

Chapter 3

Summarizing Data: Descriptive Statistics and Histograms

Key Concepts

Central tendency, Dispersion, Histogram, Kurtosis, Mean, Median, Mode, Range, Sample variance, Skewness, Standard deviation, and Standard error.

Discussion

This chapter examines the analysis tools used for summarizing single-variable (univariate) data and contains a detailed discussion of **histograms**. The concepts of central tendency and dispersion are explored. Excel output associated with the generation of the histogram provides all of the key descriptive statistics associated with a univariate data set. These are discussed in detail. The access to Excel statistical functions is explained with applied examples.

This chapter concludes with an important section on the pitfalls when working with histograms and the associated basic statistics.

Central Tendency has to do with how the distribution of data clusters around the middle. We typically refer to this position using three important statistics: **mean**, **median**, or **mode**.

- **Mean:** The mean of a data set refers to the arithmetic average of all the data values. To calculate the mean, sum all of the data and divide the sum by the number of data values.
- **Median:** The median of a data set refers to the value in the middle when the data values are arranged in ascending order. Half of the values are greater than this value and half are less than this value. The median may or may not correspond to a value in your data set; it is just a location.

- **Mode:** The mode of a data set refers to the value that occurs with the greatest frequency. If all of the data values occur with the same frequency there will not be a mode.

Dispersion refers to how spread out the data is as a distribution. We use **range** and **standard deviation** as a means of describing dispersion.

By combining the information about the **central tendency** and **dispersion**, you get a good idea of how the data is distributed even without the aid of graphs.

- **Range:** The range of a data set refers to the difference between the largest and smallest data values. The range is the simplest measure of dispersion. The range can be described by stating the smallest and the largest data values (i.e., \$25–\$100), or by stating the actual number of units between these end points (i.e., \$75).
- **Standard Deviation (σ , SD, or sd):** The standard deviation refers to the most commonly used statistic to measure the variation in any data set. This is a group statistic that summarizes the overall variation of your data. In other words, it indicates how much your data bounces around the average value. Excel provides this statistic in the same units as the original values. Figure 3.1 depicts the concept of standard deviation and provides the formula for a sample.

The concepts of a constant mean and varying spread (dispersion) are shown in Fig. 3.2. Both distributions have the same center value (the average or mean) but the

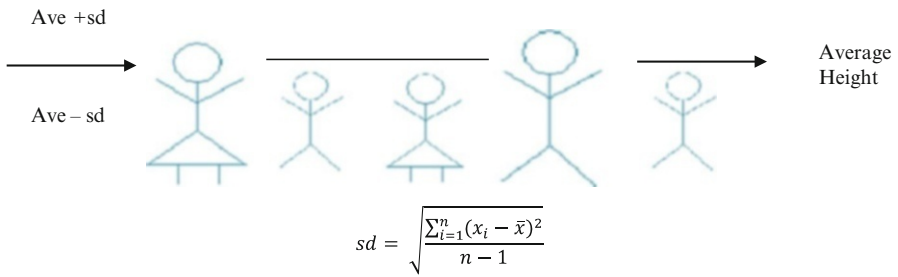


Fig. 3.1 Standard deviation

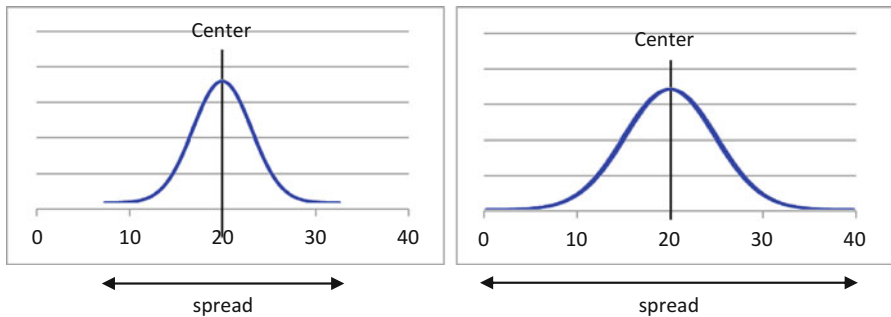


Fig. 3.2 Center and spread

Copyright © 2013, Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

second distribution is more spread out (greater dispersion). The more spread out a distribution is, the larger the standard deviation and range.

Other less used statistics that are provided in the Excel Data Analysis output include:

Kurtosis: Kurtosis refers to how peaked a distribution is or conversely how flat it is. If there are more data values in the tails, than what you expect from a normal distribution, the kurtosis is positive. Conversely if there are less data values in the tails, than you would expect in a normal distribution, the kurtosis is negative.

- Excel cannot calculate this statistic unless you have at least four data values.

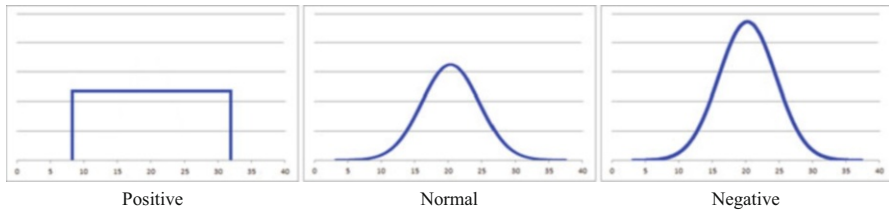


Fig. 3.3 Kurtosis

Skewness: Skewness describes the lack of symmetry in the distribution. If the data is pulled to the right creating a right hand tail, it is a positive skew. Positively skewed data has a mean greater than the median. Conversely if the data is pulled to the left creating a left hand tail, it is a negative skew and the mean is less than the median.

- ◆ Excel cannot calculate this if you input less than three data values.

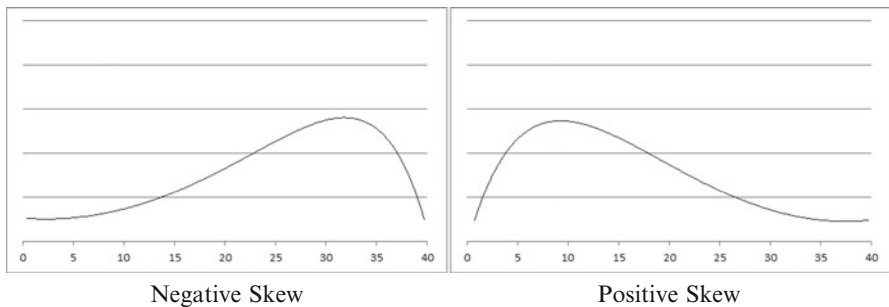


Fig. 3.4 Skewness

Standard error: This is a measure of the uncertainty about the mean. This becomes an important statistic when working with inference analysis, which is discussed in Chap. 7.

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

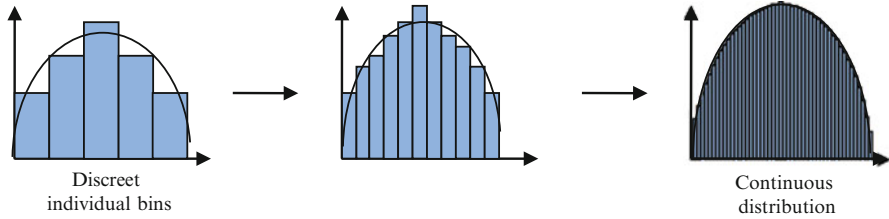


Fig. 3.5 Histogram

Sample variance: This is just the standard deviation squared.

Symbols

Σ = Greek capital letter Sigma which means “take the sum of” or “sum together” everything that follows the sign

\bar{x} or μ = average, mean

N or n = number of data values in your data set; to be accurate, we use capital N for a population (all the possible data) and lower case n for a sample (some of the data)

σ or **SD** or **sd** = standard deviation; to be accurate, we use capital **SD** or σ for the standard deviation of a population (all possible data), and lower case **sd** for a sample (some of the data)

The Histogram

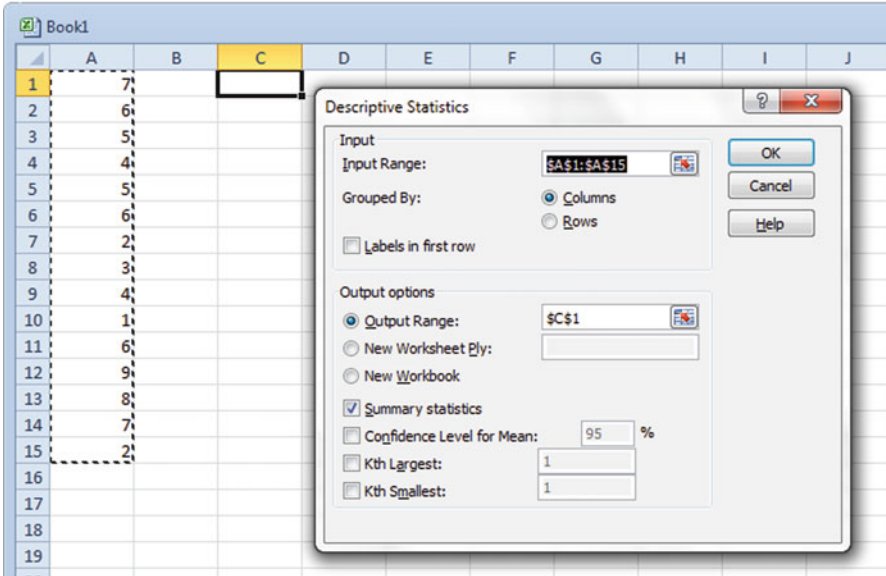
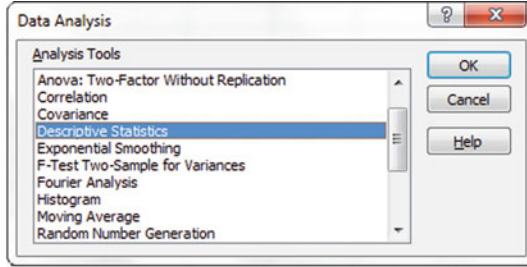
Histograms are used for charting frequency data. These graphs or charts use individual categories, bins, or classes to count how many times a variable occurs. We can also think of the histogram as more of a continuous or connected graph. See Fig. 3.5.

We see from Fig. 3.5 that if we use an infinite or unlimited number of narrower and narrower bins we would eventually approximate a continuous distribution.

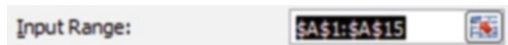
Excel

Descriptive Statistics

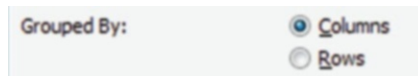
- Input the following data into Column A cells A1 through A15: 7, 6, 5, 4, 5, 6, 2, 3, 4, 1, 6, 9, 8, 7, 2
- From the **Data** tab, choose the **Data Analysis** function.
- Select the **Descriptive Statistics** option, under Analysis Tools
- Click **OK**



➤ Enter the range of data in **Input Range**:



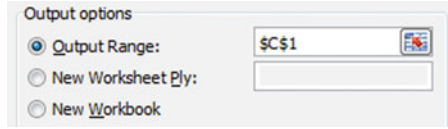
➤ Choose the option that fits your data in **Grouped By**: [columns or rows]



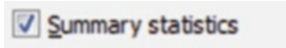
➤ Select **Output Range**:

➤ Click inside the text box.

◆ Important to do this **first** or you will create problems in your **Input Range**. Then select the output range.



➤ Click **Summary Statistics**



- **Confidence Level for the Mean:** Excel default is 95 % which should be acceptable for all of your work. However you can input any value greater than 0 and less than 100. *You can leave this unchecked.*
- **Kth largest:** If you want the third largest value in the data set you input 3. The default is 1. Values can range from 1 up to the number of data points you have in the distribution. *You can leave this unchecked.*
- **Kth smallest:** If you want the second smallest value in the data set you input 2. The default is 1. Values can range from 1 up to the number of data points you have in the distribution. *You can leave this unchecked.*
- Click **OK**

Descriptive Output Results

	A	B	C	D	E
1	7		Column1		
2	6				
3	5		Mean	5	
4	4		Standard Error	0.601585	
5	5		Median	5	
6	6		Mode	6	
7	2		Standard Deviation	2.329929	
8	3		Sample Variance	5.428571	
9	4		Kurtosis	-0.75623	
10	1		Skewness	-0.11729	
11	6		Range	8	
12	9		Minimum	1	
13	8		Maximum	9	
14	7		Sum	75	
15	2		Count	15	
16					

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

The most important output value is the **Count**; this is your check that all of the data has actually been input and used in the calculations.

But when it is difficult to read the names of the statistics, we need to adjust the column width.

Changing the Width of the Column Output

Let's make the column width large enough to show all the words

- Place your cursor on the line at the far right of the top cell in the column you want to make wider. In this case we want Column C to be wider.

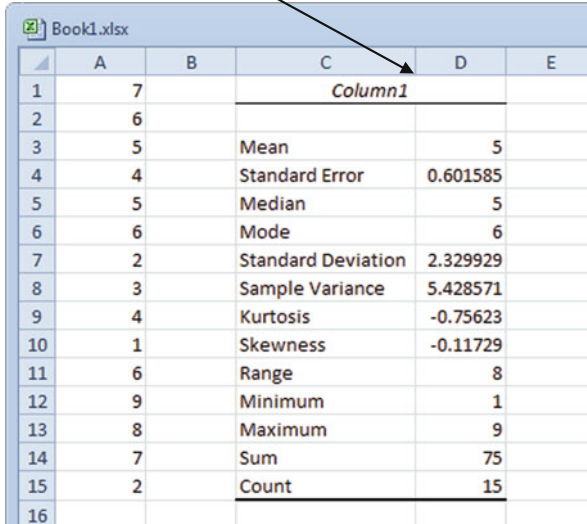
Before:

	A	B	C	D	E
1	7		<i>Column1</i>		
2	6				
3	5		Mean	5	
4	4		Standard E	0.601585	
5	5		Median	5	
6	6		Mode	6	
7	2		Standard I	2.329929	
8	3		Sample Va	5.428571	
9	4		Kurtosis	-0.75623	
10	1		Skewness	-0.11729	
11	6		Range	8	
12	9		Minimum	1	
13	8		Maximum	9	
14	7		Sum	75	
15	2		Count	15	
16					

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

- Drag this border line to the width you want for the column.

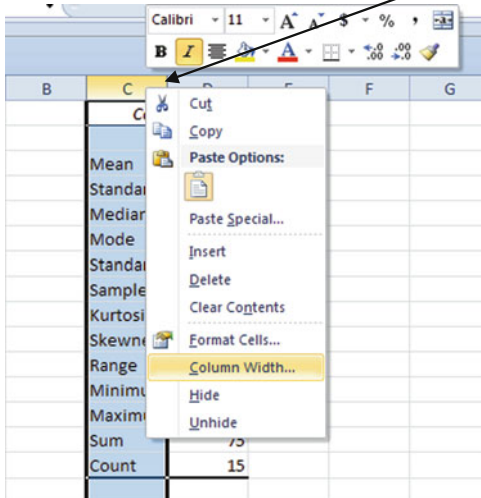
After:



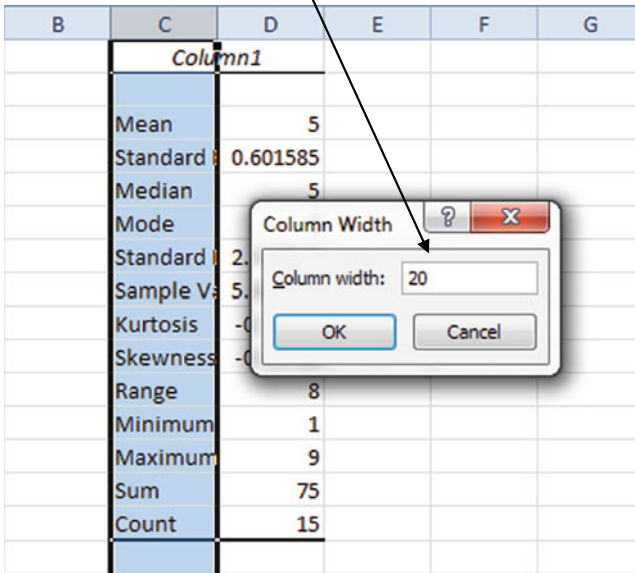
	A	B	C	D	E
1	7		<i>Column1</i>		
2	6				
3	5		Mean	5	
4	4		Standard Error	0.601585	
5	5		Median	5	
6	6		Mode	6	
7	2		Standard Deviation	2.329929	
8	3		Sample Variance	5.428571	
9	4		Kurtosis	-0.75623	
10	1		Skewness	-0.11729	
11	6		Range	8	
12	9		Minimum	1	
13	8		Maximum	9	
14	7		Sum	75	
15	2		Count	15	
16					

OR:

- Right click the top cell of the column you want to make wider. Select **Column Width**. You can make the column as wide as you need.



- Type in the desired column width in the popup box.

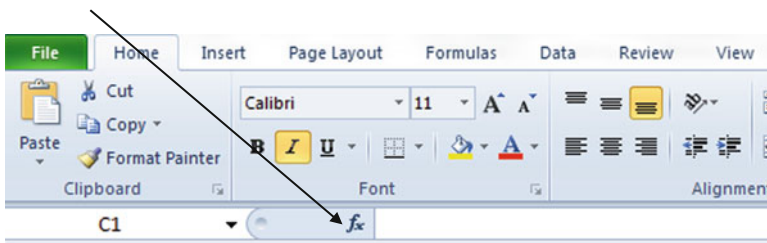


You can also auto format the column width by double clicking the line on the right side of the column label, which is above the first cell, of the column you want to make wider.

Using Excel Functions

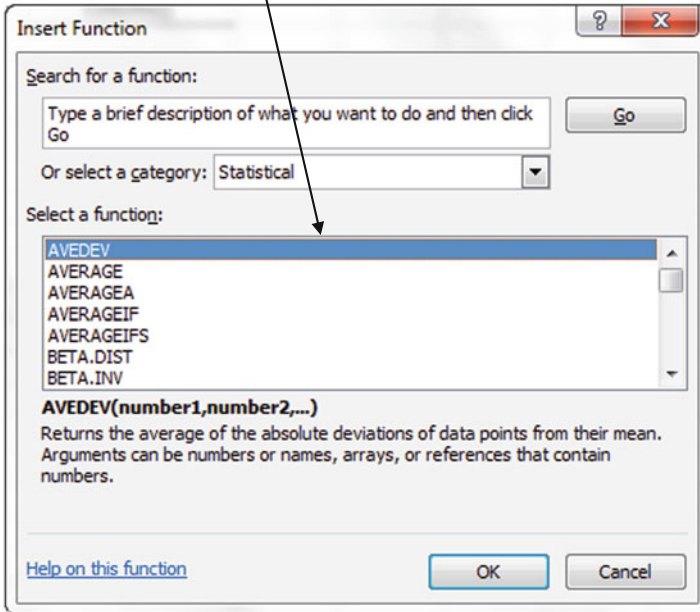
We can also calculate the descriptive statistics one by one using the separate functions in Excel.

- Click the *fx* symbol in the white function bar



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

- Input the description of what you want to do, if you do not know the name of the statistical function, **OR** click on the next box and select **Statistical**, if you do know the name of the function
- Select the function you need



- Click **OK**

Statistic	Function name	Calculates
Mean	AVERAGE	Arithmetic average of the data set
Median	QUARTILE QUARTILE.INC QUARTILE.EXC (You need to input 2 for the quartile value)	Median of the data set
Mode	MODE MODE.MULT MODE.SNGL	Mode of the data set
Standard Deviation	STDEV STDEV.P STDEV.S STDEVA STDEVPA STDEVP	Standard deviation of the data set (both population standard deviation and sample standard deviation)
Range	MIN MAX	Range of the variables
Count	COUNT	Total count
Skewness	SKEW	Calculates how skewed the data is
Kurtosis	KURT	Calculates the kurtosis of the data
Kth largest	LARGE	Displays the “K” largest number in your data set
Kth smallest	SMALL	Displays the “K” smallest number in your data set

Fig. 3.6 Functions

Be careful to choose the exact function as listed in Fig. 3.6. Excel has many functions that are only different by one letter and each of those functions will deliver a different result.

Histograms

Excel provides a table with the frequency data as well as the actual histogram graph. Remember a histogram does not have spaces between the bars because both the x and y axes are scales and show quantitative data.

Copyright © 2013, Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Example Problem

You have collected some proficiency scores for the administrative assistants in your department. The proficiency scale ranges from 0 to 1,000, with 0 being the lowest possible score and 1,000 being the highest possible score. You decide that the best way to get a sense of the data is to graph the data in a histogram. Here are the steps you would need to take.

Input the following data in column B

Data
99
150
300
780
350
500
510
360
200
175
196
450

Setting Up the Bin Ranges

Before you create the histogram you need to identify your bin/category/class ranges. As a guiding principle, charts have somewhere between 5 and 15 bins. Too few bins or too many bins confuse the story.

Arbitrarily choose a certain number of bins. Let's choose seven bins. To get the width of those bins determine the range of the data and divide by the number of bins you want. In this example the range goes from \$99 to \$780 (Fig. 3.7).

$$\begin{aligned} & 7 \text{ bins} \\ & \text{Range } \$99\text{--}\$780 \\ \text{Bin width} &= (\$780 - \$99) / 7 = \$97.28 \end{aligned}$$

Because using bin widths of \$97.28 would be confusing, we can round up or down to choose an easier bin width, such as \$100.

In a new column, type in the largest number in each of the bins to indicate the end point of each bin. In Fig. 3.7 all of the data ≤ 100 will be reported in that first bin. All of the data greater than 100 and ≤ 200 will be reported in that second bin. Remember these are the end points of the bins you input to Excel.

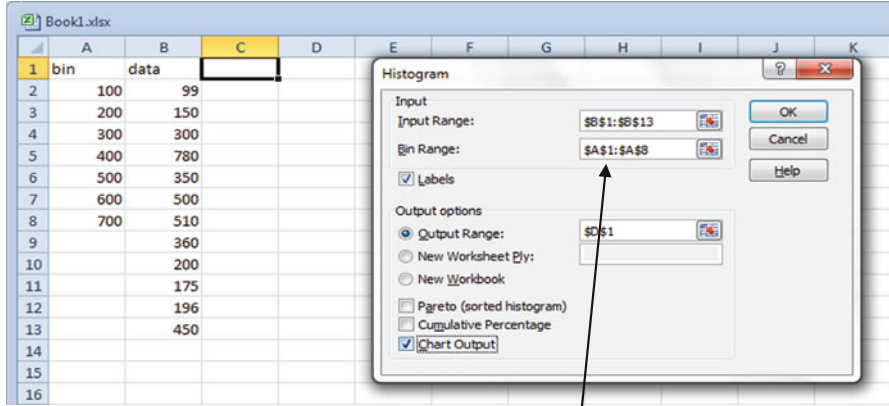


Fig. 3.7 Input Bin ranges

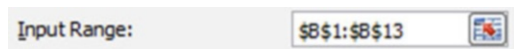
Keep track of where you have entered your bin ranges as you will need this as an input field for the histogram.

Creating the Histogram Chart

Now you are ready to create a histogram.

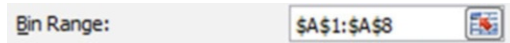
- Under the **Data** tab click on the **Data Analysis** function on the far right
- Select **Histogram** from the list of options under Analysis Tools and click **OK**

- Highlight the **Input Range** for the data



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

- Highlight the range of cells where you have typed the end points of the bins



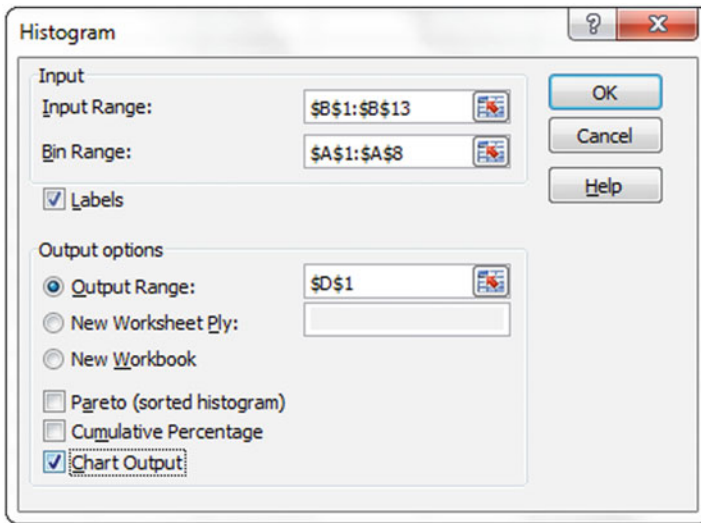
- Click on the **Output Range:** button and then click inside the box. Now highlight a cell where you want the output to start



- Check **Chart Output** or you won't get a graph



- Click **OK**

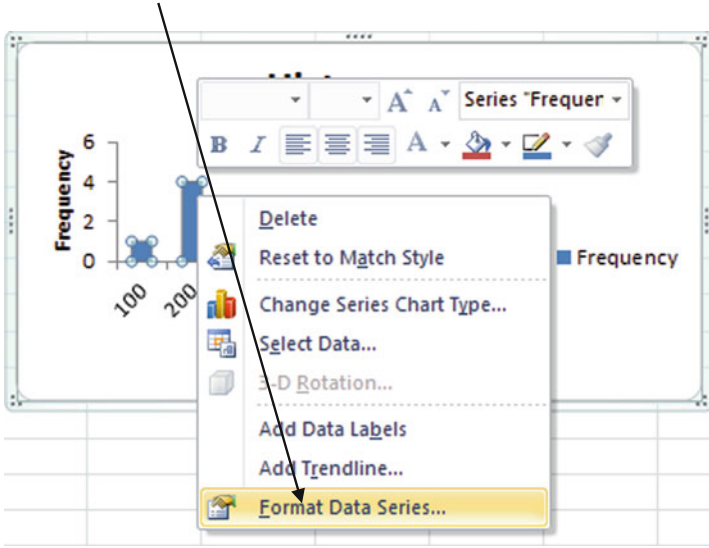


Histogram Clean Up

The actual histogram needs some cleaning up.

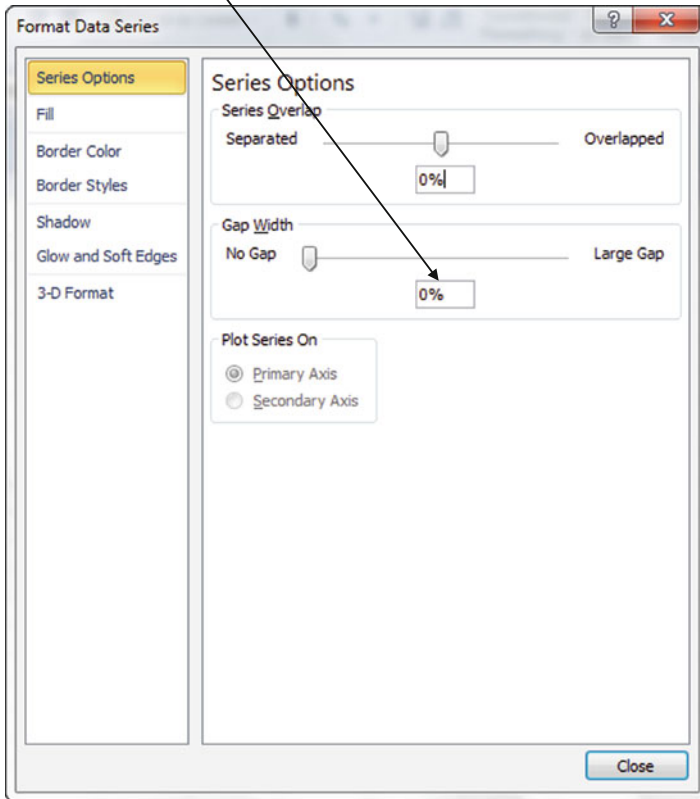
Closing Gaps Between Bars

- Right Click in any column on the graph
- Click the **Format Data Series** option from the menu



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

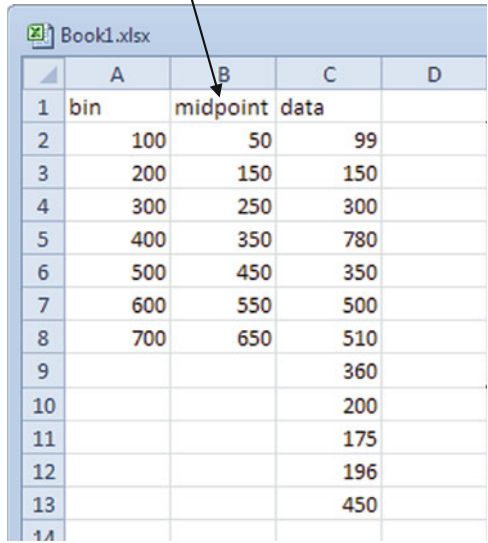
- Click on **Series Options**
- Go to **Gap Width**. Set to **0 %** and click **Close**



Change Labels on x-axis

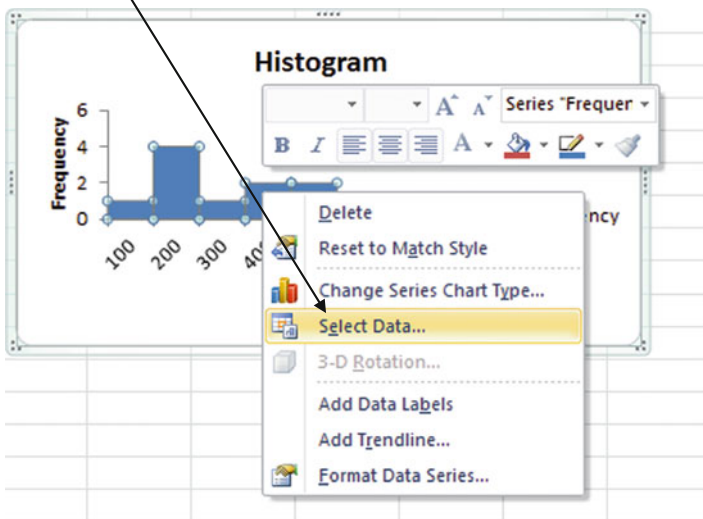
- Change the x-axis values from ***upper limits*** of the bins to ***midpoints*** of the bins

- Create the column of midpoint labels you want added to your graph
 - ◇ **NOTE:** The data labels on the x-axis indicate the upper limits of the bins. In order, to change this to the midpoints, first you need to create a column of data with the mid-points of the bins.



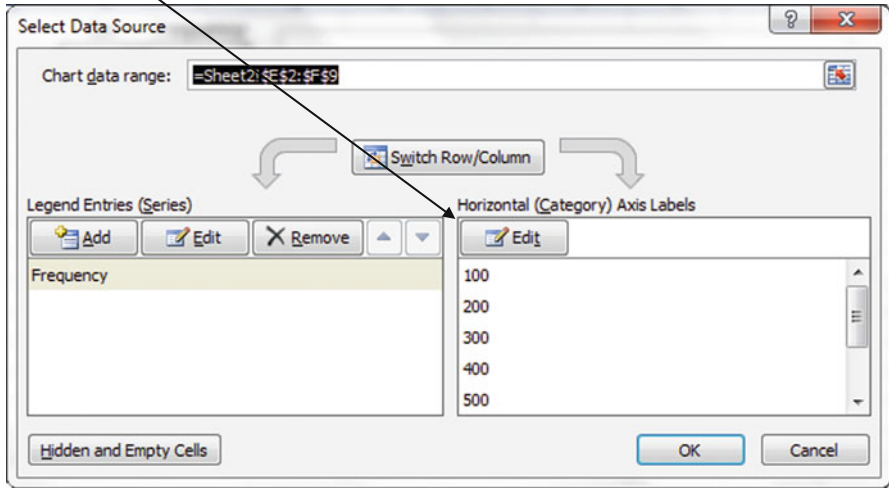
	A	B	C	D
1	bin	midpoint	data	
2	100	50	99	
3	200	150	150	
4	300	250	300	
5	400	350	780	
6	500	450	350	
7	600	550	500	
8	700	650	510	
9			360	
10			200	
11			175	
12			196	
13			450	
14				

- Right click on any column on the graph
- Choose the **Select Data** option

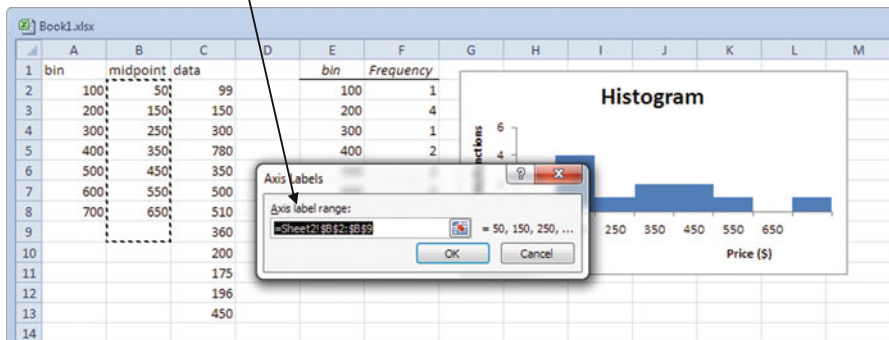


Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

➤ Under **Horizontal Axis Labels** click the **Edit** option



➤ Click inside the **Axis label range** box. Highlight the new column of midpoint labels



➤ Click **OK**

Removing More from the Chart and Labeling the Last Column

The word **More** may appear on your graph.

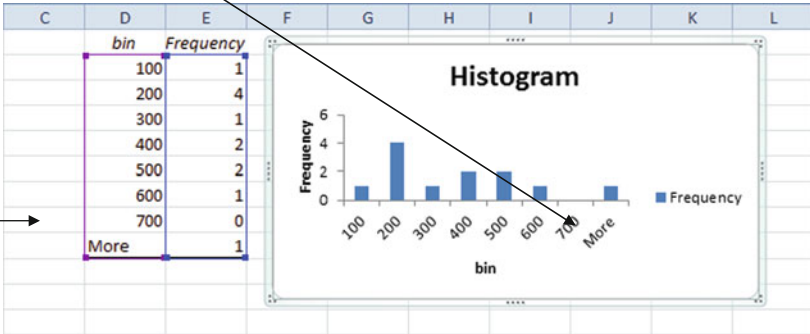


Fig. 3.8 End-point Bin values

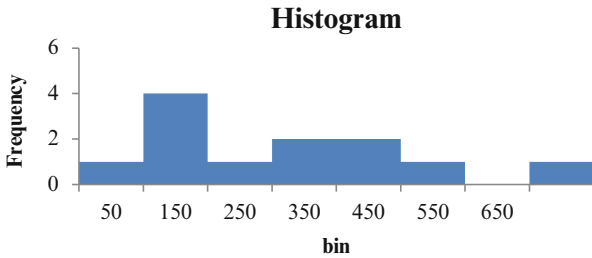
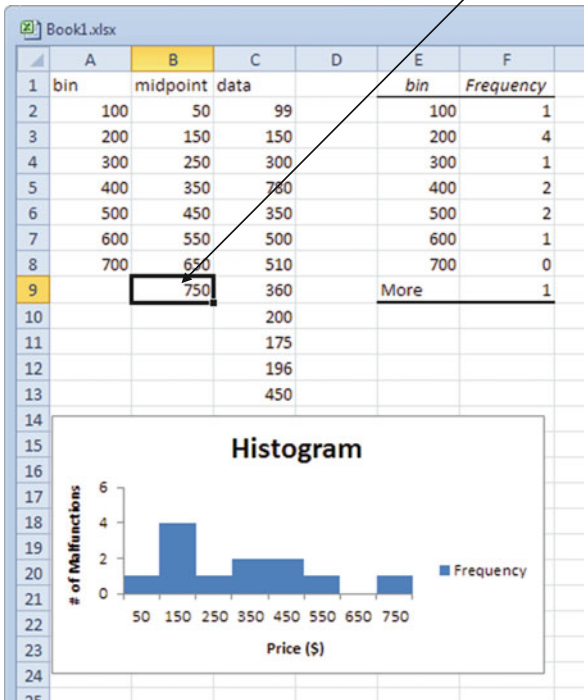


Fig. 3.9 Mid point Bin values

The last column may not be assigned a numeric label. This is because there are data values greater than the last bin value you originally input. In Fig. 3.8 the last bin end value originally selected was 700. In Fig. 3.9 this was changed to a mid-point bin value of 650. But as you can see from the data there is one value greater than 700. Excel automatically names this last bin **More** and will not assign a numeric label for it on the x-axis.

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

You can go back and add more bins to ensure all of your data is being included on your graph with the proper labels. In this case we added **750**.



Once you have added the extra bins, edit the bin values which will remove the word “More” from the graph.

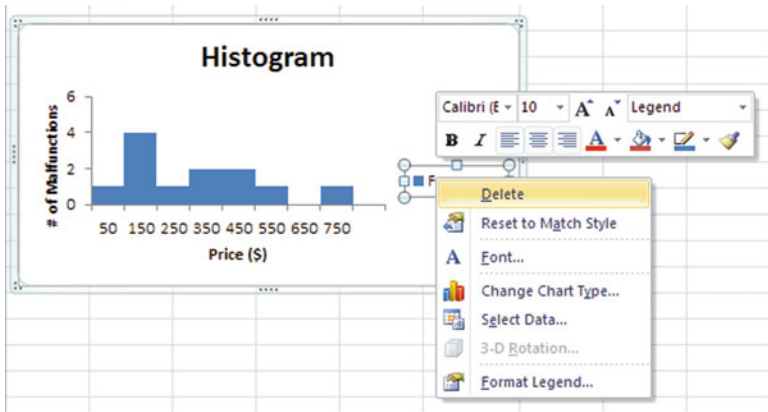


◇ Changing the bin values will only change the labels on your x axis. The frequency will still count the data using the initial points you entered, not the midpoints.

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

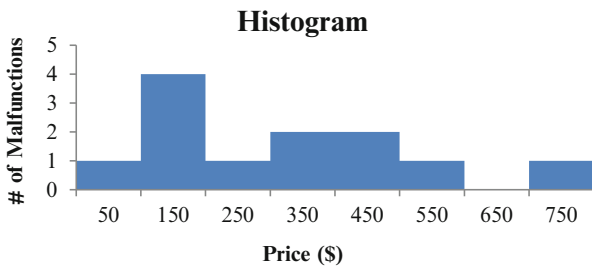
Remove Legend

- Right click on the legend and click the **Delete** option



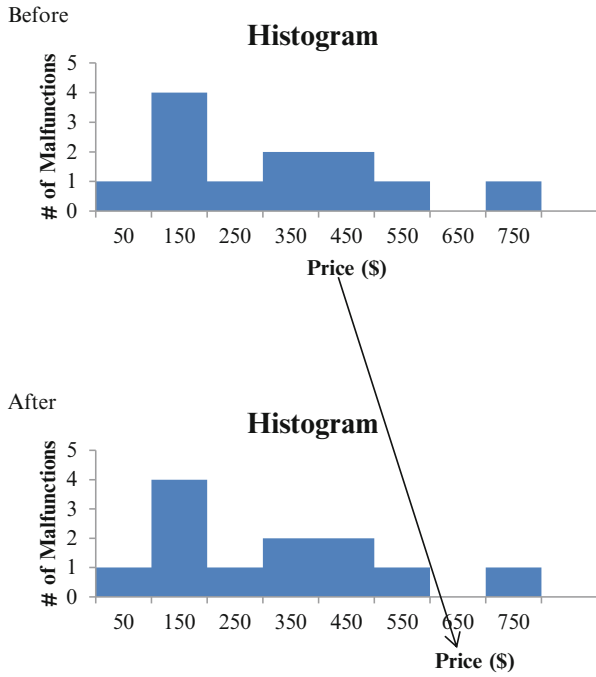
Axes Labels

- To change the axes labels click the frequency label (y) or bin label (x) and type the appropriate titles



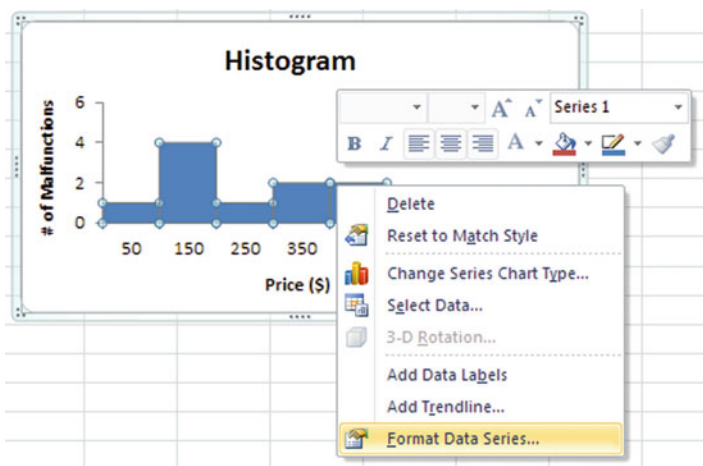
Moving Axes Labels

Left click on the axis you want to move; drag and drop



Changing the Bar Color

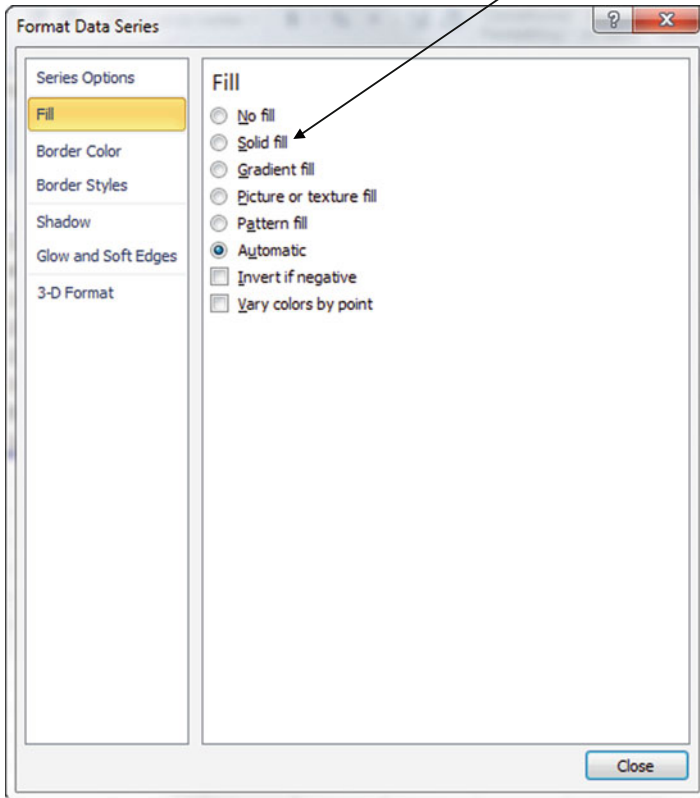
Right click on any column. Choose the **Format Data Series** option



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

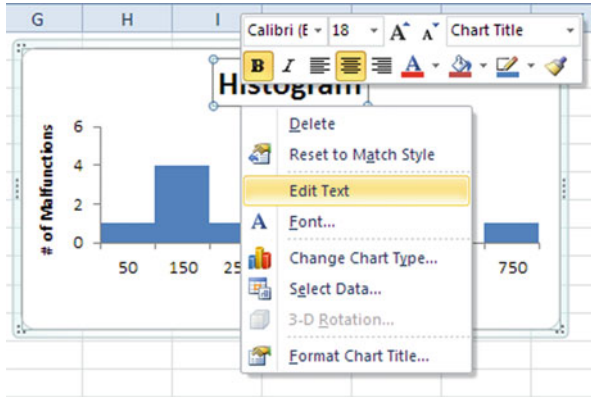
Select **Fill**.

- Change from **Automatic** to the fill you prefer for the bins; if you click on **Solid Fill** you can choose from a variety of solid colors by clicking on the color
- Click **Close**, when you are finished



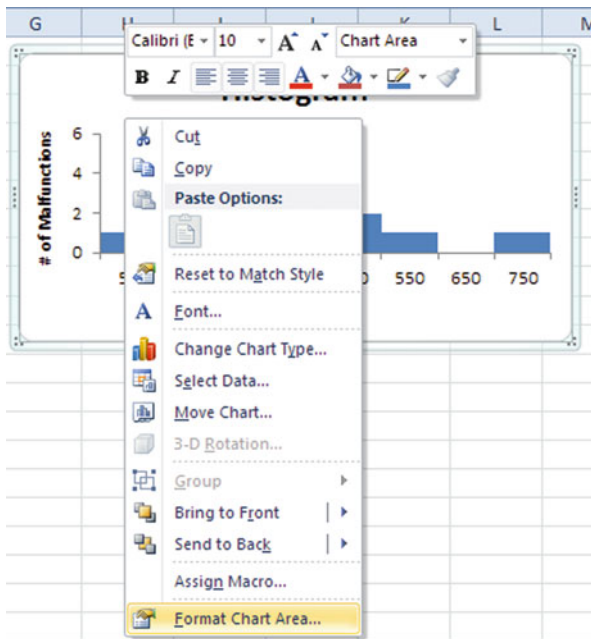
Changing Chart Title

- Right click on the chart title. Select the **Edit Text** option; you can retype the title as you choose but make sure to click outside the text box when done typing



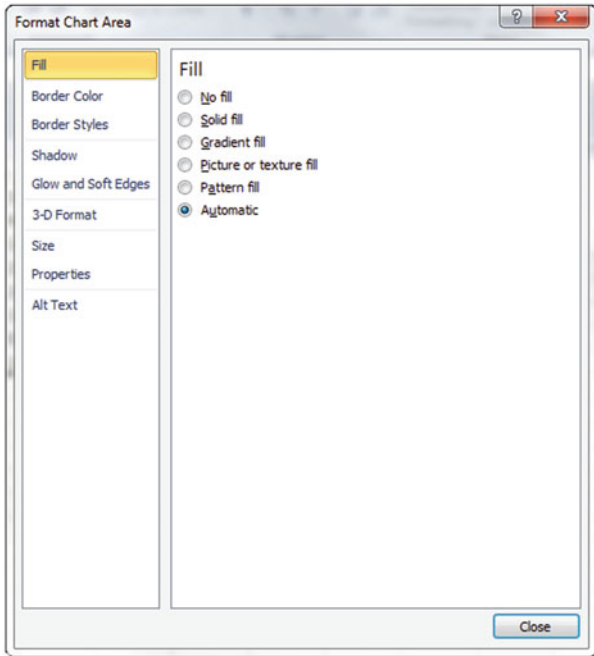
Changing Chart Background Fill

- Right click on the chart background. Click the **Format Chart Area** option



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

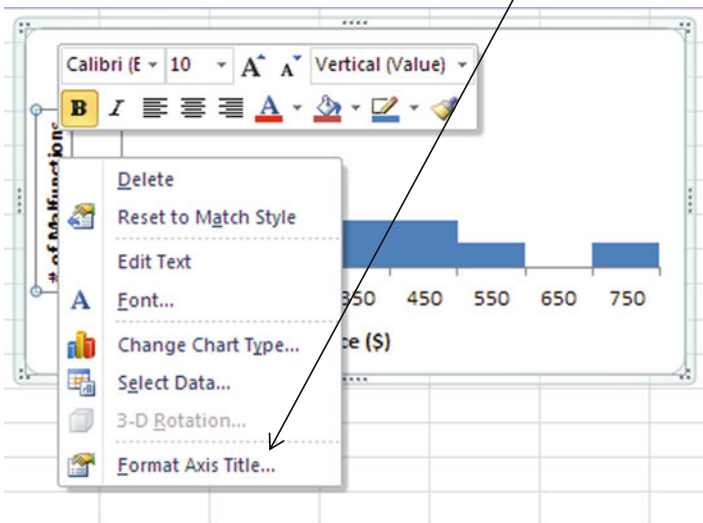
- Select **Fill** from the left hand menu
- Select the type of fill and color options you desire



- Click **Close** when you are finished

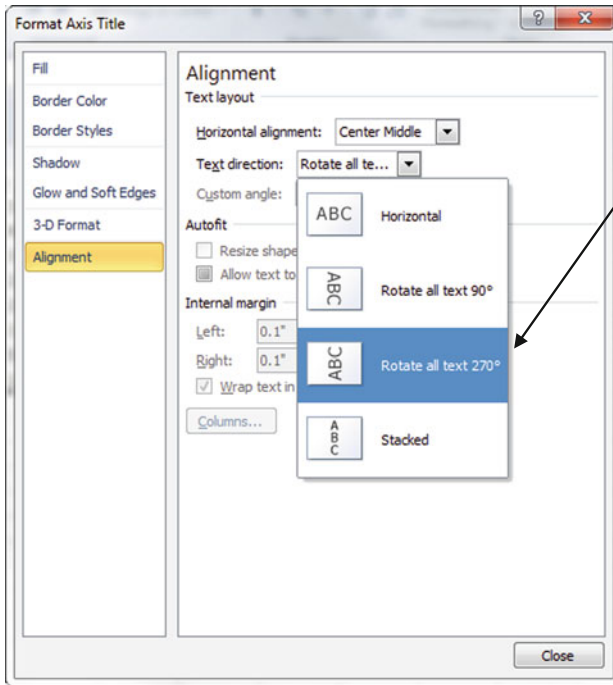
Rotating the y-axis Label from Vertical to Horizontal

- Right click on the y axis label. Click on the **Format Axis Title** option



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

- Select **Alignment** from the left hand menu; you can change the alignment and direction



Common Pitfalls

- ◆ Be very careful when calculating the mode. When there is more than one value with the same frequency (more than one mode) in a data set, Excel will report the value that appears first in your data set. If each data value appears with the same frequency Excel will report #N/A.
 - {1,1,1,2,2,2,3,3,4} Excel will report that the mode is 1
 - {1,2,3,3,3,4,5,6,7} Excel will report that the mode is 3
 - {1, 2, 3, 4, 5, 6} Excel will report #NA
- ◆ The formula used in Excel for STDEV.S to calculate the standard deviation uses a denominator of (n-1), whereas STDEV.P uses the denominator of (N). The P stands for population and the S stands for sample. When you calculate the standard deviation by hand you typically use (N) in the denominator of a population and (n-1) for a sample. These two methods will yield slightly different answers, so be careful when deciding which formula to use.

- ◆ Label all axes with variable names and units. Excel will not label the axes by default.
- ◆ Excel reports the maximum value for each interval as the bin value not the midpoint.
- ◆ Excel will include a bin labeled **More**. Rename the **More** bin as a numerical label to avoid confusion.
- ◆ Excel labels all frequency graphs as histograms. Remember bar graphs can have spaces between bins. Histograms have no spaces between bins. You need to close the gaps in the histogram to zero.
- ◆ Be careful to tell Excel if you are including labels as the first set of cells in the input. Always check the count (n) to make sure you have all of your data is included in the Excel calculations.
- ◆ If you expand your output area Excel will keep repeating the results to fill up the area. In Fig. 3.10 the data labeled as Column 1 is the same output as Column 2. There is no need to show duplicate data. You can avoid this by keeping the output area large enough for just one set of output data. To expand the columns you need to right click on the top of the column and change the width. If you simply drag the lowered left corner of the output box you will generate repeated data.

	Column1		Column2		Column3
Mean	5.333333333	Mean	5.333333333	Mean	
Standard Error	0.503952631	Standard Error	0.503952631	Standard Error	
Median	5	Median	5	Median	
Mode	5	Mode	5	Mode	
Standard Deviation	1.951800146	Standard Devi	1.951800146	Standard Deviation	
Sample Variance	3.80952381	Sample Variar	3.80952381	Sample Variance	
Kurtosis	-0.464423077	Kurtosis	-0.464423077	Kurtosis	
Skewness	0.123160466	Skewness	0.123160466	Skewness	
Range	7	Range	7	Range	
Minimum	2	Minimum	2	Minimum	
Maximum	9	Maximum	9	Maximum	
Sum	80	Sum	80	Sum	
Count	15	Count	15	Count	

Fig. 3.10 Output area. Output from dragging the lower right hand corner of the output area

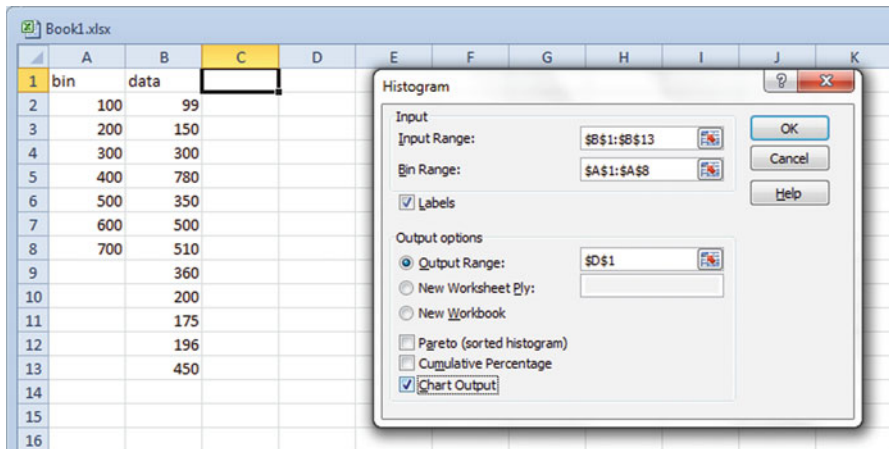
- ◆ Too many decimal places. Excel will calculate as many decimal places as you indicate in the output format. If no decimal limit is set, Excel will fill the cell and show nine decimal places. In Fig. 3.11 the input data does not have any decimal places and yet the Excel output is reported in nine decimal places. This obviously does not make any sense; Excel is reporting a greater number of significant figures than we had recorded in the data. You cannot magically create a higher level of accuracy and precision in your data than exists in the data you are using for the analysis. The rule of thumb is to allow one more decimal place in the output than exists in the input data.

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

D	E	F	G
5			
6		<i>Column1</i>	
7			
8	Mean		5.333333333
9	Standard Error		0.503952631
3	Median		5
4	Mode		5
5	Standard Deviation		1.951800146
6	Sample Variance		3.80952381
2	Kurtosis		-0.464423077
3	Skewness		0.123160466
4	Range		7
5	Minimum		2
6	Maximum		9
7	Sum		80
	Count		15

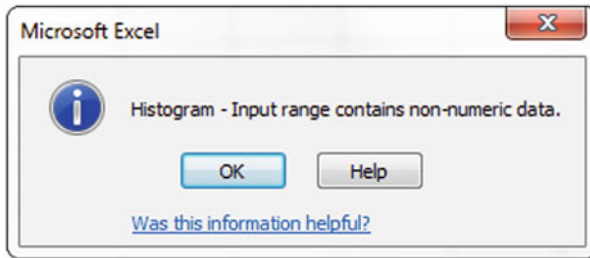
Fig. 3.11 Decimal places

◆ If you check **Labels** in the histogram input window, you need to include a label in your end point bin data; otherwise, Excel will assume the first bin end point is a label and will not use this value in the calculations.



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

- ◆ If you don't check the Labels box, and you include labels in your input range, Excel will create a warning message.



Histogram Checklist.

Do not include a legend unless it is really necessary.

Label both axes with units and variable names.

Make the graph easy to read; clean it up.

If you will be copying the graph in black and white make sure you can still differentiate data without the color. You can use different patterns rather than different colors.

Final Thoughts and Activities

Practice Problems

1. The data in the **Beverage** file represents the amount of fluid in a sample of fifty 2-l bottles. Using the data from the file, complete the following exercises.
 - (a) Construct a frequency distribution.
 - (b) From the output, what is the mean?
 - (c) From the output, what is the standard deviation?
2. The data in the **Electric** file represents the cost of electricity in June for a random sample of fifty one-bedroom apartments in a large city in the southern portion of the United States. Using the data from the file complete the following exercises.
 - (a) Construct a frequency distribution.
 - (b) Construct a histogram.
 - (c) Around what amount does the monthly electricity cost seem to center?

Discussion Boards

1. Today no one wants to be considered average but in 1835 the average man was a symbol of an egalitarian society. According to Quetelet, a French statistician in the nineteenth century, “If an individual at any given epoch in society possessed all the qualities of the average man, he would represent all that is good, great or beautiful.” How would you define the average employee in your current department?
2. Are TV commercials getting longer, or does it just seem that way? In 1990, 6 % of commercials were a maximum of 1 min long. Recently, out of a sample of 50 commercials only 16 were 1 min or shorter.
3. Class size Paradox. There are two ways to obtain a mean class size and you need to choose one for your training department’s report. In one approach you take the number of employees in all 25 seminars offered this year and average those results. The second approach is on an employee by employee basis, compile a list of class sizes for all seminars he or she attended during the year and average those results. The results for the two approaches will be different. Discuss.

Group Activity

1. The Central Intelligence Agency has specialists who analyze the frequencies of letters of the alphabet in an attempt to decipher intercepted messages. For example in Standard English text the letter “r” is used at the rate of 7.7 %. Choose another letter and using the Internet determine its usage rate. In an intercepted message sent to Iraq, a page of 2,000 characters is found to have the letter “f” occurring 42 times. Discuss if this is unusual.
2. What is meant by “Six degrees of separation” and “The Small World Problem?”
3. The Cost of Laughing Index (CLI) is developed using the same approach as is used to develop the Consumer Price Index. While standard scores and percentiles allow us to compare different values, they do not take into account time. Index numbers such as the CLI allow you to compare the current value to a value at some base time period. Describe the CLI methodology and what it is this year in the United States.

Parting Thought

Figures don’t lie, but liars can figure.

Problem Solutions

1. The data in the **Beverage** file represents the amount of fluid in a sample of fifty 2-l bottles. Using the data from the file, complete the following exercises.

(a) Construct a frequency distribution.

Bin	Frequency	Amount	
2.25	1	Mean	2.35672
2.28	1	Standard error	0.006302254
2.31	6	Median	2.36
2.34	11	Mode	2.368
2.37	15	Standard deviation	0.044563662
2.40	9	Sample variance	0.00198592
2.43	5	Kurtosis	0.080400572
More	2	Skewness	0.008832396
		Range	0.215
		Minimum	2.25
		Maximum	2.465
		Sum	117.836
		Count	50

(b) From the output, what is the mean?

Answer: Mean: 2.36 bottles but we can't have .36 bottles so just 3 bottles

(c) From the output, what is the standard deviation?

Answer: sd = 0.04 bottles

2. The data in the **Electric** file represents the cost of electricity in June for a random sample of fifty one-bedroom apartments in a large city in the southern portion of the United States. Using the data from the file complete the following exercises.

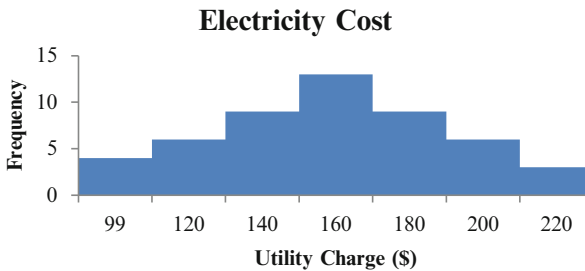
Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

- (a) Construct a frequency distribution.

Bin	Frequency
85	1
103.71	3
122.43	7
141.14	8
159.86	12
178.57	10
197.29	5
More	4

Utility charge	
Mean	150.06
Standard error	4.481837269
Median	151.5
Mode	133
Standard deviation	31.69137525
Sample variance	1004.343265
Kurtosis	-0.544163238
Skewness	0.015845641
Range	131
Minimum	85
Maximum	216
Sum	7,503
Count	50

- (b) Construct a histogram.



- (c) Around what amount does the monthly electricity cost seem to center around?

Answer: The majority of utility charges are clustered around \$160.