↓	C1	C2	C3	C4
	SEX	H.S.GPA	H.S.RANK	SIBS
1				
2				

Each row is called a **record**. When the data set consists of several variables (columns), it is a good idea to enter the data record by record (row by row.) The direction of data entry is then →. Each use of the <Enter> key moves the cursor one cell to the right. When you reach the end of a record, use

the <Ctrl> and <Enter> keys simultaneously to return to the beginning of the next record.

You could also enter data column by column (variable by variable.) In this case the direction of data entry is ↓. You change the direction of data entry by clicking on the arrow in

→	C1-T	C2	C3	C4
	SEX	H.S.GPA	H.S.RANK	SIBS
1	M	2.65	46	2
2	M	3.72	18	0
3	F	2.95	37	1
4	F	3.59	20	1
5	F	3.05	41	4
6	M	3.10	32	2
7	F	2.85	51	6
8	M	3.86	9	1

the upper left corner of the data window. Each use of the <Enter> key moves the cursor one cell down.

When you reach the bottom of a variable column, use the <Ctrl> and <Enter> keys simultaneously. The cursor will move to cell 1 of the next column.

Enter the data for the eight freshmen row by row, that is, one record at a time. Your data window should now look like the figure in the left.

The contents of your data window is called a **WORKSHEET**.

Restore the window to its intermediate size by clicking the middle box in the upper right corner of

the data window.

SAVING A WORKSHEET

Once data has been entered or changed, you should save it to your own hard drive or J: drive. If you make a mistake, if the system crashes, or if you need the data at a later time for a different project, it can be retrieved. (SEE RETRIEVING A WORKSHEET below).

Click on File>Save Current Worksheet As. You will get a dialog box similar to those on most software. Navigate to the location where you would like to save the file, give it a name (let's call it STUDENT) and click the save button. Your worksheet will be saved as STUDENT.MTW. Minitab provides the .MTW extension automatically.

WARNING: All saves and retrieves for worksheets must be done through the File menu with Open Worksheet, Save Current Worksheet, and Save Current Worksheet As... . The open and save icons on the tool bar are for opening and saving projects (discussed later in this lesson).

RETRIEVING A WORKSHEET

Data which was previously entered and saved may be retrieved from an existing file. Minitab refers to such a file as a **WORKSHEET**. It contains data and variable names only. There are no graphs, statistical summaries, or analyses in a **WORKSHEET**.

Let's close your data window and reload the data we just saved. Click the close button at the top of the data window to get a clean data window. Click on File > Open Worksheet. A dialog box similar to the previous one appears. Notice that it is at the location where we just saved our worksheet. Now click on the file STUDENTS.MTW and then on OPEN. (NOTE: If you have exited Minitab then come back to it, Minitab will be looking in the default location and you will need to navigate to the place you saved your file.)

Now close the data window again and let's retrieve a worksheet that comes with the Minitab program. It is on drive K:\, directory minitab, and subdirectory Data. The file is Trees.MTW (All Minitab worksheets have the extension .MTW).

Click on File > Open Worksheet and navigate to Applications on 'Thiel_data'(K:)\MINITAB\Data. Scroll to the right until you find the file Trees.MTW. Click on the desired file then click OPEN.

Maximize the Data window and examine it. Each column is for a different variable. You may refer to it by column number, such as C2, or by its name, such as HEIGHT. Each row is a different record. In this case, a record is a tree. So tree 4 has diameter 10.5, height 72, and volume 16.4.

For a complete description of any of the files provided by Minitab, click Help > ? Help, then on the "index" tab. Type "sam" where it says "Type in the keyword to find." Double click the line that says "Sample data sets." On the right you will now see a list of all of the available data sets. Trees is quite far down the list, so click on the "T" button at the top of the page, then click on "TREES." The description of the file will now be displayed. See the figure below

The diameter, height, and volume of 31 black cherry trees in Allegheny National Forest are recorded in this file.

Column	Name	Count	Description
C1	Diameter	31	Diameter of tree in inches (measured at 4.5 feet above ground)
C2	Height	31	Height of tree in feet
C3	Volume	31	Volume of tree in cubic feet

Restore the Data window to its original size and close it.

THE SESSION WINDOW

Comments and all of your statistical work other than graphs will appear in the session window. While the session window is active, you may edit its contents as you would in a word processor and later print it. The editing commands such as Back Space, Delete, Cut/Copy/Paste, etc. all work exactly as in most word processors.

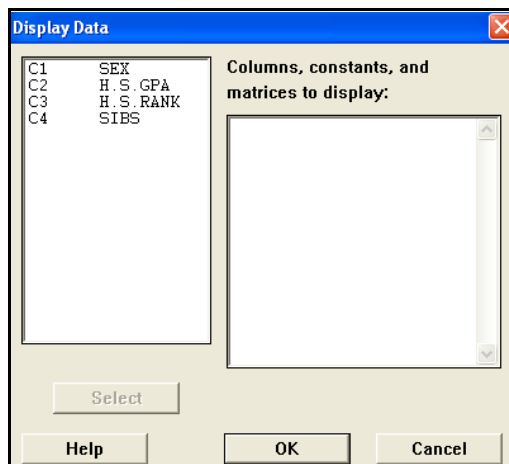
Retrieve the file STUDENTS.MTW that we saved earlier. We will now delete all text except the date/time stamp from the session window. First, activate the session window by clicking anywhere in it, then maximize it. Highlight the text to be deleted by dragging the mouse across it with the left button depressed. Click on Edit > Define from the top menu or simply hit the delete key on the keyboard. Type your name on the line below the date/time stamp, type "Lesson 1" on the next line, and "Example" on the next. Your session window should now look like the figure below.

6/23/2004 2:15:21 PM

Merv Newton
Lesson 1
Example

DISPLAY DATA

We will now have Minitab display the contents of the data window on the session window. Click on Data > Display Data..., and you will see the dialog box below on the left. The cursor should be blinking in the box labeled "Columns, constants, and matrices to display:". Highlight all four variables from the Data Window by dragging the mouse down the list in the box on the left. Click "Select" then "OK". You should see the data from the data window displayed on the session window as shown in the figure below on the right.



Results for: Students.MTW

Data Display

Row	SEX	H.S.GPA	H.S.RANK	SIBS
1	M	2.65	46	2
2	M	3.72	18	0
3	F	2.95	37	1
4	F	3.59	20	1
5	F	3.05	41	4
6	M	3.10	32	2
7	F	2.85	51	6
8	M	3.86	9	1

PRINTING THE SESSION WINDOW

You may print the session window while in Minitab. First, be sure that the Session Window is active. Click on File>Print Session Window or click on the Printer icon on the tool bar. A dialog box similar to what you see when printing from a word processor will appear. It will look different depending on the printer you are using. Check that all settings are as you want them, then click on "OK". In particular, be sure the orientation is set to Portrait. The session window should **always** be printed as Portrait and graphs should **always** be printed as Landscape. Minitab will usually make the right choice automatically, but you should always check your printed output to make sure it is correct.

SAVING THE SESSION WINDOW

You may save the session window as an ASCII file on your hard drive or J: drive so that it can be imported into a word processor such as Microsoft Word. First, be sure that the session window is active. Click on File > Save Session Window As... then proceed the same as you did in saving the data window above. Your session window will be saved with the name you specify and Minitab will assign the extension .TXT.

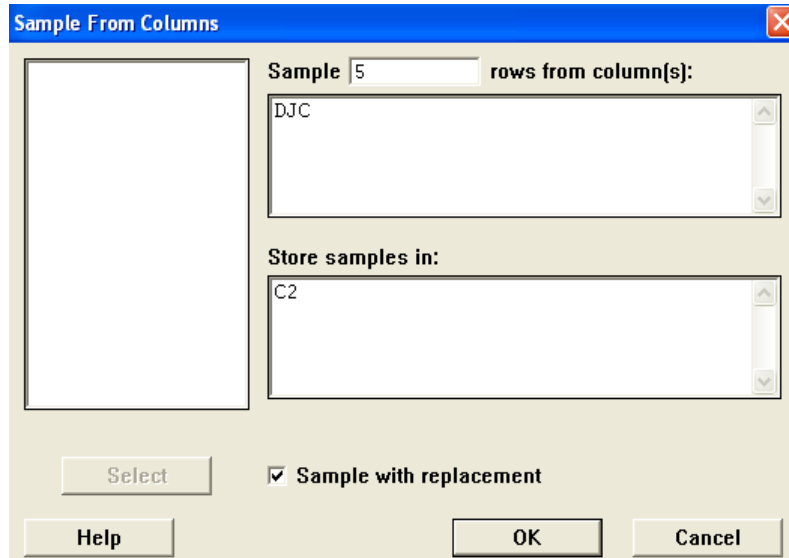
Unfortunately, the above process destroys the formatting. A better way to get your session window into a word processor is to simply copy/paste. That way the formatting will be preserved. That is how the samples of the session window above were placed into this document.

SAMPLING

As has been discussed in class, it is frequently desirable to look at a sample rather than the whole population. This is also possible in Minitab. Retrieve the worksheet K:\MINITAB\Student14\DJC20012002.mtw. The variable C2, DJC, gives the closing Dow Jones Composite index for each working day of 2001 and 2002. C1 is the date and C3 - C32 are the closing prices of the thirty stocks that make up the index. The only variable with which we will be working is C2 DJC. Click at the top of that column to highlight the column. Now click Edit > Copy Cells to send a copy of this column to the clipboard. Now click File > New > (Minitab Worksheet should already be highlighted) > 'OK' to open a new worksheet. Finally, click Edit > Paste Cells to copy the data into the new worksheet. Use File > Save Current Worksheet As to save your new worksheet in your MinitabWork folder. Call it DowJones.

Now suppose we want a sample of 5 values from DJC. Click on Calc > Random Data > Sample From Columns to open the necessary dialog box. Enter 5 into the "Sample ... rows from column(s):" box. This is the size of the desired sample. Now click in the box below that one and the box on the left will show the available variables, (only C1 DJC in this case). Click on C1 DJC to highlight it, then click Select. Now click in the box "Store samples in:" and type "C2". Minitab allows us to take a sample with or without replacement by checking or not checking the check box "Sample with replacement". In most real life applications we would want to sample without replacement, but for our purposes we will need to sample with replacement in future

lessons, so check the box. The dialog box should now look like the figure on the right. Click "OK" and look at C2 to see a sample of size 5. Click in the tan box below C2 and label this variable "Sample1". Now click Calc > Random Data > Sample From Columns again to get back into the sample dialog box and change the content of "Store sample in:" box to C3 and click "OK" to create another sample in column 3. Create 3 more samples of size 5 in C4, C5 and C6, then label these variables "Sample2", ...,



"Sample5". Finally, click Data > Display Data, select C2, ..., C6 and click OK to display these samples on your session window (See the section "DISPLAY DATA" on page 5 above). Your display should look like the figure below, but the numbers will be different since you are not likely to get the same random samples the author obtained.

Data Display					
Row	Sample1	Sample2	Sample3	Sample4	Sample5
1	10276.9	9624.6	7877.4	10281.7	10702.2
2	9834.7	8736.6	9687.1	9872.4	10059.6
3	10405.7	9269.9	10525.4	8235.8	9554.4
4	10858.3	9243.3	8571.6	8511.3	9233.0
5	10586.8	11215.9	8778.1	8847.6	10649.8

Use the Save Current Worksheet As command to save your work sheet as DJSS5 (Dow Jones Sample Size 5). You will need it again in Lesson 5.

SAVING AND OPENING PROJECTS

It is never necessary to save a project, but it is advisable to do so at frequent intervals. If your computer crashes you can reopen your project and continue from where you last saved instead of from the beginning. To save a project the first time you must click on File > Save Project As..., navigate to the desired location, give it a name, and click the Save button. Minitab will assign the extension .MPJ. For subsequent saves or to open a previously saved project you can use the save and open icons on the tool bar.

EXITING MINITAB

To exit Minitab click the close button in the upper right hand corner of the screen. A window will pop up asking you if you want to save changes to the project. If you have already saved and/or printed everything you want, click "No". If there is something you want to save; click "Cancel", save what you want to save, then repeat the exit command.

RESTARTING MINITAB

The instructions for the problems which follow tell you to start each problem with a new instance of Minitab. To do this it is not necessary to exit and restart. You can click on File > New > Minitab Project > OK. You will get the same window asking if you want to save the current project as mentioned above; click No and Minitab will start you out with a new session window and an empty data window the same as when it first comes up.

INSTRUCTIONS FOR ALL MINITAB ASSIGNMENTS

Do each problem in a new instance of Minitab. Remember that it is not necessary to close and reopen Minitab between problems; File > New > Minitab Project > OK will give you a new instance. For each problem please follow the steps listed below:

1. If a saved file is required, retrieve it.
2. Delete everything below the date/time stamp.
3. Type your name below the date/time stamp, type the lesson number on the line below your name, and the problem number on the line below the lesson number.
4. Do the work required by the problem. Your session window should look professional, something you would be proud to turn into your boss if you were hoping for a promotion. In particular, it **should not have**
 - (a) errors
 - (b) excessive blank lines
 - (c) unrequested data, analysis, comments, etc..
 - (d) things penciled in because you forgot to type them before you printed the results.
5. Print all graphs (if any) and the session window properly oriented. (See page 5.)
6. Staple the pages with the session window on top and the graphs in order.
7. Turn in your printed work on the day the assignment is due.

MINITAB ASSIGNMENT 1

See instructions on page 8.

1. Open the worksheet **Trees.MTW** in **K:\minitab\Data**.

Display all the data.

2. Consider Problem 31 (a) on page 15. Enter the data into C1 and C2. Be sure to give them reasonable names. Use "AvrgNumDrks" for C2. Display the data. Type your answer to part (a) of the problem in the session window.

3. Open the Worksheet **EMail.MTW** in **K:\minitab\Student14**.

We will only use the data in C3, the number of emails received. Copy this variable into a new worksheet and save it as **EmailsIn**. Create 5 samples with replacement of size 6 from this data. Save them in C2, ..., C6, name the columns **Sample1**, ..., **Sample5**, and display them in the session window. Save your worksheet with the samples as **EISS6**. You will need this file again in Lesson 5.

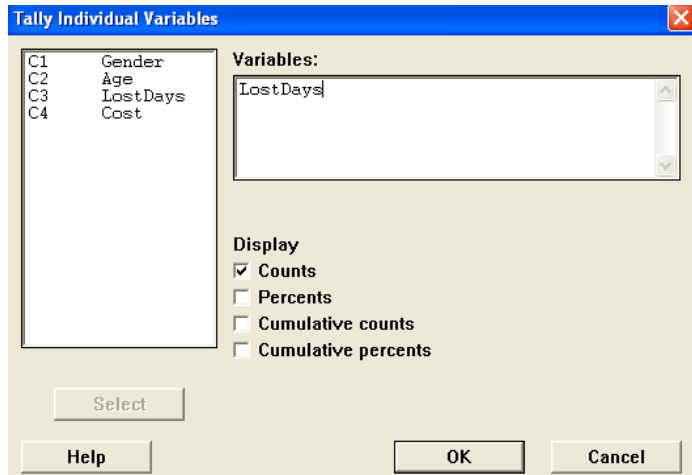
SUGGESTION: It is always a good idea to save a project at regular intervals, especially right after you have created a new worksheet - **EVEN IF YOU DON'T THINK YOU WILL EVER NEED IT AGAIN!** (See **SAVING AND OPENING PROJECTS** on page 7.) That way, if a problem arises with your computer, it may save you from having to start over from scratch.

LESSON 2 - FREQUENCY DISTRIBUTION

Open Minitab and worksheet K:\MINITAB\Student12\BACKPAIN.MTW. (See RETRIEVING A WORKSHEET on page 4 in Lesson 1.) Clear everything below the date/time stamp then type your name, Lesson 2, and Example on separate lines below the date/time stamp.

We would like to create a frequency distribution for the variable LostDays. Click on Stat > Tables > Tally Individual Variables. Select C3 LostDays into the Variables: box. Below the Variables: box we see a list of for results that can be displayed. "Counts" is what our text calls "frequency", "Percents" are what our text calls "relative frequency" except that Minitab expresses them as per cents rather than as decimal fractions. The other two are running totals of the first two. You may check any 1, 2, 3 or all four of the boxes, but for this example let us just choose the default, "Counts". Your dialog box should look like the top figure to the right. You can now click "OK" and the frequency distribution for "Lost Days" will appear in the session window as in the figure to the right.

Unfortunately, Minitab does not have an easy way to create grouped frequency distributions. In Lesson 4 we will see a roundabout way to get the grouped frequencies.



Tally for Discrete Variables: LostDays

LostDays	Count
0	125
1	72
2	8
3	14
4	10
5	5
6	5
7	8
8	3
9	3
10	5
11	1
13	1
14	5
15	1
16	1
17	1
21	1
22	3
24	1
30	1
34	1
46	1
60	1
114	1
180	1
N=	279

MINITAB ASSIGNMENT 2

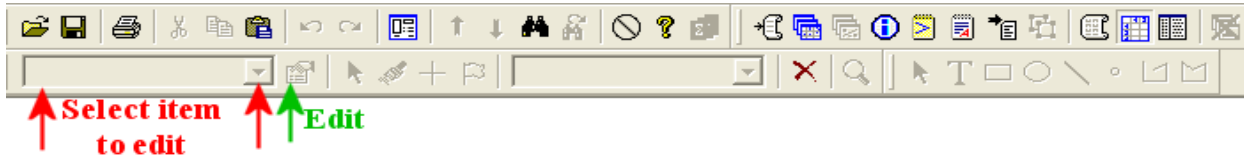
See instructions on page 8 of Lesson 1.

1. Open the Worksheet **EMail.MTW** in **K:\MINITAB\Student14**.

Create a frequency distribution for the variable C4 Out.

LESSON 3 - INTRODUCTION TO GRAPHING

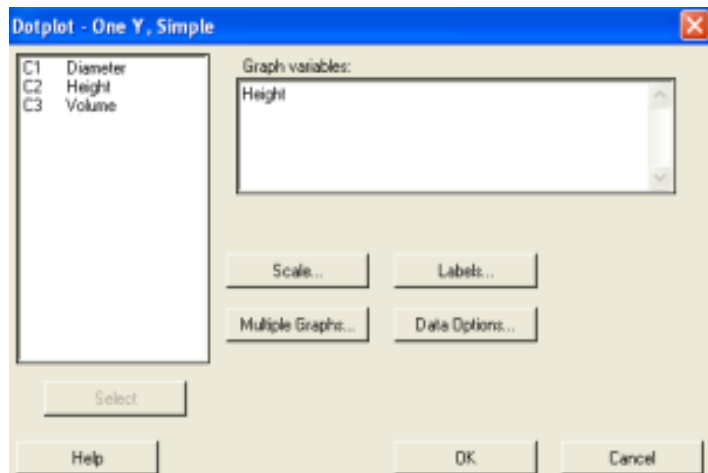
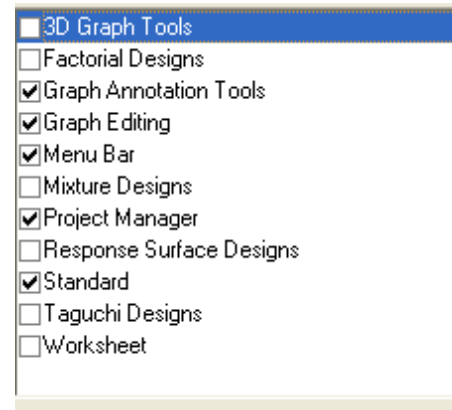
In this lesson you will learn to use Minitab to create dotplots and to separate data by variables. To start, if your tool bars do not look like the figure below,



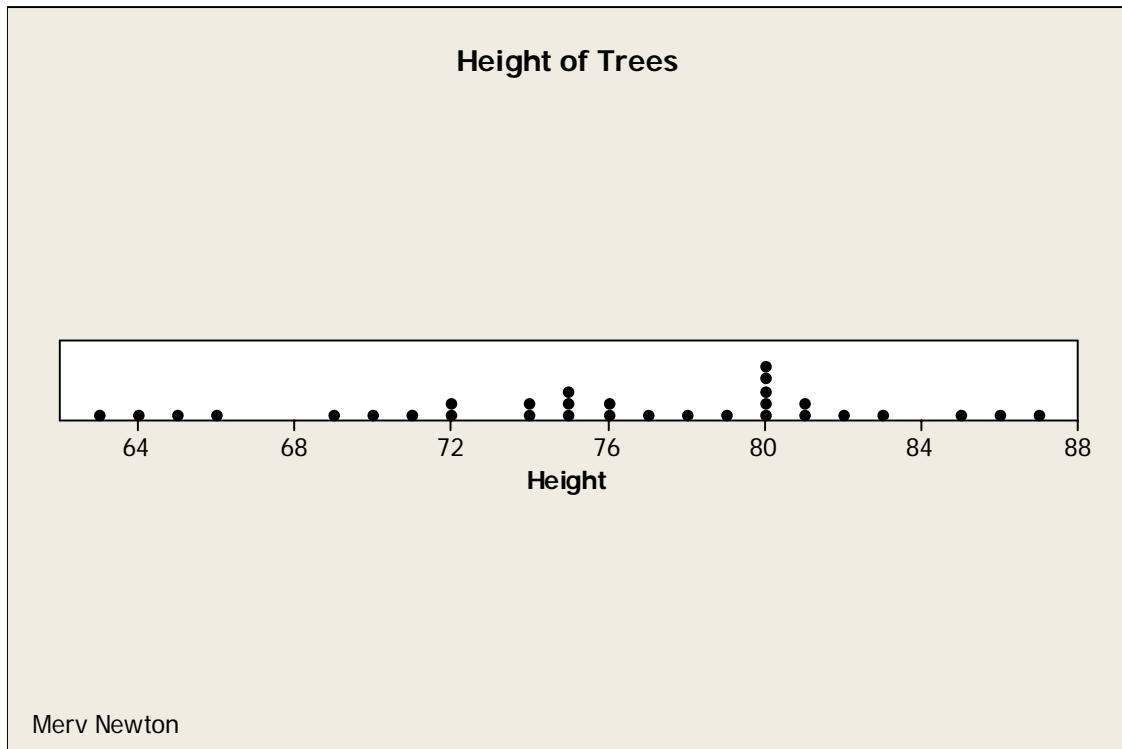
do the following to get the tools where you need them. Click on Tools > Customize then click on the Toolbars tab. In the dialog box that opens, check and uncheck as needed so that it matches the figure to the right. For this lesson we will be using the "Select item to edit" menu and the "Edit" tool indicated by arrows in the figure above.

GRAPH WINDOWS

Once data has been entered into the data window, graphs may be created using the Graph menu. If you have a worksheet open, close it and retrieve the worksheet Trees.MTW from K:\MINITAB\Data. Clear everything below the date/time stamp then type your name, Lesson 3 and Example on separate lines below the time/date stamp. You will create a dotplot of the variable Height. Click on Graph > Dotplot . Click "OK" on the first dialog box that opens, and a second dialog box will appear (shown on the right). The cursor should be in the "Graph variables:" box. Click on C2 Height in the box on the left. Click on the "Select" button and your dialog box should look as it does on the right. Now click on "Labels" and a new dialog box with several blank boxes will appear.



Type "Height of Trees" into the box called "Title:" and your name into the box called "Footnote 1:". Now click on "OK" to close the Labels dialog box then "OK" on the Dotplot dialog box to create the graph. The dotplot appears in its own graph window. It is shown at the top of the next page.



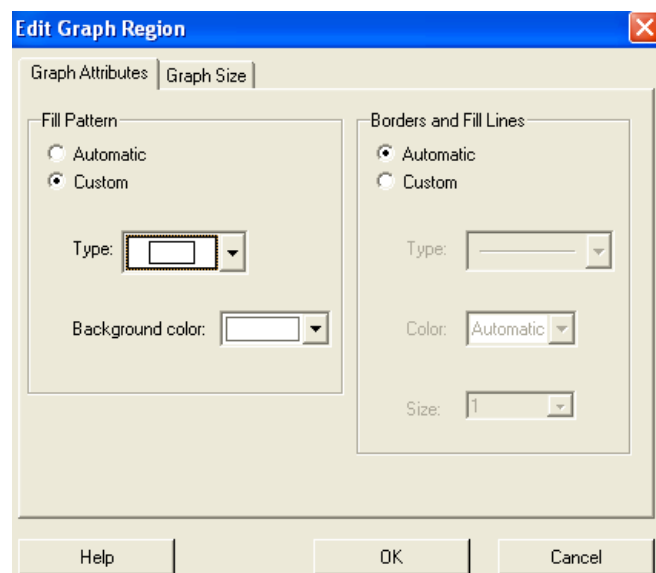
PRINTING A GRAPH

Make the graph you wish to print the active window. Click on File > Print Graph or click on the printer icon on the tool bar. The printer orientation should be Landscape, (Minitab will usually automatically set to Landscape for graphs, but check!) then print the graph.

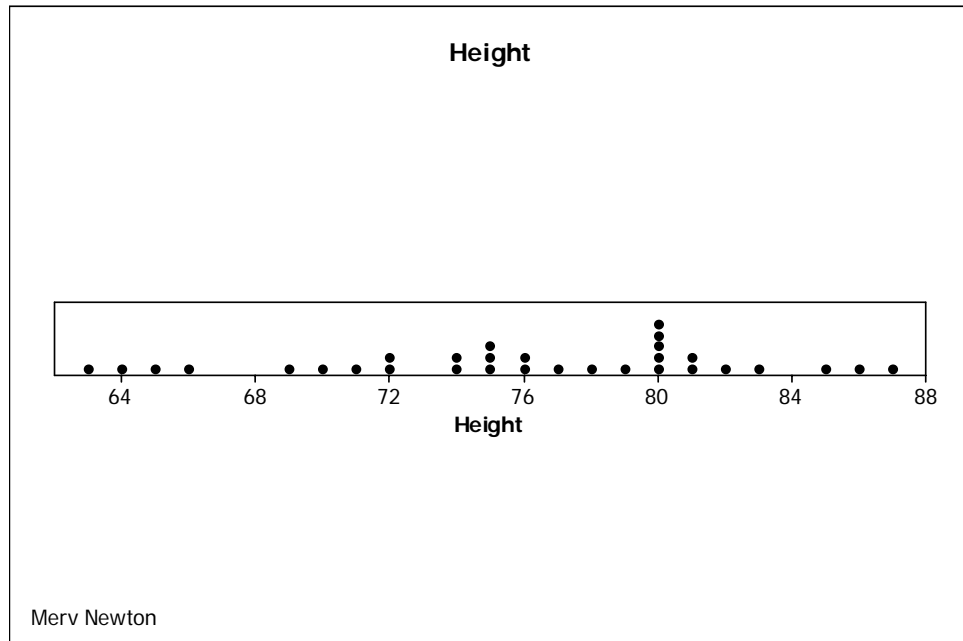
EDITING A GRAPH

It is frequently desirable to make changes to the appearance of a graph. For example, if the graph is to be printed on a black and white printer, it is better to make the graph all black and white before printing it. For this graph we want to make the tan background white.

Pull down the "Select item to edit" menu and select "Graph Region," then click on the "Edit" tool. (See the image at the top of page 12.) You will get the dialog box shown on the right. (You can also get this dialog box by double clicking anywhere in



the tan area of the graph.) Now click on the "Custom" button in the "Fill Pattern" area, then choose white from the "Background color" menu. When you click "OK" the tan will turn white. Choose "<None>" as the item to edit. You can see the result below.



In this course we will make all of our graphs black and white, but if you are doing work for other courses and have a color printer available, you may want to experiment with more colorful graphs.

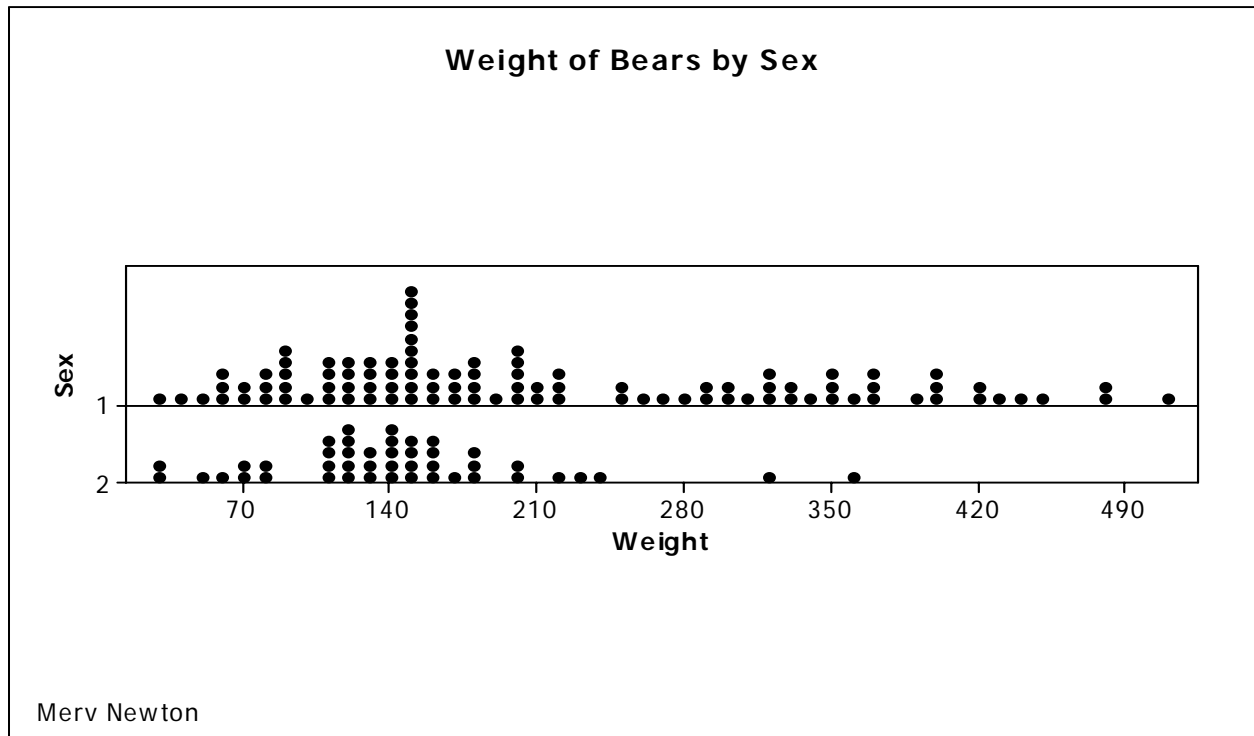
SEPARATING BY VARIABLE

Close your graph and your data window from the previous example, and open the worksheet K:\Data\Bear.MTW. We will create two dotplots, one of the weight of the male bears and another of the female bears. Click Graph > Dotplot and in this dialog box choose "With Groups" then click "OK". In the next dialog box select C10 Weight into the "Graph variables:" box, then select "C4 Sex" into the "Categorical variables for grouping:" box. The rest of the process is the same as before. You should get the graph shown at the top of the next page.

SAVING A GRAPH

There are several ways to save a graph depending on what you want to do with it later. If you wish to incorporate a graph into a report you are writing in a word processor such as Microsoft Word, you may send a copy of the graph to the clipboard and then paste it into your document.

Assuming that the graph you wish to save is the active graph window, click on Edit > Copy Graph. A bitmap of the graph is sent to the clipboard on Windows. Go to your word



processor, set the cursor to the desired location, then click on "Paste" from the Edit menu. The graph above was transferred to this document by this method.

Graphs can also be saved to disk for later use by clicking File > Save Graph As... and following the normal steps for saving any document. The default type is .mgf (Minitab graph), but if you plan to use the graph on a Web page or other document, you will want to choose a different type from the "Save as type:" menu.

MINITAB ASSIGNMENT 3

See instructions on page 8 of Lesson 1.

1. Open the worksheet K:\MINITAB\Student9\Tvhrs.mtw.

Create several dotplots of this data as follows:

- a. Create one dotplot of the variable HrsTV.
- b. Create dotplots of the variable HrsTV grouped by Gender.
- c. Create dotplots of the variable HrsTV grouped by AgeGrp.

LESSON 4 - ADVANCED GRAPHING

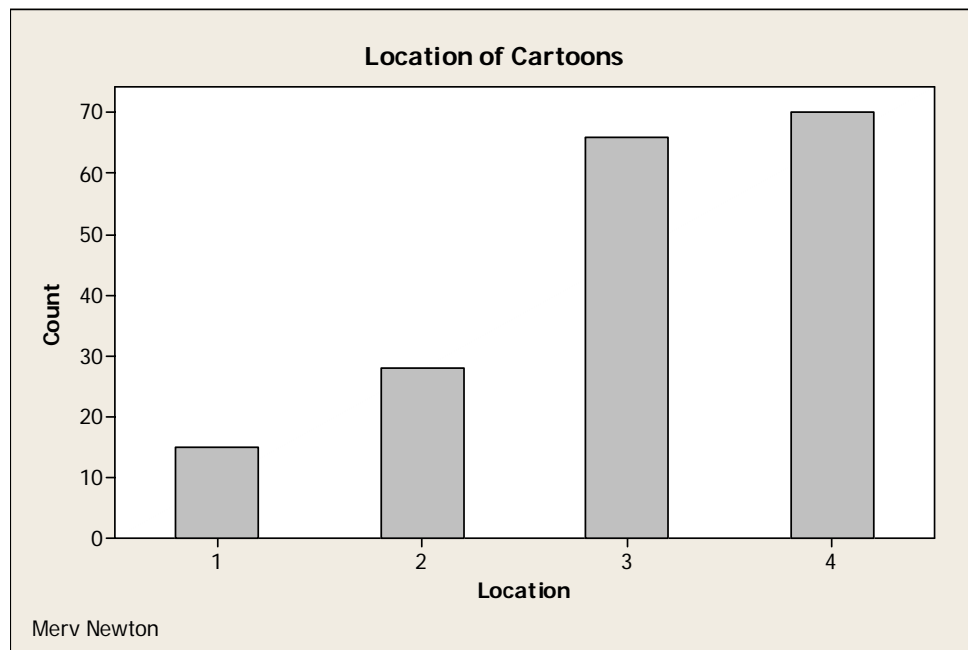
In this lesson you will learn to use Minitab to create bar graphs, histograms and polygons. Be sure your tool bars are set up as shown in Lesson 3 (See page 12).

BAR GRAPHS

Data may be entered into Minitab in more than one way. It may appear as raw data, one record at a time as we have seen in all the examples so far. Minitab also allows tabulated data. That is, data entered as a table such as a frequency table.

UNTABULATED NOMINAL VARIABLE (RAW DATA)

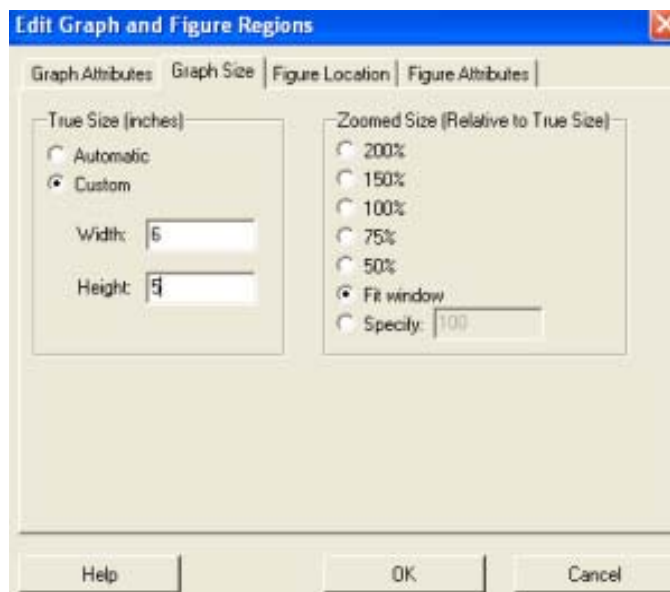
A Hollywood researcher collected data on 179 cartoons. We wish to construct a Bar Graph (Minitab calls them bar charts) of the various locations. Retrieve the worksheet K:\MINITAB\Data\Cartoon.mtw. Clear everything below the time/date stamp and type your name, Lesson 4, and Examples on separate lines below the time/date stamp. Click on Graph > Bar Chart... . Click "OK" on the first dialog box that opens. The next dialog box that appears is very similar to the dialog box for dotplots we saw in Lesson 3 (See page 12). Select C4 Location into the "Categorical variable:" box. (NOTE: Minitab uses the term "categorical variable" for what we called "nominal data" and "ordinal data" in class.) Now click on "Labels" and add a title for the graph and your name as a footnote as we did in Lesson 3. Let's use "Location of Cartoons" as the title of this graph. Now click "OK" in the Label dialog box then "OK" on the Bar Chart dialog box. You should now see the graph shown below.



First, we should make the background white as we did in Lesson 3. There may be other things we would like to change about the graph. For example, let's do the following:

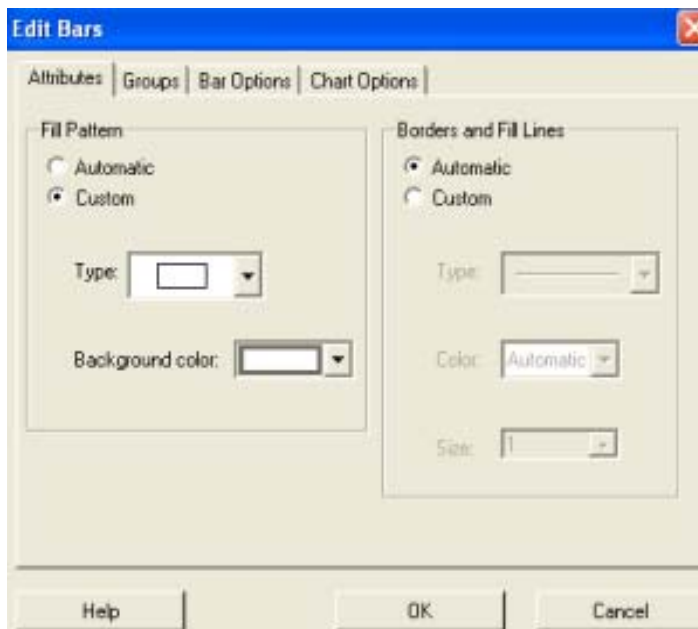
(1) make the graph conform to the 3/4 rule (right now the y-axis is only about half the x-axis), (2) make the bars white instead of gray, (Note: Gray is not necessarily bad, it is only colors that generally behave badly when printed on a black and white printer.), (3) make the spaces between the bars the same as the width of the bars.

(1) This can be done in either of two ways. After setting the background to white on the "Edit Graph Region" or the "Edit Figure Region" dialog box, click on the Graph Size tab. Click the "Custom" button in the "True Size" area of the new dialog box that opens and change the "Height:" to 5. The dialog box should now look like the figure on the right. When you click "OK", the graph window will change proportions and the vertical axis will now be about 3/4 of the horizontal axis.

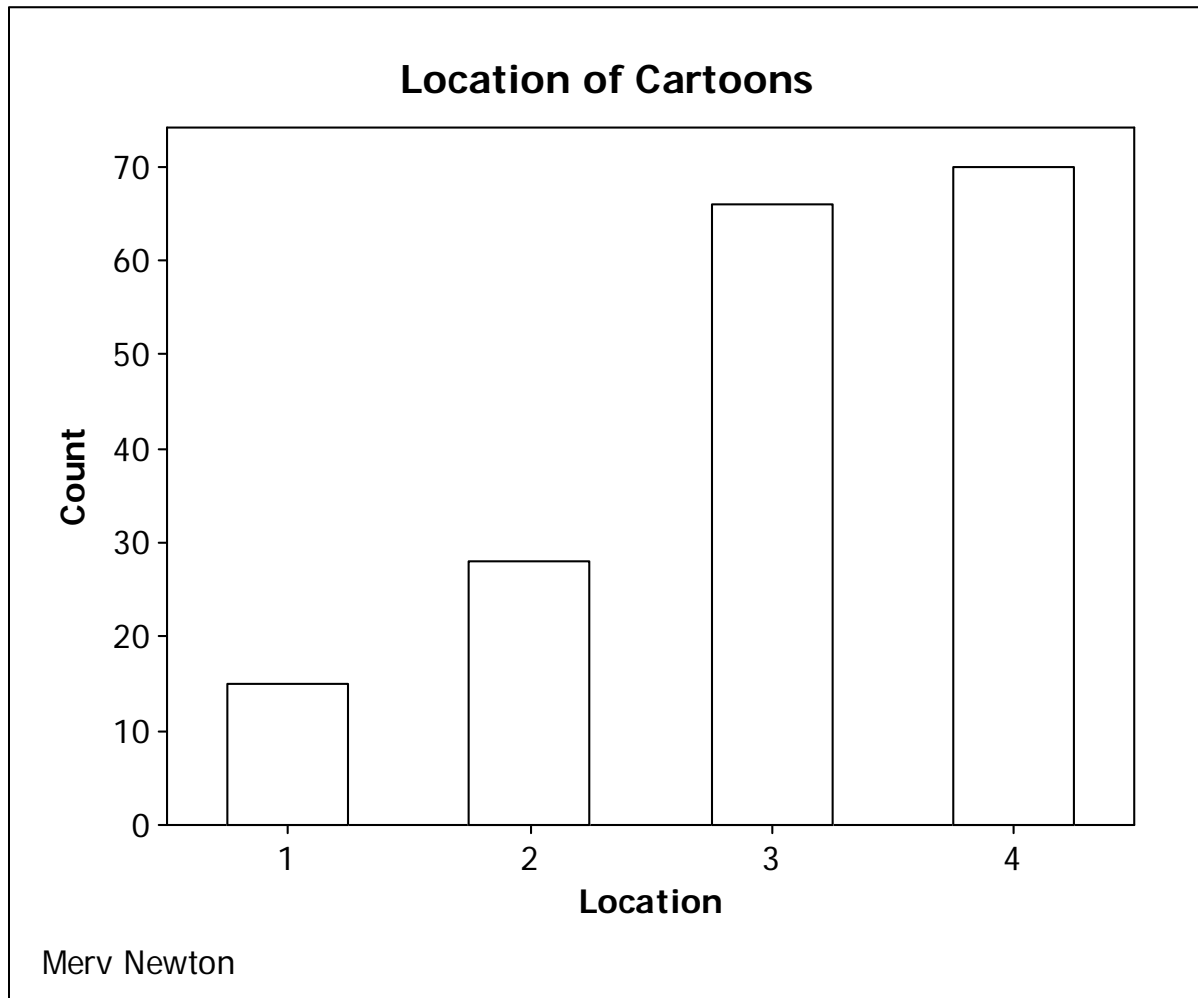


An alternative way to change the proportions of the axes without changing the proportions of the entire graph window is to select "Data Region" as the item to edit. This will cause the data area to be framed with a rectangle that has "handles" at the corners and in the middle of the sides. You can now change the proportions of the data area, and therefore of the axes, by dragging these handles with the mouse.

(2) Select "Bars" as the item to edit and click on the "Edit" tool. You will get a dialog box similar to the one shown above. Click on "Custom" in the "Fill Pattern" area then choose white from the "Background color" menu. When the dialog box looks like the figure to the right, click "OK".



(3) Select "X Scale" as the item to edit and click on the "Edit" tool. Click in the checkbox labeled "Gap between clusters:" to uncheck it. Change the number in the box to a 1 and click "OK". Your graph should now look like the image at the top of the next page.



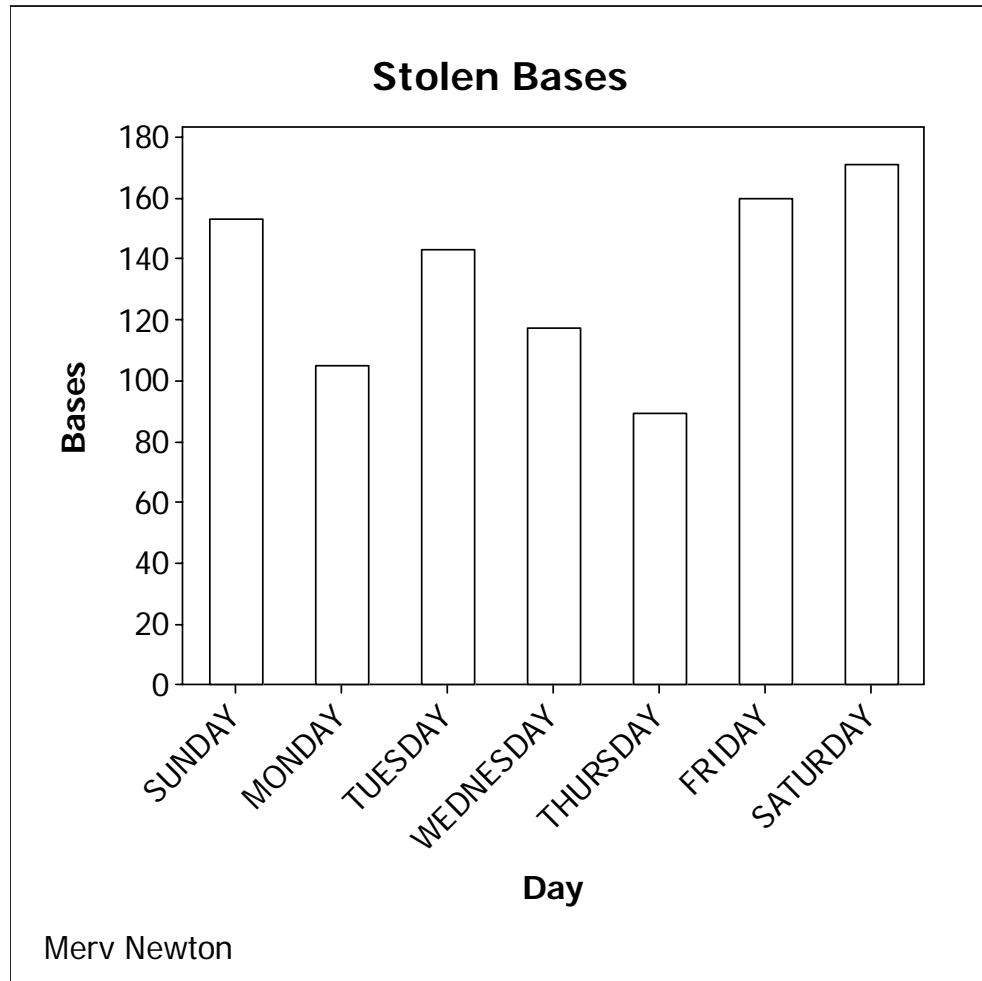
There are two alternative ways to select an object to edit on the graph. Right click on the object and a menu will appear. Click on "Edit *name of object you selected*." Try it! The other way is to double click the feature you want to edit. Also try some of the other tools not discussed in the lesson to see what fun things you can do.

If this were a homework assignment you would now print the graph as shown in Lesson 3, but for the purposes of this example, you may now close it and the data window.

TABULATED NOMINAL VARIABLE (FREQUENCY TABLE)

Early in the 1991 major league baseball season, Rickey Henderson of the Oakland Athletics became the stolen base leader, passing Lou Brock. Newspapers reported the number of bases Mr. Henderson stole for each day of the week. Retrieve the worksheet K:\Minitab\Student9\Steals.mtw. The data window contains a regular frequency table of days of the week Henderson stole bases. We wish to construct a bar graph of this information. Click on Graph > Bar Chart but this time choose "Values from a table" from the "Bars represent:" menu. Now click "OK" and on the next dialog box select C2 Bases into the "Graph variables:" box and

C1 Day into the "Categorical variable:" box. Click on "Labels" to give the graph a title (Stolen Bases) and your name as a footnote. Click "OK" and then use the editing tools to fix up the graph. Make all of the backgrounds white, make the graph conform to the 3/4 rule, and make the spaces between bars the same as the width of the bars. In this case, you will have to make the height of the window 6 to get something close to the 3/4 rule. The graph should now look as it does below.

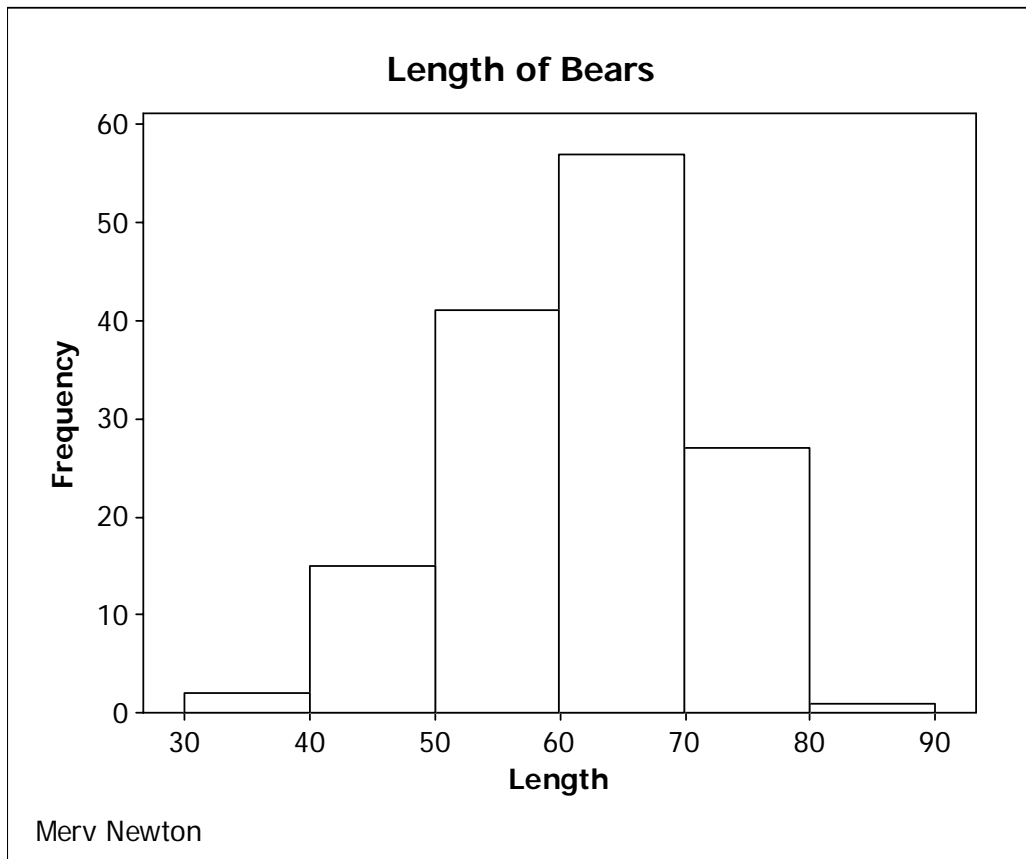


Let's make one more change. Let's make the days of the week slant in the other direction. Get back into the "X Scale" dialog box and click on the "Alignment" tab. At this point the "Text angle:" reads 45. Change it to -45 and click "OK". The names of the days should now slope down and to the right. After viewing the graph, close it and the data window.

CONTROLLING HISTOGRAMS

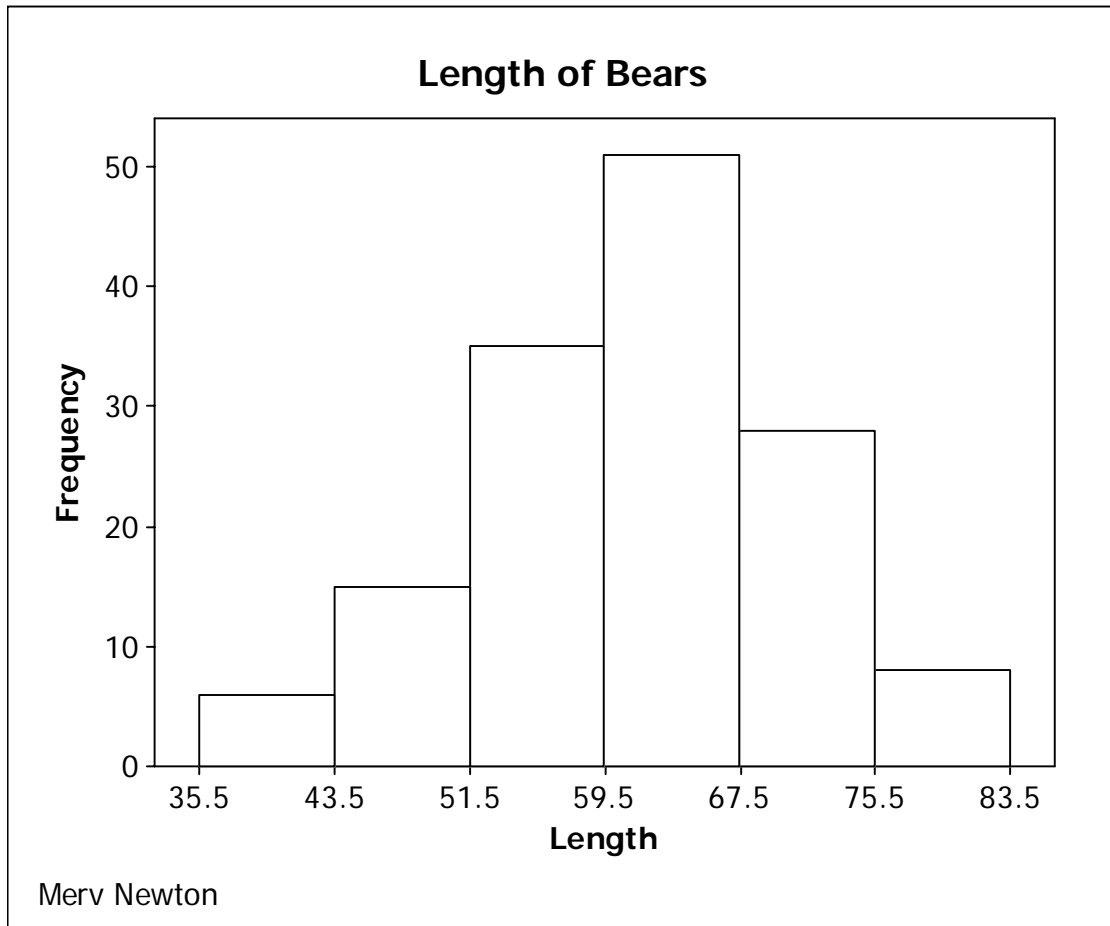
You can control the appearance of a histogram in Minitab just as you do with one constructed by hand. As an example, let's construct a histogram with 6 classes of the length of bears. Retrieve the worksheet K:\Minitab\Data\Bears.mtw then click on Graph > Histogram.

Click "OK" on the first dialog box and select Length into "Graph variables:" then click "Labels" to add a title (Length of Bears) and your name as a footnote then click "OK". What you get in this case is a histogram with 20 classes. First change the tan to white and make the graph conform to the 3/4 rule (height of 5 should do it in this case) as we did with the bar graphs. Also, change the bar color to white. To get the right number of classes, get into the "X Scale" editing dialog box and click on the "Binning" tab. For "Interval Type" click on "Cut point" and for "Interval Definition" click on "Number of intervals:" and change it to 6. Now click "OK" and you should have the graph shown below.



This graph still does not conform to our standards because the class width and class boundaries were not calculated according to our rules. To get what we want, we must define the class boundaries (what Minitab calls "cut points") ourselves. The first thing we need is the largest and smallest data values. We could search the data, but Minitab gives us an easier way. Click Sats > Basic Statistics > Display Descriptive Statistics. Select C8 Length into the "Variables:" box. Click on "Statistics". In the new dialog box that opens, click on the check boxes as needed so that only Maximum and Minimum are checked. Now click "OK" and "OK". In the session window you will see that the minimum is 36 and the maximum is 83. Our formula for the class width with 6 classes is $(83 - 36)/6 = 7.83\dots$, which rounds up to 8. (Remember, always round up unless the fraction yields an integer.) If we choose 36 as the lowest class limit, then the lowest class boundary will be 35.5, and the rest will be 43.5, 51.5, 59.5, 67.5, 75.5 and

83.5. (If you don't have a calculator handy to do the required computations, you can get one by clicking Tools > Microsoft Calculator.) Now get back into the "Binning" dialog box, click on "Midpoint/Cutpoint positions:", delete the existing cutpoints then enter the **first 2** class boundaries listed above into the box (separate with spaces, not commas) and click "OK". Notice that Minitab fills in the rest of the boundary values at equal intervals. You should now see the graph shown below, which satisfies all of our rules.

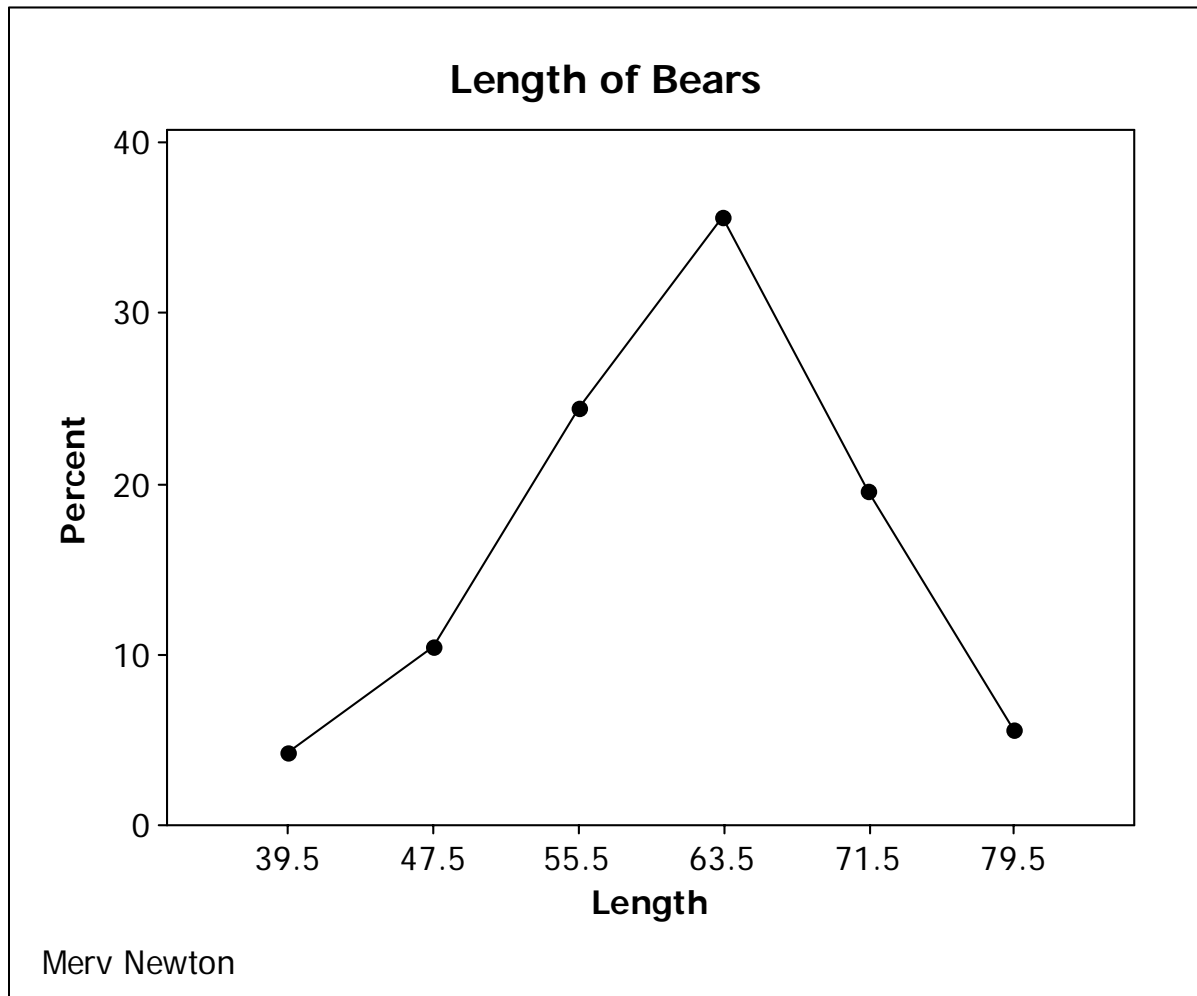


If you want a relative frequency histogram, select "Y Scale" as the item to edit, click the "Edit" tool, click on the "Type" tab, choose "Percent" and click "OK". Change the graph to relative frequency, then back to frequency. Now minimize the graph, as we will need it again later in this lesson.

POLYGONS

Now let us make a relative frequency polygon with 6 intervals of the length of the bears. Click Graph > Histogram > "OK". After selecting the variable to be graphed and entering a title and your name, click on "Data View", uncheck "Bars" and check "Symbols". Now click on the "Smoother" tab, choose "Lowess" and set "Degree of smoothing:" to zero, then click "OK" and

"OK". Select the X Scale as the item to edit, click on the Edit tool and click on the Binning tab. Leave the "Interval Type" as "Midpoint", and in the "Interval Definition" area click on "Midpoint/Cutpoint Positions:". If we start with 36 as the first class limit, the first two midpoint in this case will be 39.5 and 47.5, so delete the existing midpoints and enter these values in the box, and the rest will be computed by Minitab. Now select the "Y Scale" as the item to edit, click the edit tool, click the "Type" tab, and select "Percent". After forcing the 3/4 rule and making everything black and white, the graph will look like the figure below.



FINDING GROUPED FREQUENCIES

As promised in Lesson 2, we will now see how to find the frequencies for a grouped frequency distribution. Close the relative frequency polygon you just completed, and click on Window > Histogram of Length to bring back the histogram you minimized earlier. Now click on Editor > Add > Data Labels > "OK". After selecting "<None>" as the item to edit, you will be able to read the frequencies for the groups at the top of the columns of the histogram.

MINITAB ASSIGNMENT 4

See instructions on page 8.

1. A marketing research firm collected data on 394 individuals who bought jeans from four major stores. A subset of the data is in the worksheet K:\MINITAB\Student9\Jeansa.mtw. Create a bar graph of the Size of the jeans.
2. Do Problem 29 (a) on page 38. Enter the data with the country in C1 and the number of reactors in C2. Abbreviate United States as U.S., United Kingdom as U.K., and Other countries as Oth Contr. so they will fit better in the graph. Be sure to label the columns. Display the data and create a bar graph. See the Rickey Henderson example above.
3. The data in K:\MINITAB\Data\Grades.MTW consists of math and verbal SAT scores and corresponding QPA's.
 - (a) Create a relative frequency histogram with 8 classes of the math SAT scores.
 - (b) Create a frequency polygon with 8 classes of the math SAT scores.

LESSON 5 - MEASURES OF CENTRAL TENDENCY

In this lesson we will see how to use Minitab to find two of the three measures of central tendency that were discussed in class, the mean and the median.

Retrieve the worksheet DJSS5.MTW that you created and saved in Lesson 1. Clear everything below the date/time stamp, type your name, Lesson 5, and Examples on separate lines below the date/time stamp. Now click Stat > Basic Statistics > Display Descriptive Statistics. Select C1 DJC into the "Variables:" box then click on "Statistics". Click on the various check boxes as needed so that only "Mean" and "Median" are checked. Then click "OK" and "OK". The mean and median for DJC should now be displayed in the session window as shown in the box to the right.

Descriptive Statistics: DJC

Variable	Mean	Median
DJC	9706.6	9909.0

Now let us find the mean and median for each of our samples. Click Stat > Basic Statistics > Display Descriptive Statistics again, but this time highlight C2 through C6 and click "Select". Click on "Statistics" and check the box "N total" (to also display the sample size) then "OK" and "OK". The sample size, mean and median for all five samples will be displayed in the session window. If we think of DJC as the population, how do the means and medians of the samples compare with the those of the population?

Now let us explore the advantage of choosing larger samples. In the data window highlight the area that contains the samples of size 5 by dragging across them with the left mouse button depressed, then click Edit > Clear Cells. Now use the procedure from Lesson 1 to create samples with replacement of size 30 into columns C2 - C6. Save this worksheet as DJSS30 as you will need it again in Lesson 10. Finally, find the mean and median of these samples of size 30. If we compare these sample means to the population mean, we can see that they generally are much closer to the population mean than the means of the samples of size five.

SEPARATING STATISTICS BY VARIABLES

Just as we separated dotplots by variables in Lesson 3, we can separate statistics by variables. Close the data window and retrieve the worksheet

K:\MINITAB\Data\Bears.MTW. Use the procedure given above to find the mean length of the bears. Now click Stat > Basic Statistics > Display Descriptive Statistics again, but this time click in the box "By variables (optional):" and select C4 Sex into that box. Now click "OK". The results are shown in the figure to the right. We see that the bears of Sex 1 seem to be longer on the average than those of Sex 2. Since males of


Results for: Bears.MTW

Descriptive Statistics: Length

Variable	Total Count	Mean	Median
Length	143	61.283	61.000

Descriptive Statistics: Length

Variable	Sex	Total Count	Mean	Median
Length	1	99	62.88	64.00
	2	44	57.693	58.750

most mammals are larger than females, we might guess that Sex 1 = Male and Sex 2 = Female. To verify this, click the help icon , click the "Index" tab and type "sample" into the keyword box. Now click on "Sample data sets" and on the "Display" button at the bottom of the page. Click on the "B" button at the top of the right portion of the page, then click on "BEARS". We see from the description of the file that our guess about the variable Sex was correct. We also note that the variable Month is the month that the measurements were made. Close the help window.

Would we expect the Month that the measurements were made to effect the length of the bears? Of course not! Find the mean length of the bears separated by month. The results are in the figure to the right. There is certainly variation in the sample means, but not in any clear pattern.

Would we expect the mean length to be affected by the age of the bears? We would guess yes, for young bears, but once they become adults, the age should not matter. Find the length of bears by age. Do the results support our guess?

Descriptive Statistics: Length

Variable	Month	Total Count	Mean	Median
Length	4	13	62.23	62.00
	5	9	55.78	57.00
	6	7	57.59	58.50
	7	16	61.53	63.00
	8	34	62.04	61.25
	9	31	61.16	61.00
	10	19	63.24	64.00
	11	14	61.27	62.50

MINITAB ASSIGNMENT 5

See instructions on page 8.

1. Retrieve the worksheet EISS6.MTW that you created and saved in Minitab Assignment 1.
 - (a) Display the count, mean and median for e-mails received.
 - (b) Display the count, mean and median for the five samples.
 - (c) Delete the samples of size 6 and replace them with samples of size 36. Be sure you sample with replacement. Save this worksheet as EISS36.MTW. You will need it again in Lesson 10.
 - (d) Display the count, mean and median for these new samples of size 36.
 - (e) Is there a significant difference in the accuracy of the sample means and medians based on samples of size 6 and samples of size 36? Justify your answer. Type your answer to this part in the session window.
2. Retrieve the worksheet K:\MINITAB\Data\Bears.MTW Find the count, mean and median of the weight of all the bears together, and of the bears by sex. Do the results seem reasonable based on what we know about the sex of the bears? Explain. Type your answer in the session window.

LESSON 6 - MEASURES OF DISPERSION

The same Minitab tools we used in Lesson 5 to find measures of central tendency, can be used to find measures of dispersion. Retrieve the worksheet K:\MINITAB\Studnt14\Depth.MTW. Clear everything below the date/time stamp then type your name, Lesson 6 and Examples on separate lines below the date/time stamp.

Below is a description provided by Minitab for this file:

Thirty-six people participated in a study of depth perception under different lighting conditions. The people were divided into three groups on the basis of age. Each of the 12 members of an age group was randomly assigned to one of three 'treatment' groups. All 36 people were asked to judge how far they were from a number of different objects. An average 'error' in judgment, in feet, was recorded for each person. One treatment group was shown the objects in bright sunshine, another under cloudy conditions, and the third, at twilight.

Column	Name	Count	Description
C1-T	AgeGroup	36	Age group; Young, MiddleAge, or Older
C2-T	Light	36	Lighting condition; Sun, Clouds, or Twilight
C3	Error	36	Average error in feet

Click on Stat > Basic Statistics > Display Descriptive Statistics. Select C3 Error into the "Variables:" box. Click on "Statistics" then click on the various check boxes so that "Mean", "Standard deviation", "Median", "Interquartile range", and "N total" are checked. Then click "OK" and "OK".

Now click Stat > Basic Statistics > Display Descriptive Statistics again but this time select C1 AgeGroup into the "By variables (optional):" box and click "OK". Now repeat the procedure with C2 Light in the "By variables (optional):" box.

The results for all of these procedures are shown at the top of the next page. What do these results seem to tell us about the effect of age and light on our ability to judge distances?

The mean and the median are telling us that increased age and decreased light both seem to make us less accurate in judging distance. The standard deviation and the interquartile range appear to be telling us that how much an individual is affected by age or light tend to be less consistent as the age goes up and the light goes down. These observations are only guesses on our part at this point. We must wait for future lessons to determine if the differences we have observed are significant or not.

2/3/2005 4:36:34 PM

Merv Newton
Lesson 6
Examples

Results for: Depth.MTW

Descriptive Statistics: Error

		Total				
Variable	Count	Mean	StDev	Median	IQR	
Error	36	7.708	2.230	7.100	3.075	

Descriptive Statistics: Error

		Total				
Variable	AgeGroup	Count	Mean	StDev	Median	IQR
Error	Young	12	6.992	1.759	6.750	2.625
	MiddleAge	12	7.158	1.933	7.050	3.250
	Older	12	8.975	2.520	8.550	5.025

Descriptive Statistics: Error

		Total				
Variable	Light	Count	Mean	StDev	Median	IQR
Error	Sun	12	5.717	0.896	5.900	1.225
	Clouds	12	7.358	1.144	7.100	1.675
	Twilight	12	10.050	1.808	9.950	3.225

Now click Stat > Basic Statistics > Display Descriptive Statistics again and click on "Statistics". There are several other choices here that we have discussed. There are others that will be discussed later in the course. There are still others that are beyond the scope of this course. You can get more information about all of these choices by clicking the "Help" button on this dialog box. Please take a few minutes to look over some of these.

MINITAB ASSIGNMENT 6

See instructions on page 8.

1. The following data gives the GPA for 55 randomly selected students from Podunk University. The classes 1, 2, 3, 4 are for freshman, sophomore, junior, senior respectively. Enter this data with the class in C1 and the GPA in C2. Display the data. **Be sure to save this worksheet as you will need it again for lesson 7.** Find the mean, standard deviation, median, interquartile range, and count for GPA, first for all the students together, then by class.

(a) What does this information seem to be telling you about GPA as the classes advance? What is a likely explanation for this observation?

(b) Look particularly at the standard deviation and interquartile range for sophomores. What is strange about these? **NOTE:** We will have a better understanding for part of this after the next Minitab lesson.

Type your responses to questions (a) and (b) in the session window.

Class	1	4	1	4	2	2	3	3	2	1	4
GPA	2.91	3.18	2.12	3.10	2.78	2.99	3.95	2.14	3.45	2.12	2.35

Class	3	1	2	4	1	1	1	3	3	4	3
GPA	2.23	1.24	2.47	3.41	2.32	2.90	2.26	3.86	2.85	3.79	3.67

Class	4	2	2	4	3	3	2	1	1	2	4
GPA	4.00	2.55	2.52	3.98	2.34	3.87	3.89	1.93	3.72	3.93	2.64

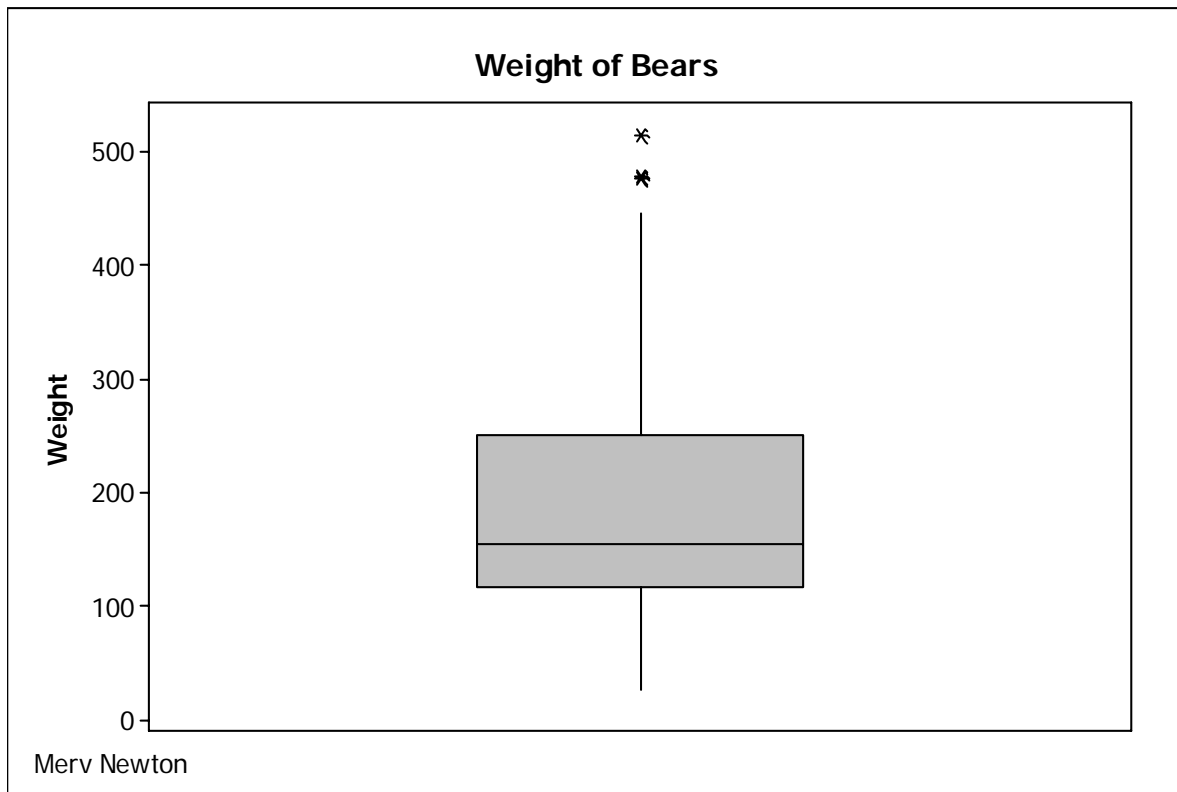
Class	1	4	2	1	4	3	3	4	3	1	3
GPA	3.88	1.98	0.45	3.90	2.94	1.99	2.91	3.99	3.55	2.93	2.76

Class	3	3	2	2	1	2	1	2	4	3	4
GPA	3.92	3.25	2.68	1.80	2.38	3.97	3.92	3.40	3.06	2.73	3.33

LESSON 7 - BOXPLOTS

Before starting this section, be sure that your toolbars are set up as indicated at the beginning of Lesson 3 (see page 12).

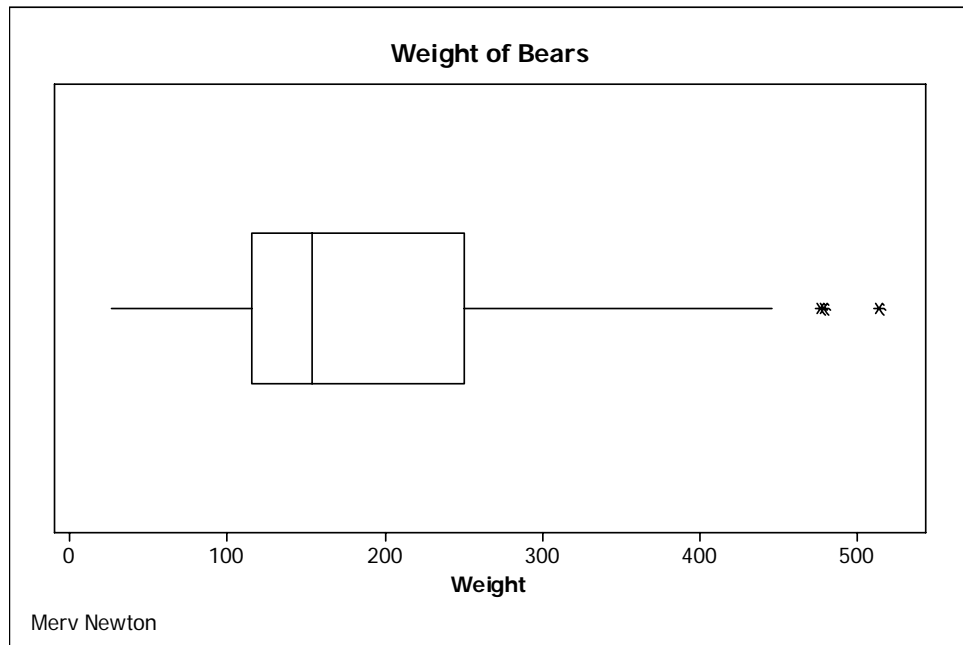
Retrieve the worksheet K:\MINITAB\Data\Bears.MTW. Delete everything below the date\time stamp then type your name, Lesson 7 and Examples on separate lines. Now let's make a boxplot of the bears' weight. Click on Graph > Boxplot, click "OK" on the first dialog box. In the second dialog box select C10 Weight into the "Graph variables:", use "Labels" to enter a title (Weight of Bears) and your name as a footnote, then click "OK". After making the tan area white, you should see a graph like the figure below.



Notice the three data points at the top of the box. Those are values that are so much larger than the rest that Minitab considers them suspected outliers. Boxplots with suspected outliers shown in this manner are what our text calls *modified boxplots*.

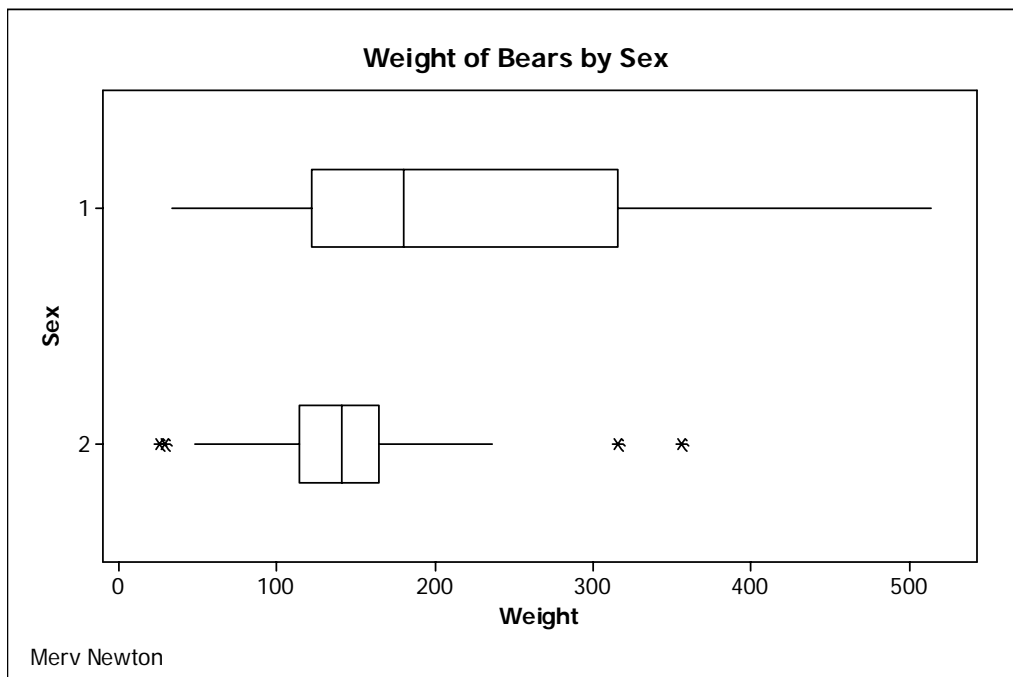
To make the box white, edit the "Interquartile Range Box" in the same way that we edited the bars in the histograms in Lesson 4.

Notice that this boxplot runs vertically instead of horizontally as we have drawn them in class. To change this select either "X Scale" or "Y Scale" as the item to edit, then click on the "Edit" tool. Now check the box in front of "Transpose value and category scales" then click "OK". Your graph should now look like the image at the top of the next page.



Although boxplots are sometimes drawn vertically in the literature, you should draw all your boxplots horizontally for this course.

To see the boxplots of the weight by sex, click on **Graph > Boxplot**, select "With groups" in the dialog box that opens, then click "OK". Select C10 Weight into the "Graph Variables:" box and C4 Sex into the "Categorical variables" box. Use "Labels" to enter a title and your name. Now click on "Scale" and check the box for "Transpose value and category scales". Click "OK" and "OK". Your graph should look like the figure below after you remove the color.



Notice that the first graph shows that when all the bears are together, the distribution of weight is significantly skewed to the right. After separation, however, only the males are skewed to the right. The distribution of the weight for the females is almost perfectly symmetric except for a few outliers.

Now let us find the mean, standard deviation, median, and interquartile range for the weight of the bears, first all together, then by sex. The results are shown below.

Results for: Bears.MTW

Descriptive Statistics: Weight

Variable	Total Count	Mean	StDev	Median	IQR
Weight	143	192.16	110.54	154.00	134.00

Descriptive Statistics: Weight

Variable	Sex	Total Count	Mean	StDev	Median	IQR
Weight	1	99	214.0	119.7	180.0	194.0
	2	44	143.05	64.48	141.00	50.50

Notice that for the bears all together and for the male bears, the standard deviation is smaller than the interquartile range, but for the female bears, the interquartile range is larger than the standard deviation. A look at the boxplots tells us why. Remember that the interquartile range measures the range of the middle 50% of the data, it is not influenced by outliers. The standard deviation, on the other hand, takes all of the data into consideration. A few outliers that are very far from the mean compared to most of the data, will result in the standard deviation as a measure of dispersion indicating greater dispersion among the data than is warranted. The bears all together have only three outliers out of 143 data points, and those are not that far out. The male bears have no outliers. The females, however, have four outliers out of only 44 data points. Almost 10% of the data are outliers, and the two on the right are very far out compared the rest of the data. This has greatly exaggerated the standard deviation. Thus, in the case of the females, the interquartile range is a more appropriate measure of dispersion.

MINITAB ASSIGNMENT 7

See instructions on page 8.

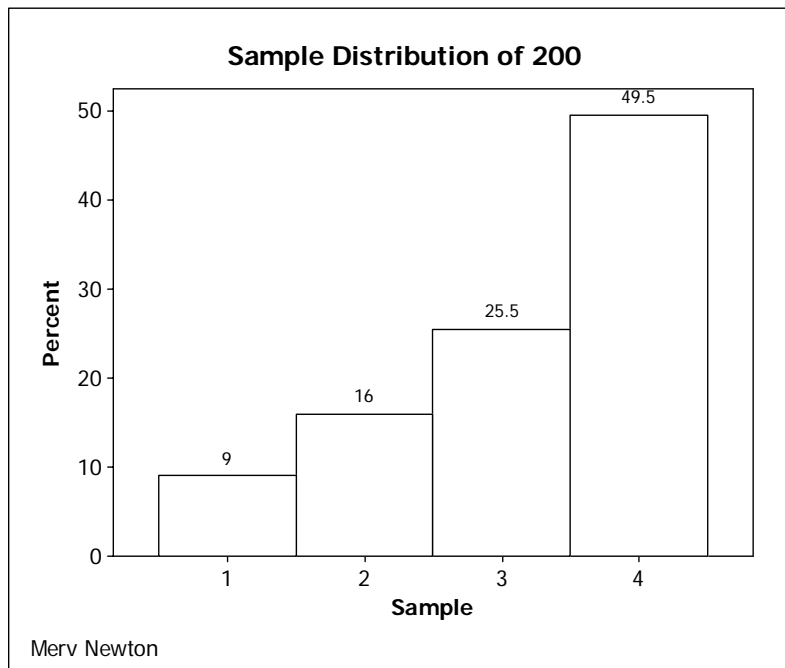
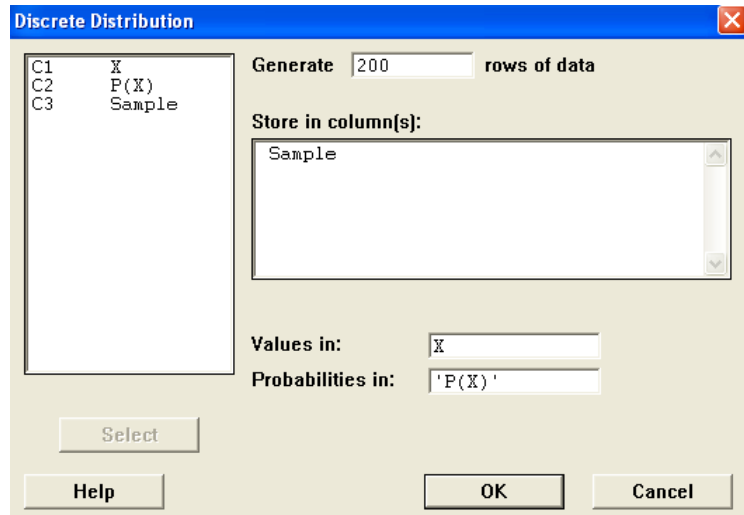
1. Using the data from Minitab Assignment 6, construct a boxplot for the GPA of all students together and boxplots for GPA by Class. How do you explain the strange behavior observed in Minitab Assignment 6 for the standard deviation and interquartile range for sophomores? Type the answer to this question in the session window.

LESSON 8 - DISCRETE DISTRIBUTIONS

In this lesson we will explore some properties of discrete distributions. As an example we will look at a random variable with four outcomes, $X = 1, 2, 3, 4$, and related probabilities $P(1) = .10$, $P(2) = .15$, $P(3) = .25$, and $P(4) = .50$. Open Minitab, clear everything below the date/time stamp and type your name, Lesson 8, and Example on separate lines below the date/time stamp. Go to the data window, label C1 X and C2 P(X), then enter the above values into the columns. Display the two columns. Save this worksheet as ProbDist.mtw as you will need it again in Lesson 10. Now label C3 Sample.

To create a sample of 200 from this distribution click on Calc >

Random Data > Discrete. Type 200 into the "Generate ___ rows of data" box. Select C3 Sample into the "Store in column(s):" box, select C1 X into the "Values in:" box, and select C2 P(X) into the "Probabilities in:" box. The dialog box should now look like the figure to the right. Click "OK". You should now have a sample of 200 in column C3. Now create a relative frequency histogram of Sample and add data labels to the tops of the bars. (See Lesson 4) You should now see a graph similar to the one shown below.



The title has been modified slightly and the percentages on your graph will probably not exactly match (because you will get a different random sample than what was obtained by the author) but they should be close. Notice also that the percentages are close, but not exactly what we expect. For example, we would expect 10% for $X = 1$, but have only 9%, and so on.

MEAN AND VARIANCE

Now let us consider the mean and variance of this distribution. Using the formula from the text and a calculator we find

$$\mu = \sum X * P(X) = 1*.10 + 2*.15 + 3*.25 + 4*.50 = 3.15$$

and

$$\sigma^2 = \sum X^2 * P(X) - \mu^2 = 1^2*.1 + 2^2*.15 + 3^2*.25 + 4^2*.5 - 3.15^2 = 1.0275.$$

Now use the basic statistics function of Minitab (see Lesson 5 and Lesson 6) to find the mean and variance of this sample. The complete session window for this example is shown in the figure to the right. Notice that the sample mean and sample variance differ a little from the mathematically expected values we calculated above, but not by much. Again, we should not expect the empirical results of any one random sample to match exactly the theoretical results.

————— 4/11/2005 3:42:13 PM —————		
Merv Newton		
Lesson 8		
Example		
Data Display		
Row	X	P(X)
1	1	0.10
2	2	0.15
3	3	0.25
4	4	0.50
Descriptive Statistics: Sample		
Variable	Mean	Variance
Sample	3.1550	0.9960

MINITAB ASSIGNMENT 8

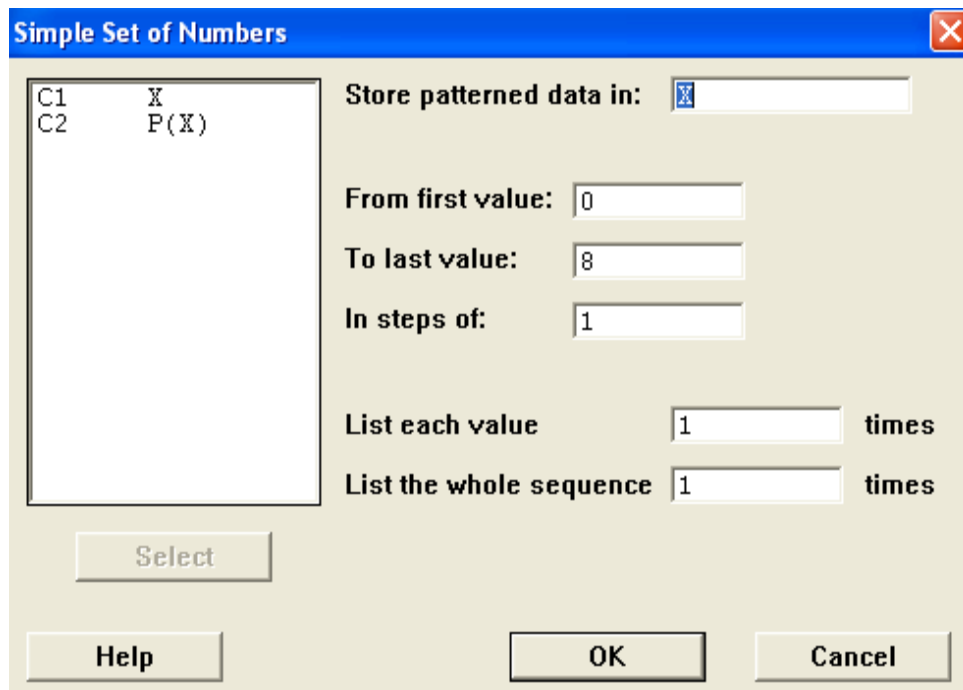
See instructions on page 8.

1. Consider Problem 34 on page 209. Do each of the following. Note that all of your answers will appear in the session window except for parts (b) and (c). The answers to parts (d), (e) and (g) will have to be typed in.
 - (a) Enter the given probability distribution into the data window and display it as was done in the example above. Save the worksheet as Walks.mtw as you will need it again for Lesson 10.
 - (b) Create a sample of size 250 from this distribution.
 - (c) Create a relative frequency histogram using the sample data from part (b). Include an appropriate title.
 - (d) How do the sample relative frequencies compare with the probabilities?
 - (e) Compute the mean and variance of this distribution (NOT of the data)..
 - (f) Use Minitab to compute the mean and variance of the sample data from part (b).
 - (g) How do your results of parts (e) and (f) compare?

LESSON 9 - BINOMIAL DISTRIBUTION

CREATING A PROBABILITY TABLE FOR $X \sim B(n, p)$

Our first task is to build a binomial table. As an example, we will build a probability table for $X \sim B(8, .3)$. First maximize the data window then label C1 as X and C2 as P(X). To enter 0, 1, ... n into C1 click on Calc > Make Patterned Data > Simple Set of Numbers. Select X into the "Store patterned data in:" box, type 0 into the "From first value:" box, and type the value of n ($n = 8$ in this case) into the "To last value:" box. Your dialog box should look like the figure below. Click "OK".



The dialog box titled "Simple Set of Numbers" has a blue title bar with a close button. On the left, a list box contains "C1 X" and "C2 P(X)". To the right, there are several input fields: "Store patterned data in:" with a dropdown menu showing "X", "From first value:" with a text box containing "0", "To last value:" with a text box containing "8", "In steps of:" with a text box containing "1", "List each value" with a text box containing "1" and the word "times" to its right, and "List the whole sequence" with a text box containing "1" and the word "times" to its right. At the bottom, there are three buttons: "Select", "Help", "OK", and "Cancel".

Next, to compute the probabilities, click on Calc > Probability Distribution > Binomial. First click the "Probability" button. Next type the value of n ($n = 8$ in this case) in the "Number of trials:" box and the value of p ($p = .3$ in this case) in the "Probability of success:" box. Now select X for the "Input Column:" and P(X) for the "Optional storage:" box. Your dialog box should now look like the figure at the top of the next page. Click "OK".

Now activate the session window and display the data for X and P(X). Your display should look like the figure on the right. Compare this table with the appropriate section of the binomial distribution table on page A-6 of the text.

Data Display

Row	X	P(X)
1	0	0.057648
2	1	0.197650
3	2	0.296475
4	3	0.254122
5	4	0.136137
6	5	0.046675
7	6	0.010002
8	7	0.001225
9	8	0.000066

Binomial Distribution

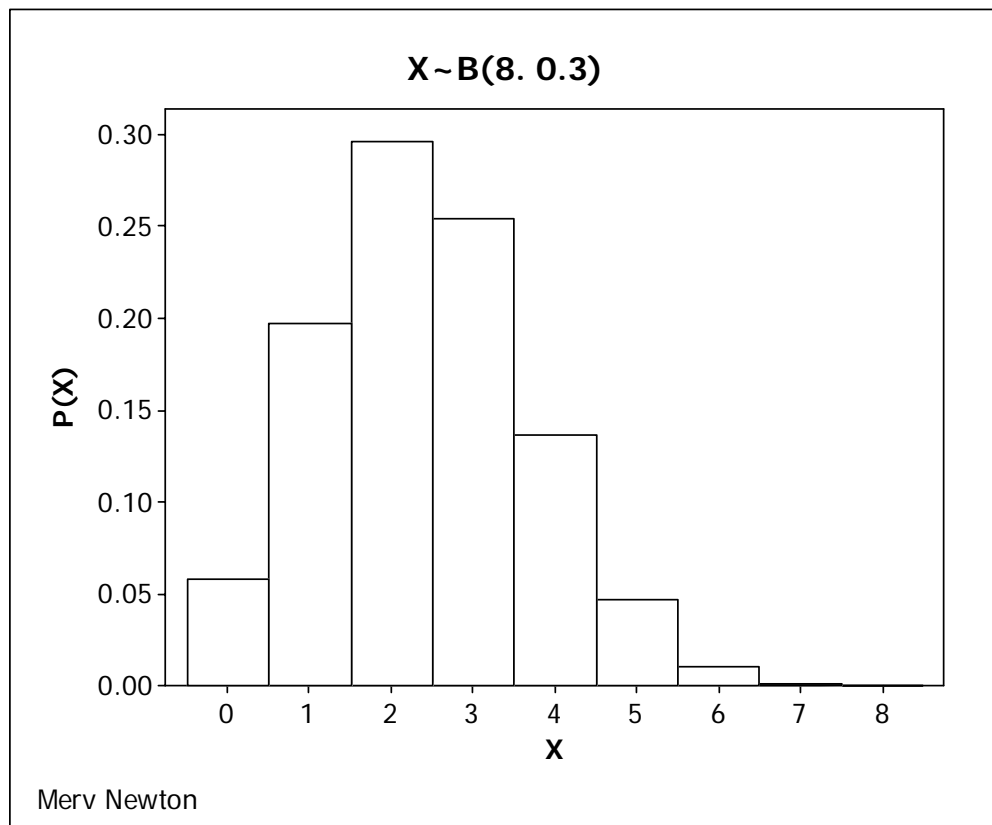
C1	X
C2	P(X)

☒ Probability
☐ Cumulative probability
☐ Inverse cumulative probability

Number of trials:
 Probability of success:

☒ Input column:
 Optional storage:

☐ Input constant:
 Optional storage:



We can also make a bar graph (similar to the probability histograms we have seen in class) of this data by using the procedure from Lesson 4 for tabular data (see TABULATED NOMINAL VARIABLE (FREQUENCY TABLE) in Lesson 4). Use X as the categorical variable and $P(X)$ as the Graph variable. Be sure to put a title and your name (as a footnote) on the graph. We will use the shorthand notation for the distribution we are plotting as the title of the graph. In this case it will be $X \sim B(8, 0.3)$. We will make one small change from the procedure we used in Lesson 4. When we edit the X-scale we will set "Gaps between clusters:" to zero. This will make the bar graph look like a histogram, which is what we want. The graph for this particular distribution is shown at the bottom of the previous page.

SAMPLE PROBLEM

Suppose there are 22 students in this class and that about 6% of Thiel's student population is in the Honors Program. We want to investigate the probabilities of having certain numbers of students from the Honors Program in this class. We will assume that this class is a random sample of Thiel's student population (not necessarily a valid assumption, but OK for the purposes of this example). We will define the variable as X = number of students in this class of 22 who are in the Honors Program. Answer the following questions. Give answers to three decimal places.

- a) What is the expected value of X ?
- b) What is the variance of X ?
- c) What is the standard deviation of X ?
- d) What is the probability that there are exactly three Honors students in this class?
- e) What is the probability that there are at least two Honors students in this class?

First close the data window, then clear everything below the date/time stamp on the session window and type your name, Lesson 9, and Example. Skip a line then type the definition of the variable and the shorthand notation for the distribution you will be using. Now use the procedure outlined above to create and display the probability table for $X \sim B(22, 0.06)$ and create a bar chart of the distribution. Next, use the table to answer the questions. Type your answers in the session window. Your output should look like the figures on the next two pages. Notice that in this case, the numbers on the X-axis have been turned 90 degrees to keep them from running together. An alternative would be to edit X scale and reduce the font size.

Merv Newton

Lesson 9

Example

X = number of students in this class of 22 who are in the Honors Program

$X \sim B(22, 0.06)$

Data Display

Row	X	P(X)
1	0	0.256337
2	1	0.359964
3	2	0.241252
4	3	0.102661
5	4	0.031126
6	5	0.007152
7	6	0.001294
8	7	0.000189
9	8	0.000023
10	9	0.000002
11	10	0.000000
12	11	0.000000
13	12	0.000000
14	13	0.000000
15	14	0.000000
16	15	0.000000
17	16	0.000000
18	17	0.000000
19	18	0.000000
20	19	0.000000
21	20	0.000000
22	21	0.000000
23	22	0.000000

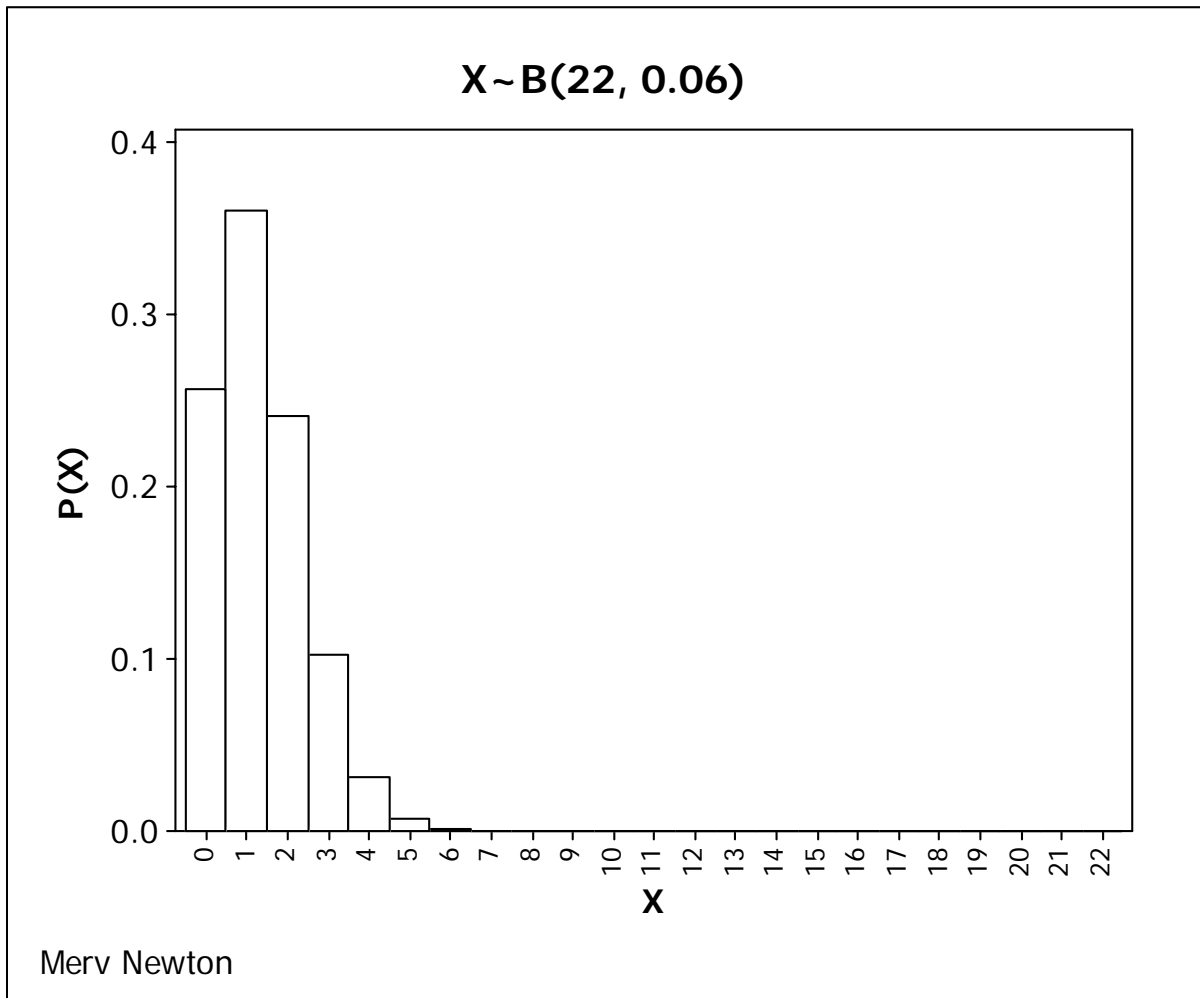
a) Expected value = $np = 22 * 0.06 = 1.320$

b) Variance = $npq = 22 * 0.06 * 0.94 = 1.241$

c) Standard deviation = $\sqrt{\text{variance}} = 1.114$

d) $P(X = 3) = 0.103$

e) $P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - [0.256 + 0.360] = 1 - 0.616 = 0.384$



MINITAB ASSIGNMENT 9

See instructions on page 8.

Do each of the following problems the same way you did the above example, including a bar graph for each distribution. Be sure to start each problem in a new instance of Minitab. Give answers to three decimal places.

1. A certain state instant lottery claims that $1/4$ of the tickets are winners. Suppose you buy 12 tickets and you are interested in counting how many winners you have bought. Describe the appropriate variable X in words and using our shorthand notation.

- a) What is the expected value of X ?
- b) What is the variance of X ?
- c) What is the standard deviation of X ?
- d) What is the probability that you bought exactly 3 winners?
- e) What is the probability that you bought at least 3 winners?

2. The manufacturer of GROW claims that it will grow hair on 85% of bald-headed men. Twenty randomly selected bald-headed men are given GROW and later checked to see how many had grown hair. Describe the appropriate variable X in words and using our shorthand notation.

- a) What is the expected value of X ?
- b) What is the variance of X ?
- c) What is the standard deviation of X ?
- d) What is the probability that exactly 15 will have grown hair?
- e) What is the probability that at least 18 will have grown hair?

LESSON 10 - CENTRAL LIMIT THEOREM

The purpose of this lesson is to illustrate the concepts involved in the Central Limit Theorem. In particular, we illustrate through repeated sampling, how to estimate the mean and standard deviation of a distribution of sample means. We will also see that the distribution of sample means is close to being a normal distribution if the samples are large enough.

Start Minitab and retrieve worksheet DJSS5.MTW that you created in Lesson 1. Delete everything below the date/time stamp and type your name, Lesson 10, and Example as usual.

Use the procedures from Lesson 5 and Lesson 6 to display the count, mean, and standard deviation of the variable DJC in the session window.

Create 10 more samples with replacement of size 5 as you did in Lesson 1. They should be stored in C7 through C16 and label them Sample6 through Sample15. Now click **Stat > Basic Statistics > Store Descriptive Statistics**. Select Sample1 through Sample15 into the "Variables: box. Click on "Statistics" and make sure that "Mean" is the only box that is checked. Now click "OK" and "OK". You will see the means of the 15 samples stored in columns C17 through C31, and Minitab has labeled them Mean1 through Mean15. Now click on **Data > Stack > Columns**. In the dialog box that opens, select Mean1 through Mean15 into the "Stack the following columns:" box, click the "Column of current worksheet:" button, and type C32 into the related input box. When you click "OK", the means will be copied into column C32. Label this column "Means5". Now use the procedure from Lessons 5 and 6 to display the count, mean, and standard deviation of Means5 in the session window.

Now close this worksheet, and retrieve DJSS30 that you created in Lesson 5. Create 10 more samples with replacement of size 30 and repeat the procedures above to compute the mean and standard deviation of the means of these 15 samples. Call the column of means in C32 "Means30" in this case.

After deleting extraneous instructions, your session window should now look like the figure to the right. Your numbers will be different since you will have produced different random samples, but they should be similar.

First, we observe that mean of the sample means from samples of size 30 is quite a bit closer to the population mean than is the mean of the sample means from samples of size 5. The Central Limit Theorem tells us that the standard deviation for the means of samples of size 5 should be the population standard deviation divided by the square root of five. Hence, $899.7 / \sqrt{5} = 402.4$. The experimental value of 423 is not exactly right, but close. For samples of size 30, the standard deviation of the means should be the population standard deviation divided by the square root of 30. Hence $899.7 / \sqrt{30} = 164.3$. The experimental value of 168.1 is pretty close!

Descriptive Statistics: DJC

Variable	Total Count	Mean	StDev
DJC	518	9706.6	899.7

Descriptive Statistics: Means5

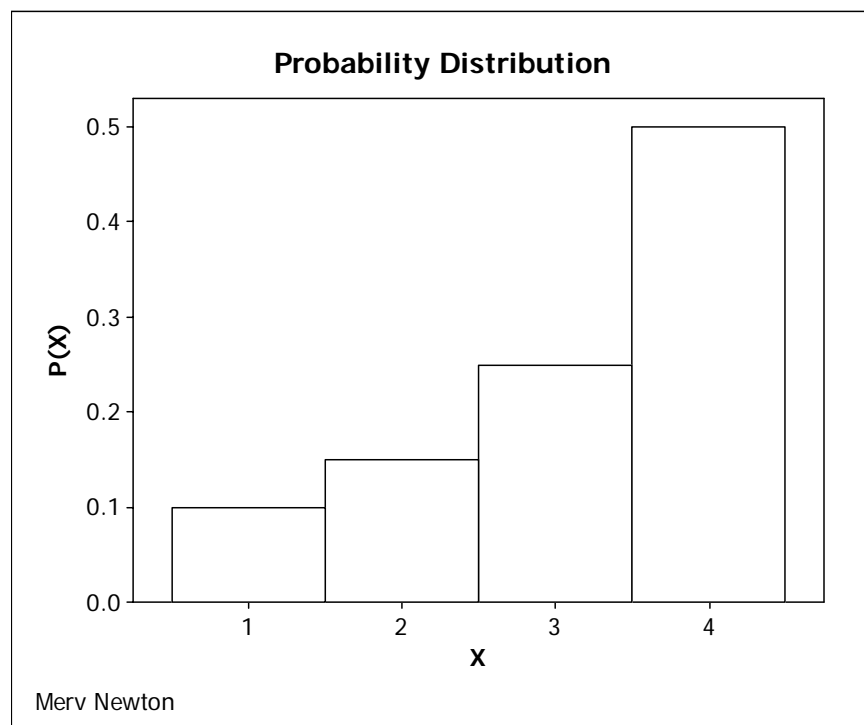
Variable	Total Count	Mean	StDev
Means5	15	9796	423

Descriptive Statistics: Means30

Variable	Total Count	Mean	StDev
Means30	15	9754.9	168.1

NORMALITY OF SAMPLE MEANS

Close your data window from the previous example and open the worksheet ProbDist.mtw that you saved from Lesson 8. First make a bar graph of this discrete distribution. (See pages 39 and 40 of Lesson 9.) The graph is shown below, and it certainly does not look normal.



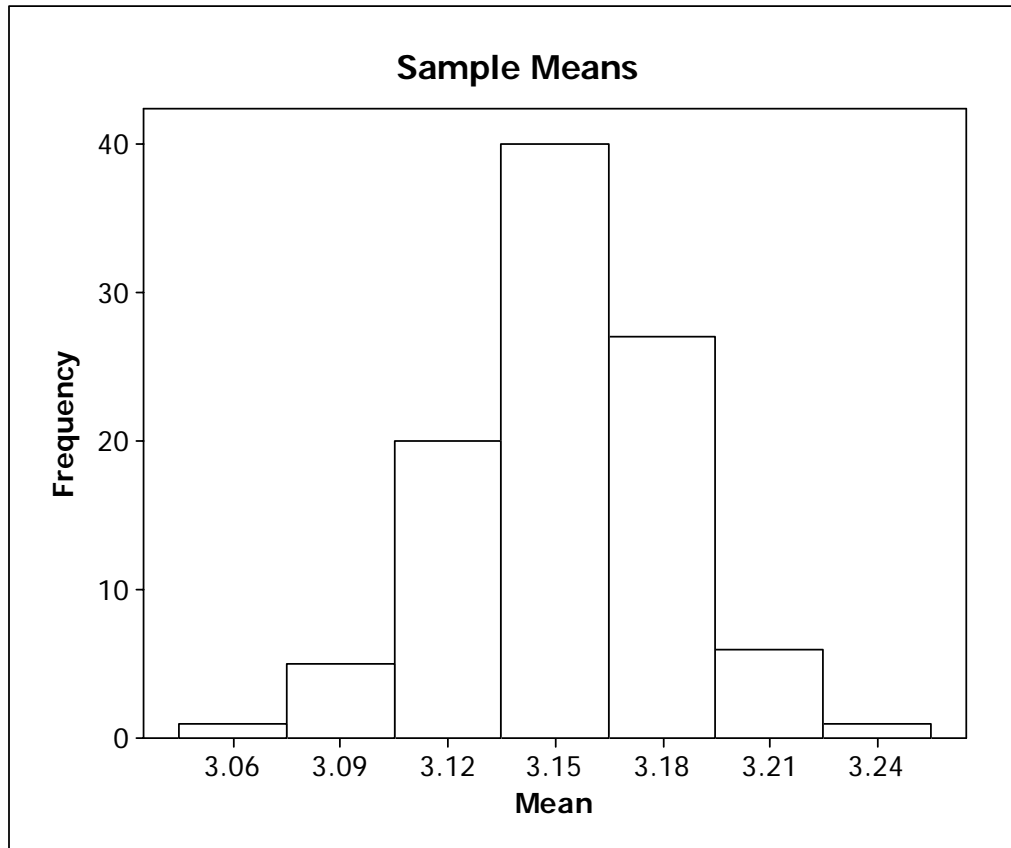
Now click on Calc > Random Data > Discrete. Type 1000 in the "Generate ___ rows of data" box and C3-C102 in the "Store in column(s):" box. Select C1 X into the "Values in:" box and P(X) into the "Probabilities in:" box. Now click "OK". You now have 100 samples of size 1000 in columns C3 through C102. Find the means of these 100 samples, (they will be stored in C103 through C202) and stack them in C203 as we did in the example above. Give C203 the label "Sample Means." Now use our basic statistics function to find the mean, and variance of the sample means. They are shown in the figure below, but your may not match exactly.

Descriptive Statistics: Sample Means		
Variable	Mean	Variance
Sample Means	3.1508	0.000930

In Lesson 8 we computed the mean and variance of this distribution to be 3.15 and 1.0275 respectively, so we should expect the mean and variance for the sample means to be 3.15 and $1.0275/1000 = 0.0010275$ respectively. We can see that the experimental values for the mean

and variance are quite close to what we would expect.

Finally, we will draw a histogram of the sample means. We will use 7 bins but leave the labels in the middle rather than at the cut points.



We can see that from a distribution that was skewed left we have obtained a probability distribution of sample means that is approximately normal by repeated sampling. The Central Limit Theorem tells us that the distribution of the means will get closer to normal the larger the sample size.

MINITAB ASSIGNMENT 10

See instructions on page 8.

1. Use worksheet EISS6.MTW that you created in Minitab Assignment 1 and EISS36.MTW that you created in Minitab Assignment 5 to recreate an experiment like the one in this lesson. Compare the mean of the means for samples of size 6 and size 36 to the population mean. Compare the standard deviation of the means for samples of size 6 and size 36 to the standard deviation that is predicted by the Central Limit Theorem. Type your observations in the session window.
2. Consider Problem 34 on page 209. You should have this distribution saved as Walk.mtw from Lesson 8. Do each of the following. Note that all of your answers, except part (a) and the graphs, will appear in the session window. In some cases you will need to type the answers into the session window.
 - (a) Create 100 samples of size 1000, find the means of the samples, and stack the sample means in a column labeled "Sample Means."
 - (b) Find the mean and variance of Sample Means.
 - (c) How do your answers to part (b) compare to the expected values? (You should have computed the mean and variance for this distribution as part of Minitab Exercise 8.
 - (d) Create a bar graph of this distribution.
 - (e) Create a histogram with 7 classes of Sample Means.
 - (f) Does your graph from part (e) look closer to being normal than the one from part (d)?

LESSON 11 - CONFIDENCE INTERVALS

Suppose we want to find a 95% confidence interval for the mean diameter of the trees in K:\minitab\data\trees.mtw. Load this worksheet then clear the session window below the date/time stamp. Type your name, Lesson 11, Example, and the definition of the variable below the date/time stamp (in this case the definition of the variable is X = the diameter of a tree). We need to find the standard deviation for the diameters before we can compute the confidence interval. You will find this under Stat > Basic Statistics > Display Descriptive Statistics as we did in Lesson 6. We note that $n = 31 > 30$, so we are justified by the Central Limit Theorem to assume that the means of the sample means will be normally distributed.

Now, to find the confidence interval, click on Stat > Basic Statistics > 1-Sample Z. Click on the box under "Samples in columns:" and select Diameter into it. Now type the value of the standard deviation (3.138 in this case) into the "Standard deviation:" box. The 1-Sample Z dialog box should now look like the figure below

1-Sample Z (Test and Confidence Interval)

☒ **Samples in columns:**

Diameter

☐ **Summarized data**

Sample size:

Mean:

Standard deviation:

☐ **Perform hypothesis test**

Hypothesized mean:

Select Graphs... Options...

Help OK Cancel

Now click "OK" and you will get a 95% confidence interval (abbreviated 95.0 % C.I. in Minitab) of (12.1437, 14.3530).

To get a 99% confidence interval get back to the 1-Sample Z dialog box as you did above and click on "Options..." . Type 99 into the "Confidence level:" box, then click "OK" to close the

options dialog box and "OK" again to find the 99% confidence interval. For a 90% confidence interval repeat but type 90 into the "Confidence level:" box. An example of the output, after excessive blank lines and other extraneous material has been removed, is shown below. Notice what happens to the length of the interval as the degree of confidence changes.

7/2/2004 1:12:00 PM

Merv Newton

Lesson 11

Example

X = the diameter of a tree

Descriptive Statistics: Diameter

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Diameter	31	0	13.248	0.564	3.138	8.300	11.000	12.900	16.000

Variable	Maximum
Diameter	20.600

One-Sample Z: Diameter

The assumed standard deviation = 3.138

Variable	N	Mean	StDev	SE Mean	95% CI
Diameter	31	13.2484	3.1381	0.5636	(12.1437, 14.3530)

One-Sample Z: Diameter

The assumed standard deviation = 3.138

Variable	N	Mean	StDev	SE Mean	99% CI
Diameter	31	13.2484	3.1381	0.5636	(11.7966, 14.7001)

One-Sample Z: Diameter

The assumed standard deviation = 3.138

Variable	N	Mean	StDev	SE Mean	90% CI
Diameter	31	13.2484	3.1381	0.5636	(12.3213, 14.1754)

The same solutions should be obtained by clicking the button for "Summarized data" and typing in the sample size (31 in this case) and mean (13.2484 in this case) in the appropriate boxes of the 1-Sample Z dialog box and clicking "OK". The results of this are shown below for a 95% confidence interval.

One-Sample Z				
The assumed standard deviation = 3.138				
N	Mean	SE Mean	95% CI	
31	13.2484	0.5636	(12.1438, 14.3530)	

Compare this to the 95% confidence interval above that was computed using the raw data. We see a minor discrepancy in the lower limit. This was no doubt the result of using a value for the mean that was rounded to 6 significant figures when we typed it in, while the computer used 8 significant figures when the computations were done from the raw data. The moral of the story is "Let Minitab use raw data whenever possible."

In some cases, however, we have no choice. Consider, for example, Problem 2 on page 284 of our text. After getting into the 1-Sample Z dialog box we must click on "Summarized data", enter 50 for "Sample size:", 23.4 for "Mean:", 7.1 for "Standard deviation:" and click "OK". The result is shown below.

7/2/2004 2:49:24 PM				
Merv Newton				
Lesson 11				
Example				
X = time a patient had to wait to see a physician (minutes)				
One-Sample Z				
The assumed standard deviation = 7.1				
N	Mean	SE Mean	95% CI	
50	23.4000	1.0041	(21.4320, 25.3680)	

MINITAB ASSIGNMENT 11

See instructions on page 8.

For each problem be sure to define the appropriate random variable as in the sample problem.

1. Create 90%, 95%, and 99% confidence intervals for the mean length of bears from the worksheet K:\minitab\data\bears.
2. Enter the data from Problem 23 on page 286 into the data window. Display the data and create 90%, 95%, and 99% confidence intervals for the mean. Be sure to save this data on your disk or J: drive as you will need it again for Minitab Assignment 12.
3. Do Problem 5 on page 284.

LESSON 12 - HYPOTHESIS TESTING; LARGE SAMPLE

Suppose the data in the worksheet K:\minitab\data\bears.mtw is from the bears in a particular state. The mean length of the bears used to be 63 inches, but because of hunting of the larger, more prized bears, the state game warden claims that the mean length is now less than 63 inches. Using the data in the above-named worksheet, test the warden's claim by the P-Value method.

Open the required file then clear the session window below the date/time stamp. Type in your name, Lesson 12, Example, and the definition of the variable (X = the length of a bear). We need to compute the standard deviation before we can do our hypothesis testing, so display the descriptive statistics for the variable Length as we did in Lesson 6. Notice that $n = 143 > 30$, so we are justified in using the normal distribution by the Central Limit Theorem.

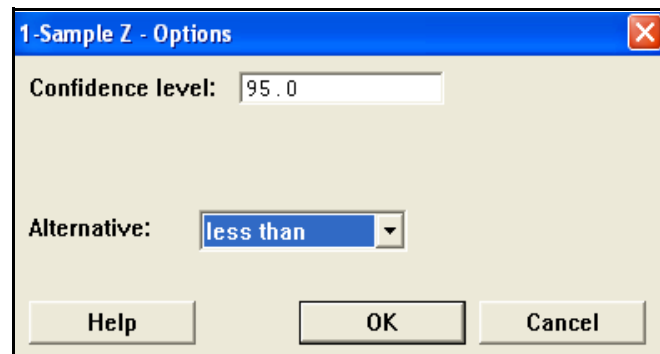
Our hypotheses in this case would be

$$H_0: \mu \geq 63$$

$$H_a: \mu < 63$$

but we will not type them since Minitab will print something equivalent to this in the Session Window when we run the tests.

At this point click on Stat > Basic Statistics > 1-Sample Z (recall that we saw this dialog box in Lesson 11). Select Length into the "Samples in columns:" box, type the standard deviation into the "Standard deviation:" box (9.352 in this case), check the "Perform hypothesis test" box and type the value of μ from the null



hypothesis into the "Hypothesized mean:" box (63 in this case). Now click on "Options" and make the appropriate choice from the "Alternative:" menu. In this case we choose "less than" because the alternative hypothesis is $<$. See image above. The value in the "Confidence level:" box is not relevant since we will be using the P-Value approach, so the reader will determine α . Click "OK" to close the options dialog box then click "OK" to perform the test.

Notice that Minitab gives us both the value of Z and the P-Value. Since we are using the P-Value approach, there is nothing more for the researcher to do. If we now take the roll of the reader, what decision would we make if we choose α to be:

- a) 10%?
- b) 5%?
- c) 1%?

Type your responses in the Session Window. The results are shown below.

7/2/2004 2:55:14 PM

Merv Newton
Lesson 12
Example

X = length of a bear

Descriptive Statistics: Length

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Length	143	0	61.283	0.782	9.352	36.000	57.000	61.000	67.500

Variable	Maximum
Length	83.000

One-Sample Z: Length

Test of mu = 63 vs < 63
The assumed standard deviation = 9.352

Variable	N	Mean	StDev	SE Mean	95% Upper Bound	Z	P
Length	143	61.2825	9.3525	0.7821	62.5689	-2.20	0.014

- a) Since $P = 0.014 < 0.10$, we would decide to reject H_0 at the 10% level of significance.
- b) Since $P = 0.014 < 0.05$, we would reject H_0 at the 5% level of significance.

As was the case with confidence intervals in Lesson 11, we may find it necessary to use "Summarized data". Consider, for example, Problem 52 (a) on page 305. Start a new instance of Minitab and add your name, etc. as usual. Define the variable X = pressure reading on a gauge (psi). Click Stat > Basic Statistics > 1-Sample Z, then click on "Summarized data". Enter 50 for "Sample size:", 30.1 for "Mean:", 0.35 for "Standard deviation:", and 30 for "Test mean:". Finally, "Alternative:" must be set to "not equal" in the "Options" dialog box. The results are shown on the right.

7/22/2004 3:18:37 PM

Merv Newton
Lesson 12
Example

X = pressure reading on a gauge (psi)

One-Sample Z

Test of mu = 30 vs not = 30
The assumed standard deviation = 0.35

N	Mean	SE Mean	95% CI	Z	P
50	30.1000	0.0495	(30.0030, 30.1970)	2.02	0.043

MINITAB ASSIGNMENT 12

See instructions on page 8.

For each of the problems type in the answers to the following questions on your output.

What decision should you make if you chose alpha to be:

(a) 0.10?

(b) 0.05?

(c) 0.01?

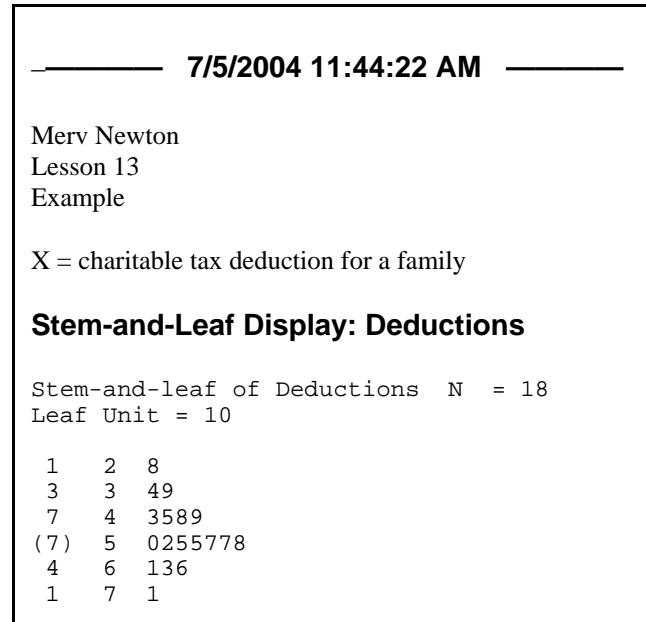
1. Using the worksheet K:\minitab\data\trees.mtw test the hypothesis that the mean diameter of the trees is not 14 inches.
2. Using the data from Problem 23 on page 286 test the hypothesis that the mean repair time is greater than 40 minutes. (Note: You should have this worksheet on you disk or J: drive from Minitab Assignment 11.)
3. Enter and display the data from Problem 37 on page 301. Test the hypothesis that the mean time for recovery is less than 6.5 days.
4. Do Problem 50(a) on page 305. Be sure to answer the questions (a), (b) and (c) above.

LESSON 13 - HYPOTHESIS TESTING; SMALL SAMPLE

We will use Problem 133 on page 345 as an example. First enter the data and name the variable (let's call it Deductions). Maximize the session Window, clear it below the date/time stamp then type your name, Lesson 13, Example, and define the variable (X = charitable tax deduction for a family).

Since the data was given to us as a stem-and-leaf graph, we can see that it is reasonably normal and that the t-test would be valid, but let's suppose the data had been given to us as a list of numbers the way we now see it in our Data Window. One way to check for normality is to create a stem-and-leaf display of the data. Click on Graph > Stem-and- Leaf and select the variable into the "Graph variable:" box. We need to tell

Minitab how far apart we want the stems to be. In this case we want the stem unit to be 100 because a data value such as 3|4 means 340, so the 3 actually is in hundreds. Type 100 in the box marked "Increment:" then click "OK". After deleting excess blank lines and other unnecessary output from the session Window, your display should look like the figure above. Notice that unlike other graphs, the stem-and-leaf graph is printed in the session Window, not in a graph window. Also notice that the key is indicated by the statement "Leaf Unit = 10," which gives us the same information as our method of indicating the key, which in this case would be 3|4 = 340.



ANDERSON-DARLING NORMALITY TEST

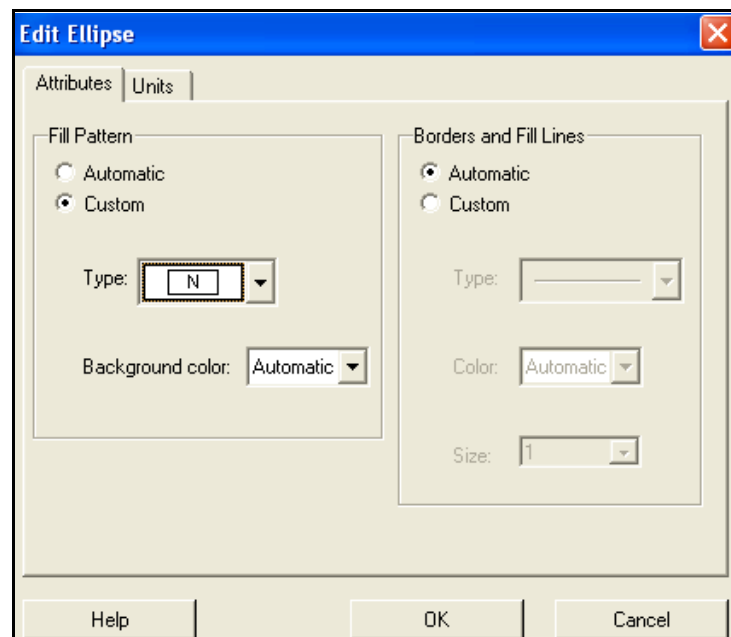
A better way to test for normality is the Anderson-Darling Normality Test. This is a test that has "The data is from a normal distribution" as its null hypothesis. We use the same level of significance on this test as we do on the t-test we would like to run on the data in question. If we fail to reject the hypothesis of normality, we go ahead with the t-test; if the P-Value of the Anderson-Darling test is smaller than our choice of alpha, we reject normality and the t-test would be invalid. To get to the Anderson-Darling test click Stat > Basic Statistics > Graphical Summary. Select C1 Deductions into the "Variables:" box and click "OK". There are several things we would like to fix. We will want everything black and white as in our previous graphs. To edit the bars in the histogram, you will need to right click in them, then click "Edit Bars." Do the same for the other areas that need to be fixed. Be sure to make the blue lines black. We will now need two new tools from the graphics tool bars, the ADD MENU and the ELLIPSE TOOL. These are shown in the figure at the top of the next page.



To put your name on the graph select either "Graph Region" or "Figure Region" from the "Item to edit menu" then select "Footnote" from the "ADD MENU", type your name in the box provided, then click OK. Your name will appear in the lower left corner of the graph. Notice that the P-Value for the Anderson-Darling test is .918 in this case. That is certainly larger than .05, (or any other alpha we would ever pick), so we will not reject normality and it is OK to use the t-test. To indicate that, choose "Subtitle" from the "ADD MENU" and type "t-test OK" in the box that is provided. The subtitle will appear right below the title, but use the mouse to drag it to the right and leave it right above the box with the Anderson-Darling test.

NOTE: If the P-Value had been less than our chosen value of alpha, we would have typed "t-test not OK" as the subtitle. We would then have had to abort the procedure since the Student's t-test may only be used if the sample data follows an approximately normal distribution.

Now click on the "ELLIPSE" tool. The cursor will become a cross hair. Place it a little above and to the left of the P-Value for the Anderson-Darling test and drag it down and to the right to create an ellipse with the P-Value inside. When you release the mouse button, the ellipse will turn gray and cover the value. To fix this, click on the "Edit" tool and change the "Type:" under "Fill Pattern" to none. The dialog box should look like the figure to the right. Click "OK" and the gray will disappear from the ellipse and the P-Value will show through.



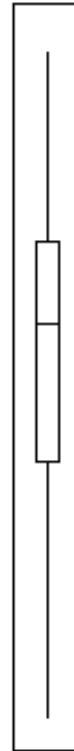
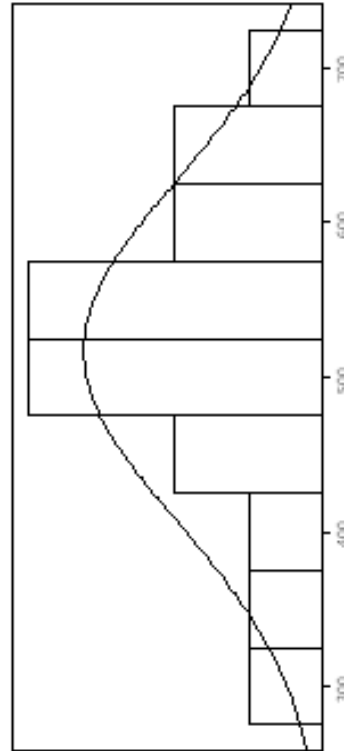
Your completed graphical summary should now look like the figure on the next page. This is the procedure we will use any time we need to verify normality of data. We will find that we will need it again in several future lessons.

If this were an assignment, we would now print the graph to turn in with the rest of the problem, but for this exercise you may now close it.

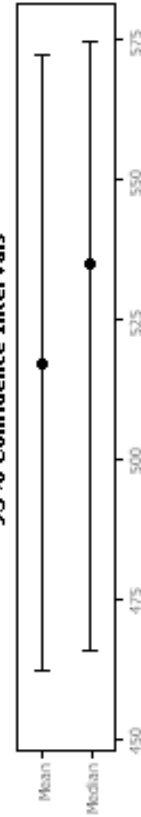
Summary for Deductions

t-test OK

Anderson-Darling Normality Test	
A-Squared	0.17
p-Value	0.918
Mean	517.22
StDev	110.87
Variance	12291.83
Skewness	-0.441861
Kurtosis	0.028052
N	18
Minimum	280.00
1st Quartile	445.00
Median	535.00
3rd Quartile	587.50
Maximum	710.00
95% Confidence Interval for Mean	
462.09	572.36
95% Confidence Interval for Median	
465.54	574.82
95% Confidence Interval for StDev	
83.19	166.21



95% Confidence Intervals



Merv Newton

ONE SAMPLE t - TEST

We will now use the t-test and the P-Value method to test the hypothesis that the mean is less than 600. Click on Stat > Basic Statistics > 1-Sample t and you see a dialog box that is almost the same as it is for the 1-Sample Z that we saw in Lesson 12, except that it is not necessary to enter a value for sigma since the t-test automatically uses the sample standard deviation. Select C1 Deductions into the "Samples in columns:" box. As with the z-test, you must check the "Perform hypothesis test" box and type the value of μ from the null hypothesis into the "Hypothesized mean:" box (600 in this case) then click on "Options" and set the "Alternative:" box to "less than" since our alternative hypothesis is $<$ in this case. Now click "OK" on the option dialog box and "OK" on the 1-Sample t dialog box. Using the P-Value approach we would normally stop here, but since we have been told to use $\alpha = .05$ we will type the decision and the conclusion based on a 5% level of significance. The results are shown in the figure below.

One-Sample T: Deductions							
Test of mu = 600 vs < 600							
					95%		
					Upper		
Variable	N	Mean	StDev	SE Mean	Bound	T	P
Deductions	18	517.222	110.869	26.132	562.682	-3.17	0.003
Decision: Since the P-Value = 0.003 < .05, reject H0.							
Conclusion: The mean tax deduction for families earning \$35,000 per year is significantly less than \$600 at the 5% level of significance. There is sufficient evidence to support the economist's claim.							

OUTLIER EXAMPLE

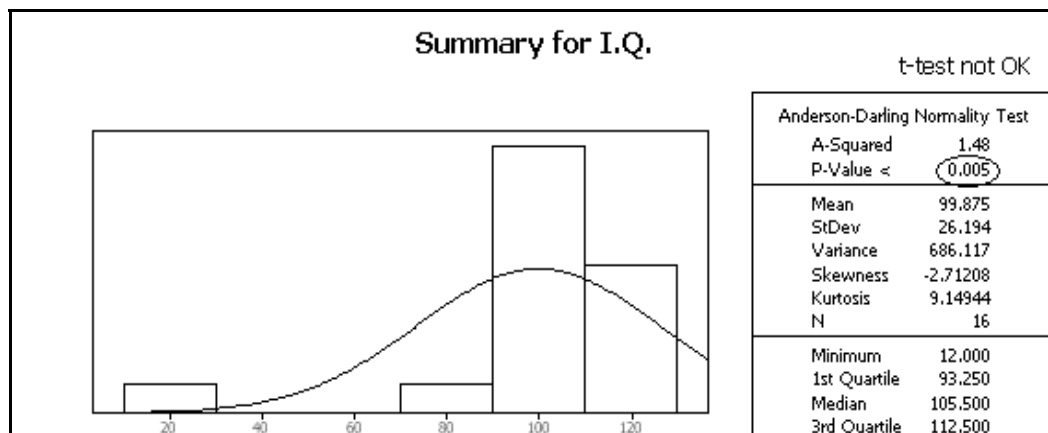
As a second example, let us consider Problem 132 on page 344. Since the problem doesn't say what this data is, let us assume it is I.Q. scores of a random sample of students from Podunk High and we want to test the principal's claim that the mean is 100 using the P-Value method with a 5% level of significance. Start a new instance of Minitab, type your name and so on as usual, and enter the data. Now create a Stem-and-Leaf display. In this case we should type 10 into the "Increment" box since we want the stems to be 10's. It should look like the figure on the right. NOTE: Since the Stem-and-Leaf Display gives all of the information that is in a data display, it is not necessary to do both.

Stem-and-Leaf Display: I.Q.

Stem-and-leaf of I.Q. N = 16
Leaf Unit = 1.0

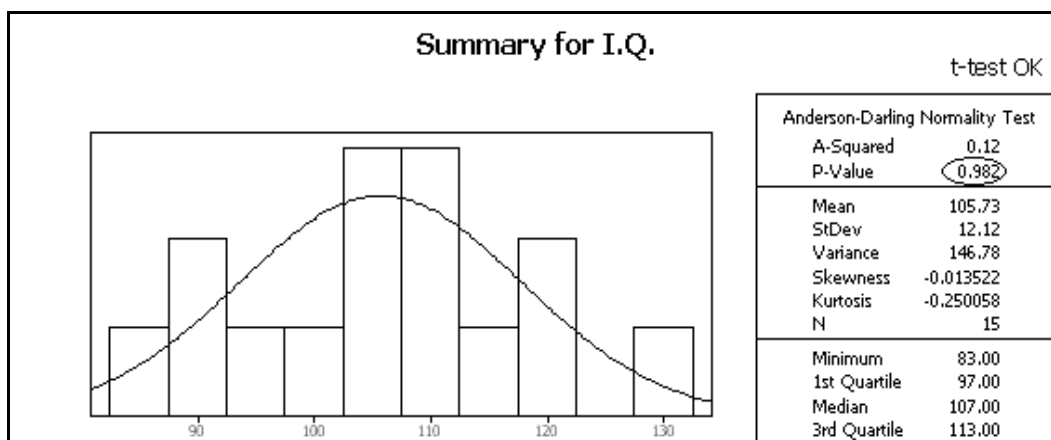
1	1	2
1	2	
1	3	
1	4	
1	5	
1	6	
1	7	
2	8	3
6	9	2278
(5)	10	34799
5	11	138
2	12	28

Except for the outlier at 12, the data looks pretty good. Let us see what the Anderson-Darling test tells us. Create a Graphical Summary for this data as we did above. The upper portion of the summary is shown below.



Notice that the outlier at 12 is also quite obvious here. Also notice that the Anderson-Darling test would reject the assumption of normality because of the small P-value, so the t-test would not be valid.

Since an I.Q. of 12 is very unlikely, we will assume that data point is a mistake and delete it, (to delete the outlier return to the Data Window, click in the cell with the 12, then click Edit > Delate cells) then try the Anderson-Darling test again.



We see that it is now OK to use the t-test so we shall do so. In this case we must make the "Test mean:" 100 then click on "Options" to set the "Alternative:" menu to "not equal". After typing in the decision and conclusion, the results are shown at the top of the next page.

One-Sample T: I.Q.

Test of $\mu = 100$ vs not = 100

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
I.Q.	15	105.733	12.115	3.128	(99.024, 112.443)	1.83	0.088

Decision: Since the P-Value = 0.088 > 0.05, fail to reject H_0 .

Conclusion: The mean I.Q. score of students at Podunk High is not significantly different from 100 at the 0.05 level of significance. There is insufficient evidence to refute the principal's claim.

SUMMARY DATA

As was the case with the z-test, we may have only summary data with which to work. Consider Problem 128 (b) on page 344. Start a new instance of Minitab, clear the text below the date/time stamp and enter your name and so on as usual. Define the variable as X = calories in a can of tomato paste. The problem tells us we may assume the data is normal, so we need not do anything with the Anderson-Darling test. Now click Stat > Basic Statistics > 1-Sample t. Click on the "Summarized data" button then fill in the boxes with the information given in the problem: "Sample size:" is 14, "Mean:" is 151.2 and "Standard deviation:" is 2.4. The test mean is 150 and we must select "greater than" from the "Alternative:" menu in the "Options" box. The results, after typing in the decision and conclusion, are shown below.

7/9/2004 3:50:04 PM

Merv Newton
Lesson 13
Example

X = calories in a can of tomato paste

One-Sample T

Test of $\mu = 150$ vs > 150

N	Mean	StDev	SE Mean	95% Lower Bound	T	P
14	151.200	2.400	0.641	150.064	1.87	0.042

Decision: Since the P-Value = 0.042 < 0.05, reject H_0 .

Conclusion: The mean number of calories in a can of tomato paste is significantly greater than 150 at the 0.05 level of significance. There is sufficient evidence to support the weight watchers group's claim.

MINITAB ASSIGNMENT 13

See instructions on page 8.

1. Use the data from Problem 62 on page 322 as the data from Problem 133 was used above.
Enter the data and create a stem-and-leaf display. Create a graphical summary as was done in the lesson above, then print it. Use the t-test and P-Value to test the partner's claim at the 5% level of significance. Type in your decision and conclusion.
2. Use the data from Problem 80 on page 325 as the data from Problem 132 was used above.
Assume that these data are the ages of a random sample of passengers who passed through Kennedy Airport on a particular day. Enter the data and create a stem-and-leaf display.
 - (a) Create a graphical summary and use the t-test (even though it is not valid in this case) to test the hypothesis that the mean of this data is 50 at the 5% level of significance.
 - (b) Now remove the outlier, create a new graphical summary and repeat the test of the hypothesis that the mean is 50 at the 5% level of significance.
3. Do Problem 129 (b) on page 344 as Problem 128 (b) was done above.

LESSON 14 - HYPOTHESIS TESTING; DEPENDENT SAMPLES

As an example, consider problem 5 on page 355. We will use the P-Value method to test the salesman's claim at the 0.10 level of significance. Enter the data into columns C1 and C2 and label them X1 and X2 respectively. Enter X as the label of C3. Clear the Session Window below the date/time stamp, type your name, Lesson 14, Example, and define the variables as

X1 = a runner's time with company shoes
X2 = a runner's time with school shoes
X = X1 - X2

To compute $X = X1 - X2$ click on Calc > Calculator. Select X into the "Store results in variable:" box and type X1-X2 into the "Expression:" box. When the dialog box looks like the figure on the right, click "OK".

Now display the three columns of data.

Since we have a small sample, we must use the t-test, but we must first make sure that X is approximately normal so that we can be sure the t-test is appropriate.

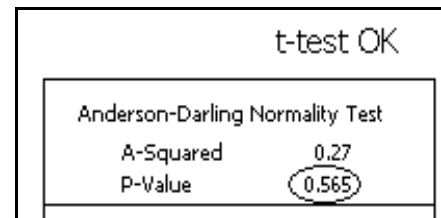
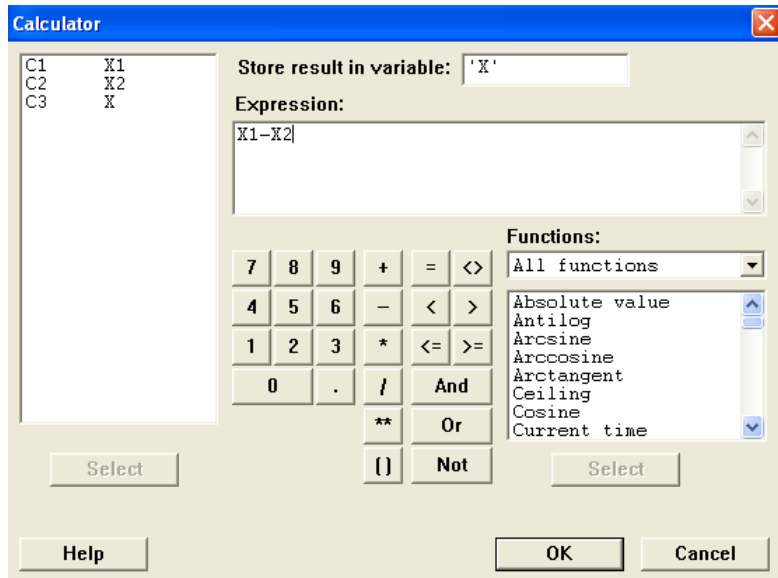
We use the Anderson-Darling Normality Test on the variable X as we did in Lesson 13. The figure on the right shows the relevant part of the graphical summary, and we can see that the t-test is OK. When doing the assignment for this lesson you will, of course, print and submit the entire graphical summary page.

The salesman's claim is that runners will run faster, that is have shorter times, with his company's shoes. What he is claiming then is that $\mu = \mu_1 - \mu_2 < 0$.

Now conduct a t-test on X the same as we did in Lesson 13. The "Hypothesized mean:" in the "1-Sample t" dialog box will be zero and the "Alternative:" in the "Options:" box will be "less than." Finally, type in the decision and the conclusion using the P-Value and $\alpha = 0.1$.

The session window which results from the above procedures is on the next page.

NOTE: Minitab does have a test called "Paired t" located under Stat > Basic Statistics. It simplifies the work done above, but it does NOT check X for normality, so the test may not be valid. That is why the procedure described above should be used unless X is known to be normal.



Merv Newton

Lesson 14

Example

X1 = a runner's time with company shoes (seconds)

X2 = a runner's time with school shoes (seconds)

$X = X1 - X2$

Data Display

Row	X1	X2	X
1	10.8	11.4	-0.6
2	12.3	12.5	-0.2
3	10.7	10.8	-0.1
4	12.0	11.7	0.3
5	10.6	10.9	-0.3
6	11.5	11.8	-0.3
7	12.1	12.2	-0.1
8	11.2	11.7	-0.5

One-Sample T: X

Test of $\mu = 0$ vs < 0

Variable	N	Mean	StDev	SE Mean	95% Upper Bound	T	P
X	8	-0.225000	0.276457	0.097742	-0.039819	-2.30	0.027

Decision: Since the P-Value = 0.027 < 0.1, reject H_0 .

Conclusion: The mean time for runners with company shoes is significantly less than for runners with school shoes at the 0.1 level of significance. There is sufficient evidence to support the salesman's claim.

MINITAB ASSIGNMENT 14

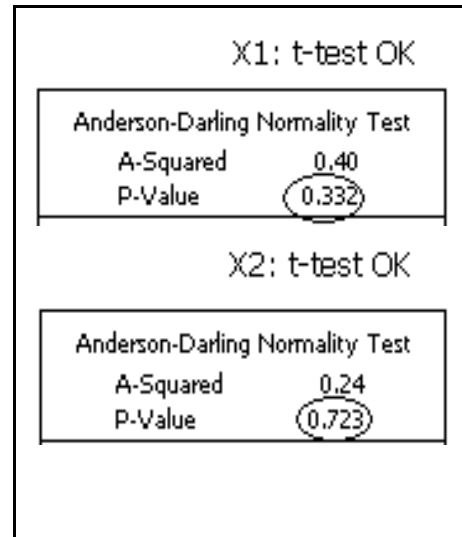
See instructions on page 8.

1. Do problem 49 (b) on page 385 using the procedure outlined above.
2. Do problem 50 on page 385 using the procedure outlined above.

LESSON 15 - HYPOTHESIS TESTING; INDEPENDENT SAMPLES

In this lesson we will see how to perform a hypothesis test for the difference of two means. We will use Problem 25 on page 371 as an example. Open Minitab, clear the area below the date/time stamp, the type in your name, Lesson 15, Example, and define the variables as usual. We will let X_1 = Final exam score of a student in Section A and X_2 = Final exam score of a student in Section B. Enter the data from Section A into C1 and label it X_1 , then enter the data for Section B into C2 and label it X_2 . Display the two data sets.

The first task is to make sure that the t-test will be valid in this case. Since the data is given as stem-and-leaf plots, it is quite clear that the two sets seem to come from approximately normal distributions, but we will use the Anderson-Darling normality test introduced in Lesson 13 to make sure. Click Stat > Basic Statistics > Graphical Summary and select both C1 and C2 into the "Variables:" box then click "OK". Two graphical summaries will be created. Make each all black and white then add your name, an appropriate subtitle, and an ellipse to each as discussed in Lesson 13. The critical portion of each is shown in the figure to the right. (But when you do a similar problem for Assignment 15 you will print the complete summaries.) We see that it is indeed valid to use the t-test in this case.



To conduct the t-test click on Stat > Basic Statistics > 2-Sample t. Now click on the "Samples in different columns" button, select X_1 into "First:" and X_2 into "Second:", then click on "Options" to make sure that "Alternative:" is set to "not equal". Finally, we click "OK" and "OK". We will use the P-Value method, but since we are told to use a 5% level of significance, we can type our decision and conclusion. The results, after clearing out superfluous "stuff", are shown at the top of the next page.

There is one very important result in that output that should call our attention. The degrees of freedom is 23. Based on the rules from our text we know that the degrees of freedom should be $14 - 1 = 13$.

Minitab uses a much more complicated formula to compute the degrees of freedom than the simple rule in our text. To see the rule used by Minitab click on Stat > Basic Statistics > Help > see also > Methods and formulas. Scroll down to about the middle of the page and you will see the formula Minitab uses to compute the degrees of freedom. The Minitab method makes the test a little more sensitive, but the book's way is certainly easier. Recommendation: If you are doing one of these problems by hand, do it the books way, but if you are using Minitab, take advantage of their more sophisticated approach.

Notice that the dialog box for the two sample t-test also makes it possible to do the test with summary data as we saw in lessons 11, 12, and 13. It works the same as in those previous cases except that you now have to enter the means, standard deviations, and sample sizes for both samples. We will not work another example.

Merv Newton

Lesson 15

Example

X1 = Final exam score of a student in Section A

X2 = Final exam score of a student in Section B

Data Display

Row	X1	X2
1	58	50
2	58	50
3	65	61
4	67	61
5	68	62
6	68	63
7	74	70
8	76	73
9	77	74
10	78	75
11	78	80
12	79	82
13	86	92
14	88	99
15	89	
16	89	
17	91	
18	92	

Two-Sample T-Test and CI: X1, X2

Two-sample T for X1 vs X2

	N	Mean	StDev	SE Mean
X1	18	76.7	11.0	2.6
X2	14	70.9	14.4	3.8

Difference = μ (X1) - μ (X2)

Estimate for difference: 5.86508

95% CI for difference: (-3.71726, 15.44742)

T-Test of difference = 0 (vs not =): T-Value = 1.27 P-Value = 0.218 DF = 23

Decision: Since P-Value = 0.218 > 0.05, fail to reject H0.

Conclusion: There is not a significant difference in the mean score on the final exam in Sections A and Section B at the .05 level of significance.

Also notice that Minitab has no "2-Sample z" test. We know from our text that the formula for the test statistic for a 2 sample z and a 2 sample t are the same, it is only the table we use to find the critical value that changes. We also know that for large samples, that is $n > 30$, the normal distribution table and the t distribution table are virtually the same. Thus, if we are faced with a two sample mean problem with both samples large, this 2-Sample t procedure from Minitab would give us the same result as using the 2 sample z procedure described in our text.

MINITAB ASSIGNMENT 15

See instructions on page 8.

1. Do problem 26 on page 371. Be sure your answer to part (a) is based on the Anderson-Darling normality test, not on the stem-and-leaf plots.

LESSON 16 - ANOVA

In this lesson we will learn how to make use of Minitab to do ANOVA problems. As an example we will do Problem 21 on page 452. We will let Minitab check that the conditions for ANOVA are satisfied and do the necessary computations, then we will use those results to write up our conclusions using the 5-step classical hypothesis testing procedure on a separate piece of paper. Enter the data and label the columns. Clear the Session Window below the time/date stamp and type your name, Lesson 16, and Example. Since the writeup will be on a separate paper, we need not type in the definition of the variables at this point. Display the data.

Recall that ANOVA requires that each data set be reasonably normal and that the largest sample standard deviation be less than twice the smallest sample standard deviation. We will use the Graphical Summary form of Basic Statistics to check these conditions. When you get into the Graphical Summary dialog box highlight all three variables by dragging the mouse down the list and select all three into the 'Variables:' box. This way you can do all three at once. The results of the Anderson-Darling Normality Test and the standard deviations are shown below.

Prof A ANOVA OK		Prof B ANOVA OK		Prof C ANOVA OK	
Anderson-Darling Normality Test		Anderson-Darling Normality Test		Anderson-Darling Normality Test	
A-Squared	0.43	A-Squared	0.53	A-Squared	0.29
P-Value	0.177	P-Value	0.104	P-Value	0.511
Mean	66.000	Mean	68.000	Mean	83.000
StDev	12.806	StDev	12.215	StDev	11.637

We can see that the standard deviation is not a problem; the largest is 12.8, which is certainly less than twice the smallest, which is 11.6. Normality, however is pretty "iffy." Since we are using a level of significance of 0.05, however, they all make it, so we can now proceed with ANOVA.

To do the analysis of variance click on Stat > ANOVA > One-way (Unstacked), select all the variables into the "Responses (in separate columns):" box, and click "OK". The session window is printed on the next page. Notice that Minitab gives us an ANOVA table for this data, but it looks a bit different from the notation in our text. First of all, the sum of squares column and the degrees of freedom column are switched. Also, the row our text calls "Between samples" is named "Factor" by Minitab, and the text's "Within samples" row is called "Error" by Minitab. Notice also that Minitab gives us both the value of F and the P-Value.

This is all you will do on the computer. Now, by hand and on a separate sheet of paper, go through the 5-step classical hypothesis testing procedure. Under "For our sample" in step 3, write in the ANOVA table using the results from Minitab, but use the format from our text. When you do the assignment for this lesson you will turn in the printed Session Window, all the graphs from the Graphical Summary, and the paper with the hypothesis testing procedure written out. A sample of the 5-step classical method for this example is on the page following the Session Window page. (NOTE: This sample is typed, but yours may be done by hand.)

Merv Newton
Lesson 16
Example

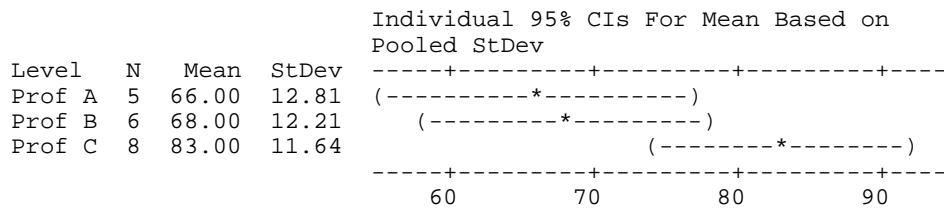
Data Display

Row	Prof A	Prof B	Prof C
1	63	67	97
2	45	45	97
3	73	76	87
4	77	80	87
5	72	70	84
6		70	74
7			74
8			64

One-way ANOVA: Prof A, Prof B, Prof C

Source	DF	SS	MS	F	P
Factor	2	1183	592	4.03	0.038
Error	16	2350	147		
Total	18	3533			

S = 12.12 R-Sq = 33.49% R-Sq(adj) = 25.17%



Pooled StDev = 12.12

Merv Newton
 April 13, 2005
 Lesson 16
 Example

Variables: X1 = A grade given by Professor A
 X2 = A grade given by Professor B
 X3 = A grade given by Professor C

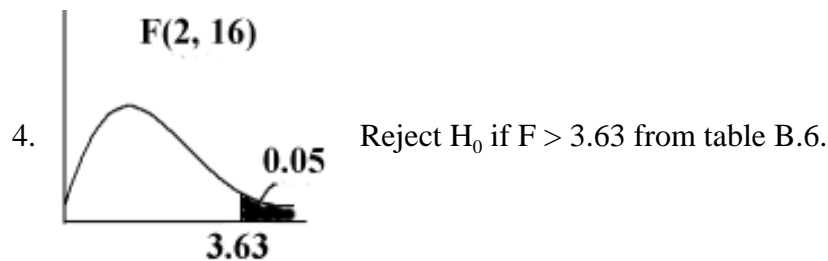
1. $H_0: \mu_1 = \mu_2 = \mu_3$.
 H_a : The means are not all equal.

2. $\alpha = 0.05$

3. Assume H_0 is true. $F = \frac{MSB}{MSW} \sim F(k - 1, n - k)$

For our sample: $k = 3, n = 19$ and $F = \frac{592}{147} = 4.03 \sim F(2, 16)$.

SOURCE	SS	df	MS	F
BETWEEN SAMPLE	1183	2	592	4.03
WITHIN SAMPLE	2350	16	147	
TOTAL	3533	18		



5. Decision: Since $F = 4.03 > 3.63$, reject H_0 .
 Conclusion: The means of the grades given by these three professors are significantly different at the 0.05 level of significance.

MINITAB ASSIGNMENT 16

See instructions on page 8.

Do each of the following problems as the example was done in this lesson. Turn in the Session Window, all the graphs from the Graphical Summary, and the 5-step classical method write-up.

1. Problem 35 on page 460.
2. Problem 36 on page 460.

LESSON 17 - REGRESSION AND CORRELATION

In this lesson we will learn to find the equation of the regression line and to plot it. We will also find the coefficient of determination, the linear correlation coefficient, and we will learn to predict values of Y for given values of X. We will use Problem 9 page 111 and Problem 37 on page 423 as an example. Notice that these two problems involve the same data, but they ask different questions about that data. Type the data into C1 and C2 in the data window, and label the columns; we will call C1 "Inflation" and C2 "Prime." Clear the session window below the time/date stamp then type your name, Lesson 17, and Example. Now display the data.

REGRESSION ANALYSIS

To find the equation of the regression line and the coefficient of determination click on Stat > Regression > Regression. Select C2 Prime into the "Response:" box and C1 Inflation into the "Predictors:" box, then click "OK". The two pieces of information we want, plus a great deal of other information we don't care about will be added to our session window. The equation of the regression line is given at the top, the item $R\text{-}Sq = 82.5\%$ is what our text calls the coefficient of determination.

The figure on the right has the unwanted information highlighted. It should be deleted. When the excess information has been removed, the session window should look like the figure below.

Regression Analysis: Prime versus Inflation

The regression equation is
Prime = 3.17 + 0.654 Inflation

Predictor	Coef	SE Coef	T	P
Constant	3.175	1.002	3.17	0.025
Inflation	0.6541	0.1347	4.85	0.005

$S = 0.816832$ $R\text{-}Sq = 82.5\%$ $R\text{-}Sq(\text{adj}) = 79.0\%$

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	15.724	15.724	23.57	0.005
Residual Error	5	3.336	0.667		
Total	6	19.060			

Regression Analysis: Prime versus Inflation

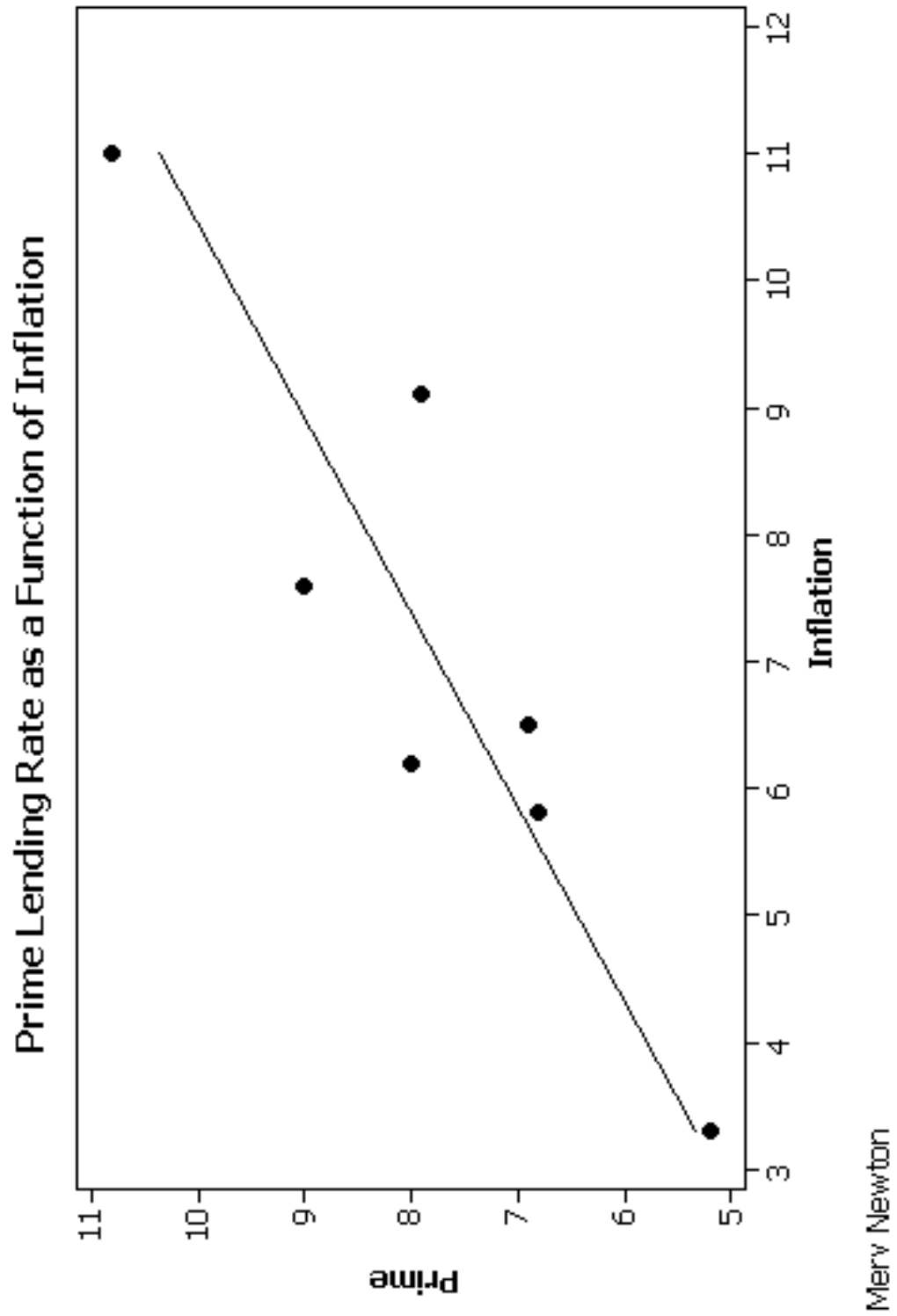
The regression equation is
Prime = 3.17 + 0.654 Inflation

$R\text{-}Sq = 82.5\%$

GRAPH

To create a scatter plot and graph the regression line, click on Graph > Scatterplot. Click on "With Regression" then "OK" in the first dialog box. In the second, select Prime into the first box in

the Y column, and Inflation into the first box in the X column. Now click on "Labels" to add a title and your name, then click "OK" and "OK". After making everything black and white as in our past graphs, it should look like the figure on the next page.



Notice that this graph violates two of our rules for graphs with two axes, the 3/4 rule and the rule to scale the y-axis from zero. It is not uncommon to violate both of these rules for scatter plots.

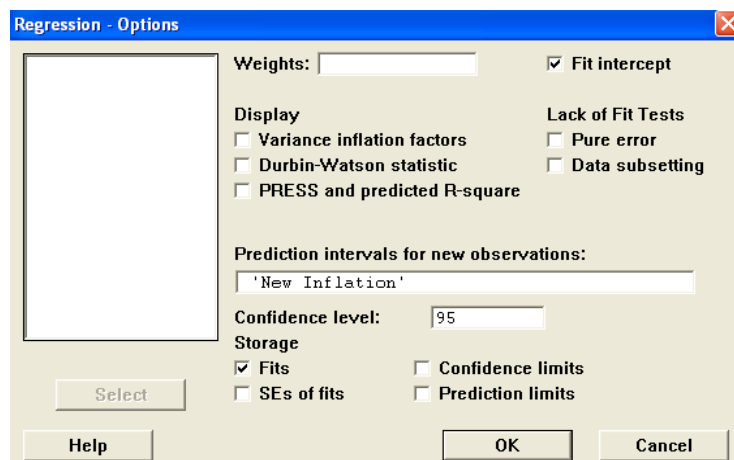
This completes what is requested for Problem 9 on page 111, but we will add one more requirement that is frequently expected when dealing with this type of problem. "Find the coefficient of determination and interpret its value." NOTE: The coefficient of determination was found along with the equation of the regression line. Write out the answers to the questions asked in the problems on a separate sheet of paper. See the Sample Write-up at the end of the lesson.

CORRELATION

To do part (a) of Problem 37 on page 423 we need to compute the linear correlation coefficient. Click on Stat > Basic Statistics > Correlation then select Inflation and Prime into the "Variables:" box and click "OK." The linear correlation coefficient, called the Pearson correlation in Minitab, (0.908 in this case) will be printed in the session window. Notice that the P-Value is also printed, so the problem can be finished by either the classical method or the P-Value method. We will use the classical approach. On a separate sheet of paper write out the 5 step classical method for this hypothesis testing problem. In the "For our sample" step, just write down the value of r provided by Minitab as in the sample write-up.

PREDICTION

Part (b) of Problem 37 asks us to predict the prime rate when the inflation rate is 7%. To do this we must give C3 a reasonable label then list in it all the values of inflation for which we want predictions. In this case there is only one value requested, so go to the data window and give C3 the title "New Inflation" and type a 7 in the column. Now click on Stat > Regression > Regression. Select Prime into the "Response:" box and Inflation into the "Predictors:" box. Now go to "Options" and select "New Inflation" into the box called "Prediction intervals for new observations:" then click on the "Fits" box under "Storage" to check it. The Options dialog box should now look like the figure on the right. Now click "OK" to get back to the Regression dialog box and click on "Results" then click on the top button, "Display nothing." Now click "OK" on the Results dialog box and on the Regression dialog box. Notice that C4 has now been given the name PFIT1 and the value 7.75328 has been added. This is the result we are seeking, but the name is rather strange. Change



the label of C4 to "New Prime." Return to the session window and do a Display Data for C3 and C4. The complete session window for this example is shown below.

4/16/2005 2:10:49 PM		
Merv Newton		
Lesson 17		
Example		
Data Display		
Row	Inflation	Prime
1	3.3	5.2
2	6.2	8.0
3	11.0	10.8
4	9.1	7.9
5	5.8	6.8
6	6.5	6.9
7	7.6	9.0
Regression Analysis: Prime versus Inflation		
The regression equation is		
Prime = 3.17 + 0.654 Inflation		
R-Sq = 82.5%		
Correlations: Inflation, Prime		
Pearson correlation of Inflation and Prime = 0.908		
P-Value = 0.005		
Data Display		
Row	New Inflation	New Prime
1	7	7.75328

Below and on the next page is a complete write-up for this example. Your write-ups for the assignment at the end of this lesson should be similar to this, except that you need not type it.

SAMPLE WRITE-UP

Merv Newton
April 16, 2005
Lesson 17
Example

x = rate of inflation for a year
y = prime lending rate for a year

Problem 9 page 111:

- (a) Line of best fit is given by $y = 3.17 + 0.654x$
- (b) Scatter plot with regression line. See attached graph.
- (c) The coefficient of determination indicates that 82.5% of the variation in the prime lending rate is explained by the variation in the rate of inflation.

Problem 37 page 423:

(a)

1. $H_0: \rho = 0$

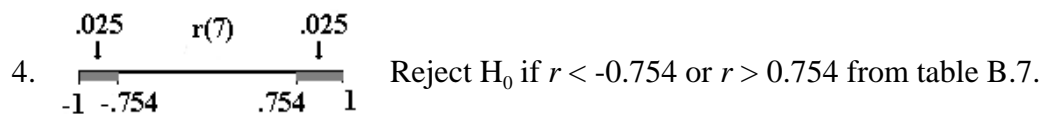
$H_a: \rho \neq 0$

2. $\alpha = 0.05$

3. Assume H_0 is true.

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}} \sim r(n).$$

For our sample $r = 0.908$



5. Decision: Since $r = 0.908 > 0.754$, reject H_0 .

Conclusion: There is a significant linear correlation between the rate of inflation and the prime lending rate at the 0.05 level of significance.

(b) If the inflation rate is 7%, then the prime lending rate should be about 7.75%.

MINITAB ASSIGNMENT 17

See instructions on page 8.

For each of the problems below enter and display the data then let Minitab do all the computations. On a separate sheet of paper write out the answers to the questions posed by the problem in the text and those added to the problem below. NOTE: These may be different from the questions posed in the sample problem in the lesson above. In each case have Minitab produce a scatter plot with the regression line included as we did in the example above. When a prediction is required, be sure to change the name of the predicted value(s) to something more reasonable than the default name given by Minitab.

1. Do Problem 49 on page 140 with the following parts added.
 - (e) Test the claim, at the 5% level of significance, that there is a linear correlation between cholesterol intake and death rate from heart disease.
 - (f) Does this prove that cholesterol causes heart disease? Explain.
2. Do Problem 12 on page 113 and Problem 38 on page 133. When doing part (b) of Problem 12 add the following parts
 - (c) Predict the number of absences for an employee earning \$45,000.
 - (d) Explain why this last prediction is not reliable.NOTE: No hypothesis testing is required for this problem.

LESSON 18 - CONTINGENCY TABLES

In this lesson we will learn how to have Minitab do the computations necessary to do a chi-square test on a contingency table. As an example we will do Problem 32 on page 490. We will let Minitab do the computations for us, then write up the result using the classical approach.

Type the contingency table into C1, C2, and C3 in the data window. Since we have no way of putting in row titles, we will dispense with the column titles also in this case. Now clear the Session Window below the date/time stamp then type your name, Lesson 18, and Example. Now click on Stat > Tables > Chi-square Test. Select C1, C2 and C3 into the box labeled "Columns containing the table:" and click on "OK". The results are shown below.

4/18/2005 11:42:55 AM				
Merv Newton				
Lesson 18				
Example				
Chi-Square Test: C1, C2, C3				
Expected counts are printed below observed counts				
Chi-Square contributions are printed below expected counts				
	C1	C2	C3	Total
1	7	7	9	23
	6.29	5.11	11.60	
	0.080	0.698	0.582	
2	19	13	38	70
	19.15	15.56	35.30	
	0.001	0.420	0.207	
3	6	6	12	24
	6.56	5.33	12.10	
	0.048	0.083	0.001	
Total	32	26	59	117
Chi-Sq = 2.120, DF = 4, P-Value = 0.714				

Notice that the expected value for each cell is typed below the observed value, but not in parentheses as is customary. Below that appears the quantity $(O - E)^2/E$ for each cell. The row and column totals for the observed values and the overall total appear where we would expect them. The chi-square value, the degrees of freedom, and the P-Value are listed at the bottom. When writing up the 5-step classical method, use what you need of this information for the "For our Sample" step, but make it look the way we learned to do it in class. An example of the write-up for this problem is on the next page.

Merv Newton
 April 18, 2005
 Lesson 18
 Example

Variables: Rows = Attitude towards racial quotas in hiring.
 Columns = Political affiliation.

1. H_0 : Attitude towards racial quotas in hiring and political affiliation are independent.
 H_a : There is a relationship between attitude towards racial quotas in hiring and political affiliation.

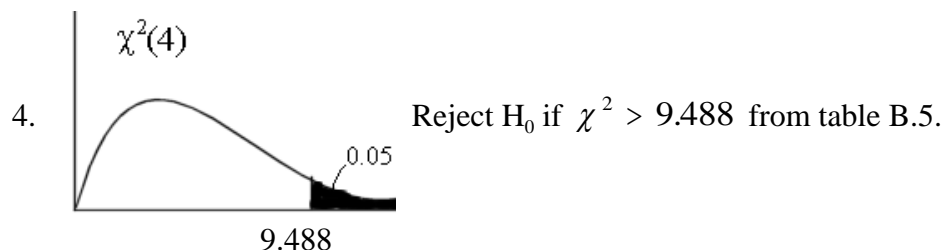
2. $\alpha = 0.05$

3. Assume H_0 is true. $\chi^2 = \sum \frac{(O - E)^2}{E} \sim \chi^2((r - 1)(c - 1))$.

For our sample:

	Democrat	Republican	Independent	
Approve	7 (6.29)	7 (5.11)	9 (11.60)	23
Disapprove	19 (19.15)	13 (15.56)	38 (35.30)	70
Undecided	6 (6.56)	6 (5.33)	12 (12.10)	24
	32	26	59	117

$$\chi^2 = 2.120$$



5. Decision: Since χ^2 is not greater than 9.488, we fail to reject H_0 .

Conclusion: There is not a significant relationship between attitude towards racial quotas in hiring and political affiliation at the 0.05 level of significance.

MINITAB ASSIGNMENT 18

See instructions on page 8.

1. Do Problem 34 on page 490 as the example was done above.
2. Do Problem 35 on page 491 as the example was done above.

LESSON 19 - SIGN TEST

In this lesson we will learn how to use Minitab to do a sign test. We will use Problem 2 on page 499 as an example. We will use the p-value method, but since the text tells us to use a 5% level of significance, we will write a decision and conclusion.

Open Minitab and clear everything below the date/time stamp. Type your name, Lesson 19 and Example on separate lines. Define the variable, X = Time for the runner to complete a run. Enter the data into C1, label the column and display the data. Now click on Stat > Nonparametrics > 1-Sample Sign. Select C1 into the "Variables:" box, click on the button for "Test median", type 30 into that box, then change the "Alternative:" to "less than" and click "OK". The results, along with the decision and conclusion that we must write, are shown in the figure below.

Merv Newton
Lesson 19
Example

X = time for the runner to complete a run.

Data Display

Time	28.5	28.0	29.5	28.5	29.0	29.5	32.0	30.5	29.0	29.5	30.5
------	------	------	------	------	------	------	------	------	------	------	------

Sign Test for Median: Time

Sign test of median = 30.00 versus < 30.00

	N	Below	Equal	Above	P	Median
Time	11	8	0	3	0.1133	29.50

Decision: Since the P-value = 0.1133 > 0.05, we fail to reject H_0 .

Conclusion: The median time for this runner is not significantly less than 30 minutes at the .05 level of significance. There is insufficient evidence to support the runner's claim.

Notice that if we had chosen to use the classical approach, the number of plus signs would be given by what Minitab has labeled "Above", 3 in this case. If we had chosen to use the classical method, our critical value for the number of +'s would have been 2 (see table B2 in our text).

MINITAB ASSIGNMENT 19

See instructions on page 8.

1. Podunk University claims that their median class size is no greater than 25. A group of students believe that the median is greater than that. The data below represents class sizes of a random sample of 15 classes at PU. Enter and display this data. Using the p-value method and a 10% level of significance, what conclusion can the students draw from this data?

32 19 26 25 28 21 29 22 27 28 26 23 26 28 29