



Student Lab Manual

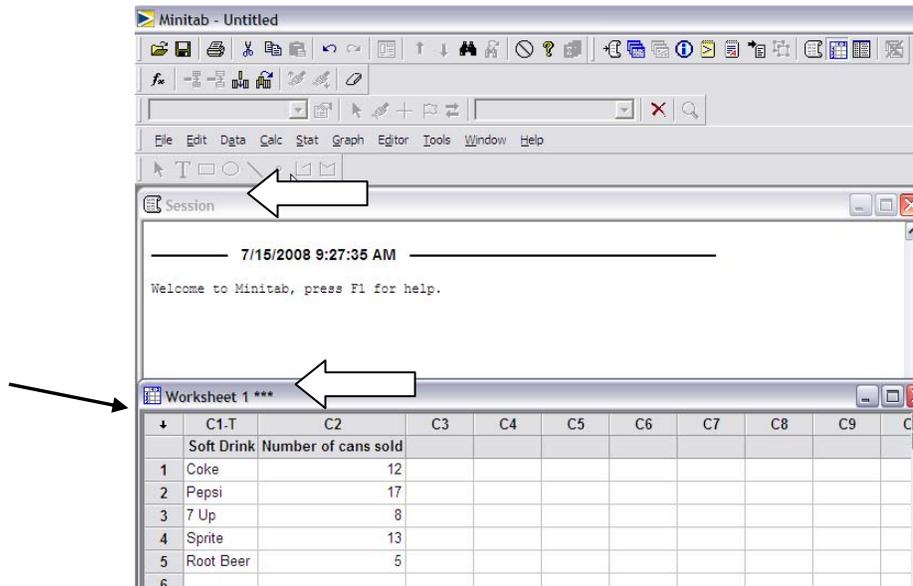
Elementary Statistics Math 119

Terrie Teegarden
San Diego Mesa College

Minitab Introduction

Getting Started

When you open a Minitab worksheet you will see two areas.



The upper area or **session window** will record the commands and the resulting analysis. You can also use it to input text such as an explanation of the analysis or the homework number.

The lower level is similar to a spreadsheet. Both text and numeric data can be input, as shown in the example above.

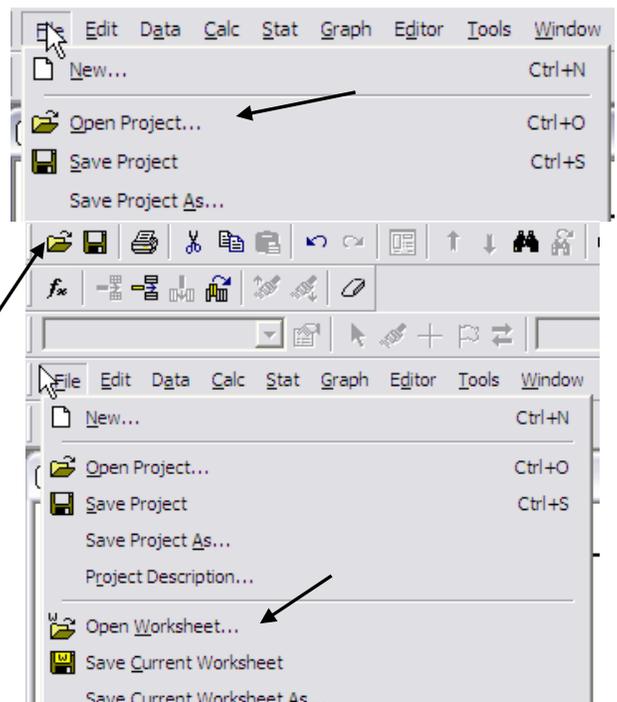
Note that the label for the first column is C1-T. The T indicates that the data in this column is nonnumeric or text. Later, it will be seen that for certain statistical applications, this column will not be available.

Under the **File menu** there are several options. You can open and save both Minitab projects and worksheets. Projects consist of data sets and all the associated tests and graphs that may have been done.

Worksheets are used for data only. Minitab comes with several data sets which can be accessed through

File → Open Worksheet

Note: Using the folder icon instead of opening the File menu, opens a project not a worksheet.



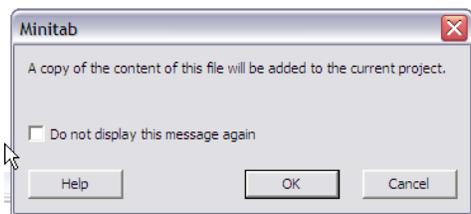
Open Files

You may have to tell the computer where to find the data files.

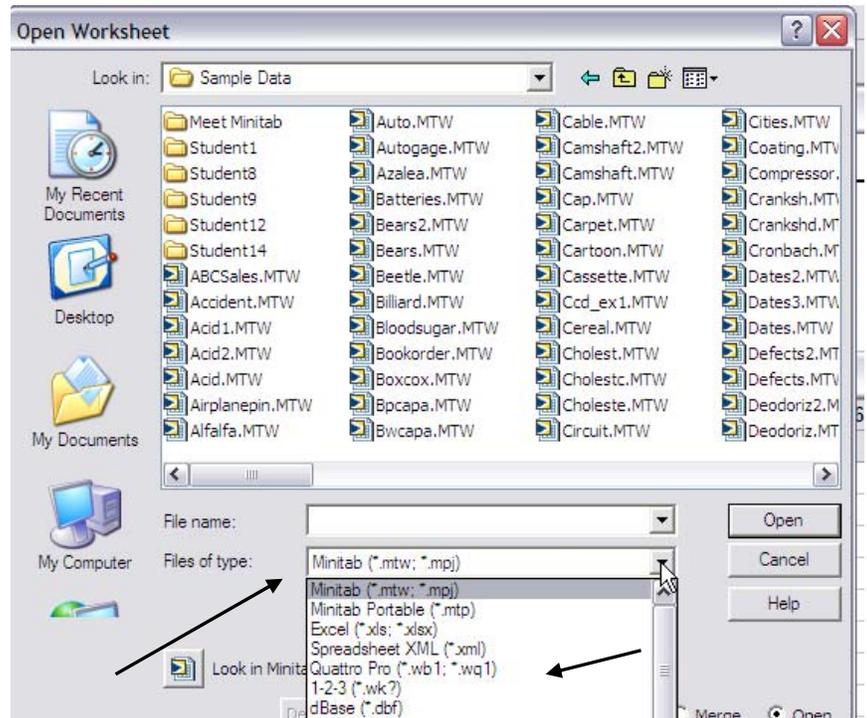
If you are in the math lab they will be in data folder on the Mesa_K server. Note that the file type is “**mtw**”.

Minitab can also open data from other formats however you must change the File type.

Selecting the data set name and clicking open, will cause the following window to appear.



Click ok and the data will appear in the open worksheet.



Save Files

When you are done you can save all your work as a project by selecting **File → Save Project as**. When you are working on homework or labs, you may wish to save your work as a project.

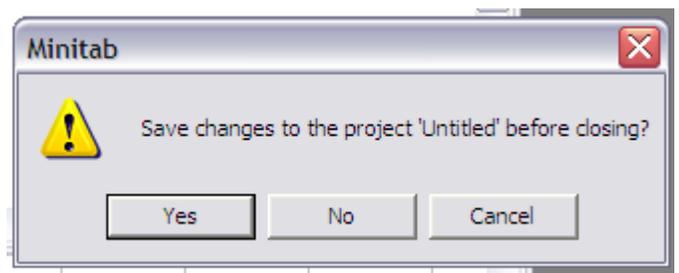
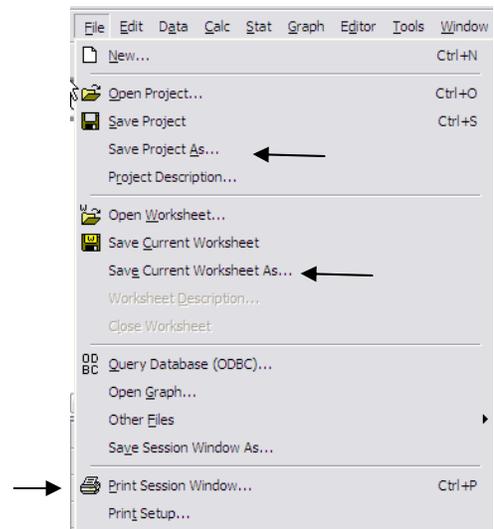
To save the data only under a different name use **File → Save Current Worksheet as**.

Print Files

Either the session window or the data can be printed. Click in the desired area of the worksheet and then use **File → Print**.

Exit

When you have finished, click on **File → Exit** and this screen will appear. If the project has not already been saved, do so at this time.



Illustrating the Data

Working with Qualitative Data

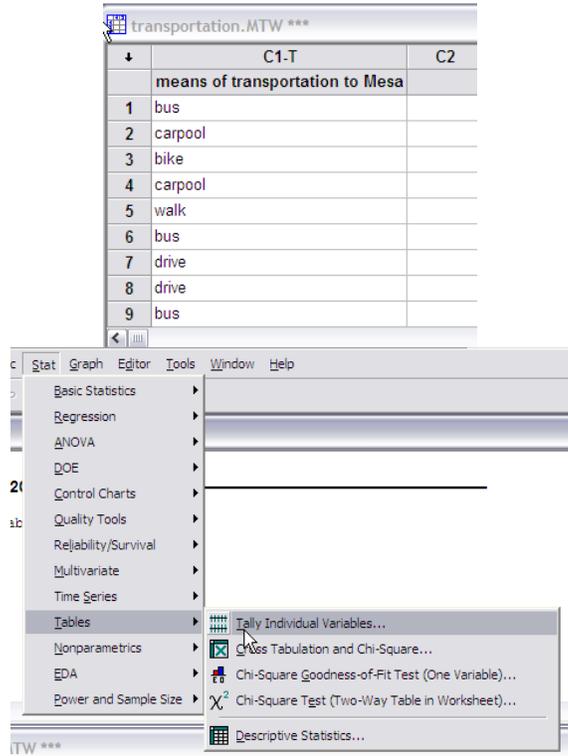
Qualitative data may be recorded in two forms, as a list or in a table as summary data.

Open **transportation.mtw** and note that the data is recorded as a list.

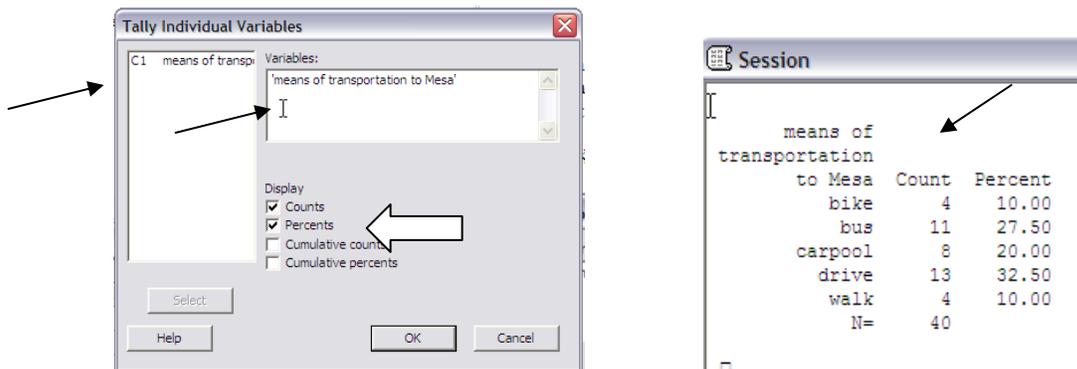
There are 40 data values and in its current form there is little information that can be derived from the data.

To summarize the data use

Stat → Tables → Tally Individual Variables



The following dialogue box will appear. Ensure that the cursor is in the box under **Variables** and the available data sets will appear in the left hand box. All columns with appropriate data will be listed. Click on **C1** and select to put **'means of transportation to Mesa'** in the box. Under **Display**, there are four options. Select **Count** and **Percent** as these are the only two types of analysis we are able to do with this data and click **OK**. The results appear in the session window of the worksheet.



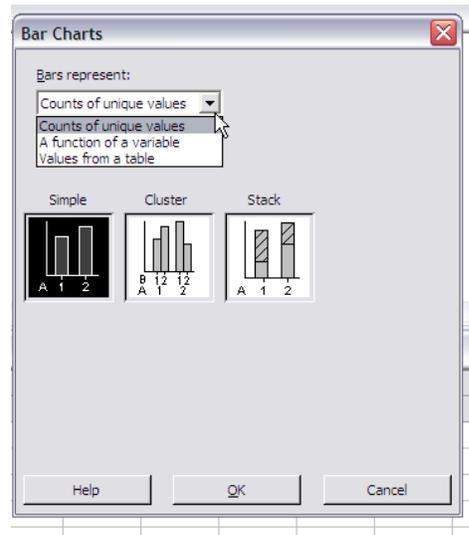
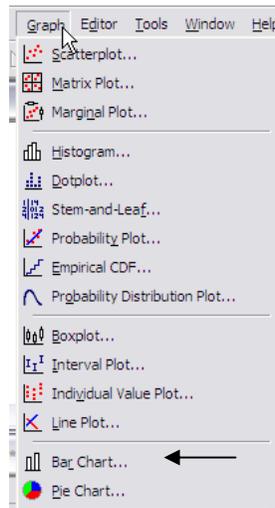
The first column lists the means of transportation in alphabetical order.

The second tells how many times each was listed in the original data set (for example, 13 people drive to Mesa), with the last number representing the total number of responses, $N = 40$.

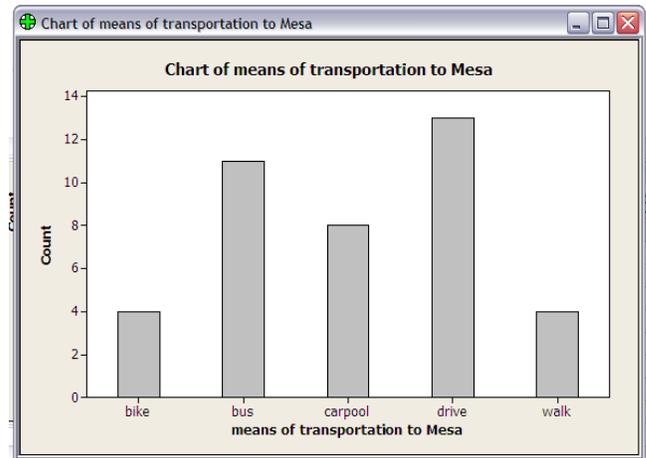
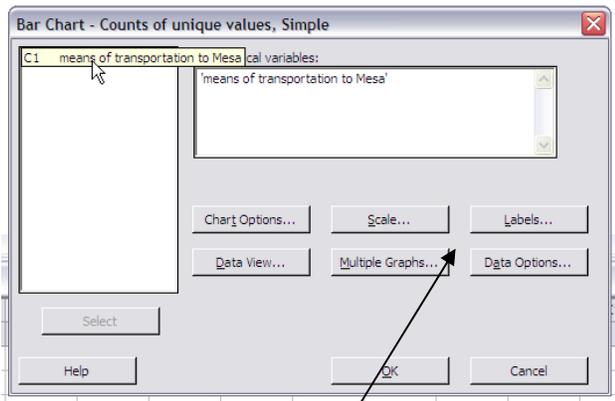
The third column gives the percent of the group that used each means of transportation. The data indicates that, 25% of the students ride the bus.

Graphing Data

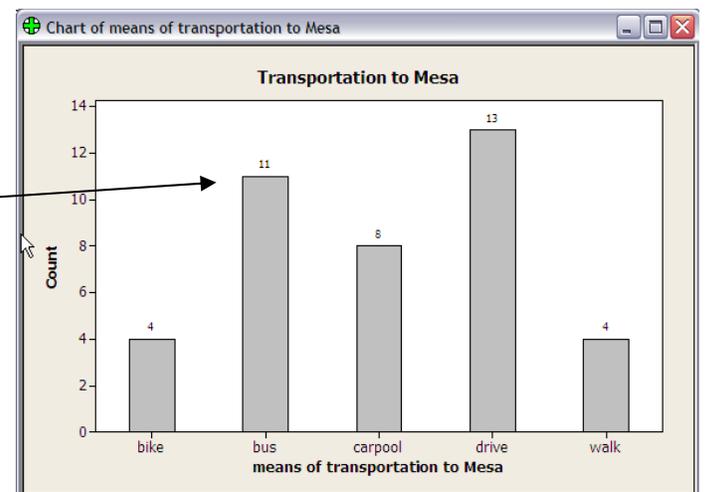
There are two ways to illustrate this data, a bar graph or a pie chart. To create a bar graph, go to **Graph** → **Bar Chart**. Each bar will represent the **counts of unique values**. Select the **Simple** graph and click **OK**.



The following dialogue box allows you to choose the categorical data set. Select C1 and then click **OK**. The bar graph appears in its own window.



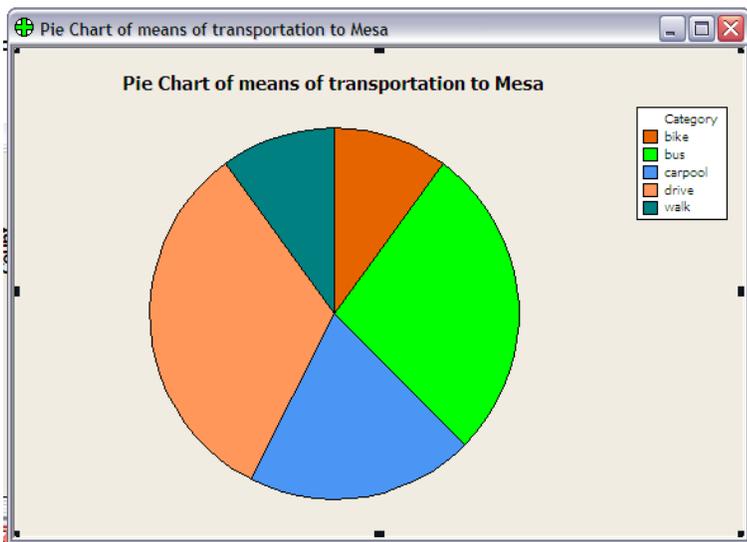
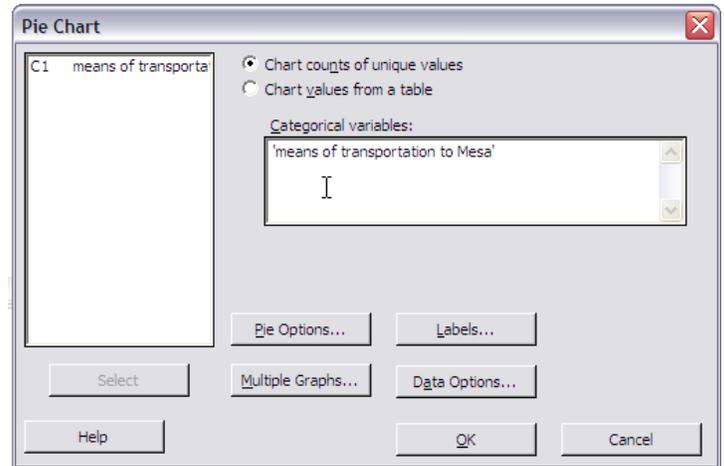
By selecting the data option from the previous screen the title can be changed and the counts can be included with each bar.



The other type of graph often used with qualitative data is a pie chart.

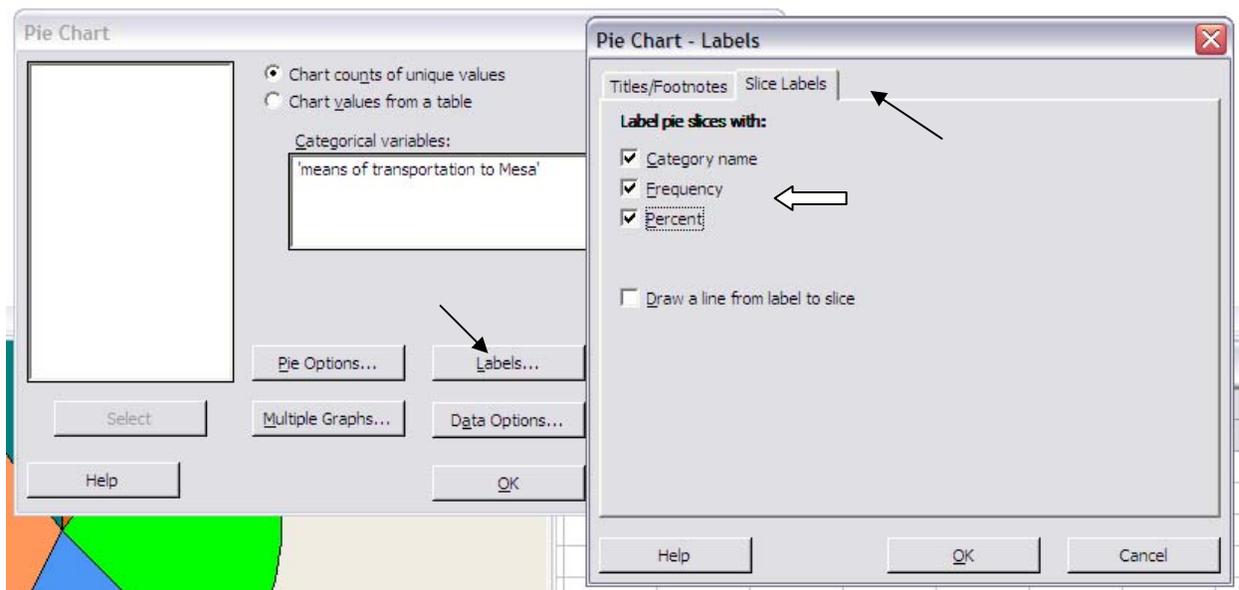
Going to the **Graph** menu, select **Pie Chart** and the window to the right appears.

Since the data is a list of values, select the top option. Place the cursor in the **Categorical variables** box and click to have the possible data columns appear in the box on the left. Select C1 and then **OK**.



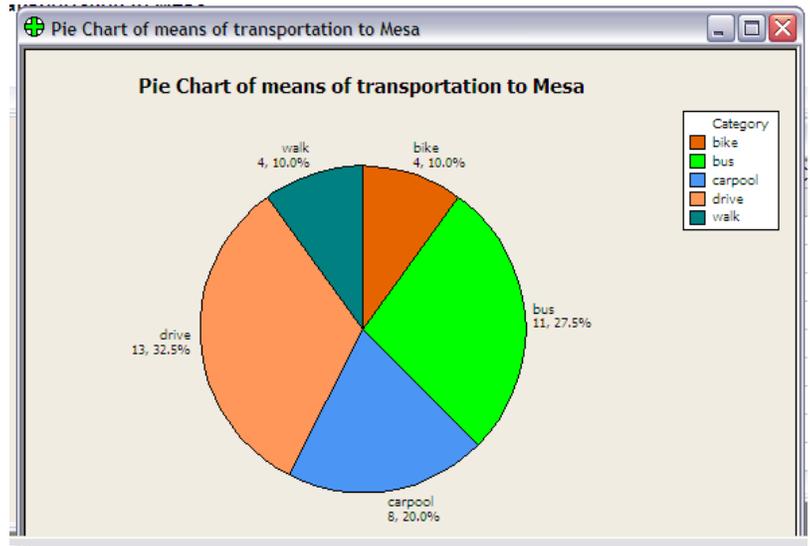
This is the standard pie chart, which does not give much information.

Go back to the pie chart dialogue box (**Graph** → **Pie Chart**) and select **labels**, then **slice labels**. Click on all the options.



This graph gives more information: the name, count and percentage for each category.

Which option you choose will depend on the information you wish to convey.



Summarized Data

Sometimes the categorical data is given in summarized or table form.

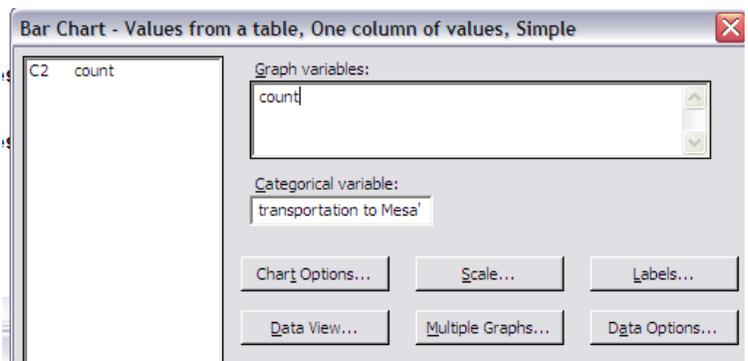
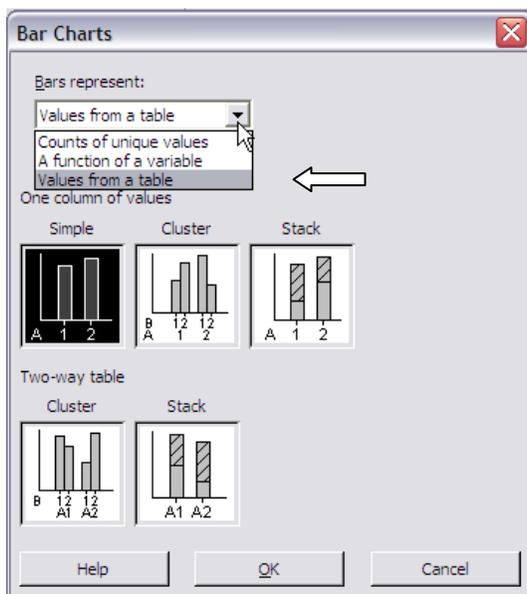
Open the worksheet **transportation2.mtw**.

	C1-T	C2	C3
	means of transportation to Mesa	count	
1	bike	4	
2	bus	11	
3	carpool	8	
4	drive	13	
5	walk	4	
6			

To create a bar or pie chart for summarized categorical data, go to **Graph → Bar Chart** and change the type of data from **Counts of unique values** to **Values from a table**.

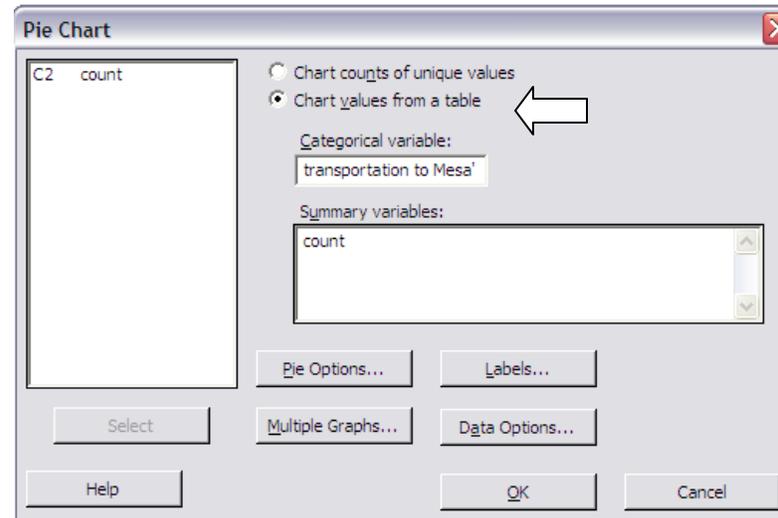
The **Categorical variable** is the 'means of transportation to Mesa' column and the **Graph variable** is the count.

Note: when you put the cursor in the **Graph variables** box, the text column is not a possible selection. This is because Minitab recognizes that text data can only be categorical.



For the pie chart, the **Chart values from a table** option is on the data selection screen.

Selecting **OK** will generate the same graphs as before. The data values are the same. It is only the form of reporting that is different.



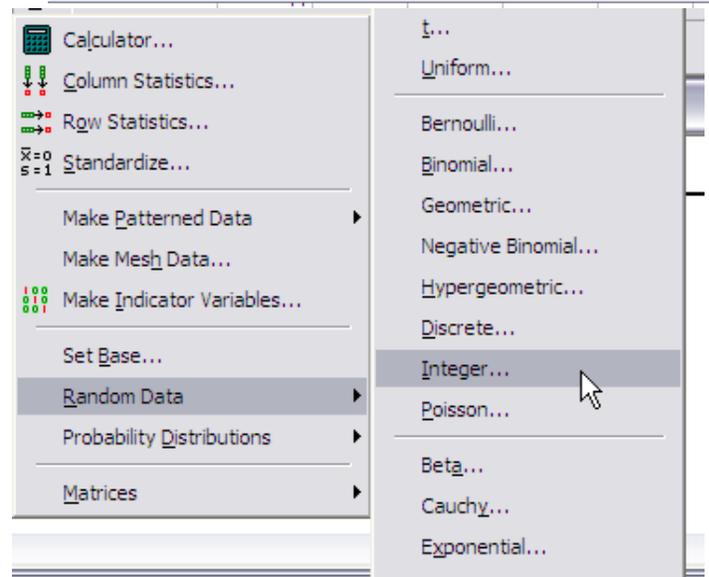
Working with Discrete Qualitative Data

Quantitative data may be discrete (countable) or continuous.

Beginning with discrete data, simulate the tossing of a die 24 times.

To do this, go to

Calc → Random Data → Integer

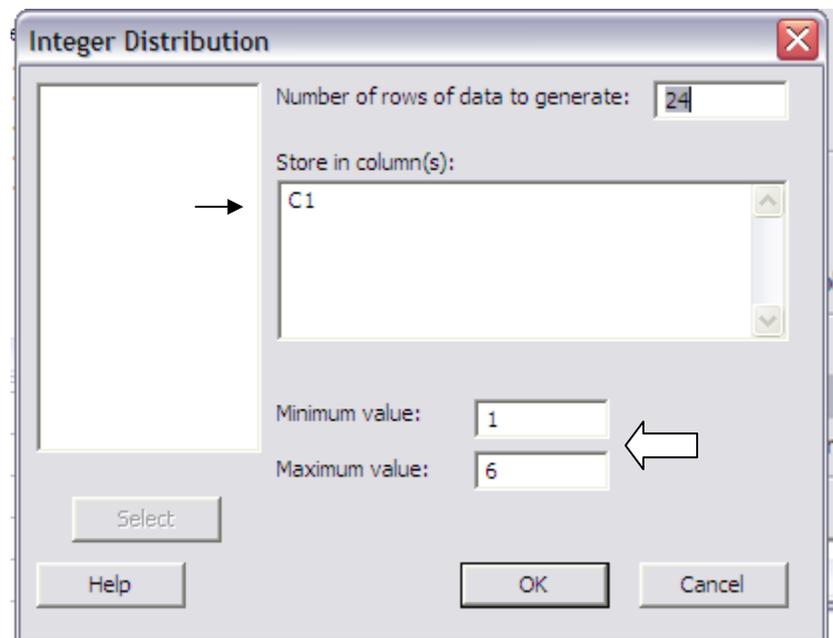


Each row represents a toss of the die, hence input 24 in that box.

Store the data in the first empty column.

The **Minimum value** for a die is 1 and the **Maximum value** is 6.

Click **OK**.

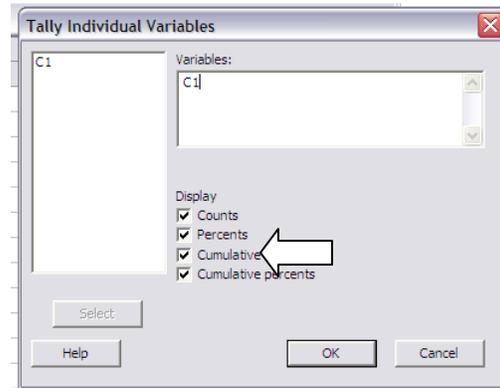


To the right is an example of the outcome. Each simulation will have its own list of data values as these are randomly generated. The data should consist of only the numbers 1, 2, 3, 4, 5, and 6.

↓	C1	C
1	3	
2	2	
3	6	
4	2	
5	1	
6	2	
7	1	
8	6	

To summarize the data use **Stat → Tables → Tally Individual Variables**.

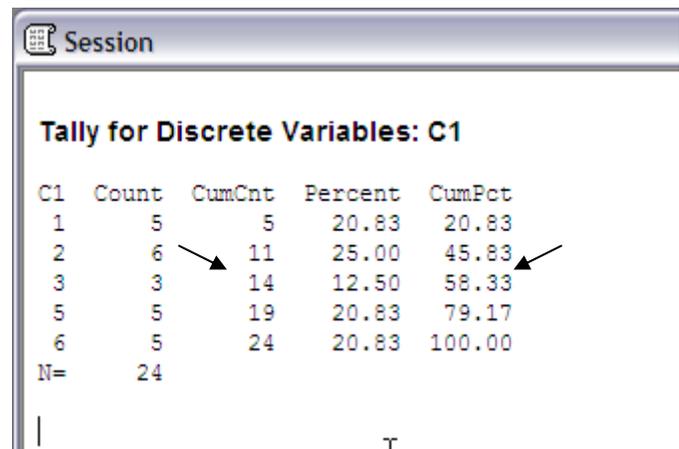
Under **Display** select all the options.



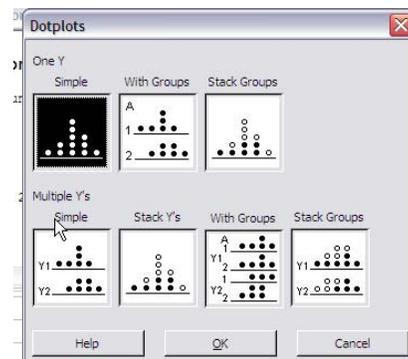
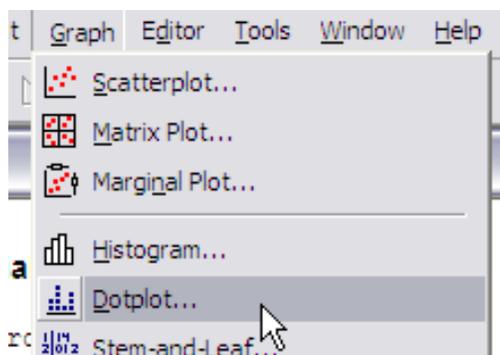
Note: For this set of data, a 4 was never tossed. In 24 rolls of a die would this surprise you?

Looking at the **CumCnt** (cumulative count) column, there are 14 tosses which resulted in a 1, 2 or 3. In other words there are 14 values of 3 or less. Similarly, 10 values are at least 5.

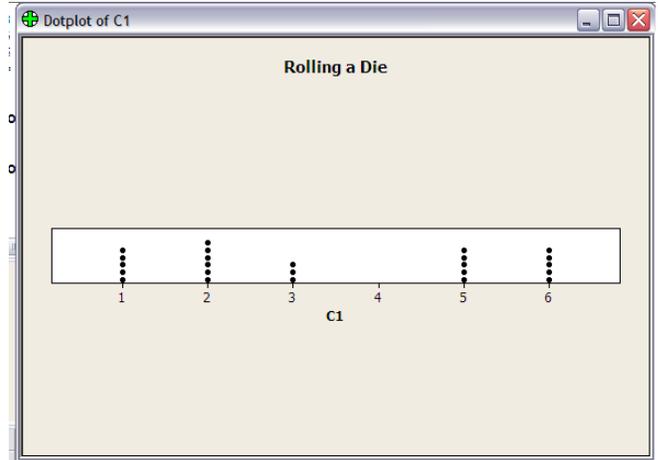
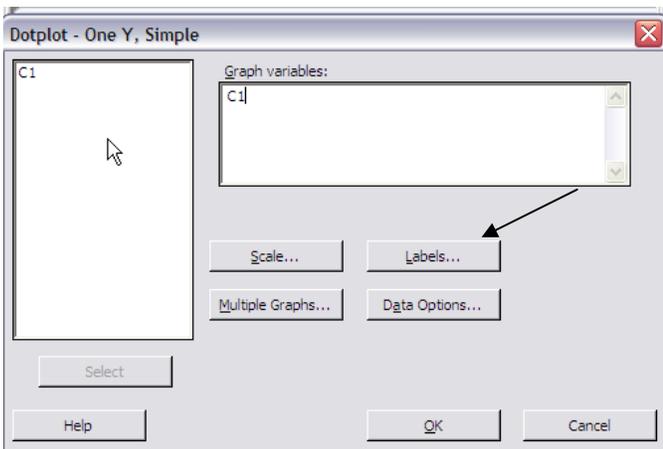
To view this data in terms of cumulative percentages look under **CumPct**. 58.33% of the data is at most 3.



To illustrate this data, the dotplot graph can be helpful. Go to **Graph → Dotplot**, then select **Simple** and click on **OK**.



Select **C1** as the **Graph variable**. To give the graph a title, click on **Labels** and input a title.



This is similar to a bar chart and works well for small data sets.

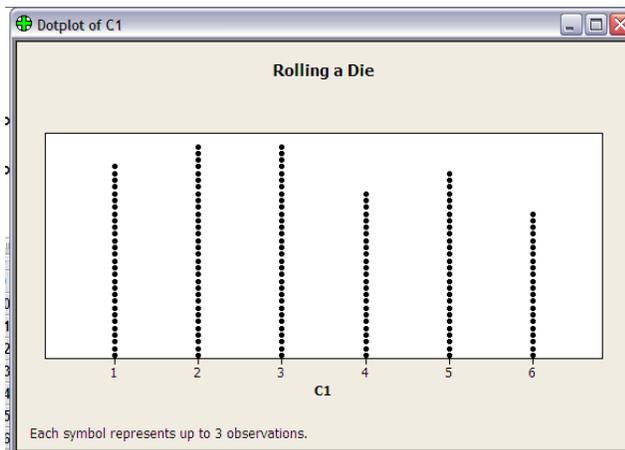
For data with high counts (20 or more) per value, a bar chart may be better. Redo the simulation with 500 tosses (rows). **Calc** → **Random Data** → **Integer**

Possible result:

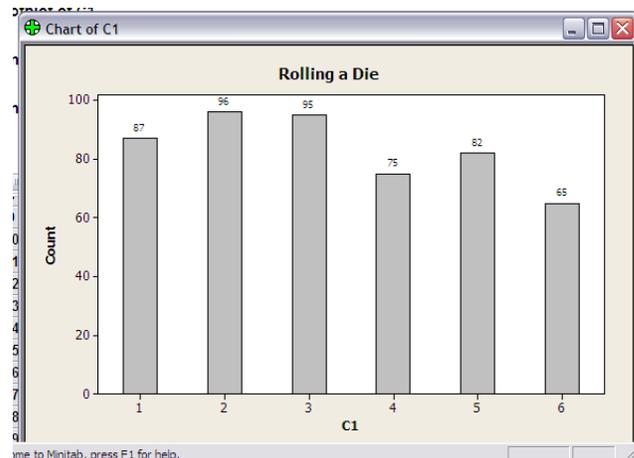
Session

Tally for Discrete Variables: C1

C1	Count	CumCnt	Percent	CumPct
1	87	87	17.40	17.40
2	96	183	19.20	36.60
3	95	278	19.00	55.60
4	75	353	15.00	70.60
5	82	435	16.40	87.00
6	65	500	13.00	100.00
N=	500			



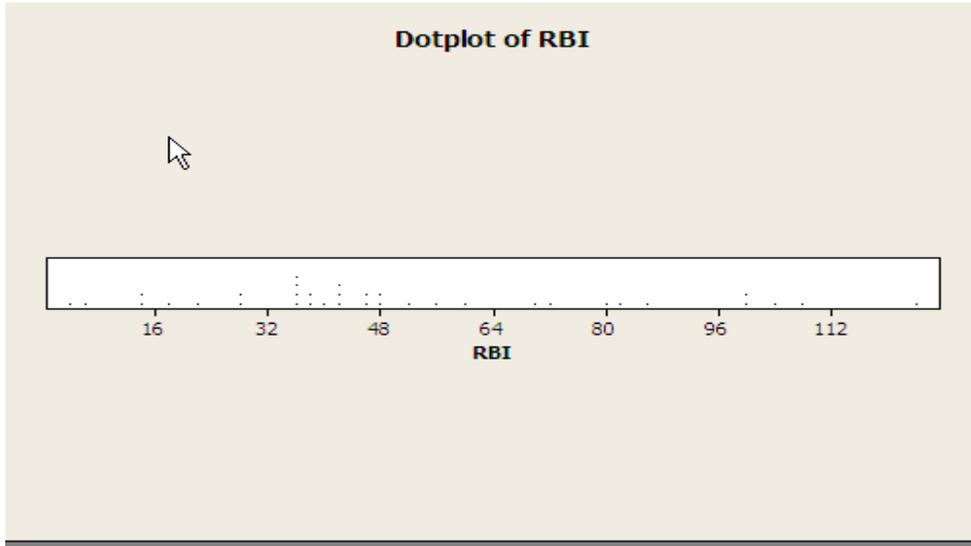
Which graph do you find easier to read?



Open the data worksheet **baseball-update.mtw** and again summarize the RBI data by selecting the appropriate column and use **Stat → Tables → Tally Individual Variables**

RBI	Count	Percent
3	1	2.86
5	1	2.86
13	1	2.86
14	1	2.86
17	1	2.86
22	1	2.86
27	1	2.86
28	1	2.86
35	1	2.86
36	3	8.57
38	2	5.71
39	1	2.86
42	3	8.57
45	1	2.86
46	1	2.86
47	1	2.86
48	1	2.86
51	1	2.86
56	1	2.86
60	1	2.86
69	1	2.86
71	1	2.86
79	1	2.86
82	1	2.86
85	1	2.86
100	2	5.71
103	1	2.86
108	1	2.86
124	1	2.86
N=	35	

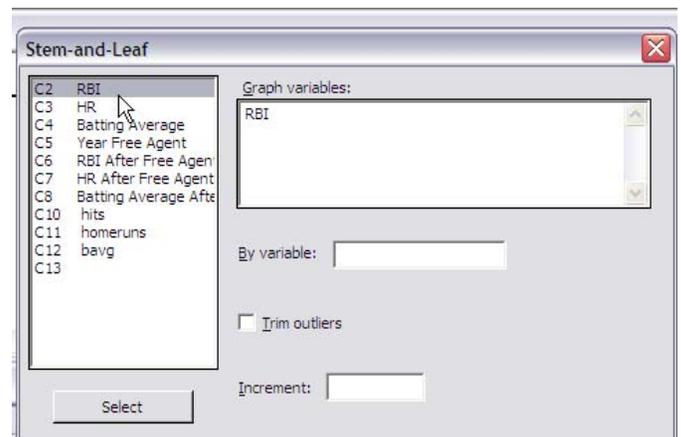
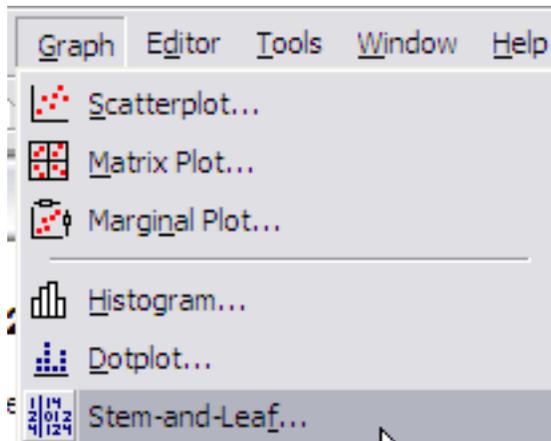
The result is not very helpful, so create a dotplot.
Graph → Dotplot

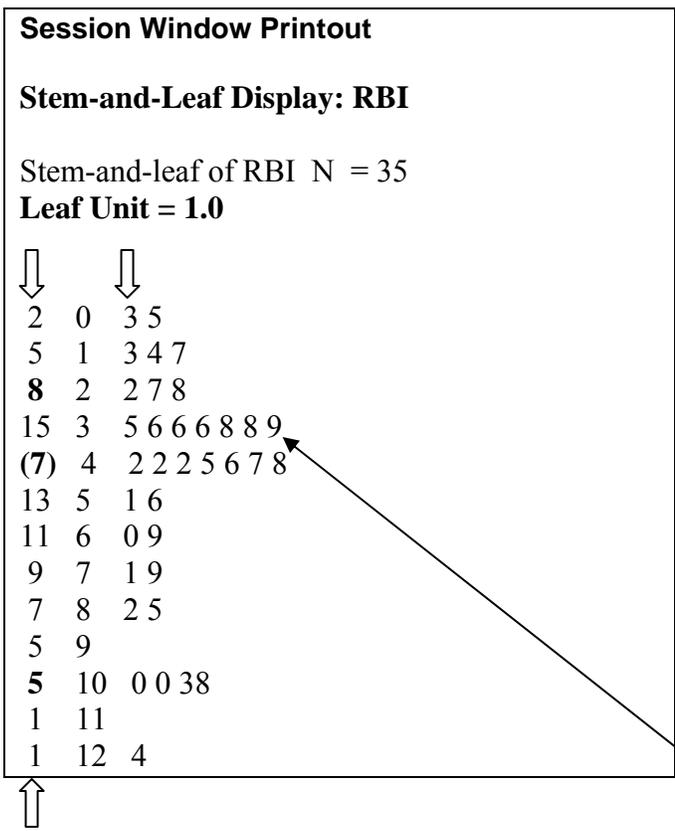


The dotplot is not very helpful. Under Graph, there is a Stem-and-Leaf plot.

Graph → Stem-and-Leaf

Add **C2 RBI** to **Graph variables** and click **OK**





There are three columns in the display. The first column is a cumulative count that begins at both the top and bottom of the column and counts toward the median. The **(7)** indicates that the median falls in that group of 7 data values.

For example, the **8** in the third row represents the total number of data values for that row and the two rows above it. Similarly the **5** in the third row from the bottom indicates that the last three rows contain a total of 5 data values.

The second column represents the stem or value in the ten's place. Since the **Leaf unit** is one, the last column represents the digit in the ones place for each data value. If the same value appears in the data several times each is recorded. If no values appear in the third column then there are no data values in that range. Note that there are no values in the nineties.

The fourth row, for example, conveys the following information:

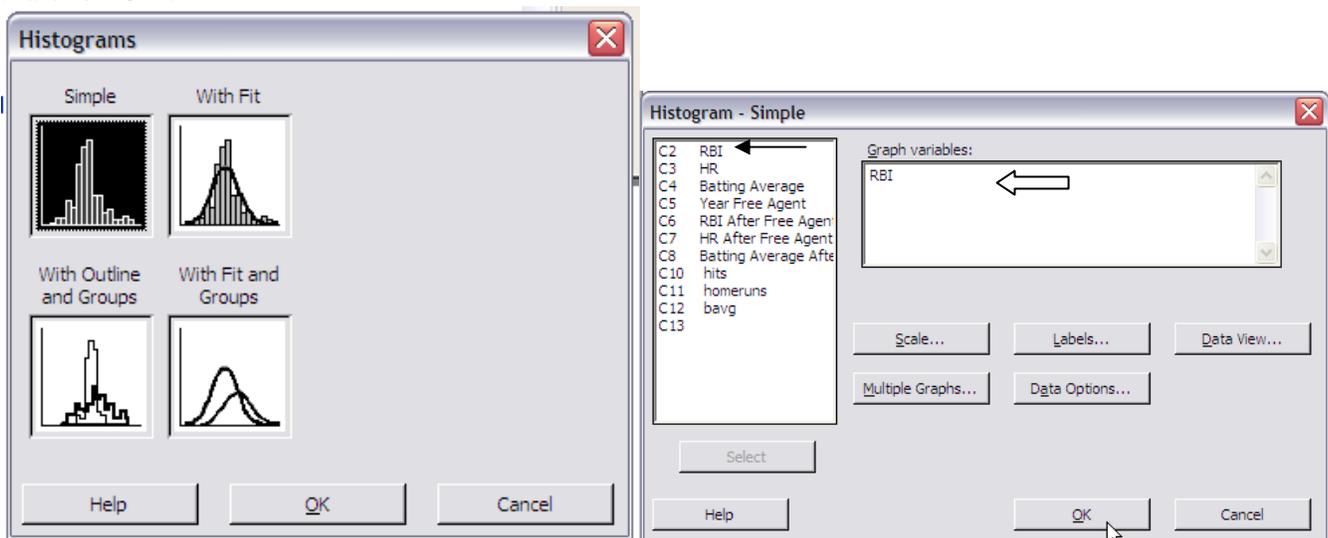
- There are a total of 15 data values in the first four columns.
- The values in this row are all in the thirties.
- The values in this row end in 5, 6, 6, 6, 8, 8, 9... therefore the data values in this row are:
- **35, 36, 36, 36, 38, 38 and 39.**

Working with Continuous Qualitative Data

For continuous data, a histogram is used in place of a bar chart. The difference is that in a histogram the bars are touching.

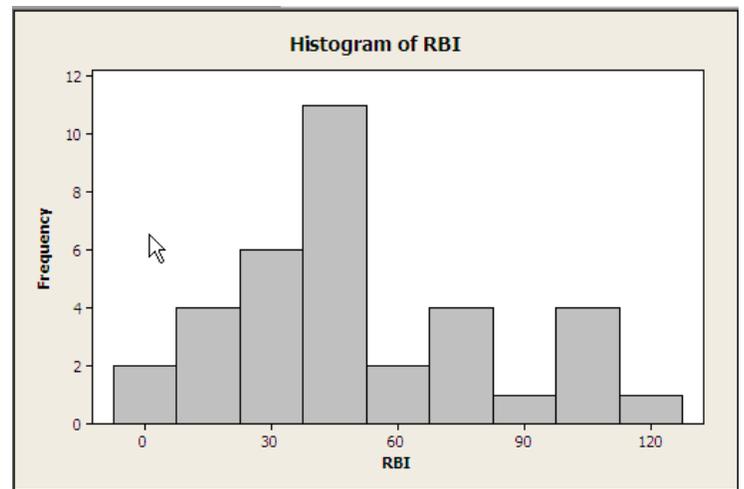
Select **Graph** → **Histogram**, select **Simple** and click **OK**.

On the next screen place the cursor in the **Graph variables** box and select **C2 RBI** from the left column and click **OK**.



On the histogram, the vertical axis represents the number of values that fall in the range for that bar (frequency). The horizontal axis represents the midpoint of the indicated bars.

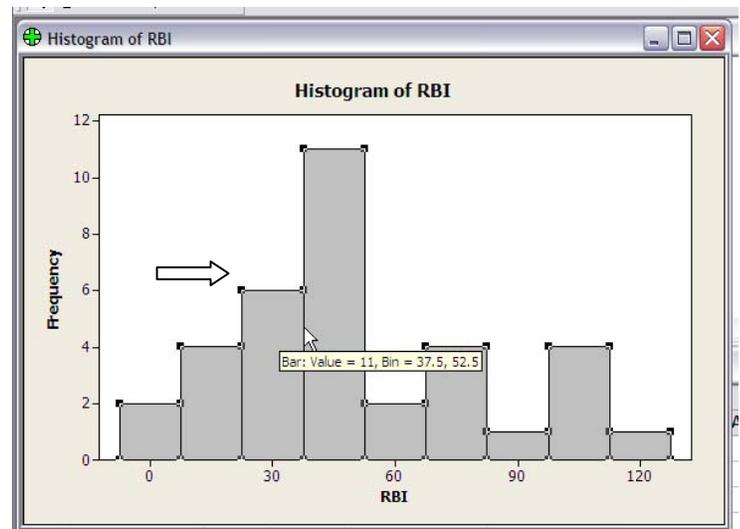
The range of values are represented in each of the bars is not indicated.



Highlighting any bar will give the number of data values and the range.

For example, the 4th bar has 11 values between 37.5 and 52.5.

It is better to label the horizontal axis with the boundary or cutpoints rather than the midpoint.

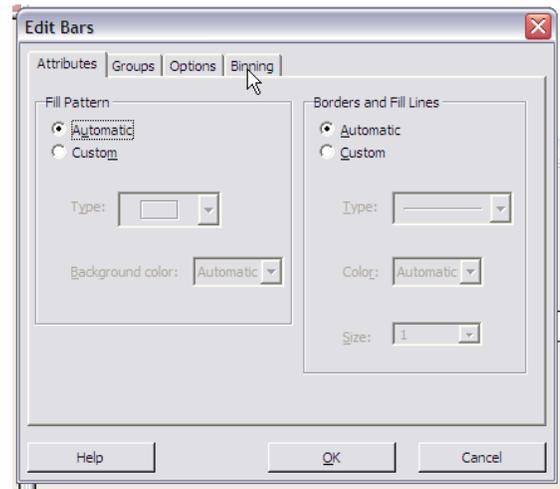
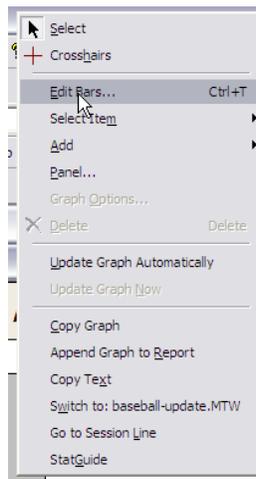


To do this, ensure that all the bars are selected and right click.

Select **Edit Bars**.

Click on the **Binning** tab and then click **OK**.

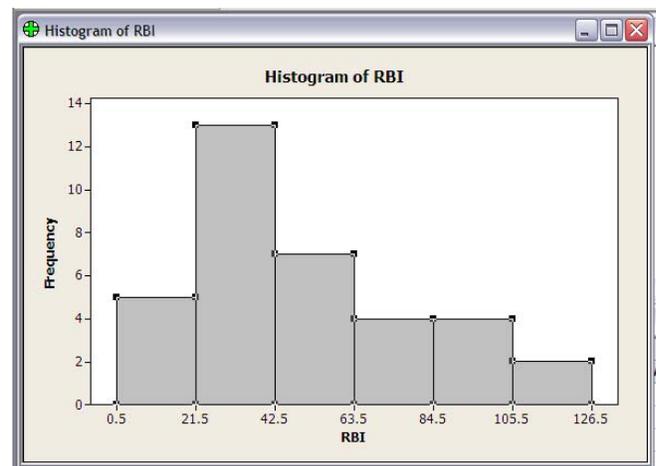
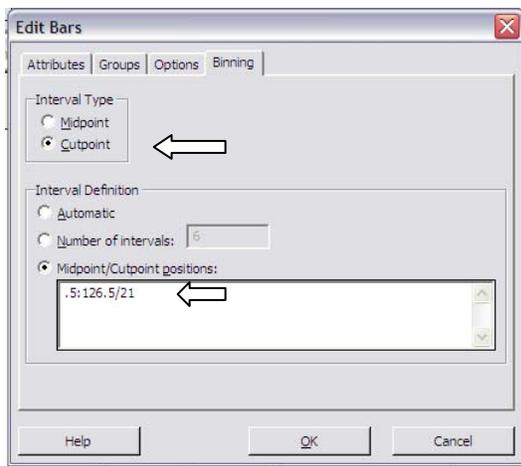
Select **Cutpoint** as the and **Midpoint/Cutpoint positions**.



Using the techniques learned in class, the lower class boundary for the first class is 0.5 and the upper class boundary for the last class is 126.5 with a class width of 21. In the box, use the following syntax:

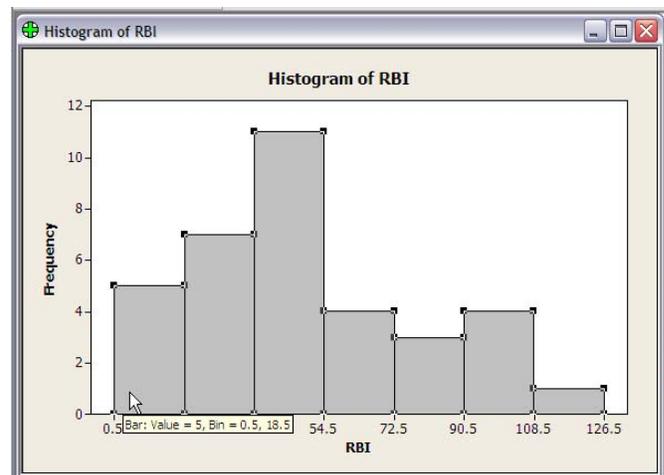
lower class boundary for the first class : upper class boundary for the last class / class width

.5:126.5/21

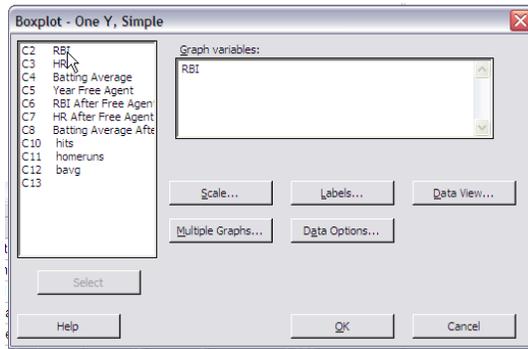
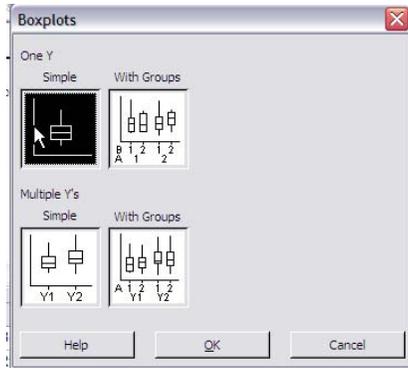


A histogram with 7 bars and a class width of 18 could also be used for this data.

Return to the **Edit Bars/Binning** screen and change the class width to obtain a new graph.



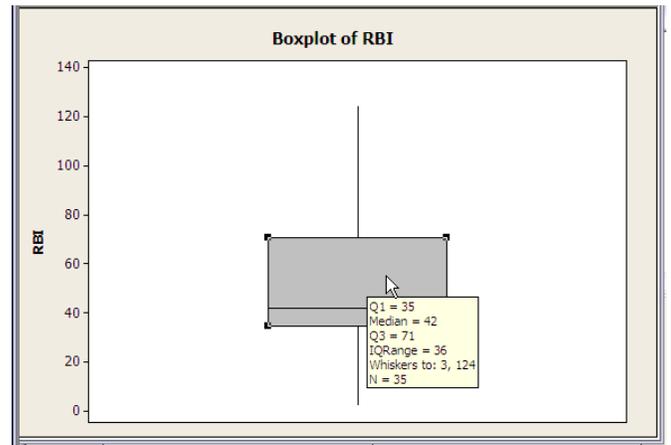
Another useful graph for analyzing this data is the boxplot. Select **Graph** → **Boxplot**. Select **Simple** and click **OK**, then select **RBI** as the **Graph variable** and click **OK** again.



The box represents the middle 50% of the data (interquartile range), the bottom and top of the box are Q1 and Q3 respectively, and the horizontal line in the center of the box is the median.

The vertical lines, called whiskers, indicate the spread of the first (lower line) and fourth (upper line) quartiles.

Click on the graph to obtain a summary of the information.

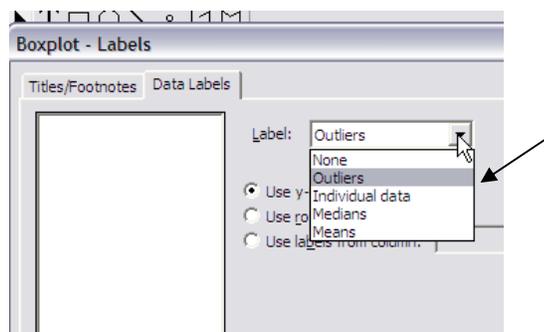
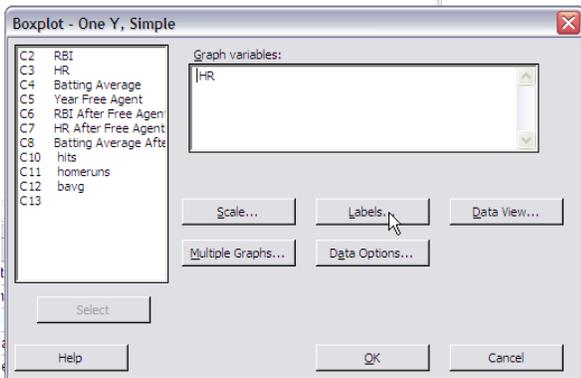
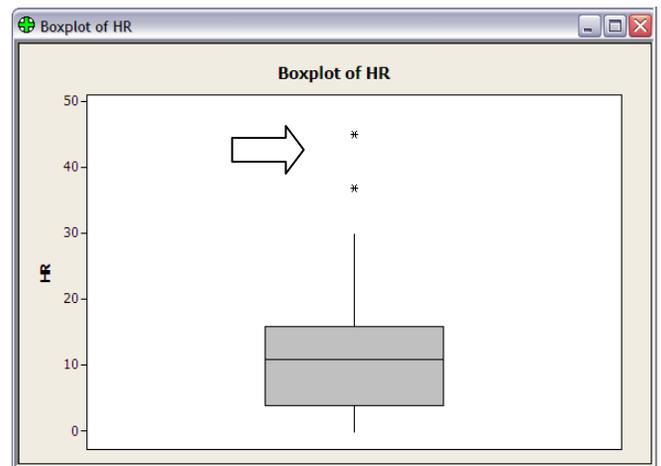


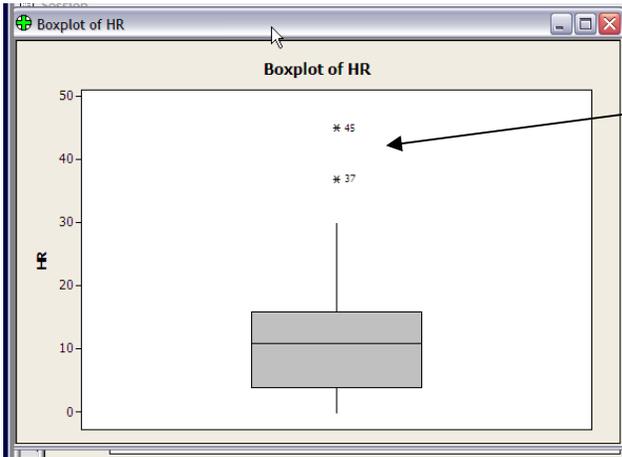
Create a **Boxplot** for home runs by selecting **C3 (HR)** as the **Graph variable**.

This boxplot has two *s, these represent the outliers.

To label the outliers, return to the boxplot window and select **Labels** → **Data Labels** → **Outliers**.

Then click **OK** and **OK**.

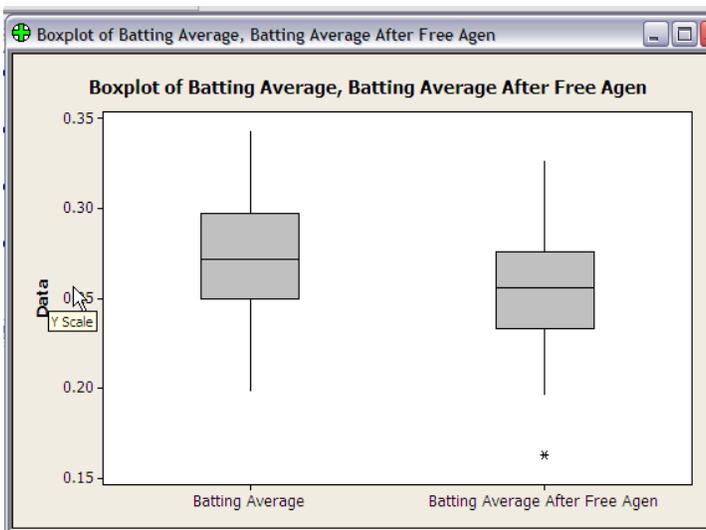
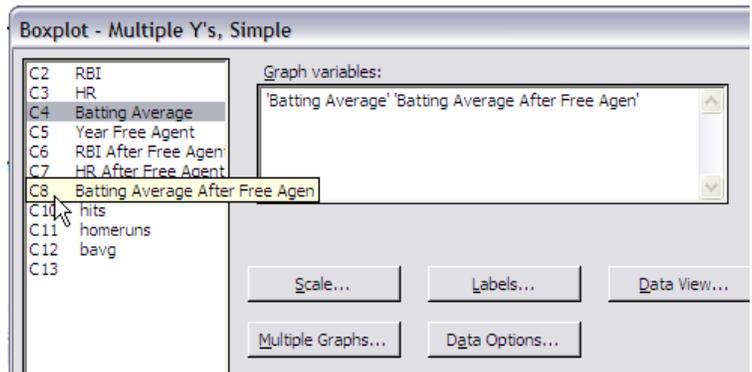
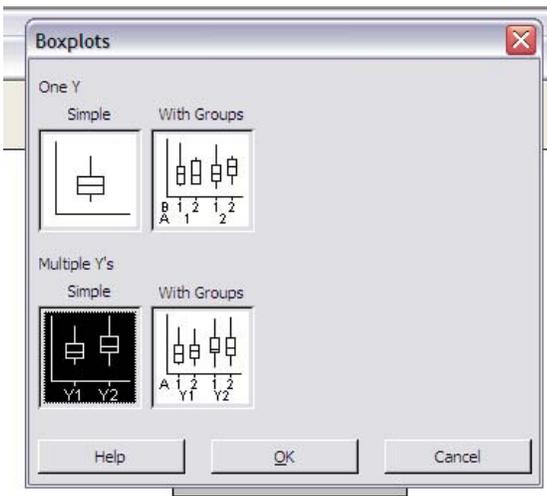




The outliers are now labeled.

Boxplots are also good for comparing two sets of data. To compare the batting averages before and after the players became free agents, create a new graph. **Graph** → **Boxplot**.

Select **Simple** under **Multiple Graphs** and click **OK**. Then select **C4** and **C8** as the **Graph variables** and click **OK** again.



Comparing the two, it appears that the players had better batting averages before they became free agents.

Hypothesis testing will determine if this difference is statistically significant or not.

Descriptive Statistics

Measures of the Centrality

Mode: Stat → Tables → Tally Individual Variables gives the count for each data value and hence the mode.

$$\text{Mode} = 15$$

Midrange: Stat → Tables → Tally Individual Variables gives both the maximum and minimum values and hence the midrange.

$$\text{Midrange} = \frac{45 + 0}{2} = 22.5$$

Median: Stat → Tables → Tally Individual Variables gives the count for each data value and the median can be determined.

$$\text{Median location} = \frac{35 + 1}{2} = 18^{\text{th}} \text{ position}$$

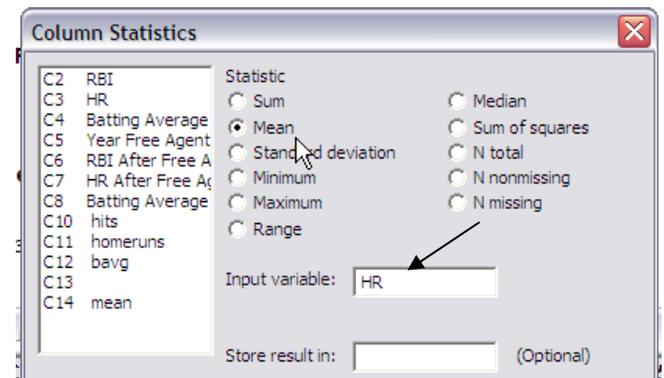
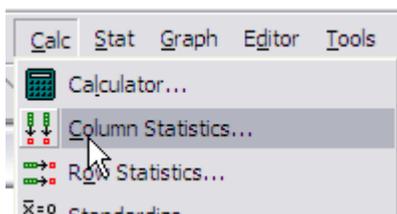
=> **Median = 11**

Tally for Discrete Variables: HR

HR	Count	CumCnt	
0	2	2	Minimum
1	3	5	
3	1	6	
4	3	9	
6	2	11	
7	3	14	
8	2	16	
10	1	17	
11	1	18	Median
13	1	19	
14	1	20	
15	4	24	Mode
16	3	27	
19	1	28	
23	1	29	
26	1	30	
28	2	32	
30	1	33	
37	1	34	
45	1	35	Maximum
N=	35		

Mean: The mean can be calculated several ways.

Calc → Column Statistics



Select the **Mean** and then **HR** as the **Input variable** column.

If the **Store result in** box is left blank the answer appears in the session window.

Session Window Printout

Mean of HR

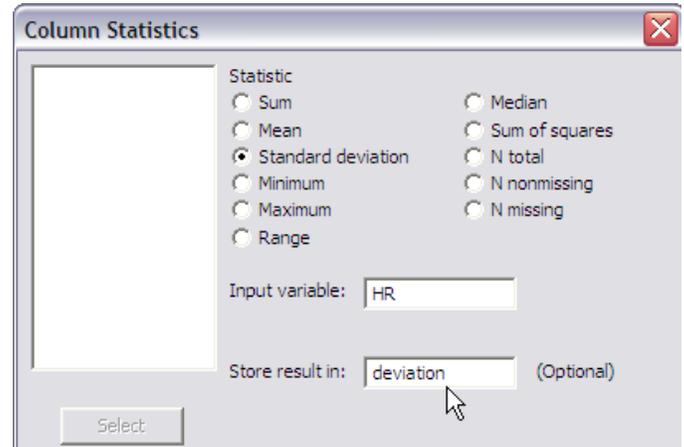
Mean of HR = 13.1143

If the mean is needed for other calculations then it should be stored in the first empty column by inputting **Mean** in the **Store result in** box.

Measures of Variation (the spread of the data)

Standard deviation: Calc → Column Statistics, select the **Standard deviation** and the **HR** as the **Input variable**.

The standard deviation is needed to calculate the variance, hence input **deviation** (or any other variable name) into the optional **Store result in** box.



The answer will appear in the session window, and will be stored as a variable, which will appear as an option when the calculator function is used in future examples.

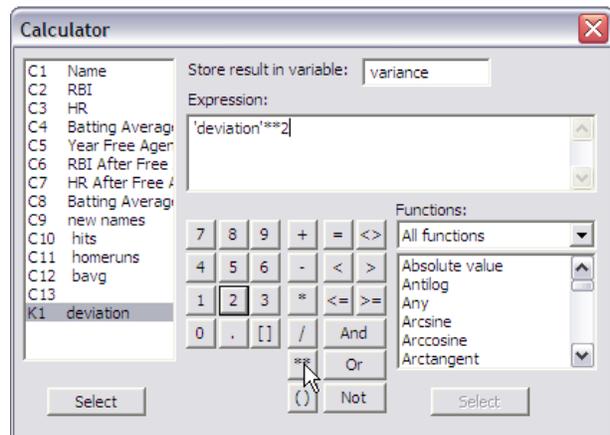
**Session Window Printout:
Standard Deviation of HR**

Standard deviation of HR = 10.9404

Variance: The variance is the square of the standard deviation.

Using **Calc → Calculator**, type any name for the **Store result in variable** (ex. variance). In the left hand column, deviation appears as the variable K1.

For **Expression**, select K1 then the ****** button (this indicates that the number is to be raised to a power) and **2** to indicate that the power is the second.

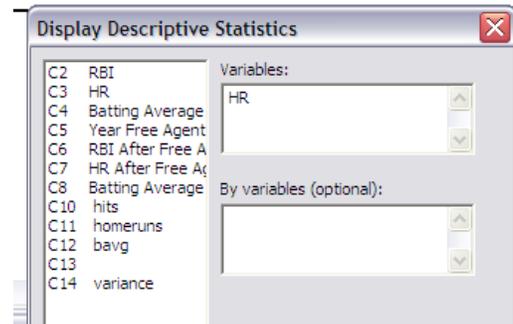
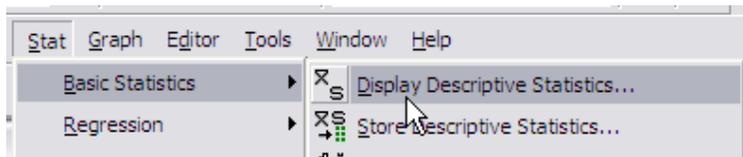


The result appears in the first empty column of the worksheet window.

C14
var
119.692

Alternatively, a data summary may be obtained by using:

Stat → Basic Statistics → Display Descriptive Statistics



Session Window Printout:

Descriptive Statistics: HR

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
HR	35	0	13.11	1.85	10.94	0.00	4.00	11.00	16.00	45.00

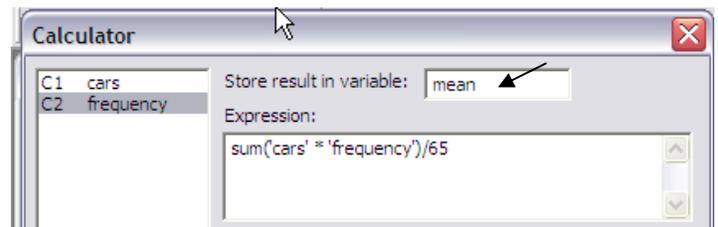
Calculations of Mean and Standard Deviation for Grouped Data

To calculate the mean and standard deviation for grouped data, the calculator feature is needed. Input the data from Chapter 2 practice in “Collaborative Statistics” by Illowski and Dean. (Note: All examples used in the following sections refer to exercises in “Collaborative Statistics.”)

# of cars	3	4	5	6	7
frequency	14	19	12	9	11

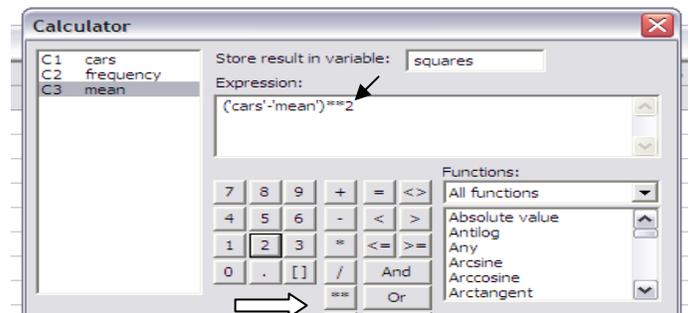
Calc → Calculator

The calculator requires a name in the **Store result in variable**. Type in **mean**.



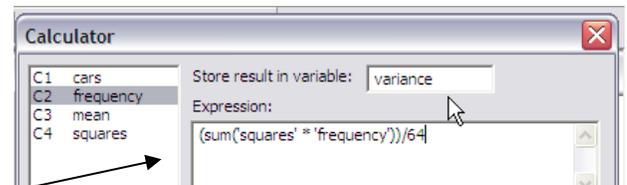
To input the formula for mean, $\bar{x} = \frac{\sum x \cdot f}{n}$,

the **Expression** box type: **sum(select C1 ‘cars’ * select C2 ‘frequency’)/65**.



The answer will appear in C3 of the worksheet.

To calculate the variance and standard deviation, first compute the square of the differences between the data values and the mean. In the **Expression** box type: **(select C1 ‘cars’-select C3 ‘mean’)**2**



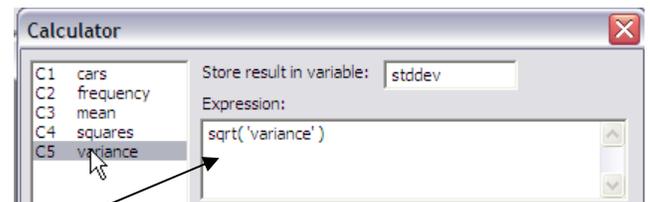
Note: The ** button is used not ^ to indicate a power.

The answer is in C4 of the worksheet.

The next step is to calculate the variance.

$$s^2 = \frac{\sum (x - \bar{x}) \cdot f}{n - 1}$$

Input **(sum(select C4 ‘squares’ * select C2 ‘frequency’)/64)**. Note the two sets of ()s.



Answer screen

The answer is in C5 of the worksheet.

The standard deviation is the square root of the variance. Input: **sqrt(‘variance’)**

The answer appears in C6 of the worksheet.

	C1	C2	C3	C4	C5	C6
	cars	frequency	mean	squares	variance	stddev
1	3	14	4.75385	3.07598	1.93846	1.39229
2	4	19		0.56828		
3	5	12		0.06059		
4	6	9		1.55290		
5	7	11		5.04521		
6						
7						

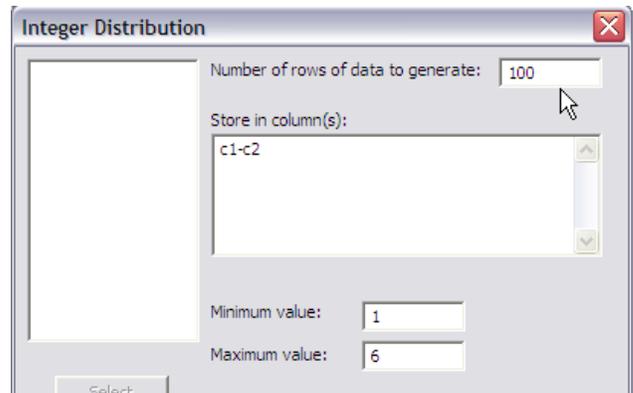
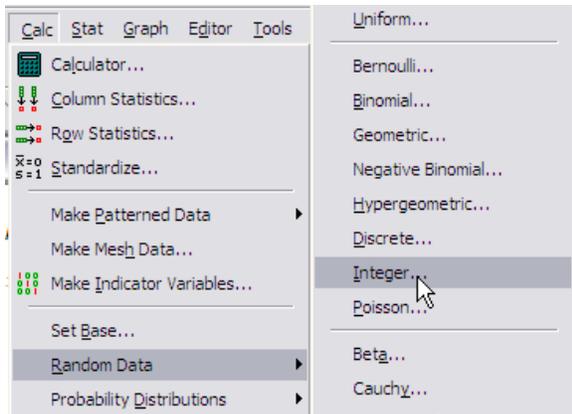
Simulations

Minitab generates sets of data from a variety of distributions, which may be used to simulate experiments.

Discrete Uniform Distribution

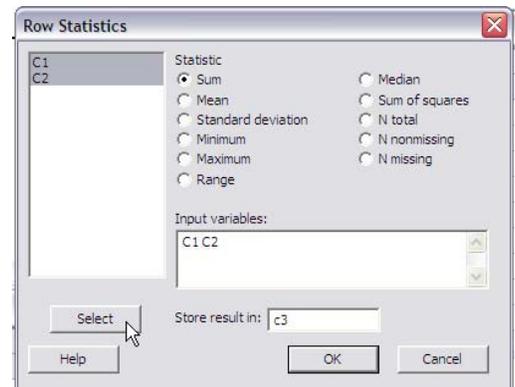
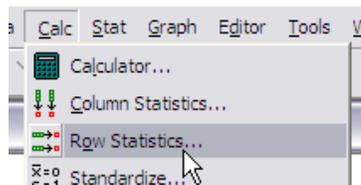
Integer function is used in situations where the outcomes are all equally likely, such as rolling dice or tossing a coin (heads = 0, tails = 1).

To simulate the tossing of two dice 100 times go to **Calc** → **Random Data** → **Integer**. For 100 tosses, generate 100 rows and store the results in columns C1 and C2. The number of rows represents the number of tosses and each column represents a die. Again the maximum value is 6 and the minimum is 1.



	C1	C2
1	5	4
2	4	4
3	1	2
4	1	5
5	2	3
6	4	1
7	1	2
8	3	4
9	5	2

To obtain the sum of the dice, use **Calc** → **Row Statistics**



Select **Sum** and **Input variables C1 and C2**.

Store result in C3.

	C1	C2	C3
	Die 1	Die 2	Sum
1	5	4	9
2	4	4	8
3	1	2	3
4	1	5	6
5	2	3	5
6	4	1	5
7	1	2	3
8	3	4	7

The columns may be named to represent the values they contain.

To investigate this data use

Stat → **Tables** → **Tally Individual Variables**

Tally for Discrete Variables: Sum

Sum	Count
2	2
3	9
4	4
5	13
6	15
7	14
8	8
9	14
10	12
11	7
12	2
N=	100

Discrete Random Variables and Their Probability Distributions

Binomial

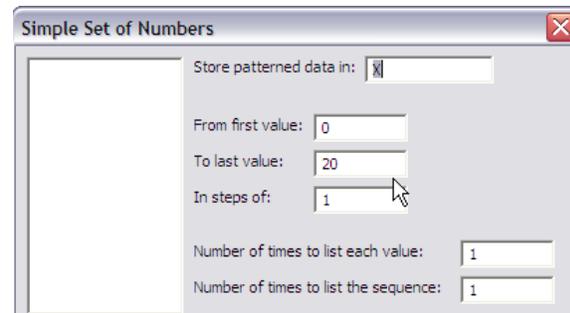
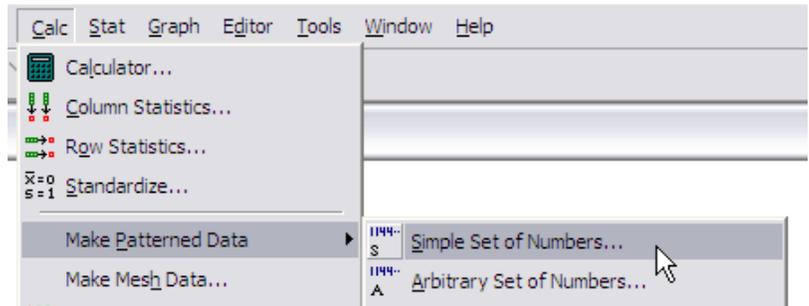
Example 4-10
 $X \sim B(20, 0.41)$

To create a probability table for a binomial distribution, begin by creating a column with all possible values for X .

Calc → Make Patterned Data → Simple Set of Numbers

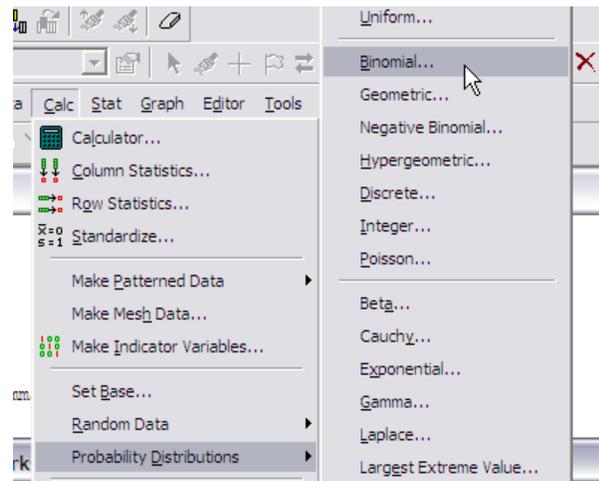
Since $n = 20$, $X = \{0, 1, \dots, 20\}$. Hence the first value is 0 and the last is 20.

The values for X appear in a column on the worksheet.



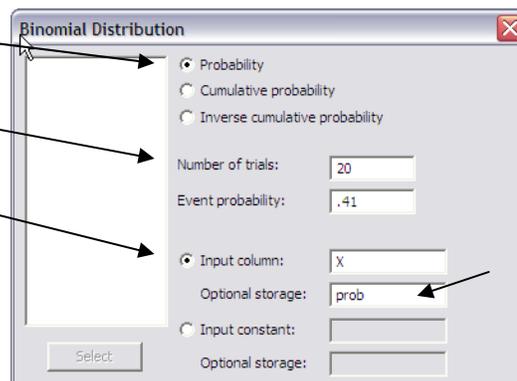
	C1
	X
1	0
2	1
3	2
4	3
5	4
6	5
7	6
8	7

Go to Calc → Probability Distributions → Binomial



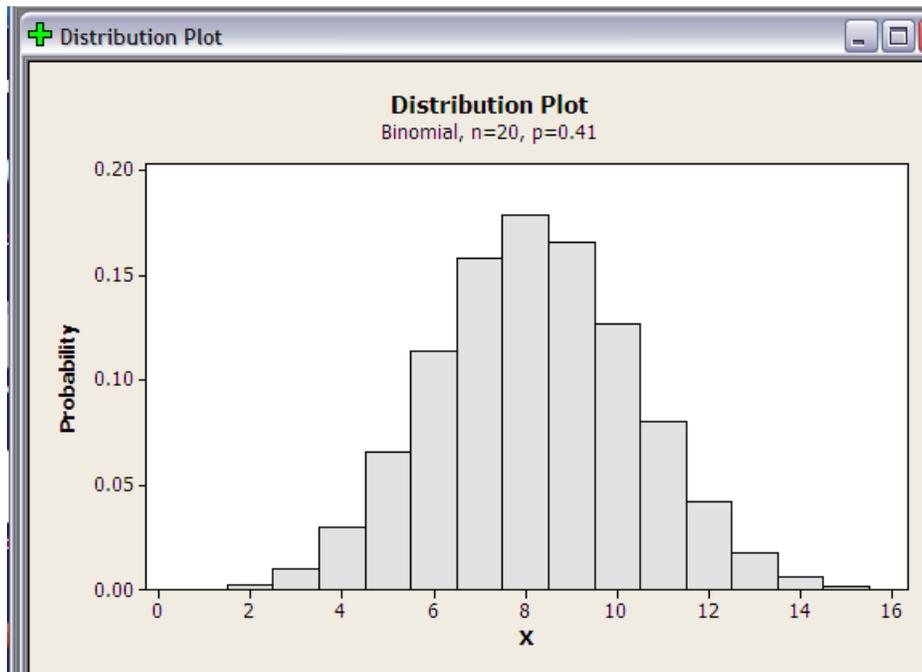
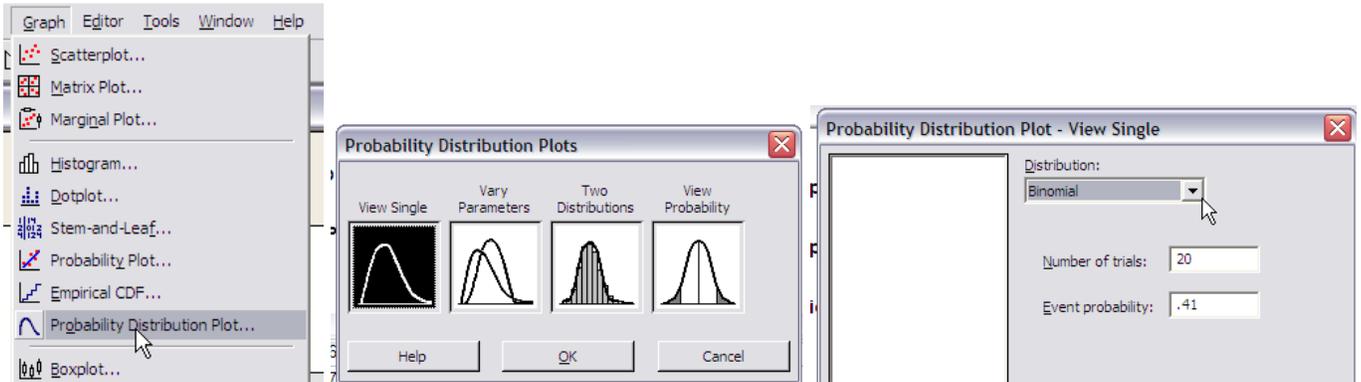
Select **Probability**, input the **Number of trials** and the **Event probability (p)**.

Select **Input column** and name the **Optional storage**. Inputting a name for the **Optional storage** places the probabilities in the worksheet with an accuracy of 6 decimal places.



	X	prob
1	0	0.000026
2	1	0.000363
3	2	0.002397
4	3	0.009994
5	4	0.029516
6	5	0.065636
7	6	0.114029
8	7	0.158481
9	8	0.178963
10	9	0.165818
11	10	0.126753
12	11	0.080075
13	12	0.041734

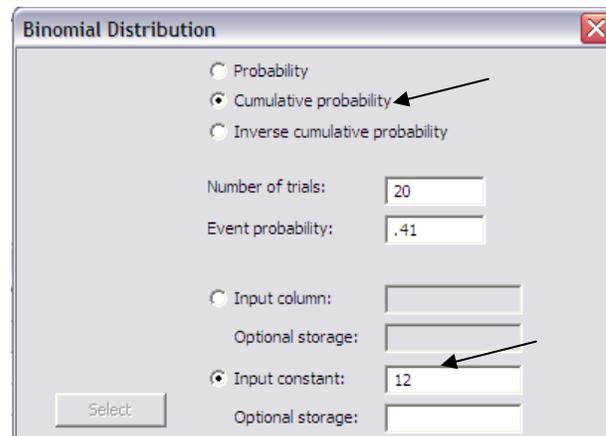
To graph the probability distribution, **Graph** → **Probability Distribution Plot**, **View Single**, select **Binomial** from the **Distribution** list and input the **Number of trial** and the **Event probability**.



To determine $P(X \leq 12)$, the probabilities in the generated table for ($X = 0$ to 12) may be summed.

Another option is to return to the **Calc** → **Probability Distributions** → **Binomial** and select **Cumulative probability** and **Input constant**.

The **Input constant** is 12 since Minitab gives the cumulative probability up to and including the input constant value.



The result is printed in the session window.

Note: the summary gives the distribution, n, p and x values. Always check this information for any possible input errors.

To determine the $P(X > 12)$, remember that this is the same as $1 - P(X \leq 12)$
 $\Rightarrow 1 - .973785 = .026215$

Session Window Printout:

Cumulative Distribution Function

Binomial with n = 20 and p = 0.41

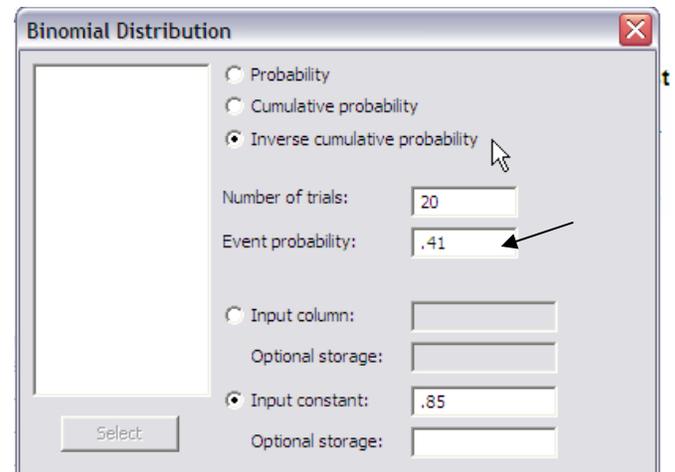
x	P(X <= x)
12	0.973785

To determine $P(X < 12)$, rewrite the expression with \leq in the new expression. In other words,

$$P(X < 12) = P(X \leq 11)$$

The **Input constant** would be **11** since that is the largest value X would take.

To determine the data value such that the probability of that value or fewer is 85%, return to the **Binomial Distribution** window and select **Inverse cumulative probability** and **Input constant**. The input constant is now the probability (.85) and the result will be the data value.



Note: Since 85% is not possible for this discrete distribution, the two closest values are given.

To ensure that the probability is at least 85%, then $x = 10$.

Session Window Printout

Inverse Cumulative Distribution Function

Binomial with n = 20 and p = 0.41

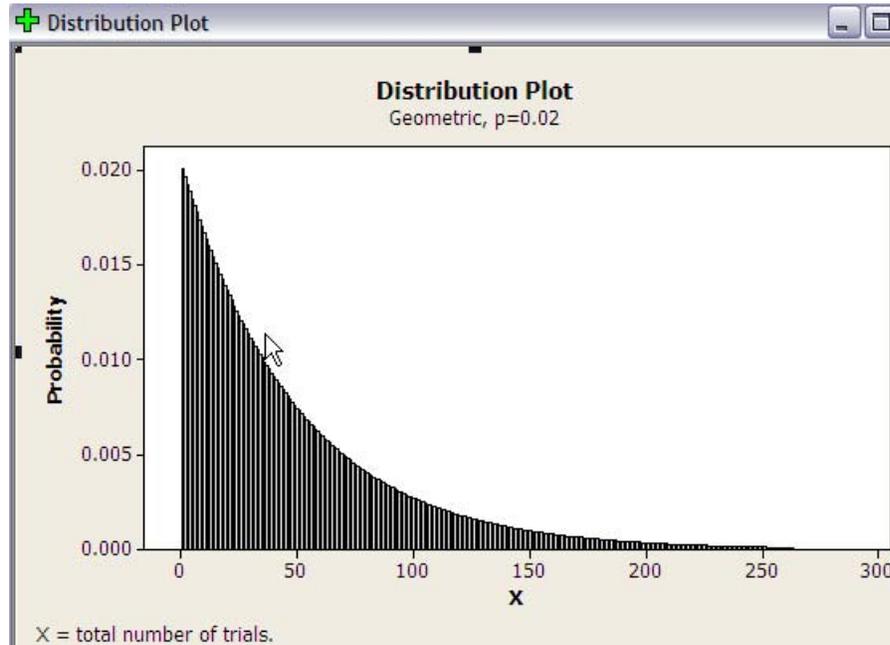
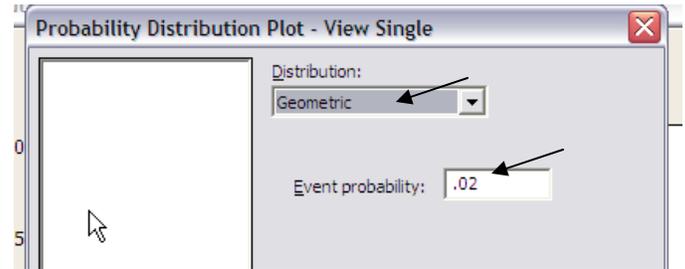
x	P(X <= x)	x	P(X <= x)
9	0.725223	10	0.851976

Geometric
Example 4-15
 $X \sim G(0.02)$

A graph for the Geometric function is created in the same manner as for the Binomial.

Calc → Probability Distributions

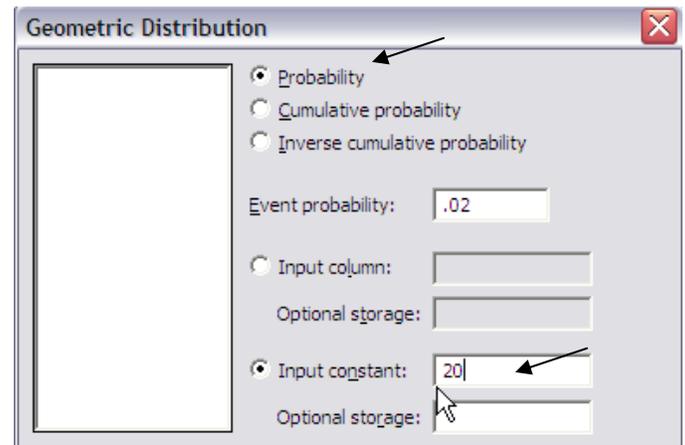
Select **Geometric** as the distribution and input the **Event probability**.



To determine the probability of a particular value, for example $P(X = 6)$, select

Calc → Probability Distributions
→ Geometric

Use the **Probability** option, and **Input constant 20**.



Session Window Printout:
Cumulative Distribution Function

Geometric with $p = 0.02$

x	$P(X \leq x)$
20	0.332392

* NOTE * X = total number of trials.

Continuous Random Variables and Their Probability Distributions

Continuous random variables and their probabilities are found in the same way as discrete random variable.

Exponential Distribution

Example 5-5

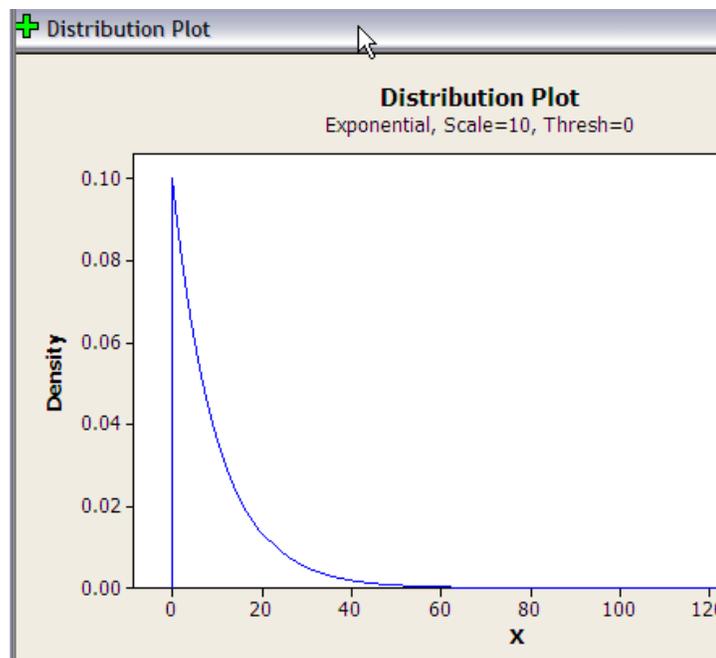
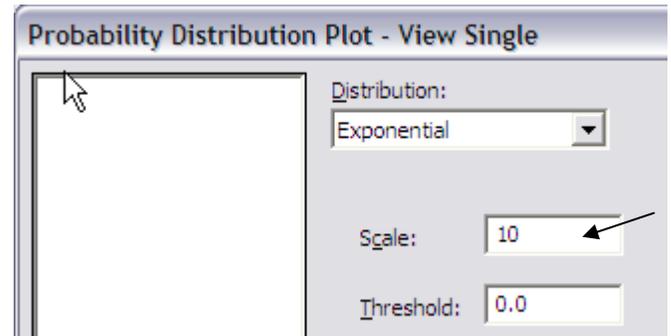
$X \sim \text{Exp}(0.1)$

Note: $\mu = 10$, hence $m = 0.1$

A graph for the Exponential function is created in the same manner as for the Binomial.

Graph → **Probability Distribution Plot**

View Single, select **Exponential** from the distribution list and input the **Mean** where it asks for the **Scale**.



To calculate $P(X > 7) = 1 - P(X < 7)$

Calc → **Probability Distributions**
→ **Exponential**

Select **Cumulative probability**, input the mean (Scale) and **Input constant 7**.

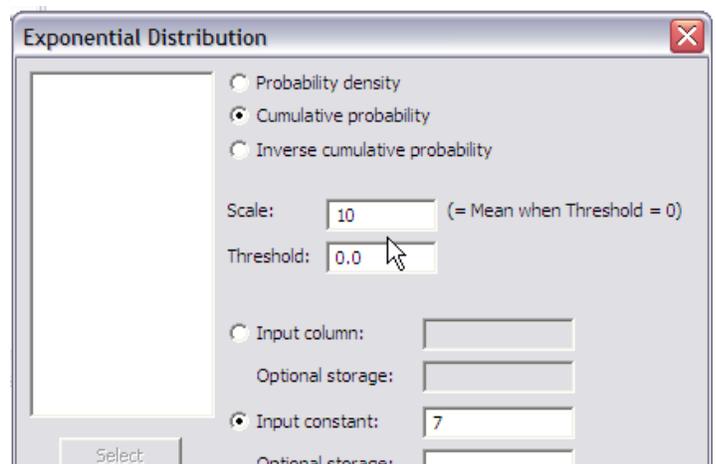
Session Window Printout
Cumulative Distribution Function

Exponential with mean = 10

x $P(X \leq x)$

7 0.503415

$P(X > 7) = 1 - 0.503415 = 0.496584$

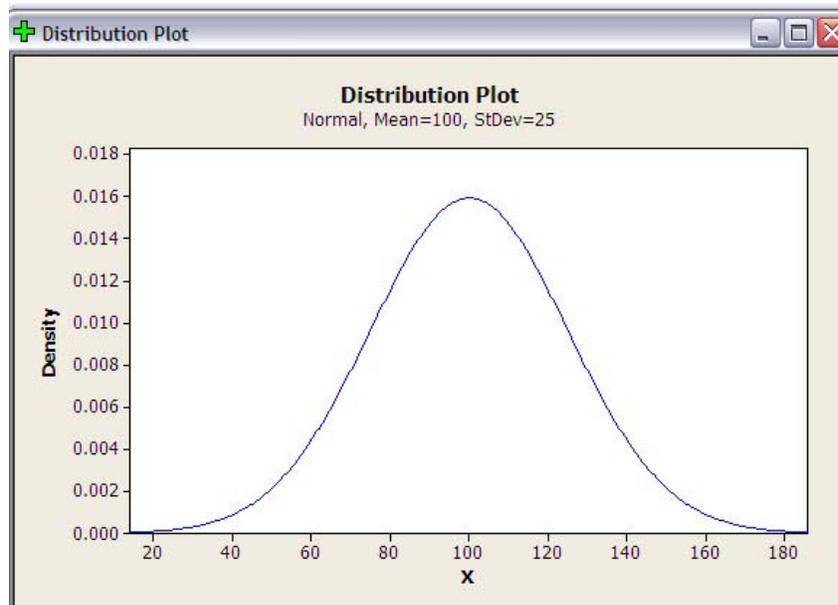
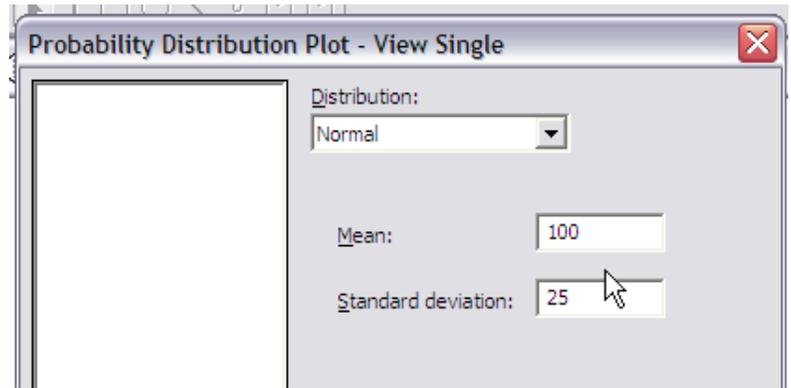


Normal Distribution

$$X \sim N(100, 25)$$

A graph for the **Normal** function is created in the same manner as for the **Exponential**.

Graph → **Probability Distribution Plot View Single**, select **Normal** from the distribution list and input the **Mean** and **Standard deviation**.



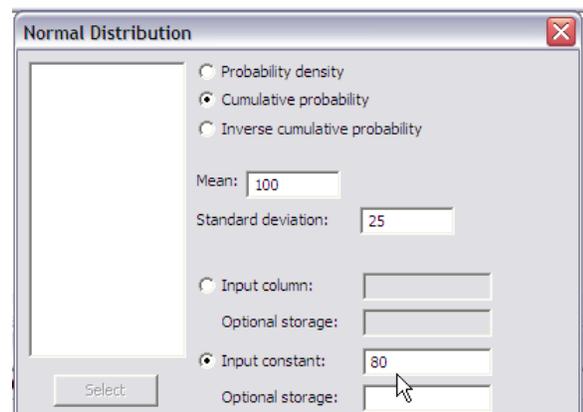
Ex. 1

Calculate $P(80 < X < 130)$

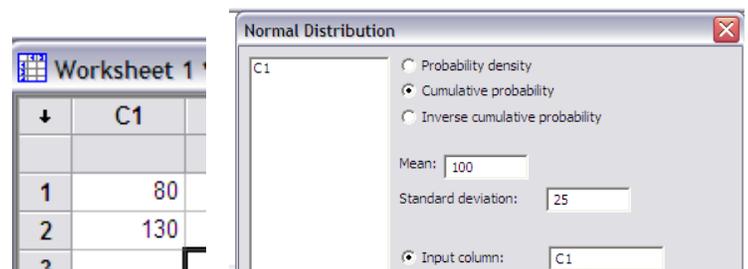
Calc → **Probability Distributions**
→ **Normal**

Select **Cumulative probability**, input the **Mean** and **Standard deviation**.

Input constant 80 and then repeat for **130**.



Alternately, input the values 80 and 130 into a column on the worksheet and use **Input column**.



Session Window Printout
Cumulative Distribution Function

Normal with mean = 100 and standard deviation = 25

x	P(X <= x)
80	0.211855
130	0.884930

Hence

$$P(80 < X < 130) = P(130) - P(80)$$
$$= 0.884930 - 0.211855 = 0.673075$$

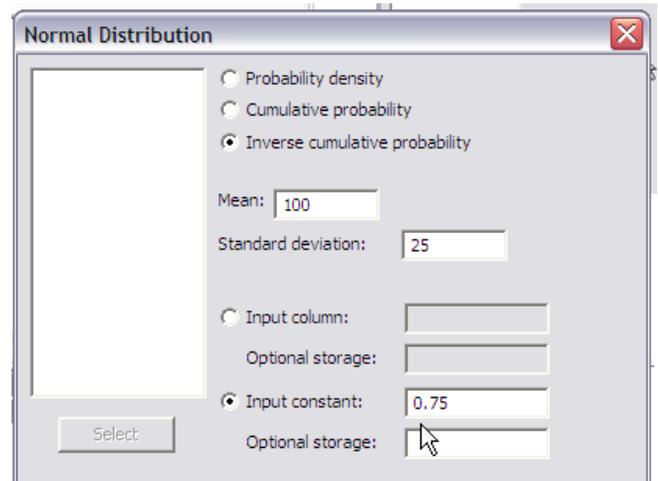
Ex. 2 Find x such that $P(X < x) = 75\%$

Calc → **Probability Distributions** → **Normal**

Select **Inverse cumulative probability**, input the **Mean** and **Standard deviation**.

Input constant is **0.75**.

Note: The input constant for probabilities must be input as the decimal form.



Session Window Printout:

Inverse Cumulative Distribution Function

Normal with mean = 100 and standard deviation = 25

P(X <= x)	x
0.75	116.862

Hence, $P(X < 116.862) = 75\%$

Confidence Intervals

Single Population Mean, Population Standard Deviation given, Normal Distribution

Example 8 – 2

Given $\bar{x} = 68$, $\sigma = 3$, $n = 36$; find a 90% confidence interval for the population mean.

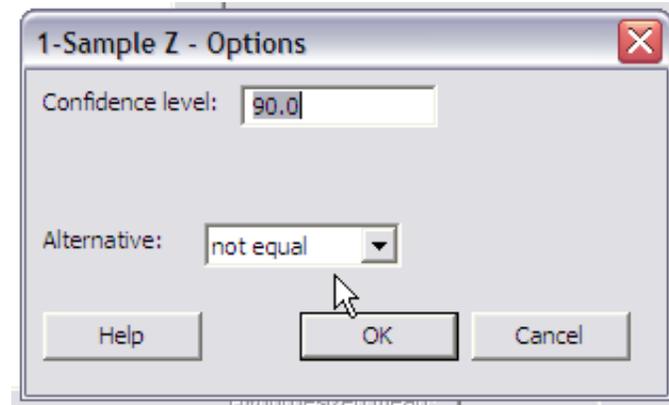
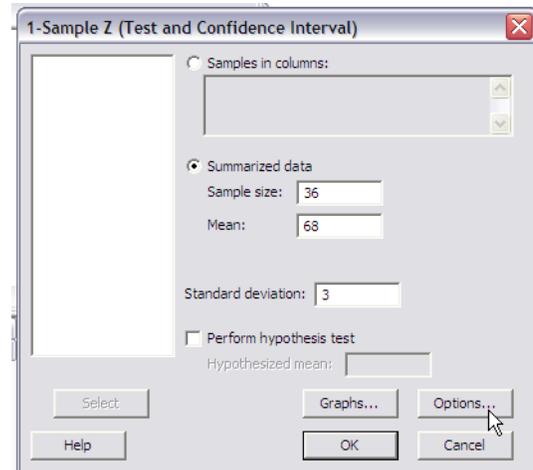
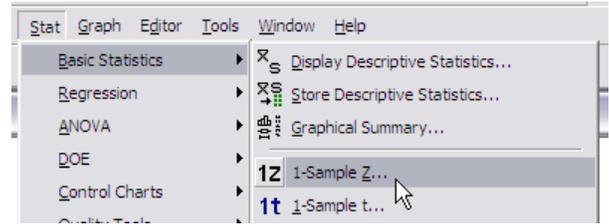
Stat → Basic Statistics → 1-Sample Z

Select **Summarized data**, input **Sample size**, **Mean** and population **Standard deviation**.

Select **Options**

Input the desired percentage in **confidence level** and ensure that the **Alternative** is **not equal**. **OK** → **OK**

The result is in the session window.



Session Window Printout

One-Sample Z

The assumed standard deviation = 3

N	Mean	SE Mean	90% CI
36	68.000	0.500	(67.178, 68.822)

$$SE\ Mean = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.500$$

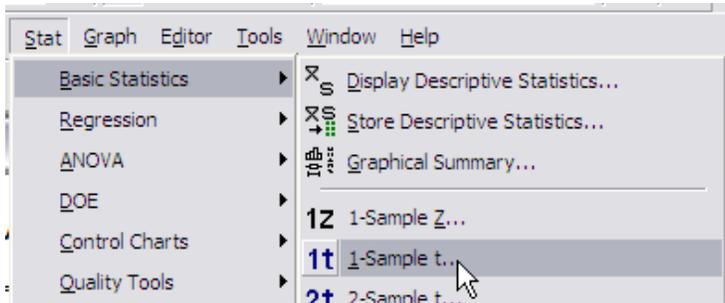
The 90% confidence interval for this example is **(67.178, 68.822)**.

Single Population Mean, Population Standard Deviation Unknown, Normal Distribution

Input the following data from the example 8 - 4 into a column in the worksheet:

8.6, 9.4, 7.9, 6.8, 8.3, 7.3, 9.2, 9.6, 8.7, 11.4, 10.3, 5.4, 8.1, 5.5, 6.9

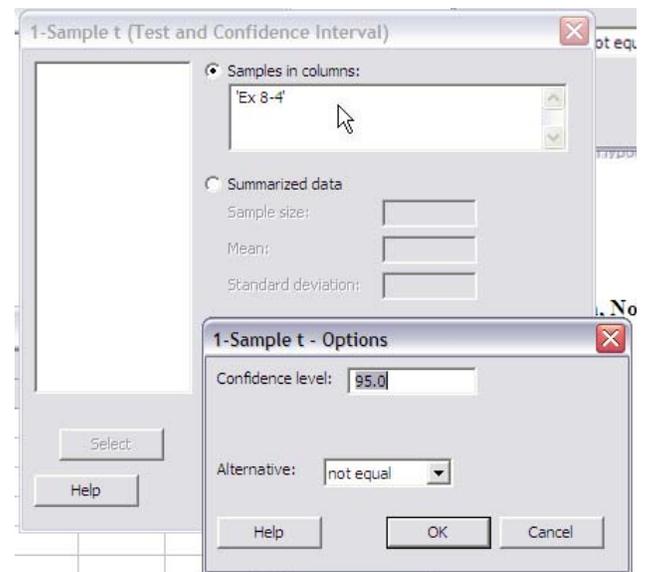
Then go to **Stat → Basic Statistics → 1-Sample t**



Since the raw data is given, select **Samples in columns** and the column with the data.

Again choose **Options** and ensure that the **Confidence level** is correct and that the **Alternative** option is **not equal**. **OK → OK**

The summary information includes the sample mean and sample deviation as well and the confidence interval.



Session Window Printout One-Sample T: Ex 8-4

Variable	N	Mean	StDev	SE Mean	95% CI
Ex 8-4	15	8.227	1.672	0.432	(7.301, 9.153)

Single Population Proportion

The underlying distribution for a proportion is binomial.

Example 8 – 5

$n = 500, x = 421$

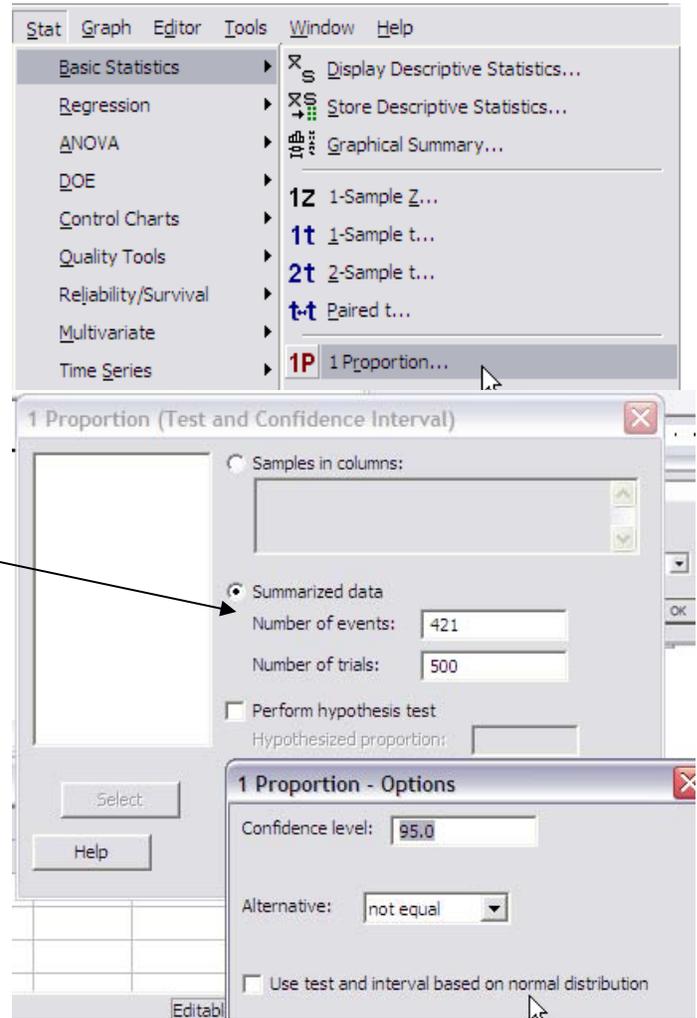
Find a 95% confidence interval for the population proportion.

Stat → Basic Statistics → 1-Proportion

Select **Summarized data**. Input the **number of successes** in the **Number of events** box.

Input the **Number of trials**. Use **Options** to ensure that 95 is the **Confidence level** and **Alternative** is **not equal**.

OK → OK



Session Window Printout Test and CI for One Proportion

Sample	X	N	Sample p or p'
1	421	500	0.842000

95% CI
(0.807022, 0.872869)

Hypothesis Testing

Test for Single Mean with Population Standard Deviation Known

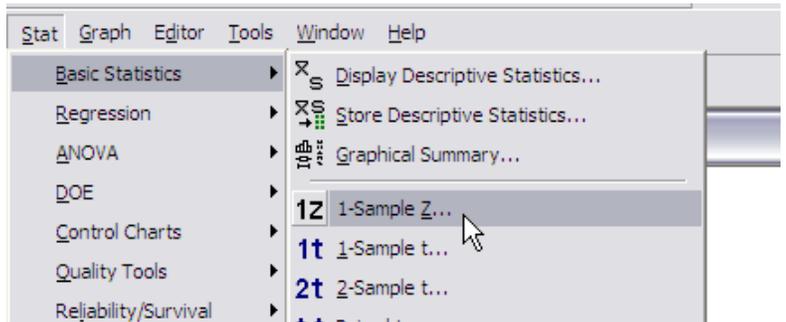
Example 9 – 11

$$H_0: \mu \geq 16.43 \quad H_a: \mu < 16.43$$

$$\bar{x} = 16, \sigma = 0.8, n = 15 \text{ and } \alpha = 0.05$$

To access the hypothesis testing options, return to the same screen as used for confidence intervals.

Stat → Basic Statistics → 1-Sample z



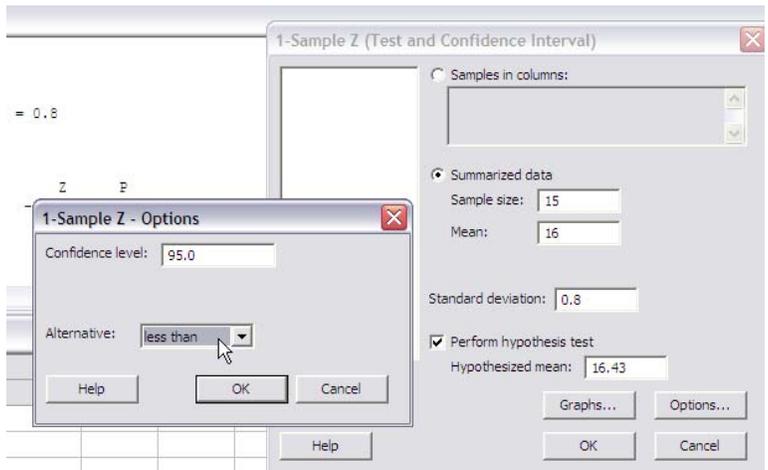
Check **Perform hypothesis test**.

Input the **Hypothesized mean**.

Click **Options**.

It is not necessary to enter a **Confidence level**, although one may appear by default. Ensure that **Alternative: less than** is selected.

OK → OK



Session Window Printout One-Sample Z

Test of $\mu = 16.43$ vs < 16.43
The assumed standard deviation = 0.8

95% Upper					
N	Mean	SE Mean	Bound	Z	P
15	16.000	0.207	16.340	-2.08	0.019

p-value $< \alpha$, hence reject H_0

There is significant statistical evidence to support the claim that the goggles helped Jimmy's swim time.

NOTE: The result in the session window includes the hypothesis being tested, indicating that a left tail test is used. The sample mean and standard deviation are given. In addition, both the Z_{test} value (Z) and the p-value are given.

Test for Single Mean with the Population Standard Deviation Unknown

Example 9 – 13

$$H_0: \mu \leq 65 \quad H_a: \mu > 65$$

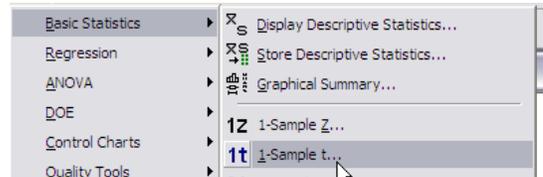
Data values: 65, 65, 70, 67, 66, 63, 63, 68, 72 and 71

$$\alpha = 0.05$$

Input the data values into a column on the worksheet.

Access the hypothesis test

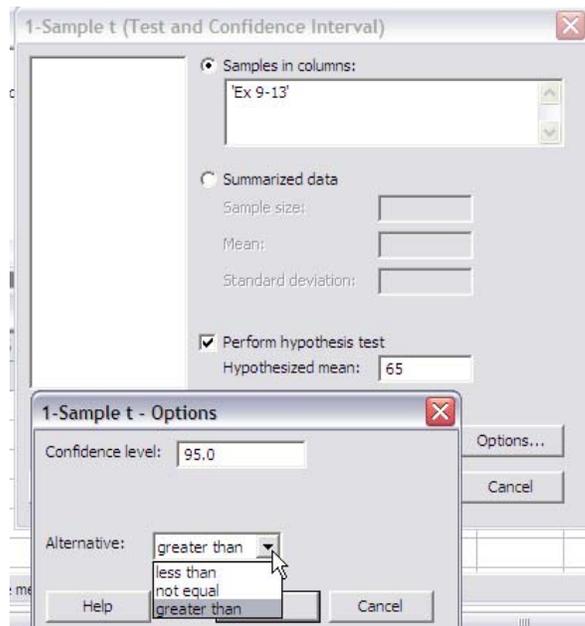
Stat → Basic Statistics → 1-Sample t



Select the column with the data. Check **Perform hypothesis test**. Input the **Hypothesized mean**.

Click **Options**. The **Confidence level** is not necessary. Ensure that **Alternative: greater than** is selected.

OK → OK



Session Window Printout

One-Sample T: Ex 9-13

Test of $\mu = 65$ vs > 65

Variable	N	Mean	StDev	SE Mean	95% Lower Bound
Ex 9-13	10	67.00	3.20	1.01	65.15

T (t _{test})	P
1.98	0.040

Since the p-value $< \alpha \Rightarrow$ reject H_0

There is significant statistical to support the claim that the average test score is greater than 65.

Test for Single Population Proportion

Example 9 – 14

$$H_0: p = 0.5 \quad H_a$$

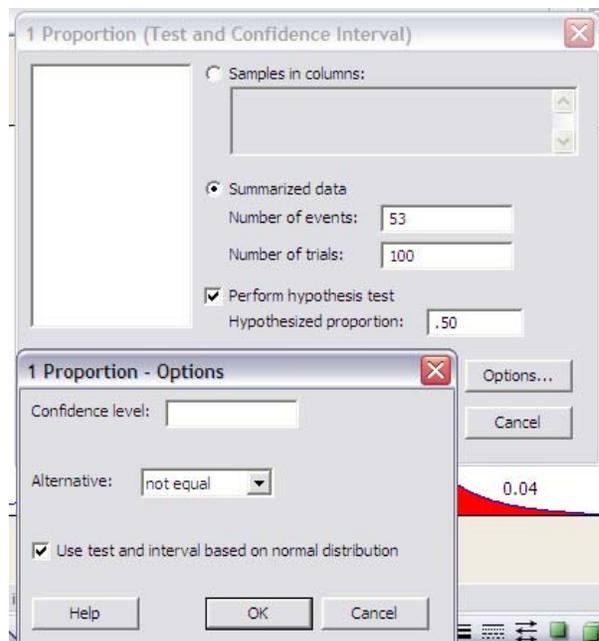
$$p \neq 0.5 \quad x = 53, n = 100 \quad \alpha = 0.01$$

To access the hypothesis testing options, return to the same screen as used for confidence intervals.

**Stat → Basic Statistics
→ 1-Proportion**

Input the given values. Check the **Perform hypothesis test** and input the **Hypothesized proportion**

Select **Options** to ensure that the **Alternative** is **not equal**.



The result gives the hypothesis tested, indicating that a two tail test is used. In addition, a 95% confidence interval, the Z_{test} value and the p-value are given.

Test and CI for One Proportion

Test of $p = 0.5$ vs $p \text{ not } = 0.5$

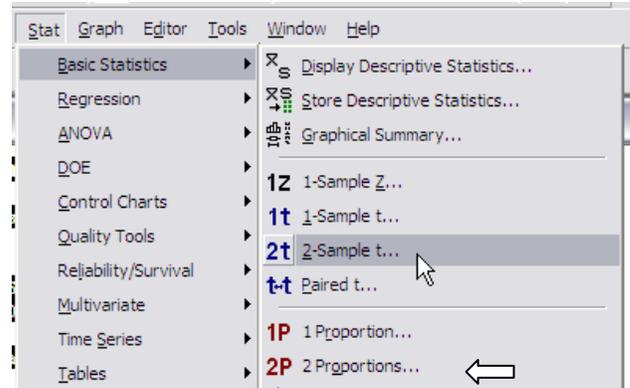
Sample	X	N	Sample p	95% CI	Z-Value	P-Value
1	53	100	0.530000	(0.432178, 0.627822)	0.60	0.549

Using the normal approximation.

Tests for Two Mean or Two Proportions

Hypothesis testing for two means or two proportions is done using the same methods as for a single mean.

Under **Stats** → **Basics Statistics** select the appropriate test: 2-Sample t, Paired t, or 2 Proportions.



Two Independent Means

Example 10-3

$$H_0: \mu_1 \leq \mu_2 \quad H_a: \mu_1 > \mu_2$$

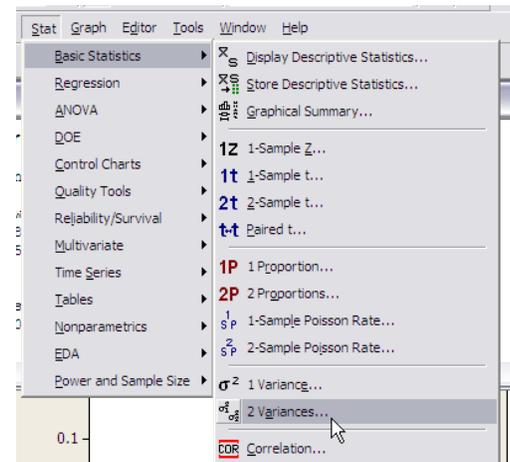
$$\bar{x}_1 = 3; s_1 = 0.33; n_1 = 20;$$

$$\bar{x}_2 = 2.9; s_2 = 0.36; n_2 = 20$$

First test for equal variances

$$H_0: \sigma^2_1 = \sigma^2_2 \quad H_a: \sigma^2_1 \neq \sigma^2_2$$

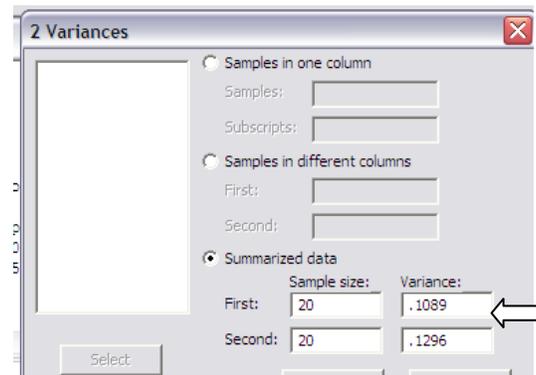
Stats → **Basic Statistics** → **2 Variances**



Input the **Variance** (standard deviation squared) for each data set. When running this test several answers are given. Using only the results from the F-Test (assuming both data sets are from normal distributions), the $p\text{-value} > \alpha$, hence fail to reject the null. There is significant statistical evidence to assume the variances are equal.

F-Test (Normal Distribution)

Test statistic = 0.84, p-value = 0.708



Select **Stats** → **Basic Statistics** → **2-Sample t** and input the required summarized data. Check **Assume equal variances**.

Click **Options** and ensure the **Alternative** is **greater than**

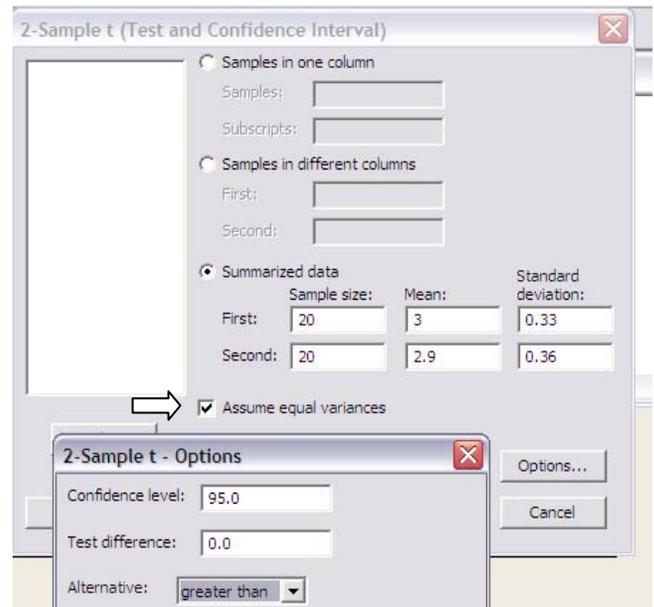
Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
1	20	3.000	0.330	0.074
2	20	2.900	0.360	0.080

Difference = $\mu_1 - \mu_2$
 Estimate for difference: 0.100
 95% lower bound for difference: -0.084

T-Test of difference = 0 (vs >):
 T-Value = 0.92 P-Value = 0.183 DF = 38
 Both use Pooled StDev = 0.345

p-value > α => fail to reject H_0



There is not significant statistical evidence to support the claim that wax 1 lasts longer than wax 2.

Suppose that the standard deviation for the second sample of wax is 0.56 not 0.36, and repeat the test.

Ex 10-3

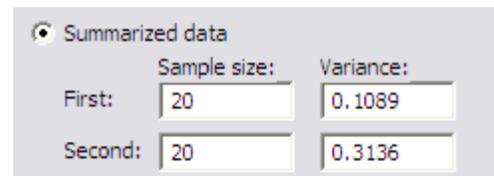
$H_0: \mu_1 \leq \mu_2$ $H_a: \mu_1 > \mu_2$
 $\bar{x}_1 = 3; s_1 = 0.33; n_1 = 20;$ $\bar{x}_2 = 2.9; s_2 = 0.56; n_2 = 20$

First test for equal variances
 $H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \neq \sigma_2^2$

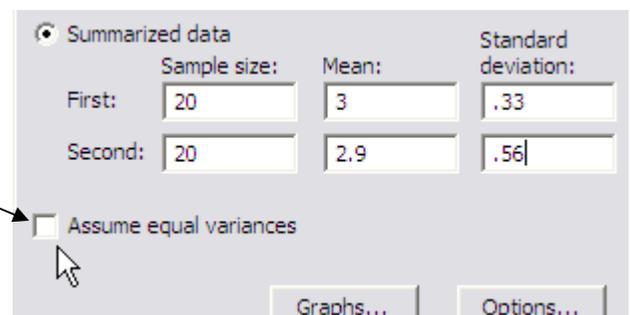
Stats → **Basic Statistics** → **2 Variances**

Input the **Variance** for each data set.

Using only the F-Test (assuming both data sets are from normal distributions), the p-value < α , hence reject the null. There is significant statistical evidence to assume the variances are **not** equal, therefore the **Assume equal variance** box should **NOT** be checked.



F-Test (Normal Distribution)
Test statistic = 0.35, p-value = 0.026



Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
1	20	3.000	0.330	0.074
2	20	2.900	0.560	0.13

Difference = $\mu(1) - \mu(2)$

Estimate for difference: 0.100

95% lower bound for difference: -0.147

T-Test of difference = 0 (vs >): T-Value = 0.69 P-Value = 0.248 DF = 30

The p-value is $> \alpha$, hence fail to reject H_0 .

There is not significant statistical evidence to support the claim that wax 1 lasts longer than wax 2.

Two Sample Means – Paired Data

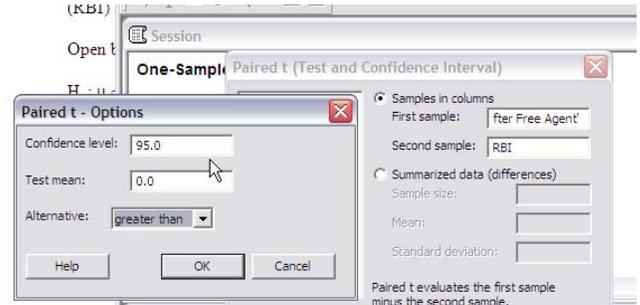
Do baseball players have more runs batted in (RBI) after they become free agents? In other words, do the players perform better after they obtain big salaries?

Open **baseball-update.mtw**.

$H_0: \mu_{\text{free agent}} \leq \mu_{\text{RBI}}$ $H_a: \mu_{\text{free agent}} > \mu_{\text{RBI}}$

Select **Stats** → **Basic Statistics** → **Paired t** and input the free agent column in as the first sample and the RBI column as the second.

Note that the first sample is always the mean on the left of the inequality and the second sample is on the right hand side.



Session Window Printout

Paired T-Test and CI: RBI After Free Agent, RBI

Paired T for RBI After Free Agent - RBI

	N	Mean	StDev	SE Mean
RBI After Free Agent	35	39.94	29.23	4.94
RBI	35	51.06	30.63	5.18
Difference	35	-11.11	31.66	5.35

95% lower bound for mean difference: -20.16

T-Test of mean difference = 0 (vs > 0): T-Value = -2.08 P-Value = 0.977

There is NOT significant statistical evidence to support the claim that players perform better after they become free agents.

Two Sample Proportion Test

Example 10-4

$$H_0: p_A = p_B \quad H_a: p_A \neq p_B$$

Medication A $x_A = 20, n_A = 200$

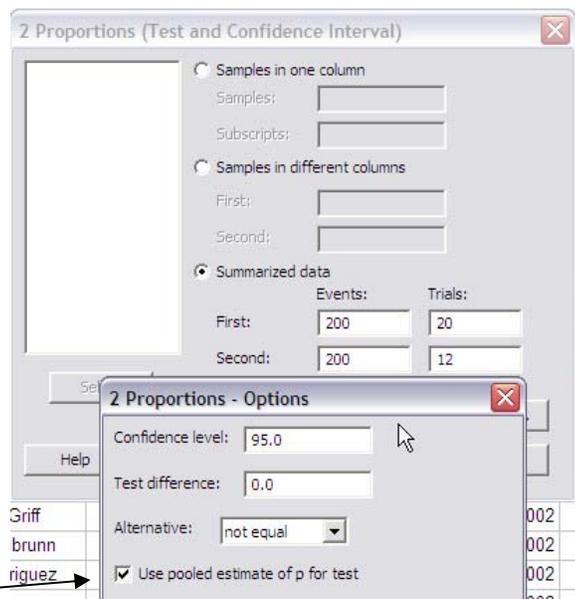
Medication B $x_B = 12, n_B = 200 \quad \alpha = 1\%$

Stats → Basic Statistics → 2 Proportions

Input the **x value** in the **Trials** box and **n** in the **Events** box.

Select **Options** and ensure that the **Alternative** is **not equal**.

Check the **Use pooled estimate of p for test** box.



Session Window Printout

Test and CI for Two Proportions

Sample	X	N	Sample p
1	20	200	0.100000
2	12	200	0.060000

Difference = p (1) - p (2)

Estimate for difference: 0.04

95% CI for difference: (-0.0130278, 0.0930278)

Test for difference = 0 (vs not = 0): Z = 1.47 P-Value = 0.140

There is NOT significant statistical evidence to support the claim that there is a difference between the two medications.

Chi-Squared Goodness of Fit Test

Example 11-4

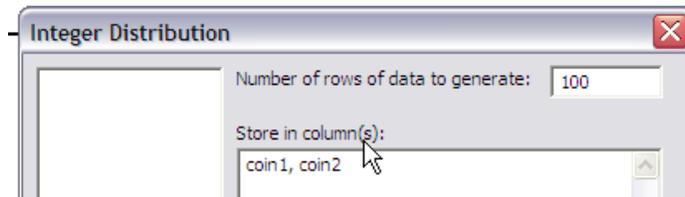
To test if two coins are fair, toss them 100 times and determine the number of heads.

Simulate this by using:

Calc → Random Data → Integer

Input 100, **number of trials**, as the **number of rows of data to generate**.

OK -> OK



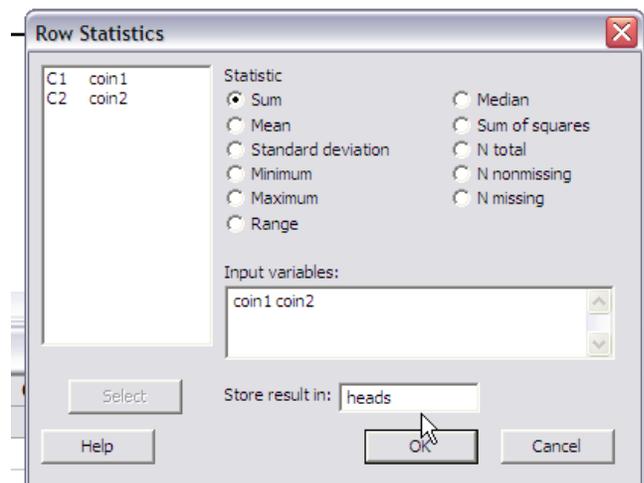
On the worksheet the columns titles 'coin1' and 'coin2' should appear with a combination of 1s and 0s.

Input variables coin1 and coin2 as the names of the columns.

If tails = 0 and heads = 1, then the sum of the columns = # of heads for the two coins.

Using **Calc → Row Statistics**, select **Sum** and input the two columns in variables box, saving the result as heads. To determine the number of heads:

Stats → Tables → Tally Individual Variables



Session Window Printout Tally for Discrete Variables: heads

heads	Count
0	14
1	62
2	24
N=	100

C4	C5	C6
value	Observed	prob
0	14	0.25
1	62	0.50
2	24	+ 0.25

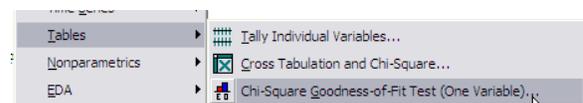
Input this information and the associated probabilities into the worksheet.

To test:

H_0 : The coins are fair

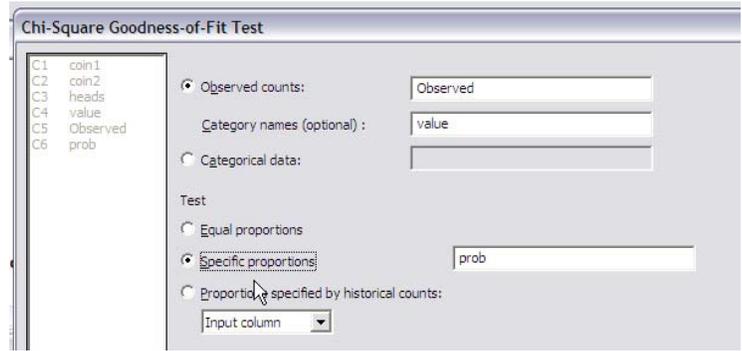
H_a : The coins are unfair

**Stats → Tables →
Chi-Square Goodness of Fit Test (One Variable)**



Input the information in the following boxes:

heads → C4 'value' → Category names
 counts → C5 'observed' → Observed counts
 select **Specific proportions** → C6 'prob'



Session Window Printout
Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: Observed

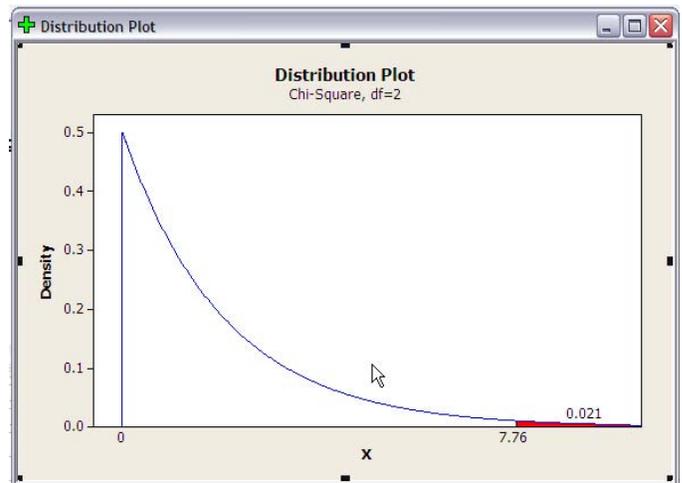
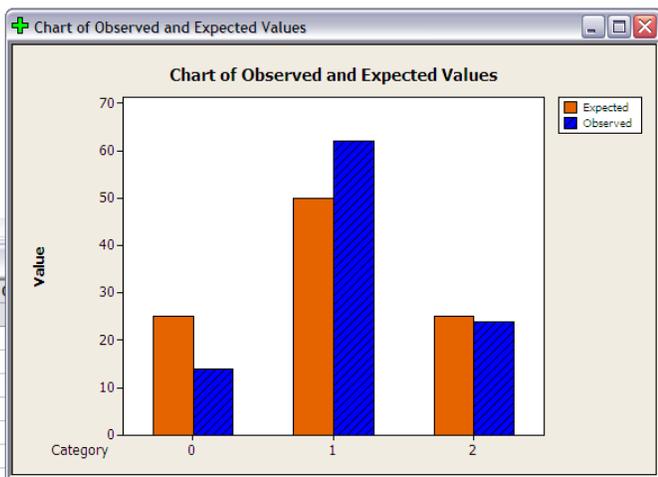
Using category names in value

Category	Observed	Test		Contribution to Chi-Sq
		Proportion	Expected	
0	14	0.25	25	4.84
1	62	0.50	50	2.88
2	24	0.25	25	0.04

N	DF	Chi-Sq	P-Value
100	2	7.76	0.021

Since the p-value < .05, fail to reject the null.

There is significant statistical evidence to support the claim that the coins are fair.



Chi-Squared Test of Independence

Example 11-3

Input the following information into a worksheet:

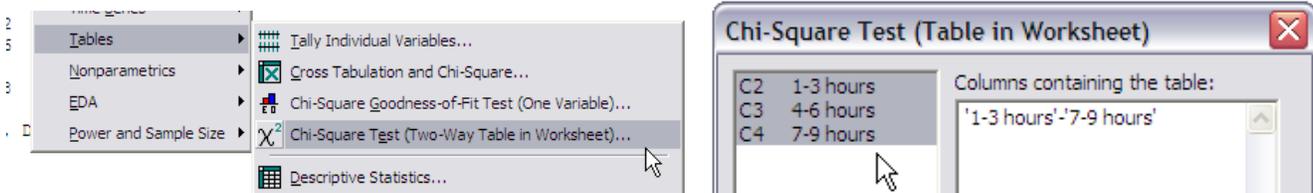
	C1-T	C2	C3	C4
	type of volunteer	1-3 hours	4-6 hours	7-9 hours
1	CC students	111	96	48
2	4-year students	96	133	61
3	nonstudents	91	150	53

H₀: Type of volunteer and number of hours are **independent**

H_a: Type of volunteer and number of hours are **dependent**

Stats → Tables → Chi-Square Test (Two-Way Table in Worksheet)

Select all the data columns.



Session Window Printout

Chi-Square Test: 1-3 hours, 4-6 hours, 7-9 hours

Expected counts are printed below observed counts

Chi-Square contributions are printed below expected counts

	1-3 hours	4-6 hours	7-9 hours	Total	
1	111	96	48	255	Observed
	90.57	115.19	49.24		Expected
	4.607	3.197	0.031		$\chi^2 = \frac{(O - E)^2}{E}$
2	96	133	61	90	
	103.00	131.00	56.00		
	0.476	0.030	0.447		
3	91	150	53	294	
	104.42	132.81	56.77		
	1.726	2.225	0.250		
Total	298	379	162	839	

Chi-Sq = 12.991, DF = 4, P-Value = 0.011 < 0.05 => reject H₀.

There is significant statistical evidence to conclude that the type of volunteer and the number of hours worked are dependent.

Linear Regression and Correlation

Example 12 – 6

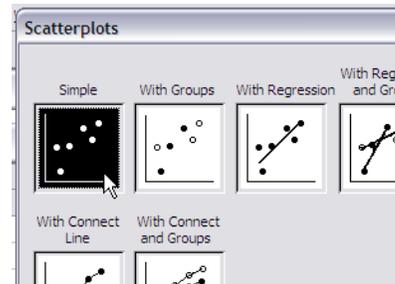
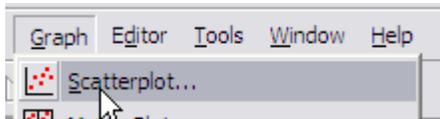
Is there a correlation between a student's score on Test 3 and his/her score on the Final Exam?

Input the data in a worksheet:

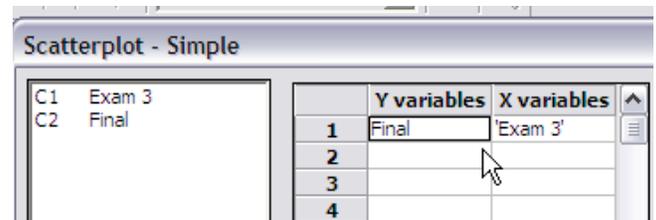
Exam 3	65	67	71	71	66	75	67	70	71	69	69
Final Exam	175	133	185	163	126	198	153	163	159	151	159

To view the scatter plot:

Graph → Scatterplot

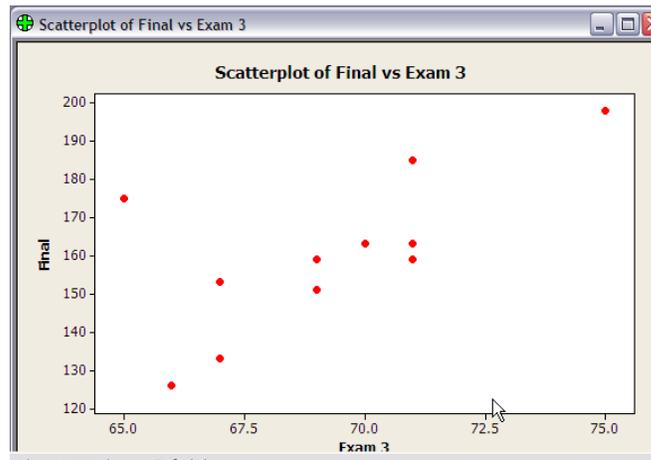


Select the **Simple**, then input **Final** for the **Y variable** (dependent or predicted variable) and **Exam 3** for the **X variable** (independent or predictor variable).



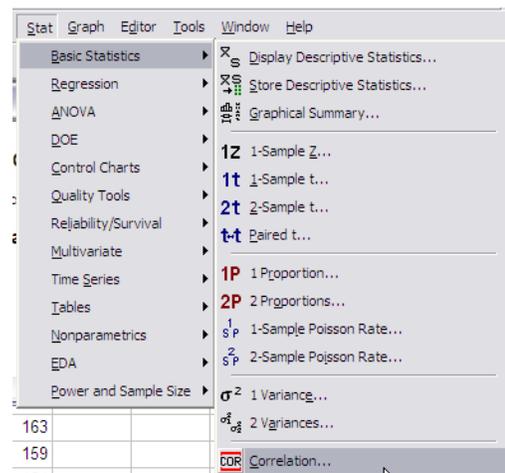
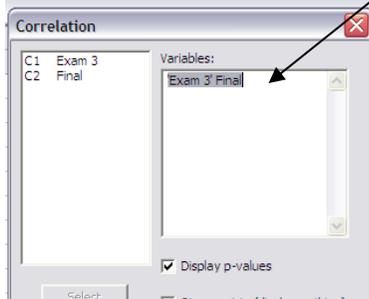
Does there appear to be a relationship between the two values?

To determine if the relationship is statistically significant, a test for correlation is used.



Stat → Basic Statistics → Correlation

Select the two variables. The default **Display p-values** will be checked.



Session Window Printout
Correlations: Exam 3, Final

Pearson correlation of Exam 3 and Final = 0.663 (**r – value**)

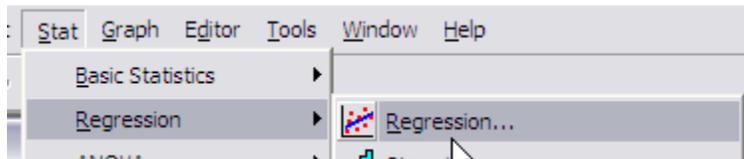
P-Value = 0.026

P-Value = 0.026 < 0.05 => significant correlation

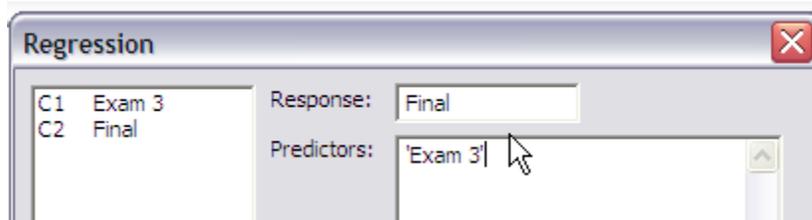
Significant correlation implies there is a relationship between the student’s score on Exam 3 and the score on the Final Exam.

To calculate the regression line:

Stat → Regression → Regression



The **Response** is the dependent variable and the **Predictors** are the independent variable(s).



Session Window Printout

Regression Analysis: Final versus Exam 3 (This is the printout. Comments are in bold.)

The regression equation is

Final = - 174 + 4.83 Exam 3 => **y = -175 + 4.83x**

Predictor	Coef	SE Coef	T	P
Constant	-173.5	125.8	-1.38	0.201
Exam 3	4.827	1.816	2.66	0.026

S = standard deviation of errors, s_e

S = 16.4124 *R-Sq* = 44.0% *R-Sq(adj)* = 37.7%

R-Sq = coefficient of determination, r²,

hence the Person correlation coefficient, r = √0.44 = 0.663 (see previous test)

Analysis of Variance

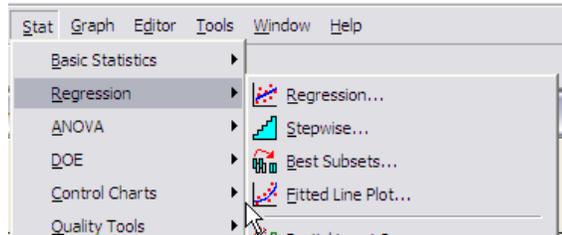
Source	DF	SS	MS	F	P
Regression	1	1902.4	1902.4	7.06	0.026
Residual Error	9	2424.3	269.4		
Total	10	4326.7			

Obs	Exam 3	Final	Fit	SE Fit	Residual	St Resid
1	65.0	175.00	140.27	9.07	34.73	2.54R

R denotes an observation with a large standardized residual.

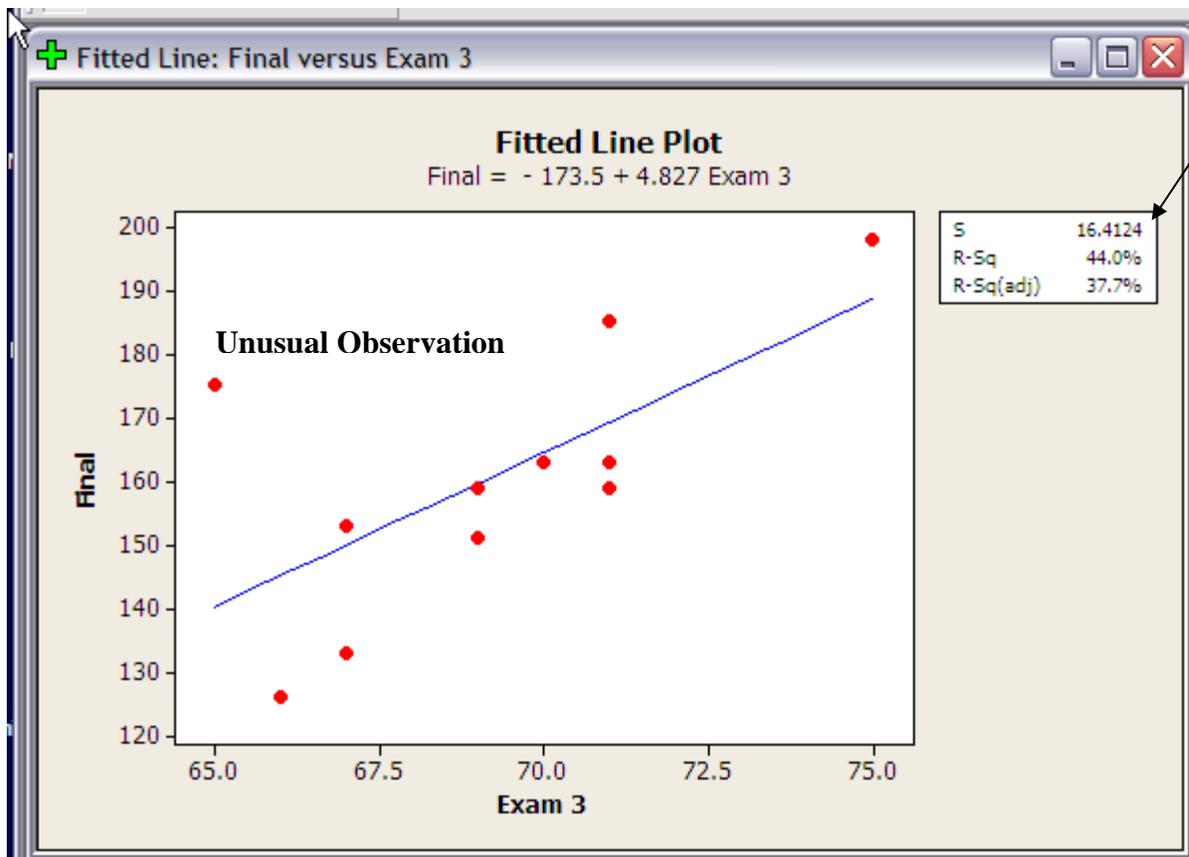
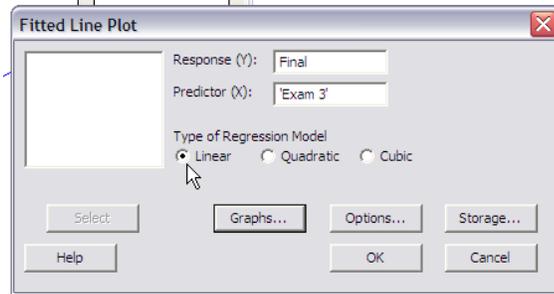
Unusual Observations: From the scatter plot there is one data value appear out of place (65, 175). This value is reported here with the predicted value (Fit) of 140.27.

To see the data and regression line together:
Stat → Regression → Fitted Line Plot



Input **Final** as the **Response (Y)** and **Exam 3** as the **Predictor (X)**.

Select **Linear** for linear regression.



Note that this graph gives the regression equation as well as the s_e and R^2 values.

To find a best fit and a 95% confidence interval for the Final Exam score for a student who

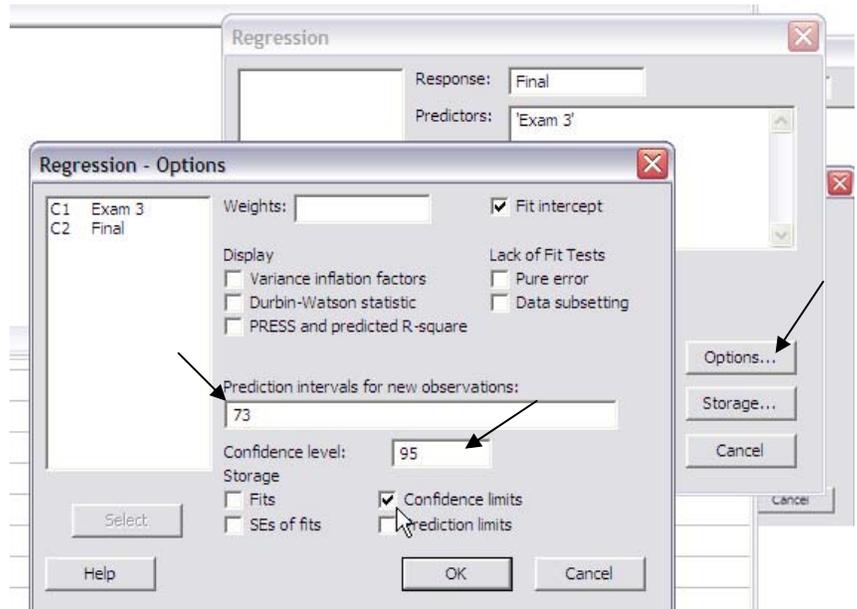
scored 73 on Exam 3, return to the regression window.

Stat → Regression → Regression

Click on **Options**.

Input **73** as the **new observation** and **95** and the **confidence level**.

Check the **Confidence limits**.



Session Window Printout

Predicted Values for New Observations

New

Obs	Fit	SE Fit	95% CI	95% PI
1	178.89	8.52	(159.61, 198.16)	(137.05, 220.72)

Values of Predictors for New Observations

New

Obs	Exam 3
1	73.0

Note that the best single value is **178.89 (Fit)** and the **95% confidence interval is (159.61, 198.16)**.

Quick Reference Guide

Bar Chart	Graph → Bar Chart
Binomial Distribution	Calc → Probability Distributions → Binomial
Binomial Distribution, Graph	Graph → Probability Distribution Plot, View Single, select Binomial
Boxplot	Graph → Boxplot
Boxplot, Lable Outliers	Graph → Boxplot → Labels → Data Labels → Outliers
Chi-Square Test of Independence	Stat → Tables → Chi-Square Test (Two-Way Table in Worksheet)
Chi-Square Test, One Variable	Stat → Tables → Chi-Square Goodness of Fit Test (One Variable)
Correlation	Stat → Basic Statistics → Correlation
Data and Regression Line	Stat → Regression → Fitted Line Plot
Data Summary	Stats → Basic Statistics → Display Descriptive Statistics
Dotplot	Graph → Dotplot
Exit	File → Exit
Exponential Distribution	Graph → Probability Distribution Plot, View Single, select Exponential
Exponential Distribution, Graph	Calc → Probability Distributions → Exponential
Geometric Distribution	Calc → Probability Distributions Select Geometric
Geometric Distribution, Graph	Calc → Probability Distribution → Geometric
Histogram	Graph → Histogram
Independent Means, Two	Stat → Basic Statistics → 2-Sample t
Mean	Calc → Column Statistics
Median	Stat → Tables → Tally Individual Variables
Midrange	Stat → Tables → Tally Individual Variables
Mode	Stat → Tables → Tally Individual Variables
Normal Distribution	Graph → Probability Distribution Plot, View Single, select Normal
Normal Distribution, Graph	Calc → Probability Distributions → Normal
Open	File → Open Worksheet
Paired Date, Two Sample Means	Stat → Basic Statistics → Paired t
Pie Chart	Graph → Pie Chart
Print	File → Print
Probability Table	Calc → Make Patterned Data → Simple Set of Numbers

Proportion, Single Population	Stat → Basic Statistics → 1-Proportion
Proportion, Two Populations	Stat → Basic Statistics → 2 Proportions
Regression	Stat → Regression → Regression
Save, Project	File → Save Project as
Save, Worksheet	File → Save Current Worksheet as
Scatterplot	Graph → Scatterplot
Simulate Data	Calc → Random Data → Integer
Single Population Mean, Standard Deviation Given	Stat → Basic Statistics → 1-Sample Z
Single Population Mean, Standard Deviation Unknown	Stat → Basic Statistics → 1-Sample t
Standard deviation	Calc → Column Statistics
Stem-and-Leaf	Graph → Stem-and-Leaf
Sum, Column	Calc → Column Statistics
Summarize Data	Stat → Tables → Tally Individual Variables
Variance	Calc → Calculator
Variance, Two	Stat → Basic Statistics → 2 Variances