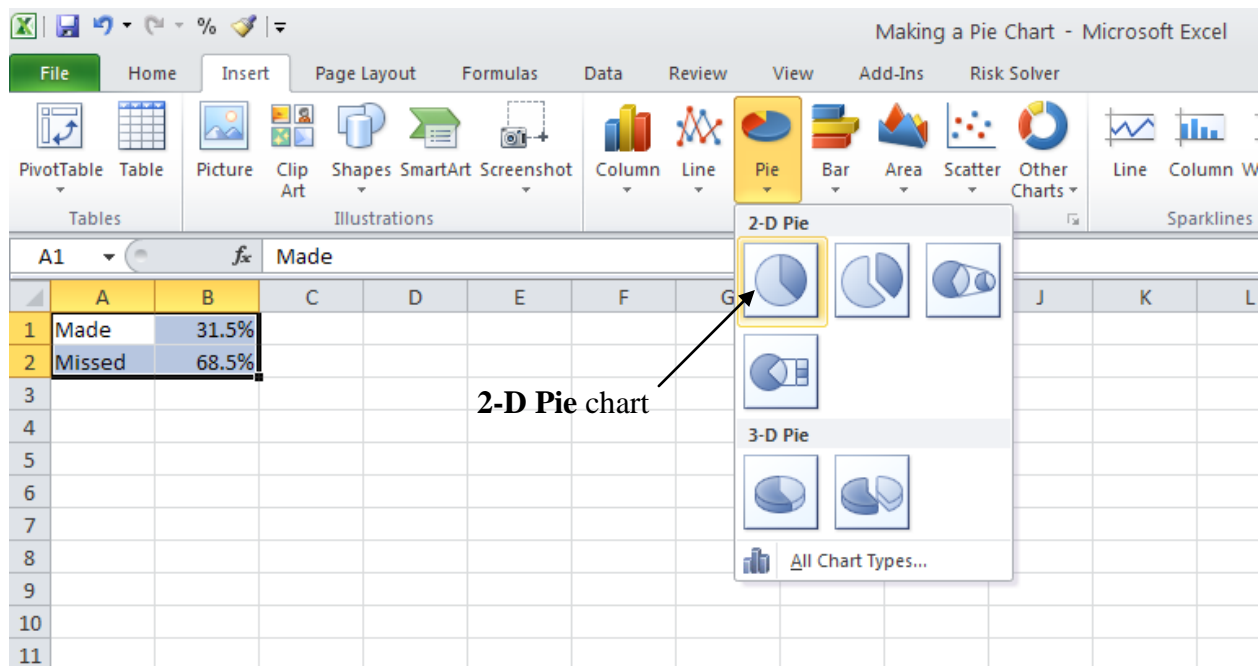# Appendix B: Using Excel 2010

Chapter 1: Making Pie Charts

Follow these steps for creating a pie chart using LeBron James' 2007–2008 regular season three-point shooting *PERFORMANCE* (31.5% made and 68.5% missed).

1. Enter the following labels and data using the indicated format.

	A	B
1	Made	31.5%
2	Missed	68.5%

2. Highlight the information in cells A1 to B2, open the **Insert** menu, and click on the **Pie** chart icon. Then, select the **2-D Pie** chart in the upper-left.
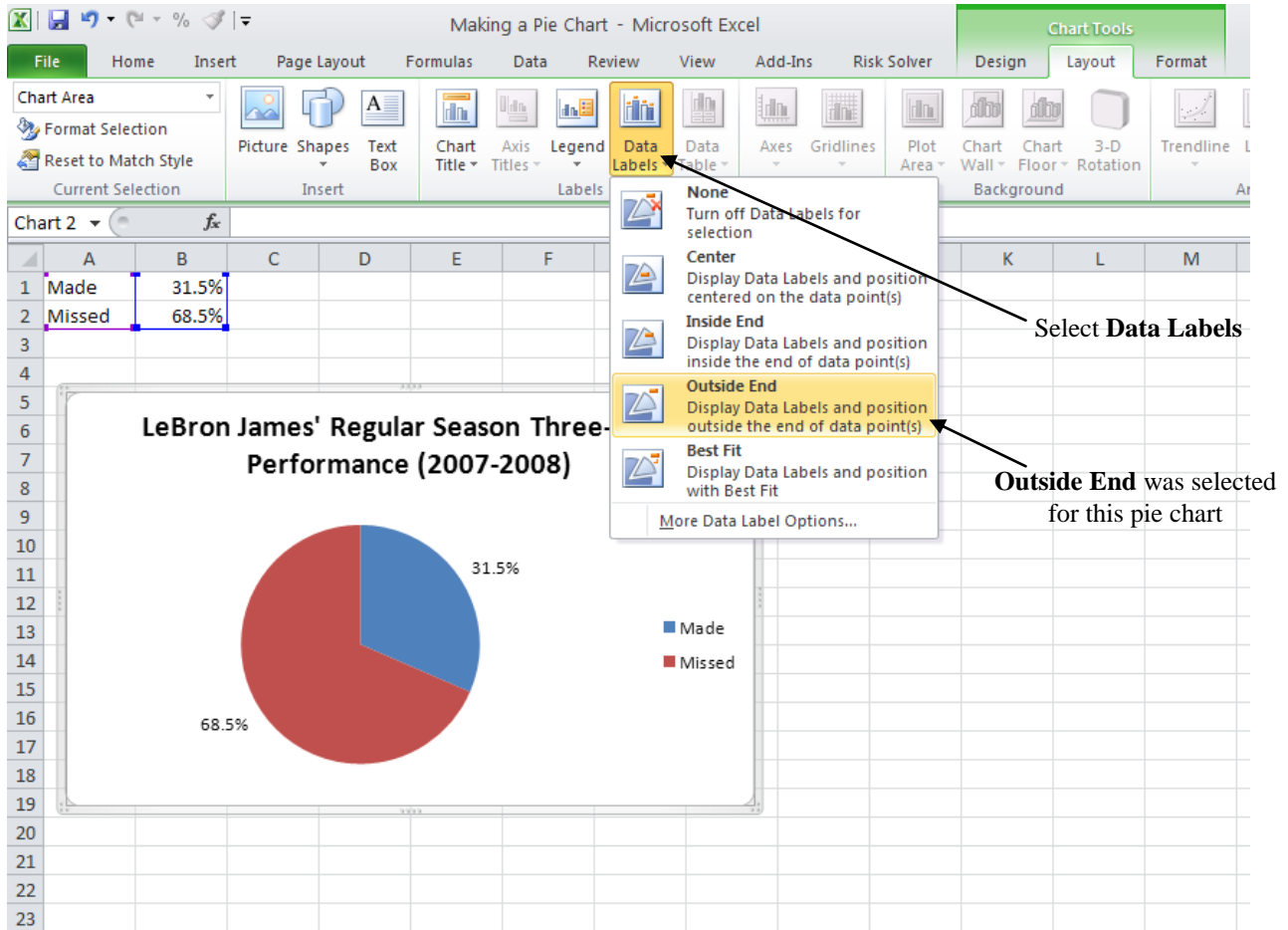

2-D Pie chart

Appendix B: Using Excel

3. The pie chart that appears will be very basic with no labels. To add labels, consider using one of the premade **Chart Layouts** by highlighting the pie chart and clicking the **Design** tab of the **Chart Tools** that appear at the top of the window. Or, select the **Layout** tab of the **Chart Tools** to add individual features yourself. For example, after choosing the **Layout** tab, there are buttons for **Chart Title**, **Legend**, and **Data Labels**. Select the **Chart Title** tab and type *LeBron James' Regular Season Three-Point Performance (2007-2008)*.

Layout tab

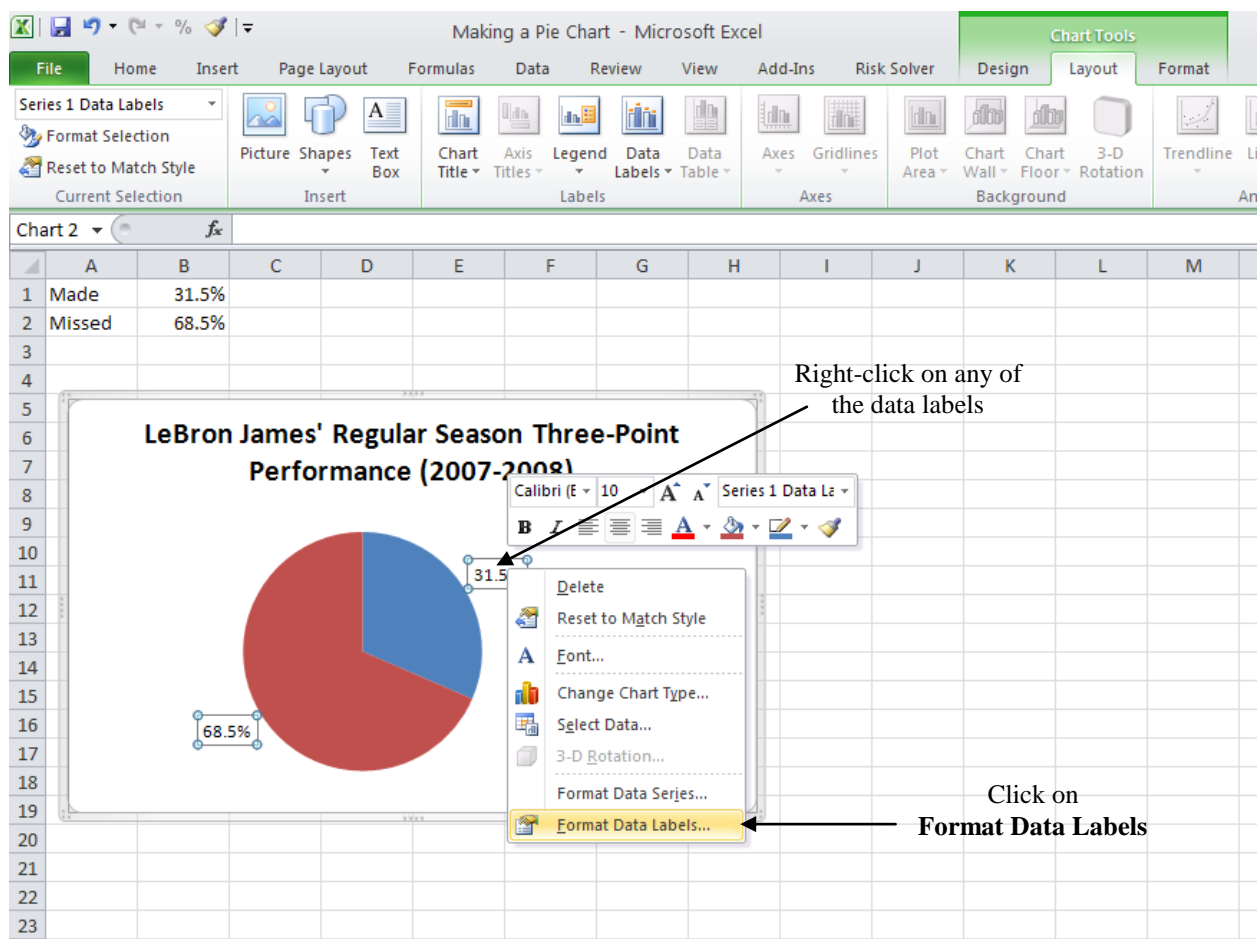Use these to add **Chart Title**, **Labels**, etc.
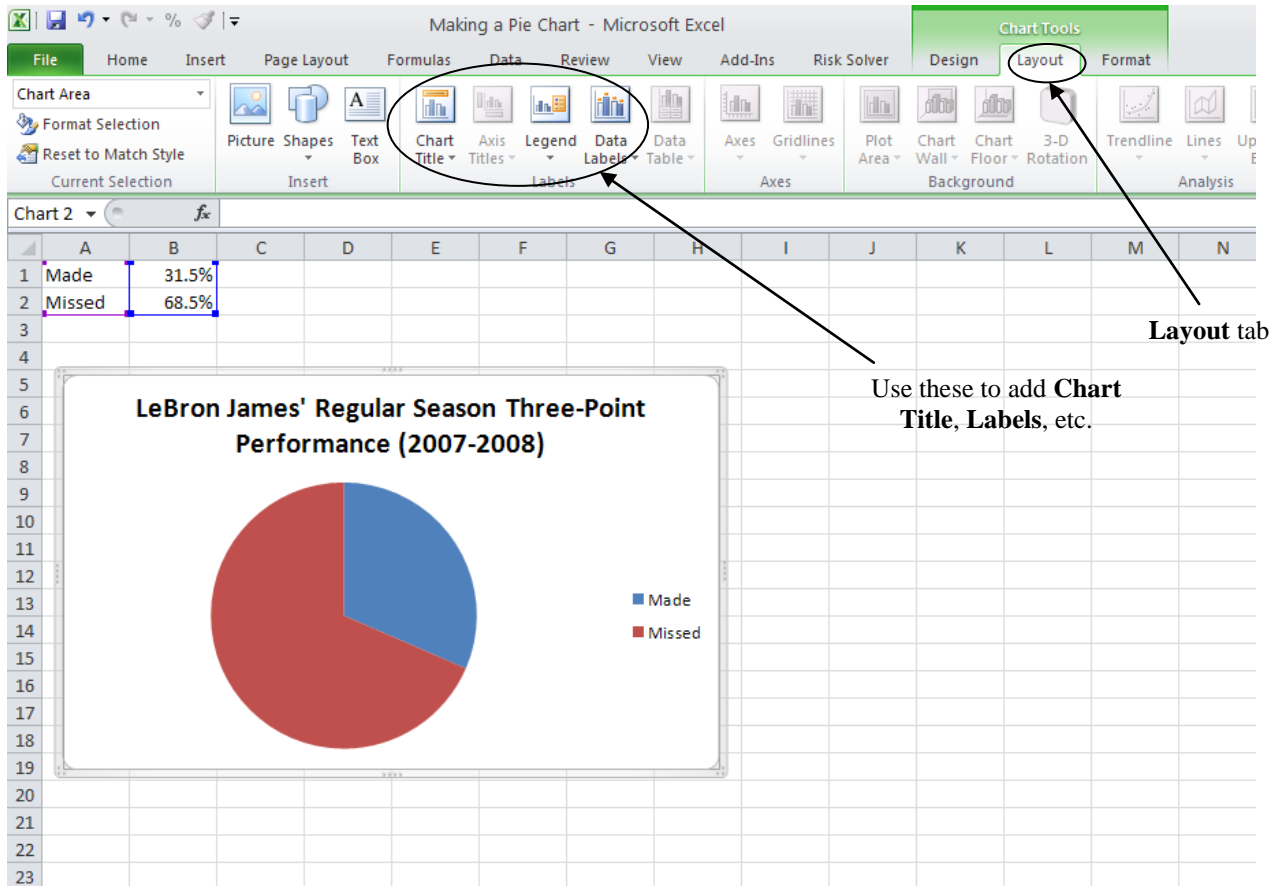
Appendix B: Using Excel

4. To display the data value, select the **Layout** tab of the **Chart Tools**, and click on the **Data Labels**. For this pie chart, **Outside End** was selected to display the data label outside the end of the data points.

Appendix B: Using Excel

5. Right-click on any of the data labels and click on **Format Data Labels**.

Appendix B: Using Excel

6. Select from a variety of Label Options. For this pie chart, the **Category Name** and (display) **Value** boxes were checked. When counts (and not percentages) are used to construct the pie chart, EXCEL will display the corresponding percentages when the **Percentage** box is checked.

Check **Category Name** and **Value** boxes
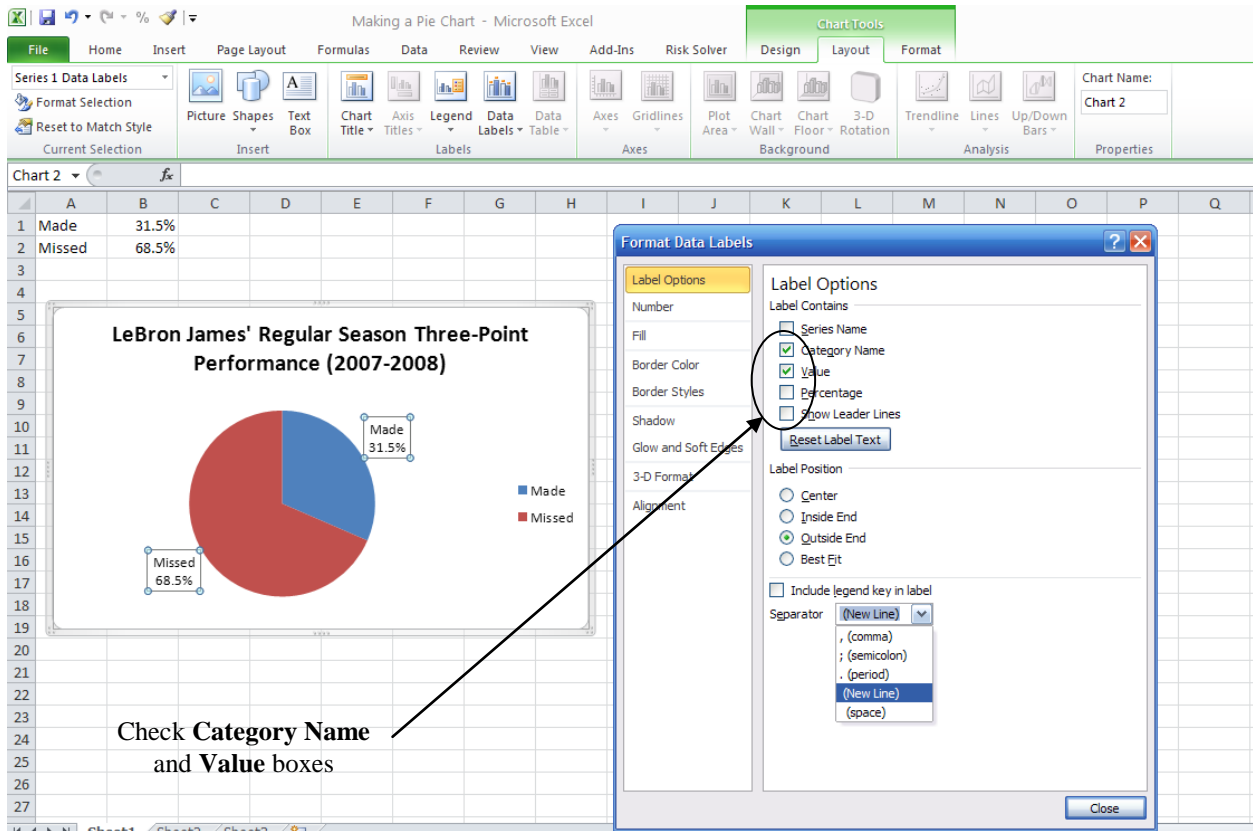
7. To eliminate the legend (if desired), select the **Layout** tab of the **Chart Tools**. Select the **Legend** tab and choose **None**.
8. Here is an example of a well-labeled pie chart.
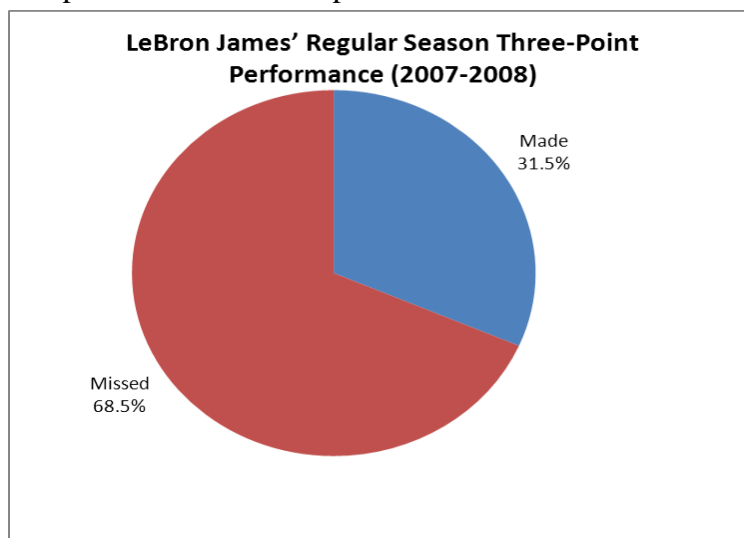
Chapter 1: Making Segmented Bar Charts

Follow these steps for creating a segmented bar chart using LeBron James' 2007–2008 regular season *PERFORMANCE* (31.5% made, 68.5% missed) and playoff *PERFORMANCE* (25.7% made, 74.3% missed).

1. Enter the following labels and data using the indicated format.

	A	B	C
1		Regular	Playoffs
2	Made	31.5%	25.7%
3	Missed	68.5%	74.3%

2. Highlight the information in cells A1 to C3, open the **Insert** menu, and click on the **Column** icon. Then, select the **100% Stacked Column** chart in the upper-right.
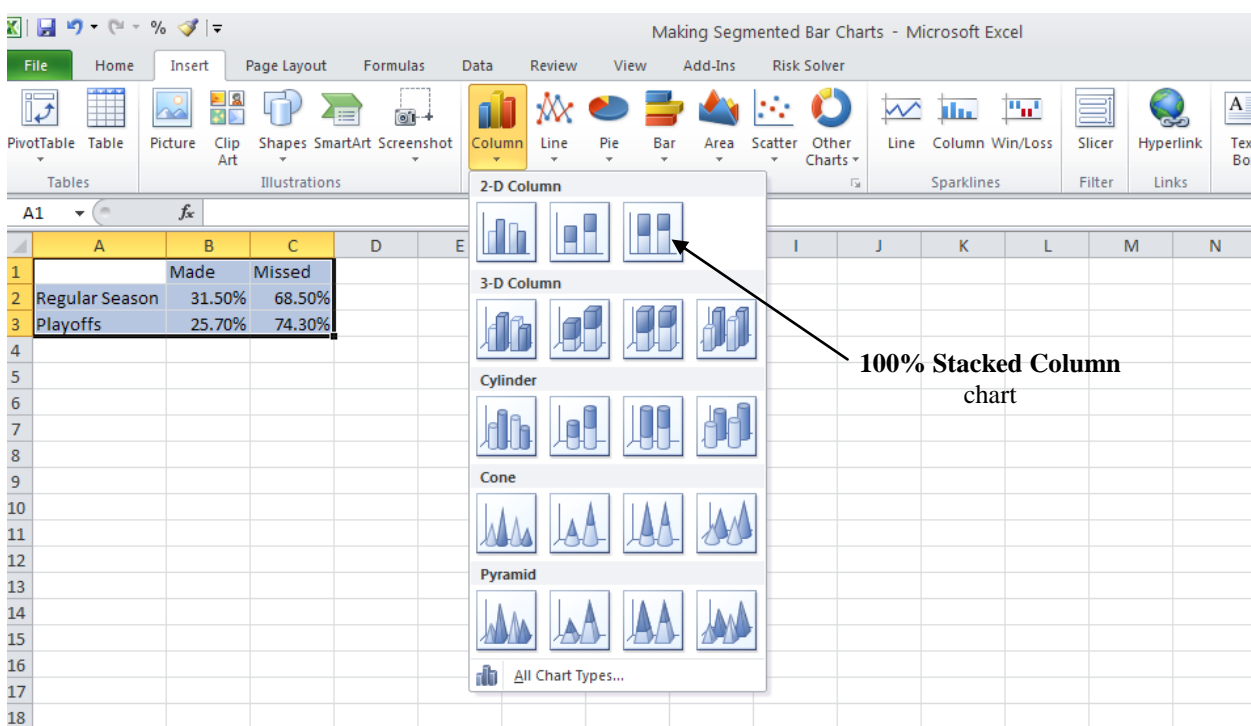

3. The segmented bar chart will identify the columns for *Made* and *Missed* as well as a legend identifying columns for *Regular Season* and *Playoffs*. To add titles and data labels, select the **Layout** tab of the **Chart Tools** to add individual features yourself. For example, after choosing the **Layout** tab, there are buttons for **Chart Title**, **Axis Titles**, **Legend**, and **Data Labels**. Select the **Chart Title** tab and type *LeBron's Performance*. To label the vertical axis, select **Axis Titles**, then **Primary Vertical Axis Title**, then **Rotated Title** and type *Percent*. To add the data labels to the top of a particular set of columns (as pictured), select **Data Labels**, then **Outside End**.

Appendix B: Using Excel
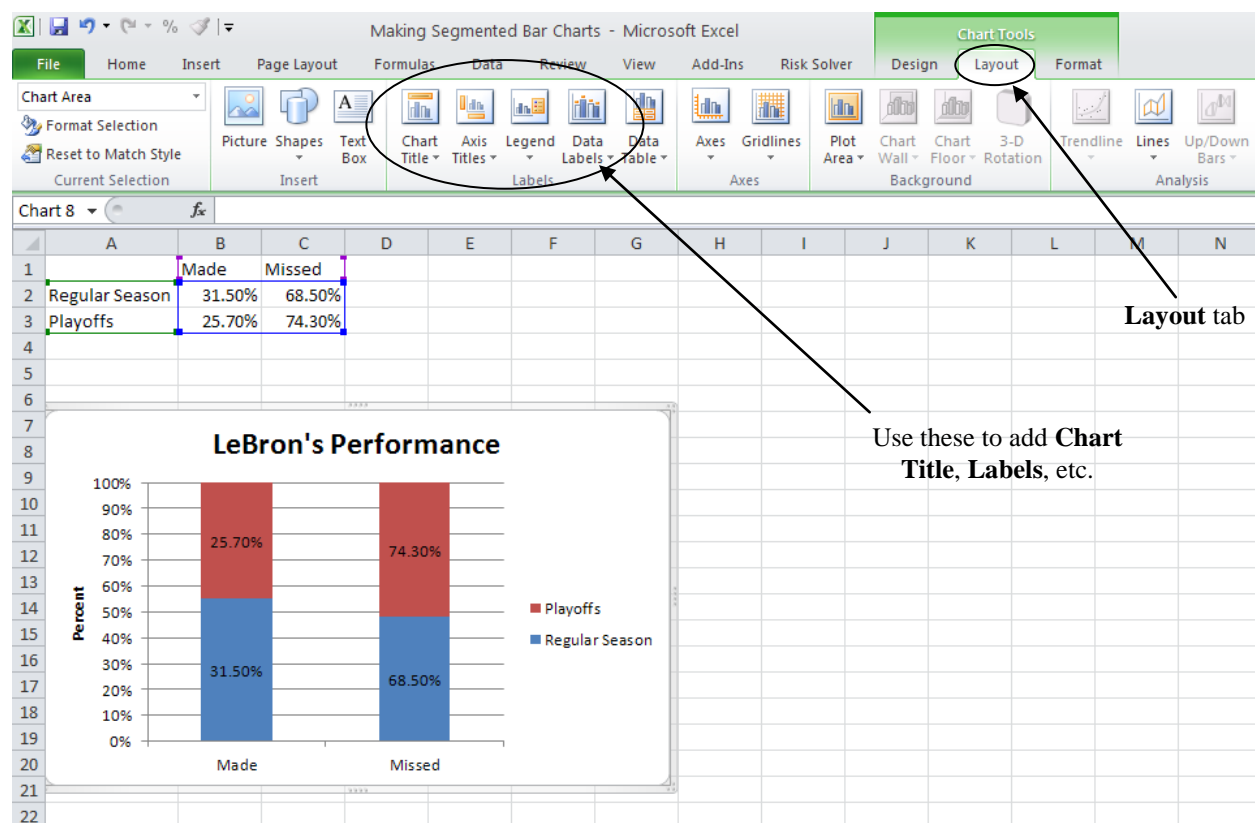
4. To eliminate the vertical axis major grid lines (if desired), right-click on any one of the lines and choose **Delete**. If a different background color is desired, right-click on the background and choose **Format Plot Area**. Select **Fill**, then the desired options.

5. Here is an example of a well-labeled segmented bar chart.
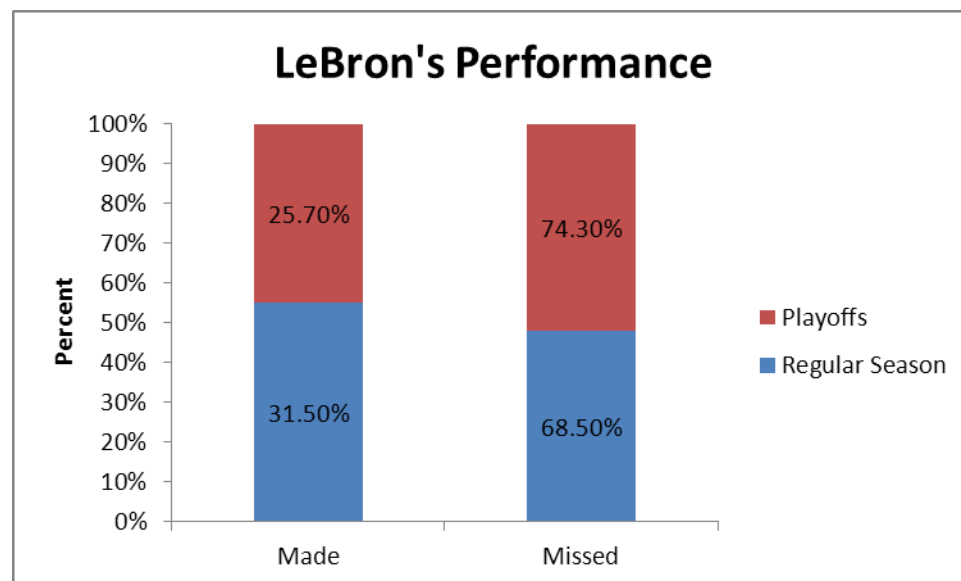
Chapter 1: Making Comparative Bar Charts

Follow these steps for creating a comparative bar chart using LeBron James' 2007–2008 regular season PERFORMANCE (31.5% made, 68.5% missed) and playoff PERFORMANCE (25.7% made, 74.3% missed).

1. Enter the following labels and data using the indicated format.

	A	B	C
1		Made	Missed
2	Regular Season	31.5%	68.5%
3	Playoffs	25.7%	74.3%

2. Highlight the information in cells A1 to C3, open the **Insert** menu, and click on the **Column** icon. Then, select the first chart sub-type (**Clustered Column**).
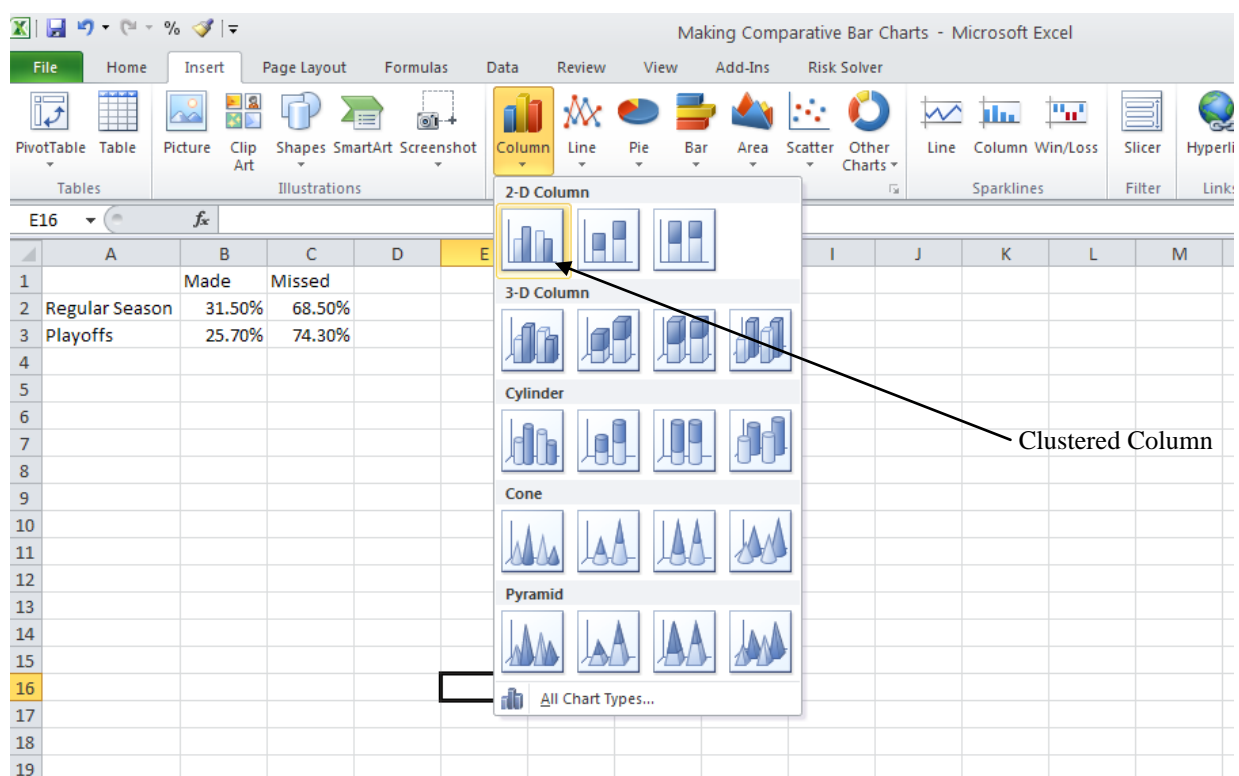

3. The comparative bar chart will identify the columns for *Made* and *Missed* as well as a legend identifying columns for *Regular Season* and *Playoffs*. To add titles and data labels, select the **Layout** tab of the **Chart Tools** to add individual features yourself. For example, after choosing the **Layout** tab, there are buttons for **Chart Title**, **Axis Titles**, **Legend**, and **Data Labels**. Select the **Chart Title** tab and type *LeBron's Performance*. To label the vertical axis, select **Axis Titles**, then **Primary Vertical Axis Title**, then **Rotated Title** and type *Percent*. To add the data labels to the top of a particular set of columns (as pictured), select **Data Labels**, the **Outside End**.

Appendix B: Using Excel
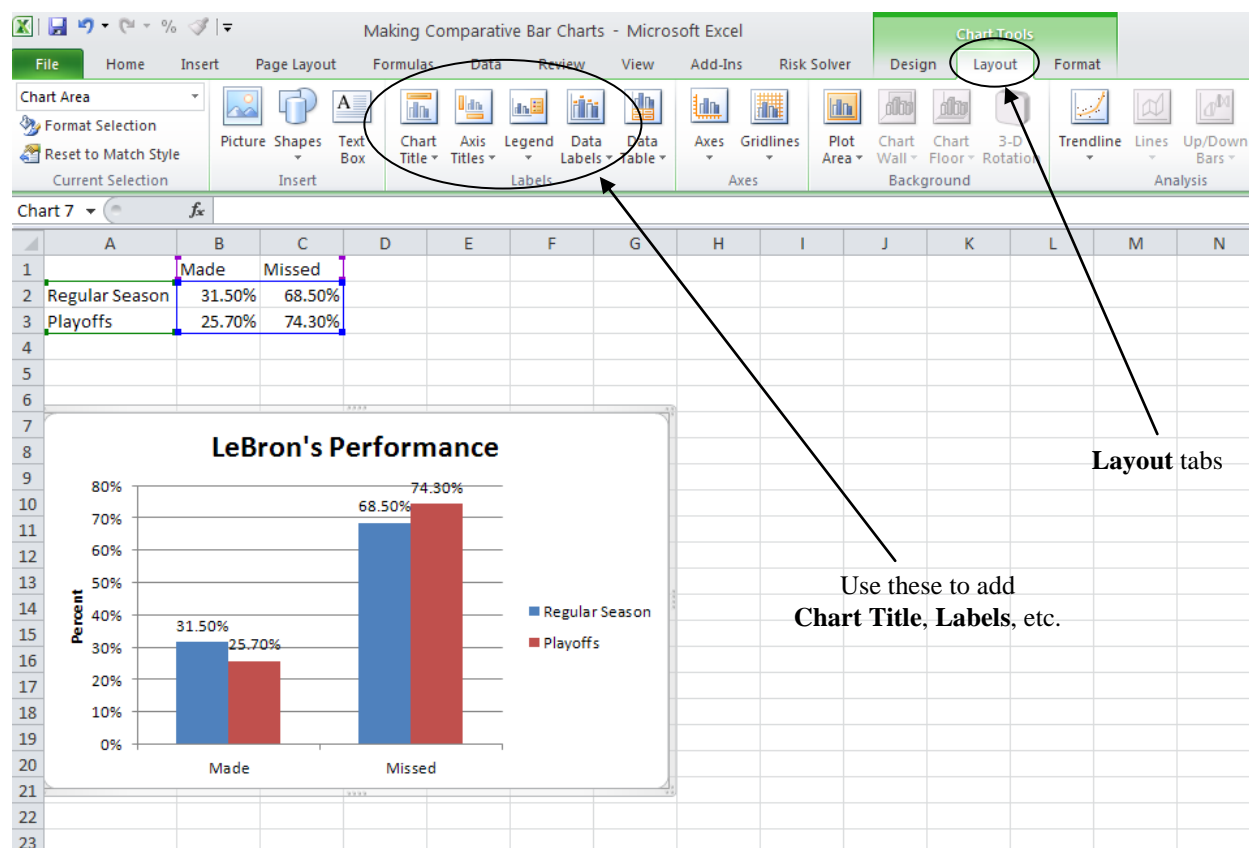
Use these to add
Chart Title, **Labels**, etc.

Layout tabs

4. To eliminate the vertical axis major grid lines (if desired), right-click on any one of the lines and choose **Delete**. If a different background color is desired, right-click on the background and choose **Format Plot Area**. Select **Fill**, then the desired options.

5. Here is an example of a well-labeled comparative bar chart.

Appendix B: Using Excel
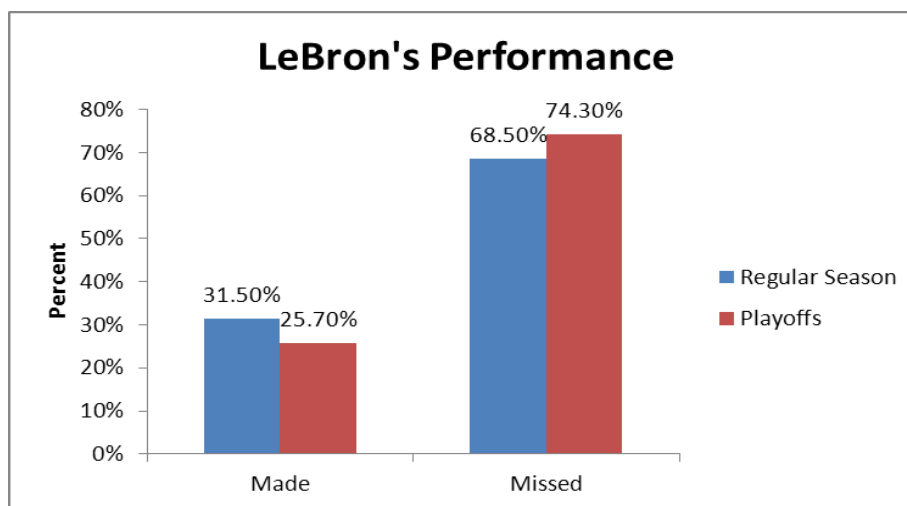
Chapter 1: Simulating Athletic *PERFORMANCE*

Follow these steps for simulating LeBron James' three-point shooting *PERFORMANCE*, assuming his *ABILITY* to make a shot is 31.5%.
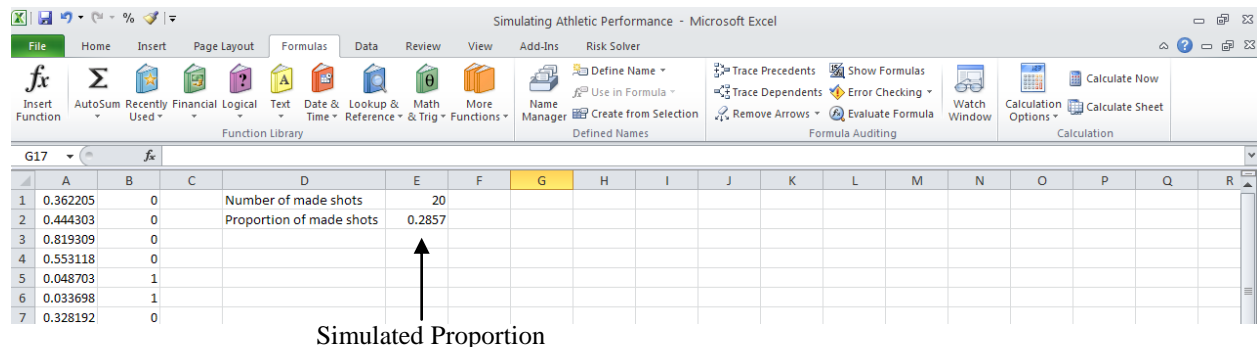
1. In cell A1, enter the formula ⎡=rand()⎤. This will generate a random number from 0 to 1.
2. In cell B1, enter the formula ⎡=if(A1<=0.315,1,0)⎤. If the random number in column A is less than or equal to 0.315 (LeBron's assumed *ABILITY*), the result will be a 1 (made shot). Otherwise, the result will be a 0 (missed shot).
3. To simulate 70 shots, copy the equations from A1 and B1 down to cells A70 and B70. To do this, highlight cells A1 and B1, drag your mouse down to cells A70 and B70 so that the cells A1 to B70 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down** as in the screen shot below.

Fill Down option

Highlight cells A1 to B70
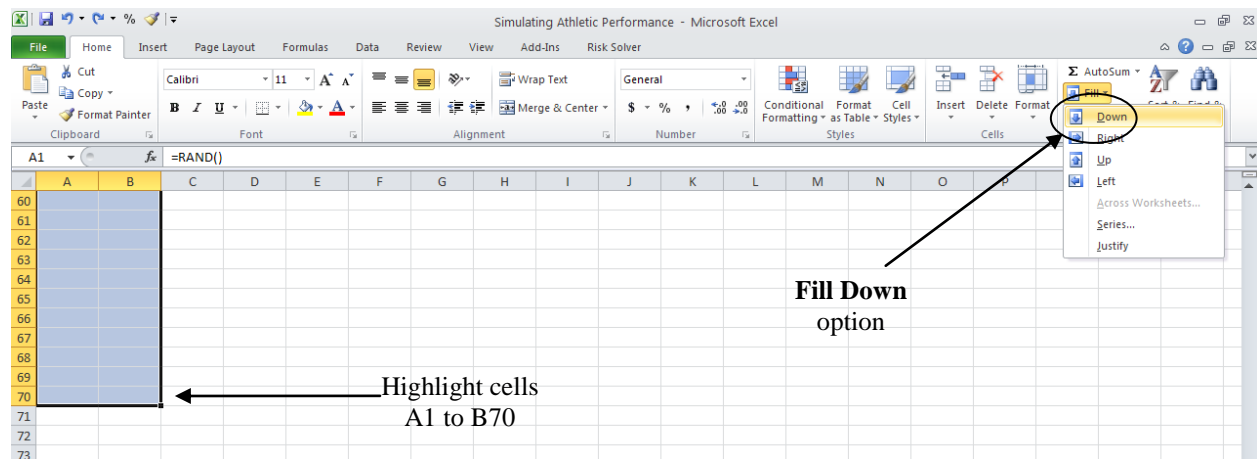
4. To display the simulated *number* of made shots and the simulated *proportion* of made shots in 70 attempts enter the following labels and formulas using the indicated format.
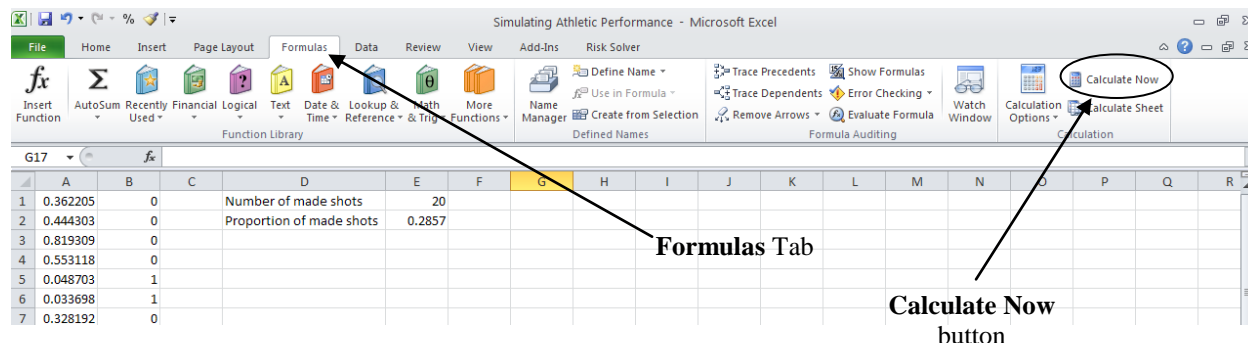
	D	E
1	Number of made shots	=sum(B1:B70)
2	Proportion of made shots	=sum(B1:B70)/70

The results for this particular simulation are shown in the screen shot below.


Simulated Proportion

5. To generate more simulated outcomes, repeatedly press the **F9** key. Or, choose the **Formulas** tab and repeatedly press the **Calculate Now** button.

Formulas Tab

Calculate Now button

Chapter 2: Simulating a Difference in Proportions

Follow these steps for simulating the difference in the 2004–2008 Arizona Cardinals' *PERFORMANCE* at home (23 wins and 17 losses) and on the road (10 wins and 30 losses).

1. In cell A1, enter the formula $\boxed{\text{=rand()}}$. This will generate a random number from 0 to 1.
2. Copy the formula in A1 down to cell A80 because there were a total of 80 games. To do this, highlight cell A1, drag your mouse down to cell A80 so that cells A1 to A80 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down** as in the screen shot below.

Fill Down option

Highlight cells A1 to A80
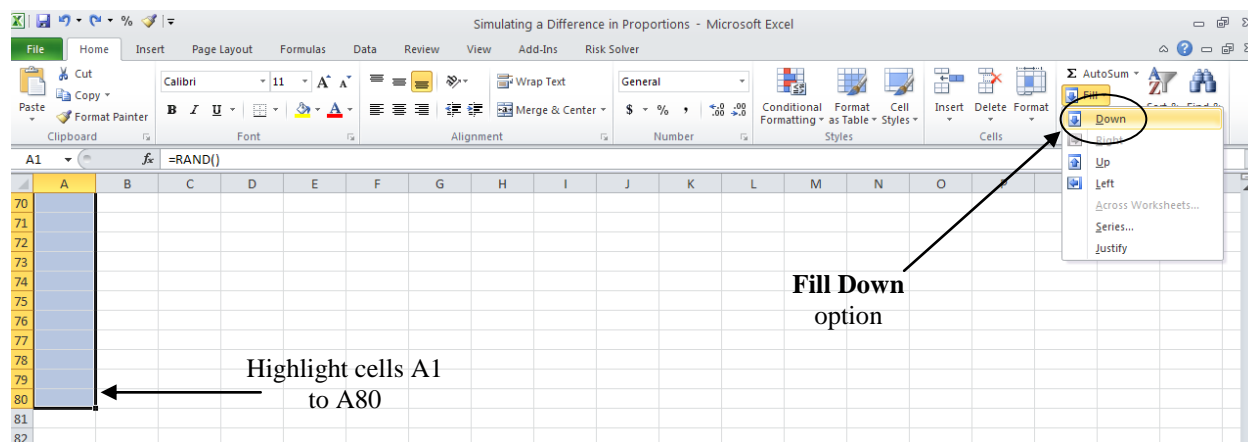
3. In cells B1 to B40, enter 23 1's and 17 0's to represent the 23 wins and the 17 losses that the Cardinals had at home. Then, in cells B41 to B80, enter 10 1's and 30 0's to represent the 10 wins and 30 losses that the Cardinals had on the road.

Appendix B: Using Excel

4. To display the proportion of wins at home, the proportion of wins on the road, and the observed difference in these proportions, enter the following labels and formulas using the indicated format.

	D	E
1	Proportion of wins at home	=SUM(B1:B40)/40
2	Proportion of wins on the road	=SUM(B41:B80)/40
3	Difference in proportions	=E1-E2

The results are shown in the screen shot below.

5. To simulate the difference in proportions, we will sort columns A and B according to the random numbers in column A. This will shuffle the 0's and 1's in column B. To do this, click on cell A1 and highlight cells A1 to B80. Then, select the **Data** tab and choose the **Sort Smallest to Largest** icon $\begin{array}{c} A \\ Z \end{array}\downarrow$. The value in cell E3 is now the *simulated difference in proportions*.

Data tab

Simulated difference in proportions

Sort Smallest to Largest icon

6. To generate more simulated differences, repeatedly press the **Sort Smallest to Largest** icon. *Note*: *Don't be confused when you notice that the random numbers in column A are not in ascending order. Each time Excel sorts the columns, it generates new random numbers for column A.*

Chapter 3: Simulating the Longest Streak

Follow these steps for simulating the longest streak for the 2008 Green Bay Packers, who won 6 games and lost 10 games.

1. In cell A1, enter the formula $\boxed{\text{=rand()}}$. This will generate a random number from 0 to 1.
2. Copy the formula in A1 down to cell A16 because there were a total of 16 games. To do this, highlight cell A1, drag your mouse down to cell A16 so that cells A1 to A16 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down** as in the screen shot below.
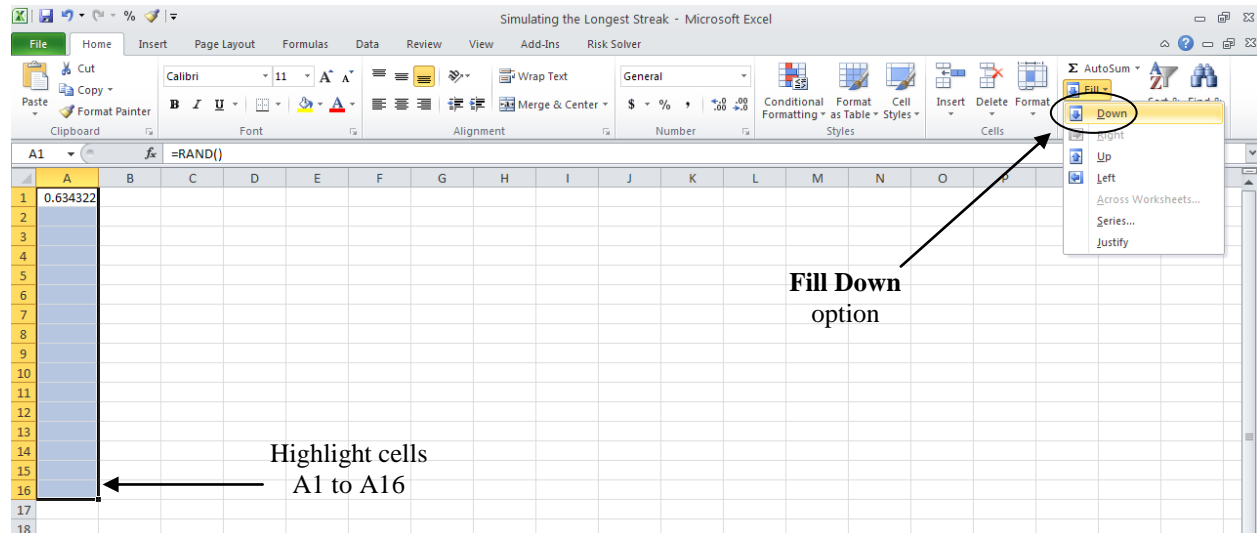
3. In column B, enter 6 1's for the Packer's 6 wins and 10 0's for the Packers 10 losses.
4. In cell C1, enter the number 1. In cell C2, enter the formula $\boxed{\text{=if(B1=B2,1+C1,1)}}$. Then, copy this formula down to cell C16 like you did in step 2. This will display the length of the current streak after each game.
5. To display the length of the longest streak, enter the following labels and formulas using the indicated format.

|	E	F
1	Longest streak	=max(C1:C16)

The results are shown in the screen shot below.

Appendix B: Using Excel

Length of longest streak

Length of current streak after each game

6. To simulate the length of the longest streak, we will sort columns A and B according to the random numbers in column A. This will shuffle the 0's and 1's in column B. To do this, click on cell A1 and highlight cells A1 to B16. Then, select the **Data** tab and choose the **Sort Smallest to Largest** icon $\begin{array}{c} A \\ Z \end{array} \downarrow$. The simulated length of the longest streak will now be displayed in cell F1.

Data tab

Sort Smallest to Largest icon

Simulated length of longest streak

7. To generate more simulated longest streaks, repeatedly press the **Sort Smallest to Largest** icon. *Note*: *Don't be confused when you notice that the random numbers in column A are not in ascending order. Each time Excel sorts the columns, it generates new random numbers for column A.*

14

Chapter 3: Simulating the Number of Streaks

Follow these steps for simulating the number of streaks of length 4 or more for Mario Lemieux, who scored in 49 of his 76 games and didn't score in 27 of his 76 games in the 1988–1989 regular season.

1. In cell A1, enter the formula =rand() . This will generate a random number from 0 to 1.
2. Copy the formula in A1 down to cell A76 because there were a total of 76 games. To do this, highlight cell A1, drag your mouse down to cell A76 so that cells A1 to A76 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down** as in the screen shot below.
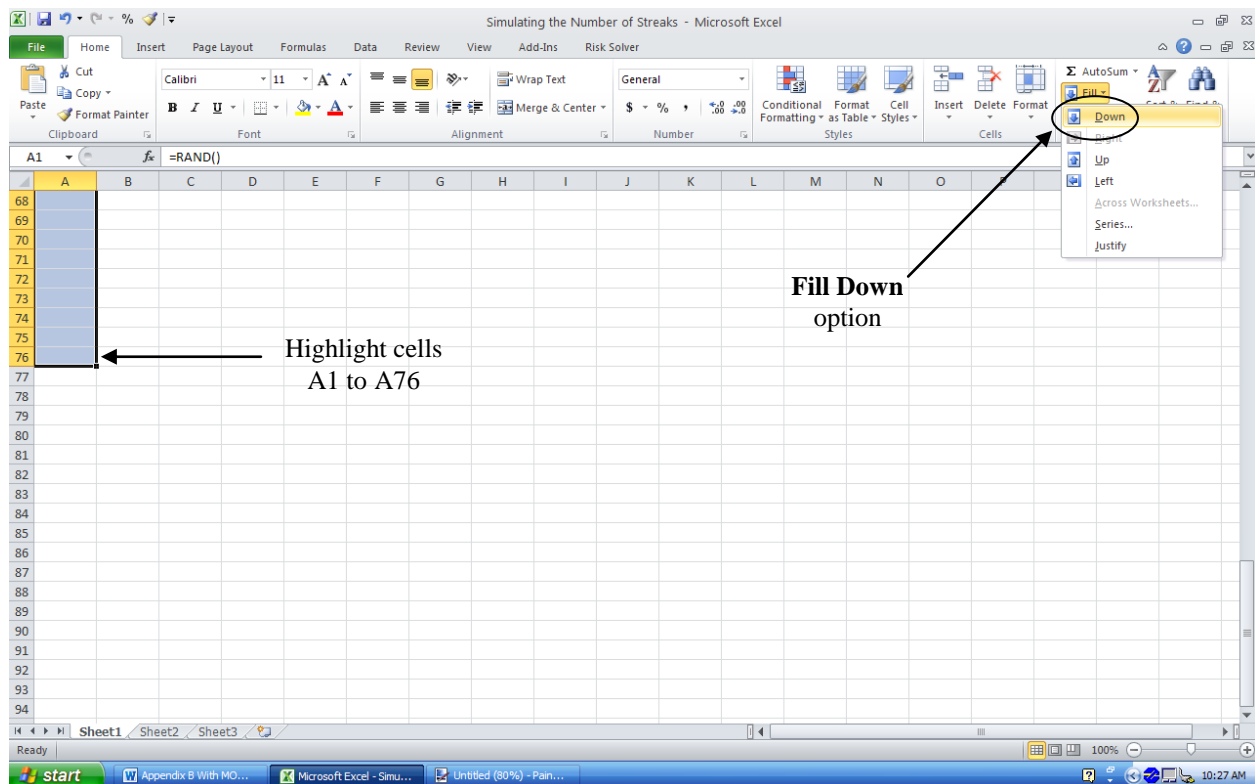

3. In cells B1 to B49, enter 49 1's for the 49 games where he scored a goal. Then, in cells B50 to B76, enter 27 0's for the 27 games where he didn't score a goal.
4. In cell C1, enter the number 1. In cell C2, enter the formula =if(B1=B2,1+C1,1) . Then, copy this formula down to cell C76 like you did in step 2. This will display the length of the current streak after each game.
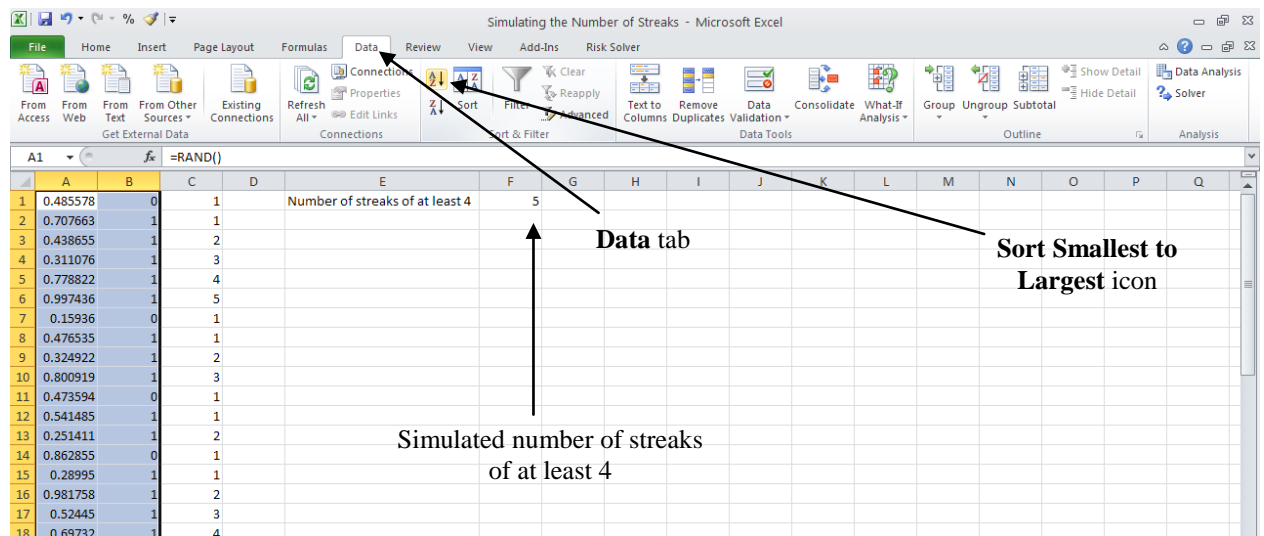
Appendix B: Using Excel

5. To display the number of streaks of at least 4, enter the following labels and formulas using the indicated format.

	E	F
1	Number of streaks of at least 4	=countif(C1:C76,4)

If you are looking for streaks of a different length, change the last number in the formula to the number you are looking for. The results are shown in the screen shot below. In this example there are only 2 streaks of at least 4 (one of 49 and one of 27).

Number of streaks of at least 4

Length of current streak after each game
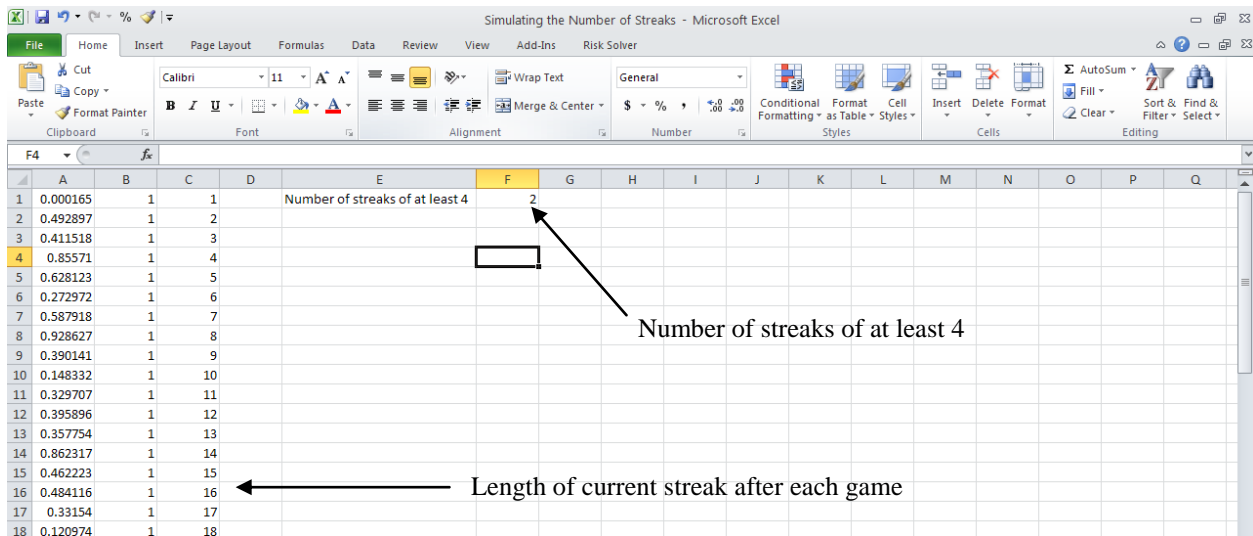
6. To simulate the number of streaks, we will sort columns A and B according to the random numbers in column A. This will shuffle the 0's and 1's in column B. To do this, click on cell A1 and highlight cells A1 to B76. Then, select the **Data** tab and choose the **Sort Smallest to Largest** icon A↓/Z. The simulated number of streaks of at least 4 will now be displayed in cell F1.

Data tab

Sort Smallest to Largest icon

Simulated number of streaks of at least 4

7. To generate more simulated number of streaks, repeatedly press the **Sort Smallest to Largest** icon. *Note: Don't be confused when you notice that the random numbers in column A are not in ascending order. Each time Excel sorts the columns, it generates new random numbers for column A.*

Chapter 4: Calculating Summary Statistics*

Follow these steps for calculating summary statistics for the number of runs scored by the 14 AL teams in 2008 (782, 845, 811, 805, 821, 691, 765, 829, 789, 646, 671, 774, 901, 714).
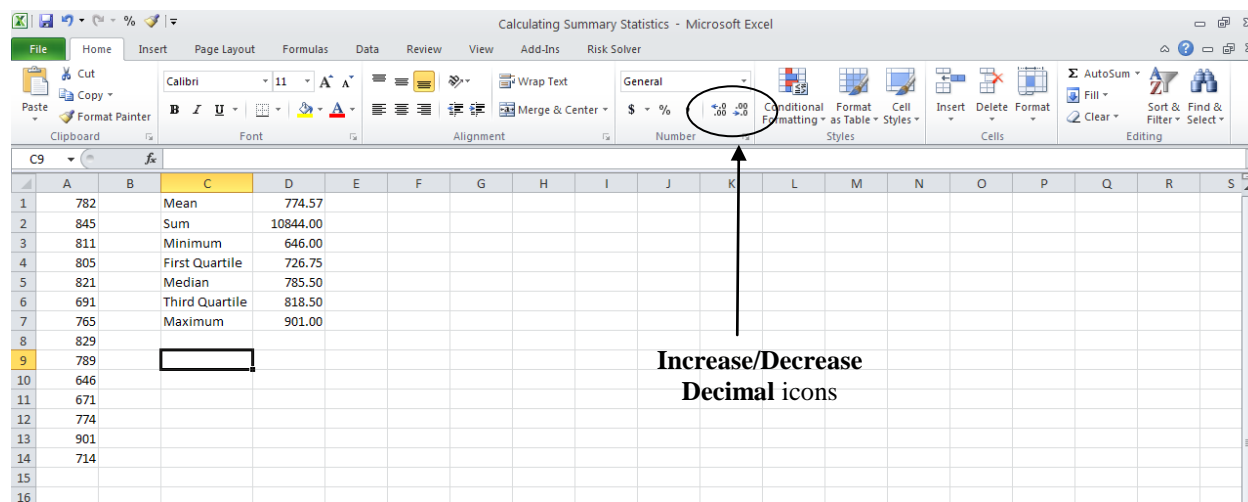
1. Enter the 14 run totals in cells A1 to A14.
2. Enter the following labels and formulas using the indicated format.

	C	D	
1	Mean	=average(A1:A14)	
2	Sum	=sum(A1:A14)	
3	Minimum	=min(A1:A14)	
4	First Quartile	=quartile(A1:A14,1)	**
5	Median	=median(A1:A14)	
6	Third Quartile	=quartile(A1:A14,3)	**
7	Maximum	=max(A1:A14)	

The results are shown in the screen shot below.

3. To display more (or fewer) decimal places, highlight cells D1 to D7 and use the **Increase/Decrease Decimal** icons found in the **Home** tab as shown above.

*As a reminder, Excel is not a preferred tool for making graphs of numerical data unless you have special add-ins or macros.

**The method Excel uses to calculate the quartiles is slightly different than the method we describe in the chapter.

Appendix B: Using Excel

Chapter 5: Simulating a Difference in Means (or Medians*)

Follow these steps for simulating the difference in the mean distance kicked for the 5 warm footballs (61, 58, 63, 55, 60) and the 5 cold footballs (60, 52, 53, 57, 59).

1. In cell A1, enter the formula =rand() . This will generate a random number from 0 to 1.
2. Copy the formula in A1 down to cell A10 because there were a total of 10 kicks. To do this, highlight cell A1, drag your mouse down to cell A10 so that cells A1 to A10 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down**.
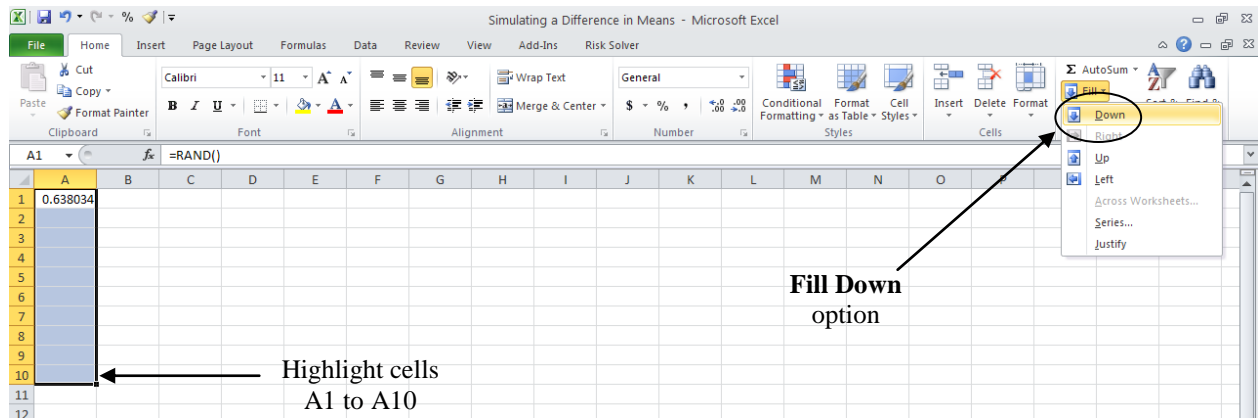

3. In cells B1 to B10, enter the 10 distances for the 10 kicks.
4. To display the average distance of kicks with the warm ball, the average distance of kicks with the cold ball, and the observed difference in these mean, enter the following labels and formulas using the indicated format.

|	D	E
1	Average distance of kicks with the warm ball	=average(B1:B5)
2	Average distance of kicks with the cold ball	=average(B6:B10)
3	Difference in means	=E1-E2

The results are shown in the screen shot below.

18
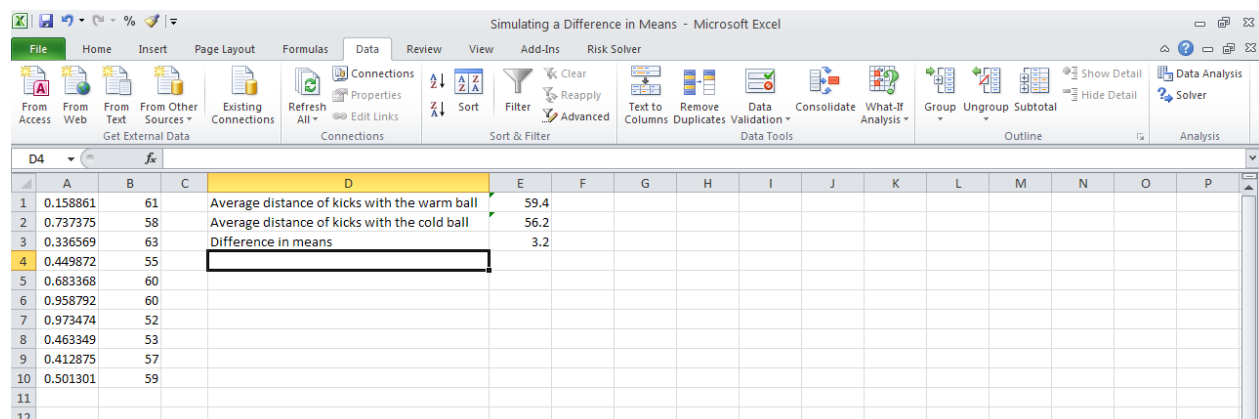
5. To simulate the difference in means, we will sort columns A and B according to the random numbers in column A. This will shuffle the distances in column B. To do this, click on cell A1 and highlight cells A1 to B10. Then, select the **Data** tab and choose the **Sort Smallest to Largest** icon $\boxed{\begin{array}{c} A \\ Z \end{array} \downarrow}$. The value in cell E3 is now the *simulated difference in means*.
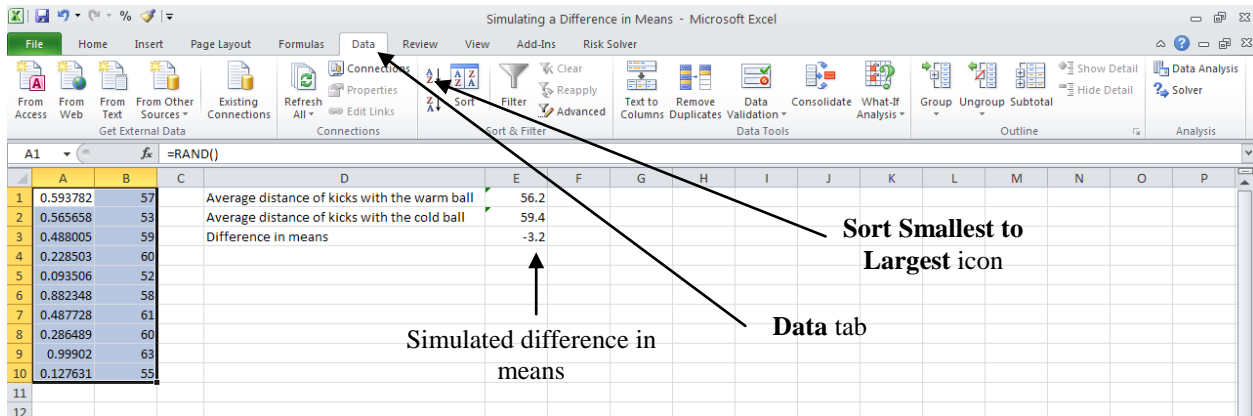
6. To generate more simulated differences, repeatedly press the **Sort Smallest to Largest** icon. *Note*: *Don't be confused when you notice that the random numbers in column A are not in ascending order. Each time Excel sorts the columns, it generates new random numbers for column A.*

*To simulate the difference in medians, change the labels and formulas in cells D1 to E3 using the format below.

	D	E
1	Median distance of kicks with the warm ball	=median(B1:B5)
2	Median distance of kicks with the cold ball	=median(B6:B10)
3	Difference in medians	=E1-E2

Chapter 6: Simulating a Mean Difference

Follow these steps for simulating the mean difference in the number of home runs hit by Major League Baseball from 2006 to 2008 using the observed differences (–1, –92, 8, –19, 18, –1, –30, –25, 3, –3, 26, –7, –4, 0, –16, 18, –32, –28, –30, –50, –2, 12, –7, –69, –48, –10, –10, 11, –73, –47)

1. In cell A1, enter the formula =rand() . This will generate a random number from 0 to 1.
2. Copy the formula in A1 down to cell A30 because there were a total of 30 differences. To do this, highlight cell A1, drag your mouse down to cell A30 so that cells A1 to A30 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down**.
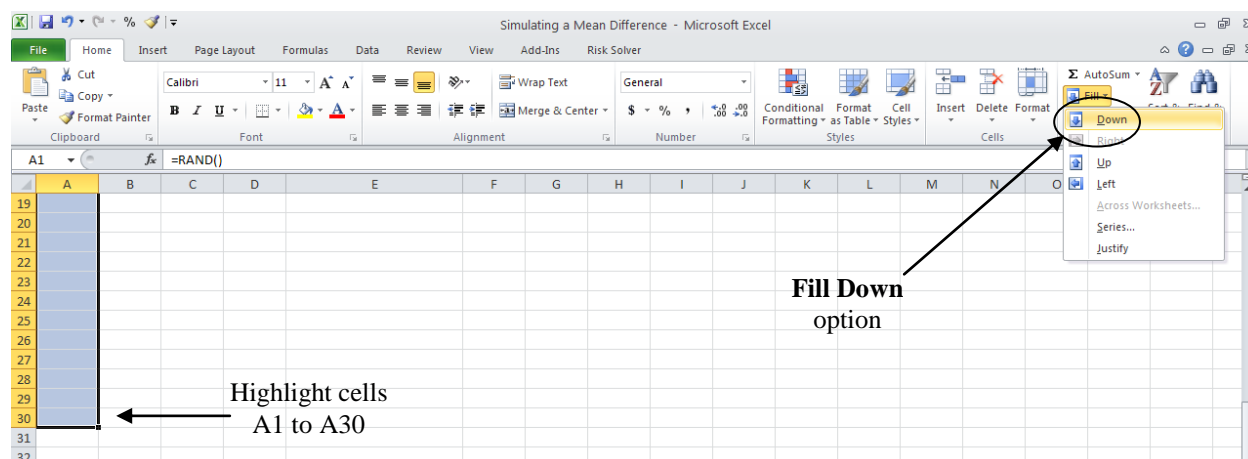

3. In cells B1 to B30, enter the 30 observed differences for the 30 MLB teams.
4. In cell C1, enter the formula if(A1<0.5,B1,-B1) . Then, copy the formula in C1 down to C30 like you did in step 2. This will make the simulated difference positive half the time and negative half the time, with the sign determined at random.
5. To display the mean difference, enter the following label and formula using the indicated format.
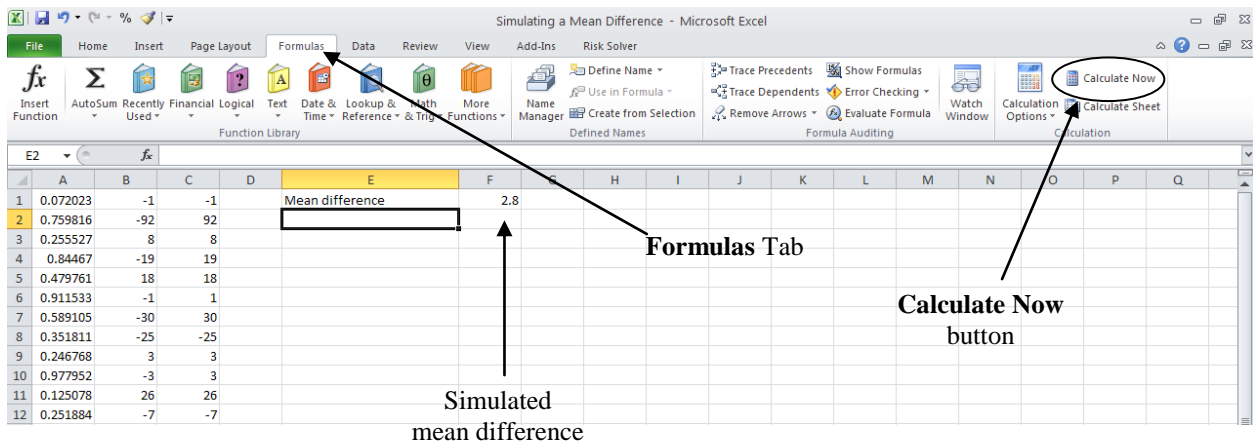
	E	F
1	Mean difference	=average(C1:C30)

The result is shown in the screen shot below.

Appendix B: Using Excel
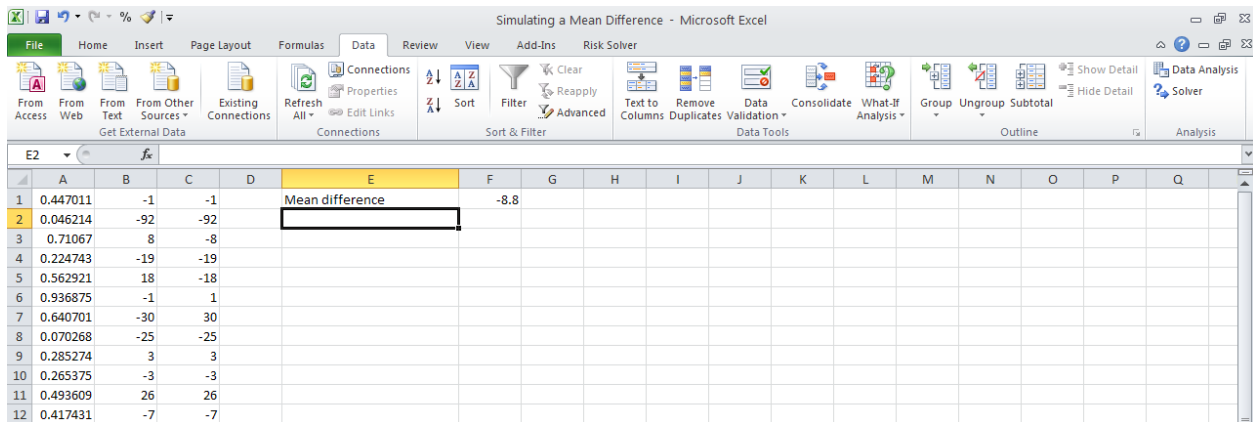
6. To generate more simulated mean differences, repeatedly press the **F9** key. Or, choose the **Formulas** tab and repeatedly press the **Calculate Now** button.

# Chapter 7: Calculating the Mean Absolute Deviation and Standard Deviation

Follow these steps for calculating the mean absolute deviation (MAD) and the standard deviation (SD) for the Chicago White Sox's 10 runs allowed values (4, 2, 4, 6, 5, 6, 6, 12, 1, 7).

1. Enter the 10 run totals in cells A1 to A10.
2. Enter the following labels and formulas using the indicated format.

	C	D
1	MAD	=avedev(A1:A10)
2	SD	=stdev(A1:A10)

The results are shown in the screen shot below.

Appendix B: Using Excel
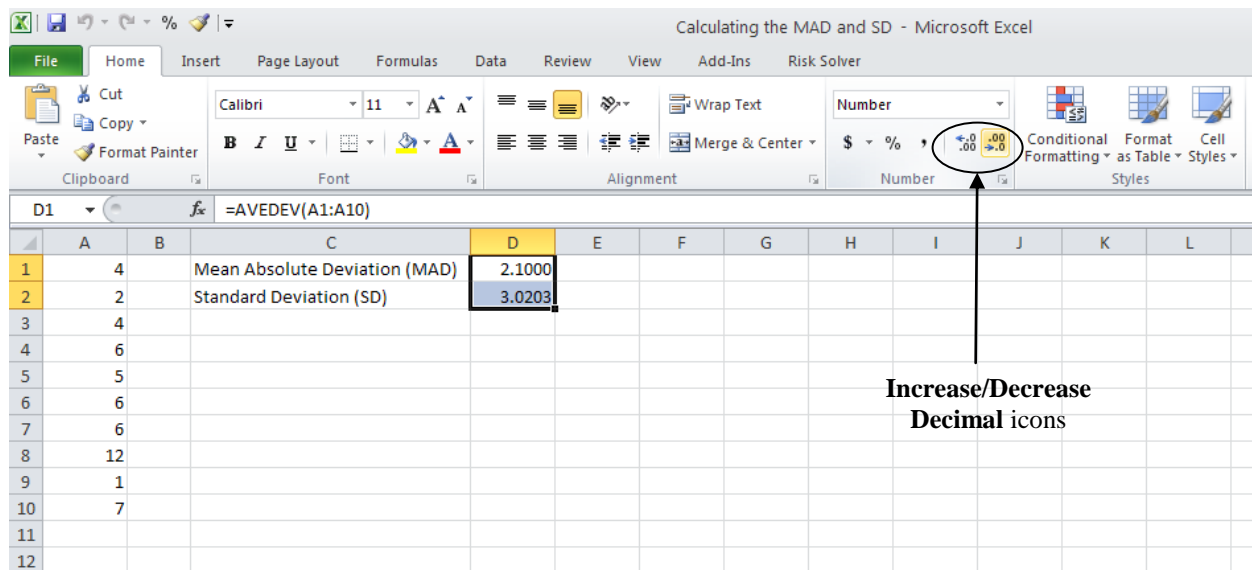
3. To display more (or fewer) decimal places, highlight cells D1 to D2 and use the **Increase/Decrease Decimal** icons found in the **Home** tab as shown above.

Chapter 7: Simulating a Difference in Standard Deviations

Follow these steps for simulating the difference in the standard deviations of distances for the 10 swings with the new club and the 10 swings with the old club. Note that you will need to use the *deviations* from the mean of the new club (–2, 9, 1, –19, 21, –16, –7, 14, –12, 11) and the *deviations* from the mean of the old club (4, –9, –2, 4, 6, –13, 10, 8, –1, –7).

1. In cell A1, enter the formula =rand(). This will generate a random number from 0 to 1.

2. Copy the formula in A1 down to cell A20 because there were a total of 20 swings. To do this, highlight cell A1, drag your mouse down to cell A20 so that cells A1 to A20 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down**.
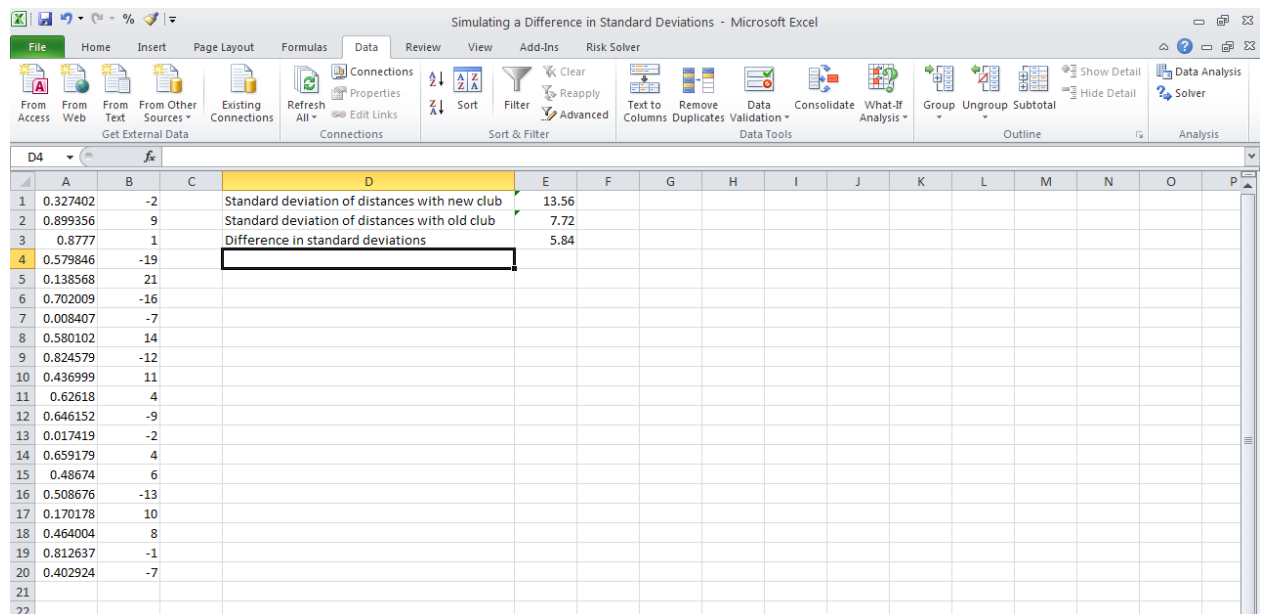
Appendix B: Using Excel


Fill Down option

Highlight cells A1 to A20
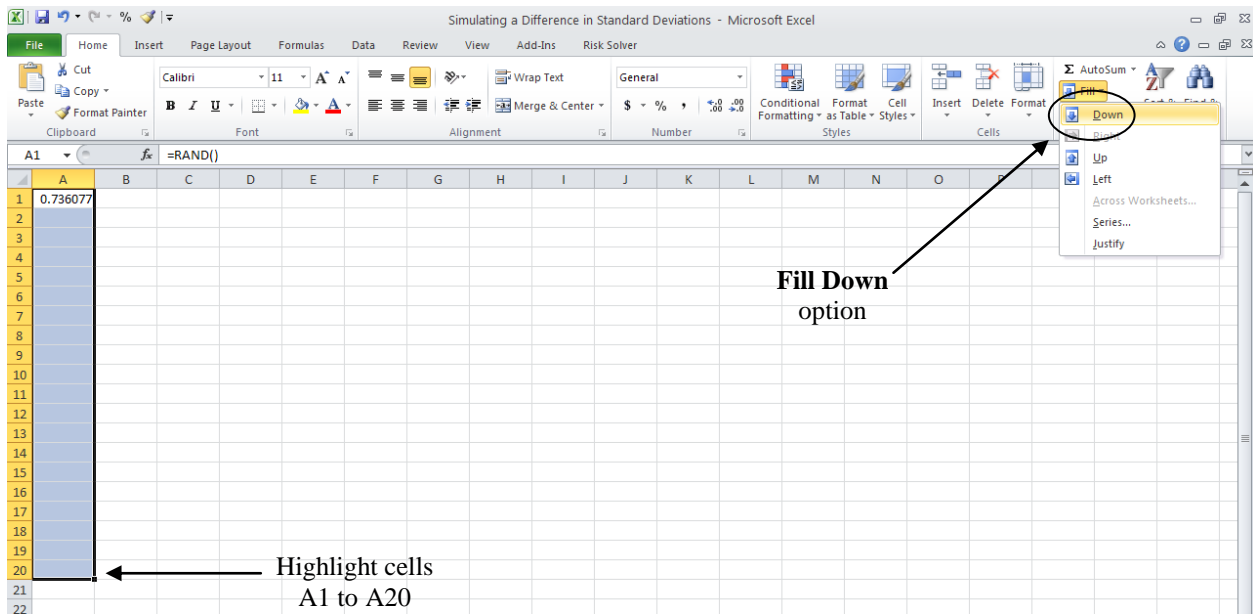
3. In cells B1 to B20, enter the 20 distances for the 20 swings.
4. To display the standard deviation of distances with the new club, the standard deviation of distances with the old club, and the observed difference in these standard deviations, enter the following labels and formulas using the indicated format.

	D	E
1	Standard deviation of distances with new club	=stdev(B1:B10)
2	Standard deviation of distances with old club	=stdev(B11:B20)
3	Difference in standard deviations	=E1-E2

The results are shown in the screen shot below.

Appendix B: Using Excel

5. To simulate the difference in standard deviations, we will sort columns A and B according to the random numbers in column A. This will shuffle the distances in column B. To do this, click on cell A1 and highlight cells A1 to B20. Then, select the **Data** tab and choose the **Sort Smallest to Largest** icon $\boxed{\begin{smallmatrix} A \\ Z \end{smallmatrix} \downarrow}$. The value in cell E3 is now the *simulated* difference in standard deviations.
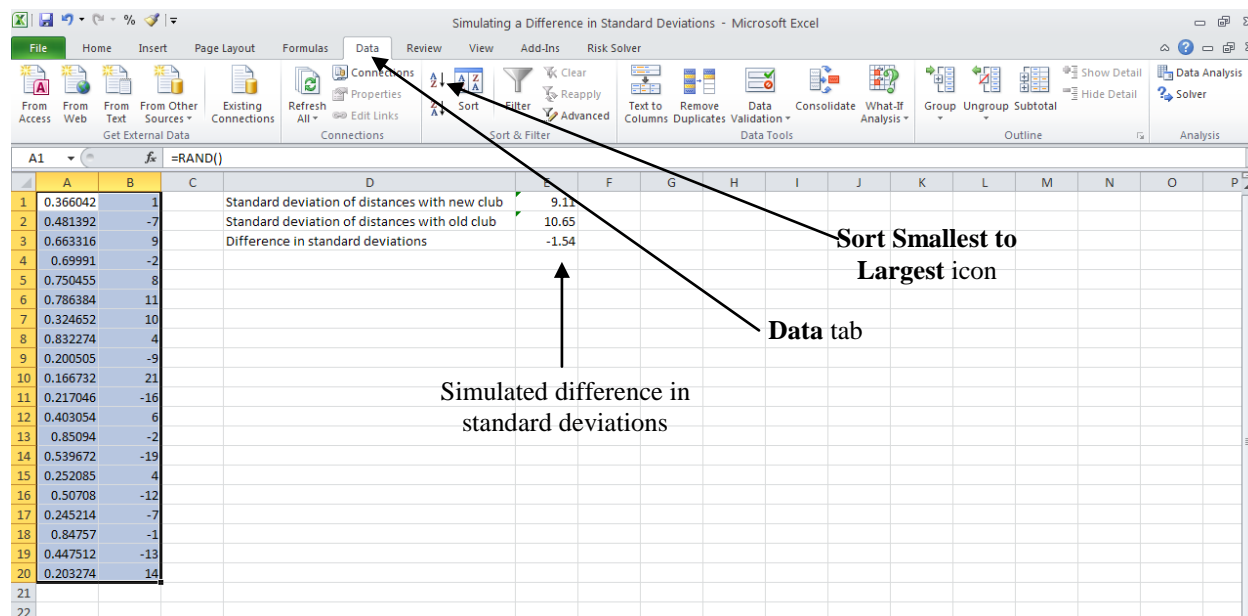
6. To generate more simulated differences, repeatedly press the **Sort Smallest to Largest** icon. *Note*: *Don't be confused when you notice that the random numbers in column A are not in ascending order. Each time Excel sorts the columns, it generates new random numbers for column A.*

*To simulate the difference in mean absolute deviation, change the labels and formulas in cells D1 to E3 using the format below.

	D	E
1	Mean absolute deviation of distances with new club	=avedev(B1:B10)
2	Mean absolute deviation of distances with old club	=avedev(B11:B20)
3	Difference in mean absolute deviations	=E1-E2

Chapter 8: Normal Distribution Calculations

The process for calculating the area to the left of a boundary in a Normal distribution is to use the function normdist(boundary, mean, SD, cumulative). The value of cumulative should always be 1, which tells Excel to find the area to the left (and not just the height) of the Normal curve at that boundary. For example, suppose Anne's scoring PERFORMANCES are approximately Normally distributed with a mean of 13.1 points and a standard deviation of 4.2 points.

1. To find the proportion of games where she scores less than 10 points, enter the formula =normdist(10,13.1,4.2,1) in any cell. Press **Enter** to display the answer 0.2302 as shown in cell B3 of the screen shot below.

2. To find the proportion of games where she scores *more* than 20 points, enter the formula =1−normdist(20,13.1,4.2,1) in any cell. Press **Enter** to display the answer 0.0502 as shown in cell B5 of the screen shot below.

3. To find the proportion of games where she scores between 10 and 20 points, enter the formula =normdist(20,13.1,4.2,1)−normdist(10,13.1,4.2,1) in any cell. Press **Enter** to display the answer 0.7196 as shown in cell B7 of the screen shot below.

The process for calculating percentiles in a Normal distribution is to use the function **norminv(area to left, mean, standard deviation)**. For example, suppose Anne's scoring PERFORMANCES are approximately Normally distributed with a mean of 13.1 points and a standard deviation of 4.2 points. To find the value of the 25th percentile, enter the formula =norminv(0.25,13.1,4.2) in any cell. Press **Enter** to display the answer 10.27 as shown in cell B9 of the screen shot below.

## Chapter 10: Making a Scatterplot

Follow these steps for constructing a scatterplot using average driving distance (246.8, 265.2, 254.3, 263.7, 255.5, 268.1, 269.2, 249.3, 248.6, 245.7) and driving accuracy (0.824, 0.718, 0.757, 0.714, 0.733, 0.660, 0.654, 0.748, 0.811, 0.770) for the top 10 money winners on the 2009 LPGA tour.

1. Enter the average driving distance data in cells A1 to A10 and the driving accuracy data in cells B1 to B10.

2. Highlight the data in cells A1 to B10, open the **Insert** menu, choose the **Scatter** icon, then the **Scatter with only Markers** in the upper left.
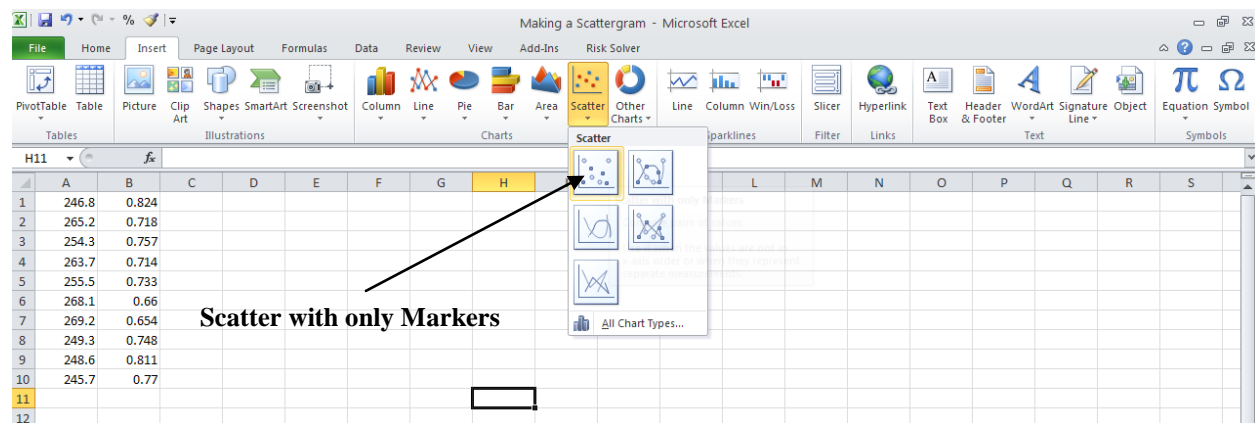

3. Click on the scatterplot and select the **Layout** tab of the **Chart Tools**. Select the **Chart Title** tab and type *2009 LPGA Top 10*. Then select the **Axis Titles** tab, label the **Primary Horizontal Axis Title** (*x* axis) *Average Driving Distance (yards)* and the **Primary Vertical Axis Title** (*y* axis) *Driving Accuracy*.

Appendix B: Using Excel
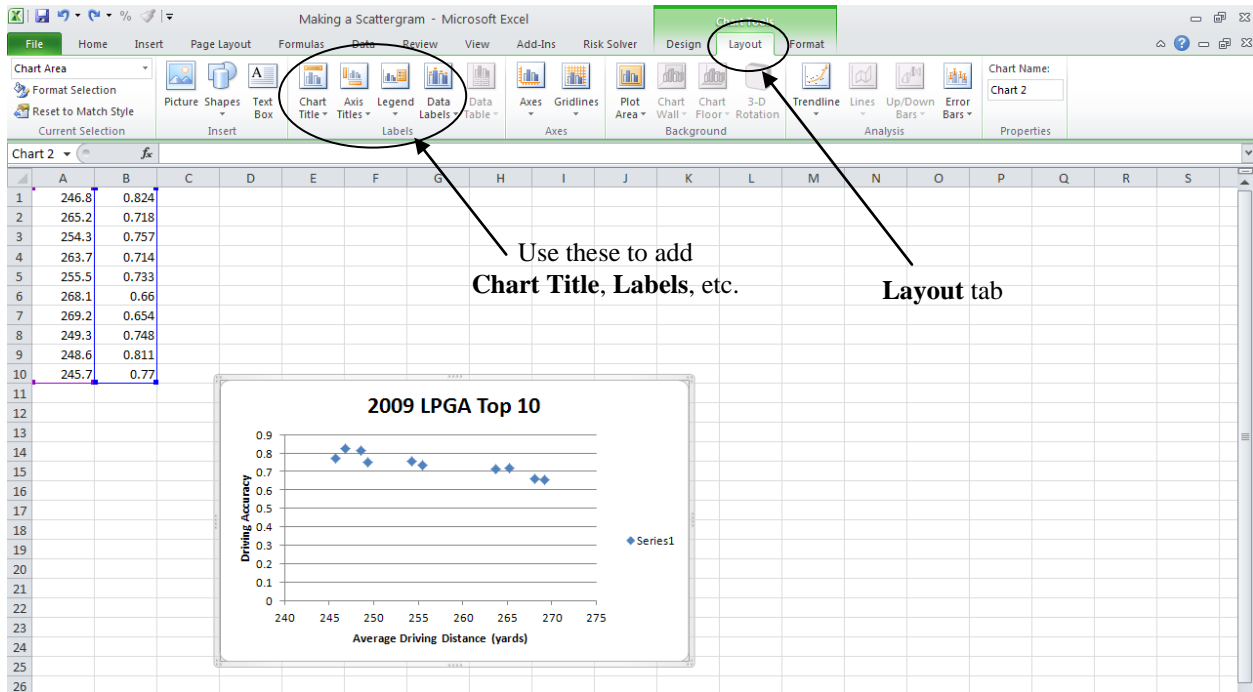
4. To eliminate the legend, select the **Layout** tab of the **Chart Tools**. Select the **Legend** tab and choose **None**. To eliminate the vertical axis major grid lines, right-click on any one of the lines and choose **Delete**.

5. To display a better scale for the vertical axis, right click anywhere on the vertical axis, then click on **Format Axis**. Under **Axis Options**, change the **Minimum** from **Auto** to **Fixed**, then enter 0.6 for the minimum. Click **Close**.

Appendix B: Using Excel
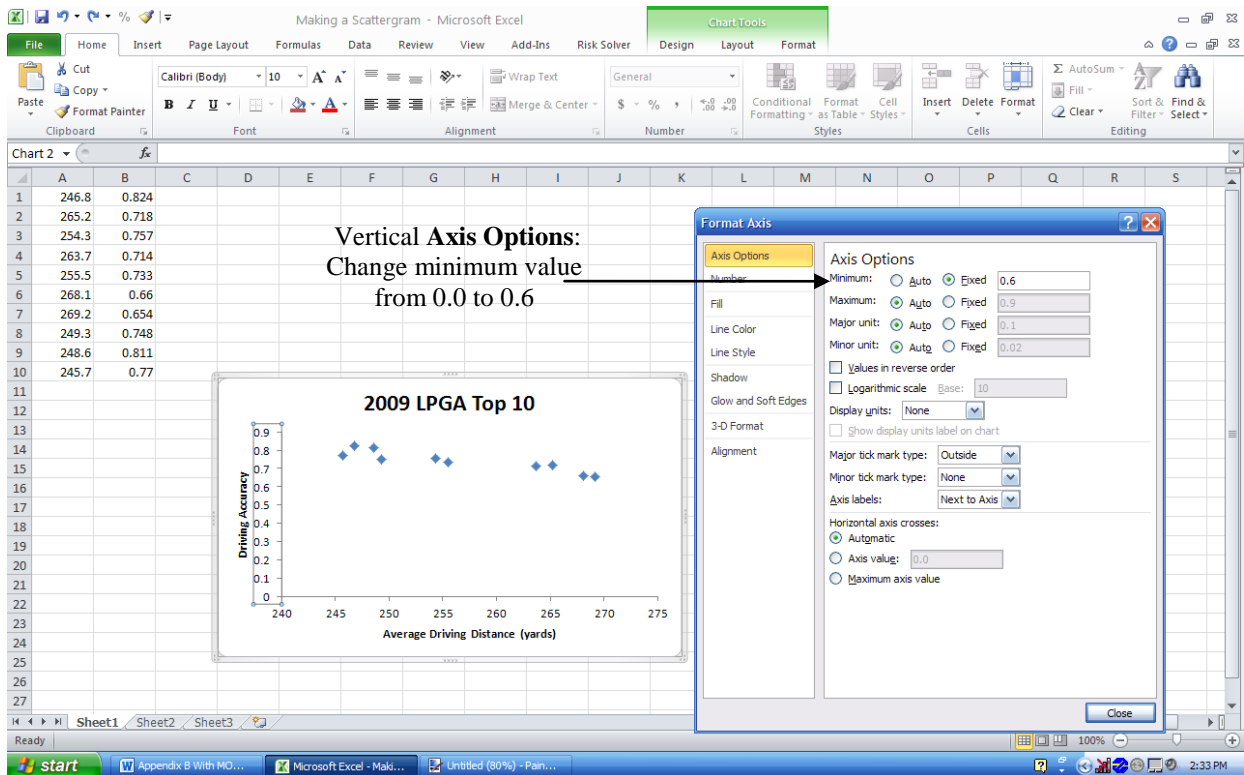
6. If a different background color is desired, right-click on the background and choose **Format Plot Area**. Select **Fill**, then the desired options.
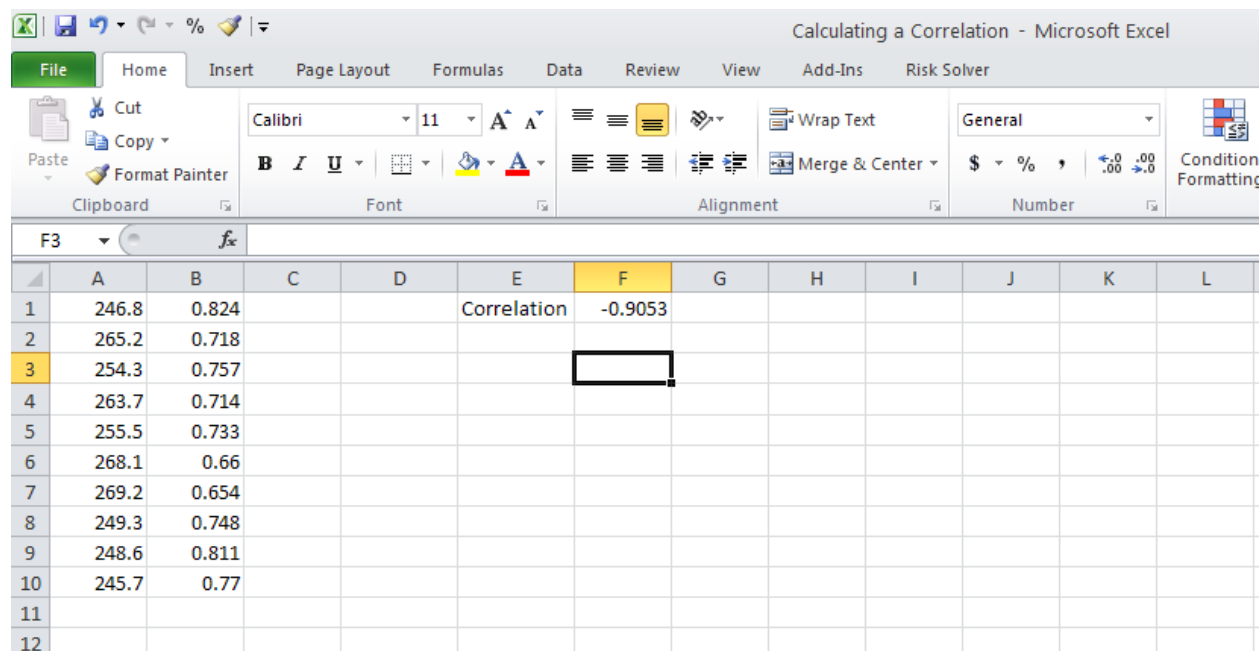7. Here is an example of a well-labeled scatterplot.
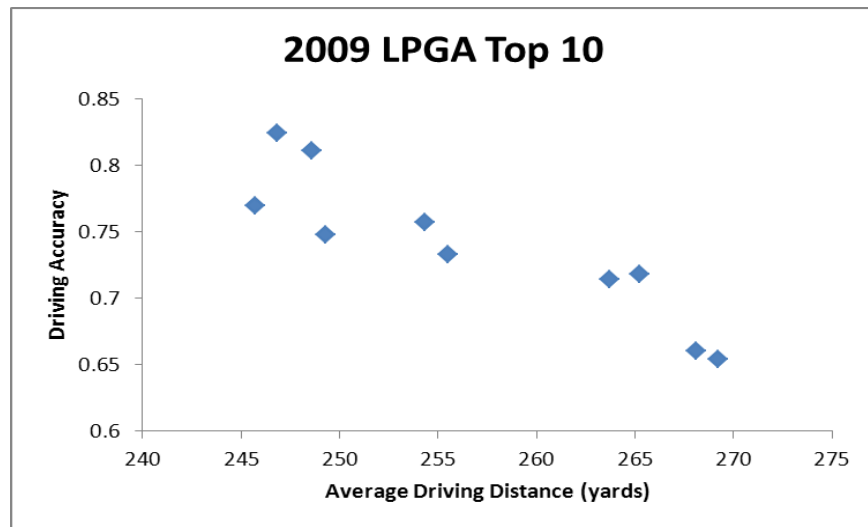
Chapter 10: Calculating the Correlation

Follow these steps for calculating the correlation between average driving distance and driving accuracy for the top 10 winners on the 2009 LPGA tour. Refer to the data in the *Chapter 10: Making a Scatterplot* section.

1. Enter the average driving distance data in cells A1 to A10 and the driving accuracy data in cells B1 to B10.
2. To display the correlation, enter the following label and formula using the indicated format.

	E	F
1	Correlation	=correl(A1:A10,B1:B10)

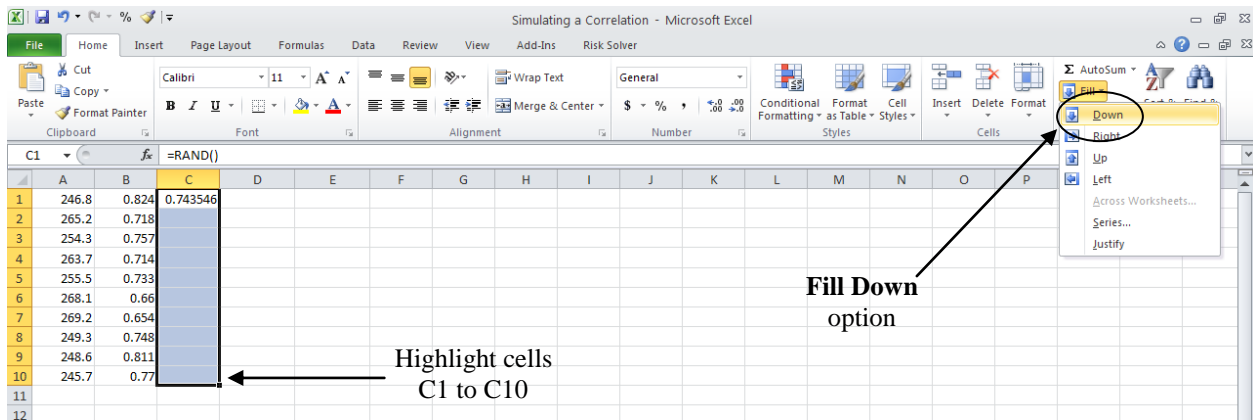The result is shown in cell F1 of the screen shot below.

Chapter 10: Simulating the Correlation

Follow these steps for simulating the correlation between average driving distance and driving accuracy for the top 10 winners on the 2009 LPGA tour. Use the data in *Chapter 10: Making a Scatterplot* section.

1. Enter the average driving distance data in cells A1 to A10 and the driving accuracy data in cells B1 to B10.
2. In cell C1, enter the formula =rand().
3. Copy the formula in C1 down to cell C10 because there were a total of 10 golfers. To do this, highlight cell C1, drag your mouse down to cell C10 so that cells C1 to C10 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down**.
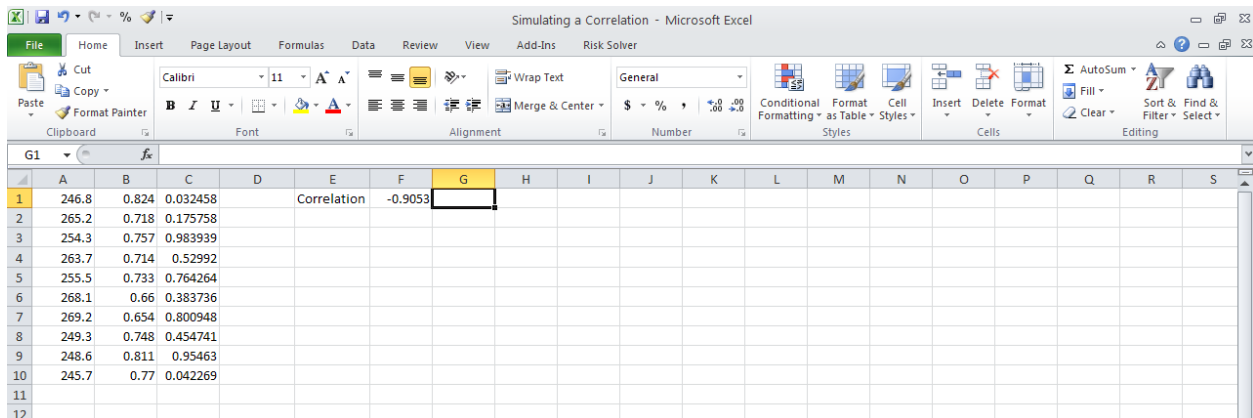
Appendix B: Using Excel

4. To display the correlation, enter the following label and formula using the indicated format.

	E	F
1	Correlation	=correl(A1:A10,B1:B10)

The result is shown in cell F1 of the screen shot below.

5. To simulate the correlation, we will sort columns B and C according to the random numbers in column C. This will shuffle the driving accuracy values in column B. To do this, click on cell C1 and highlight cells C1 to B10. Then, select the **Data** tab and choose the **Sort Smallest to Largest** icon $\begin{array}{c} A \\ Z \end{array}\downarrow$. The value in cell F1 is now the *simulated* correlation.

Appendix B: Using Excel
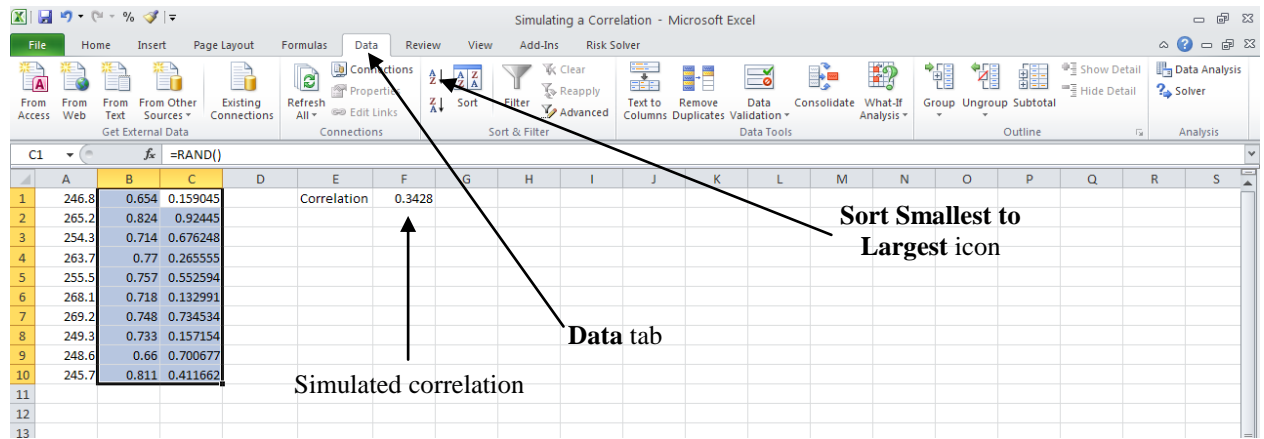
6. To generate more simulated correlations, repeatedly press the **Sort Smallest to Largest** icon. *Note*: *Don't be confused when you notice that the random numbers in column C are not in ascending order. Each time Excel sorts the columns, it generates new random numbers for column C.*

Chapter 10: Making a Timeplot

Follow these steps for making a timeplot using Emmitt Smith's rushing yards for ages 21–35 (937, 1563, 1713, 1486, 1484, 1773, 1204, 1074, 1332, 1397, 1203, 1021, 975, 256, 937).

1. Enter his ages from 21 to 35 in cells A1 to A15 and his rushing yards in cells B1 to B15.
2. Highlight the data in cells A1 to B15, open the **Insert** menu, choose the **Scatter** icon, then the **Scatter with Straight Lines and Markers** as indicated in the screen shot below.
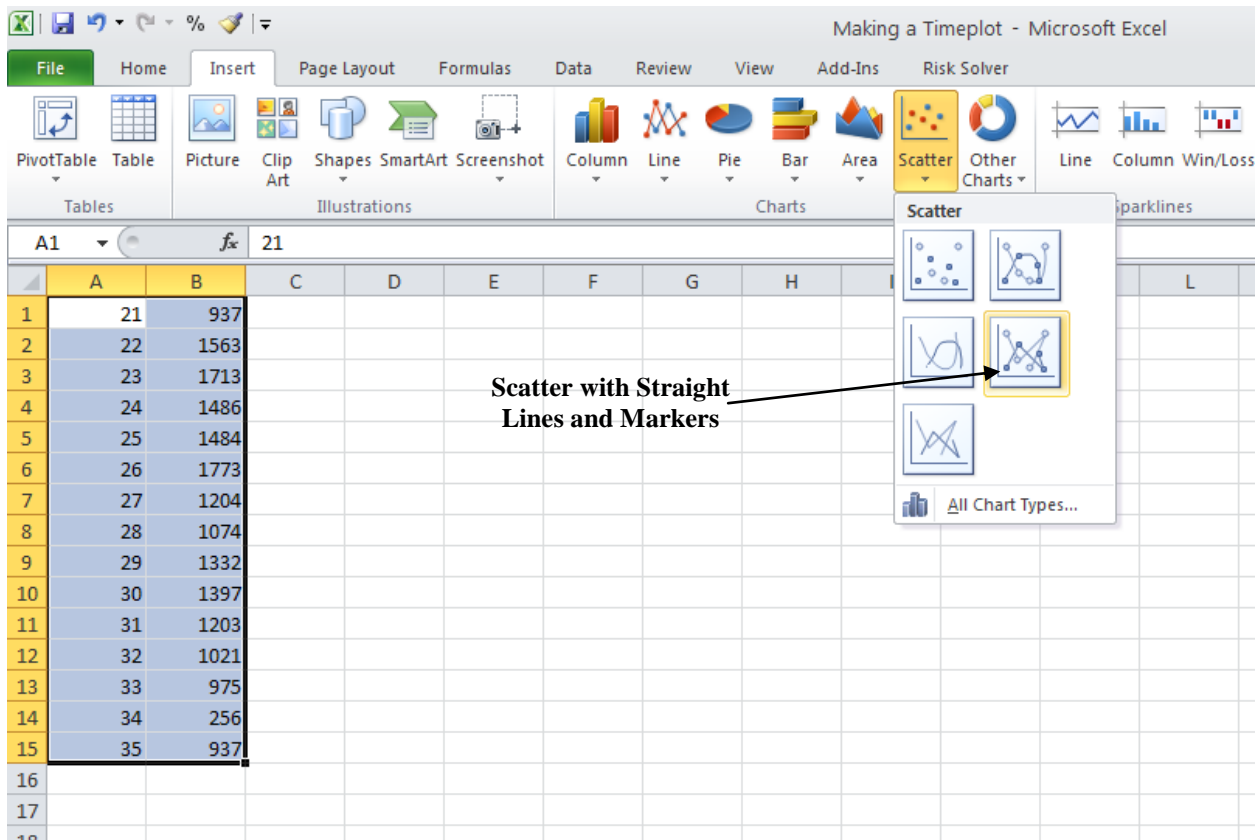

3. Click on the scatterplot and select the **Layout** tab of the **Chart Tools**. Select the **Chart Title** tab and type *Emmitt Smith's Career*. Then select the **Axis Titles** tab, label the **Primary Horizontal Axis Title** (*x* axis) *Age* and the **Primary Vertical Axis Title** (*y* axis) *Rushing Yards*.
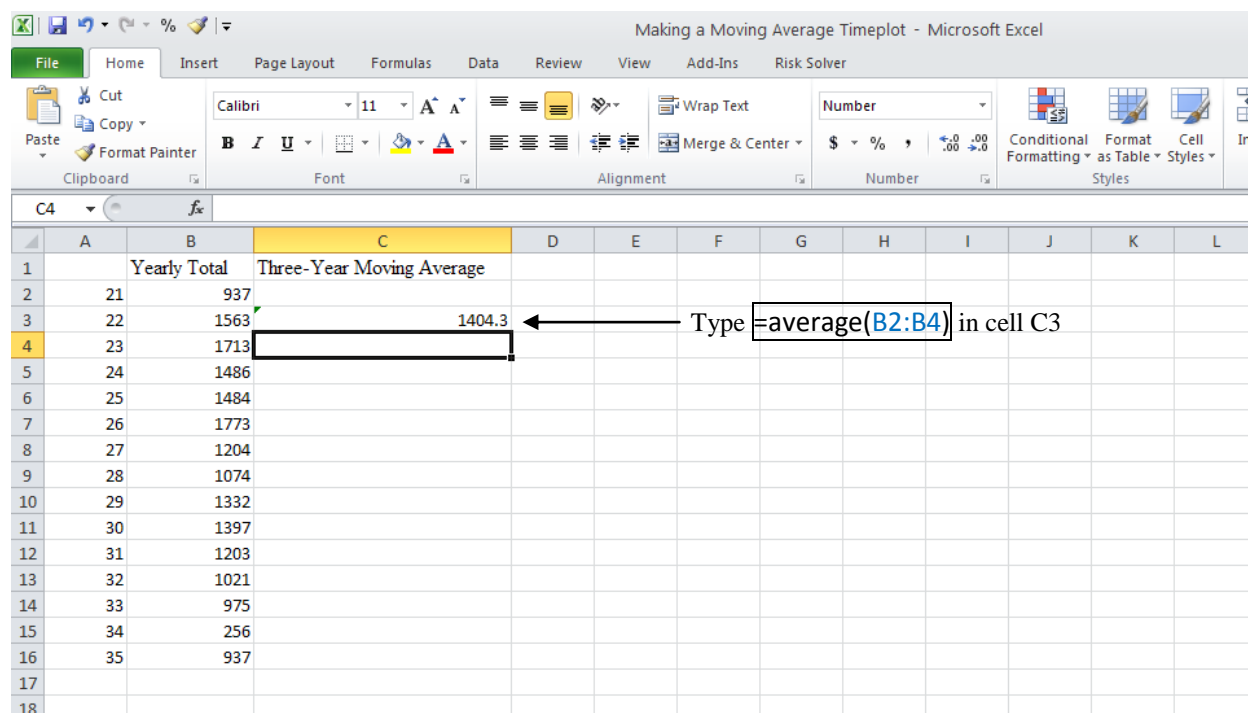
Appendix B: Using Excel
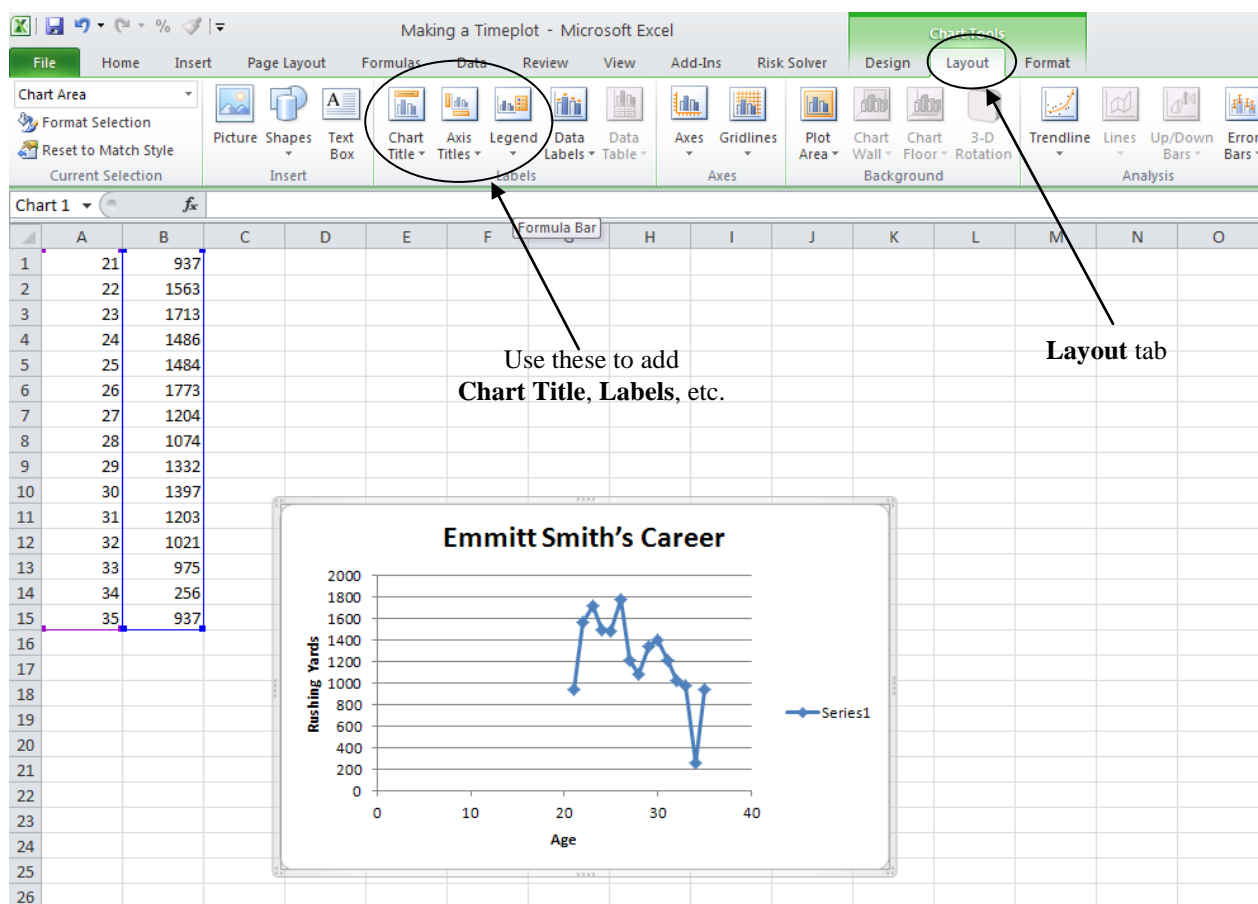
4. To eliminate the legend, select the **Layout** tab of the **Chart Tools**. Select the **Legend** tab and choose **None**. To eliminate the vertical axis major grid lines, right-click on any one of the lines and choose **Delete**.

5. To display a better scale for the horizontal axis, right click anywhere on the horizontal axis, then click on **Format Axis**. Under **Axis Options**, enter 20 for the minimum, 36 for the maximum, and 2 for the major unit. To display a better scale for the vertical axis, right click anywhere on the vertical axis, then click on **Format Axis**. Under **Axis Options**, change the major unit to 400.

6. If a different background color is desired, right-click on the background and choose **Format Plot Area**. Select **Fill**, then the desired options.

7. Here is an example of a well-labeled timeplot.

Appendix B: Using Excel

Emmitt Smith's Career
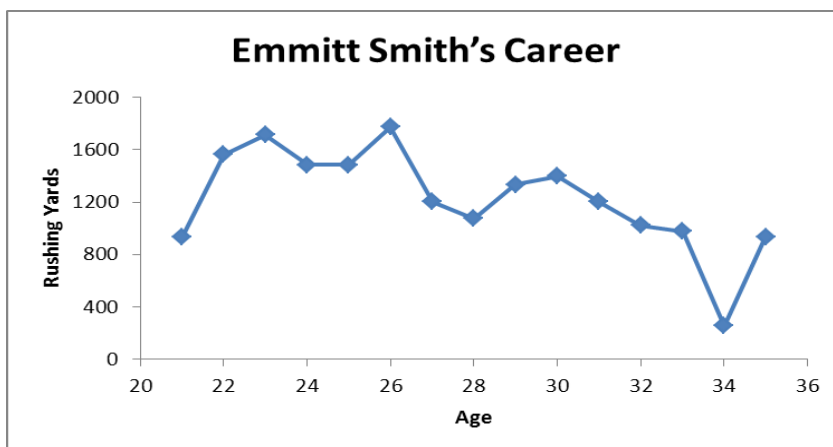
Chapter 10: Making a Moving Average Timeplot

Follow these steps for making a three-year moving average timeplot using Emmitt Smith's rushing *PERFORMANES*. Use the data from the previous section.

1. In cell B1, type *Yearly Total*. In cell C1, type *Three-Year Moving Average*. The column headings are necessary for the legend when two timeplots are displayed together (see step 6). Enter Smith's ages from 21 to 35 in cells A2 to A16 and his rushing yards in cells B2 to B16.

2. In cell C3 (the age 22 row), enter the formula =average(B2:B4) .

Type =average(B2:B4) in cell C3

Appendix B: Using Excel

3. Copy the formula in C3 down to cell C15 because we can only calculate a three-year average for his age 22 to 34 seasons. To do this, highlight cell C3, drag your mouse down to cell C15 so that cells C3 to C15 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down**.
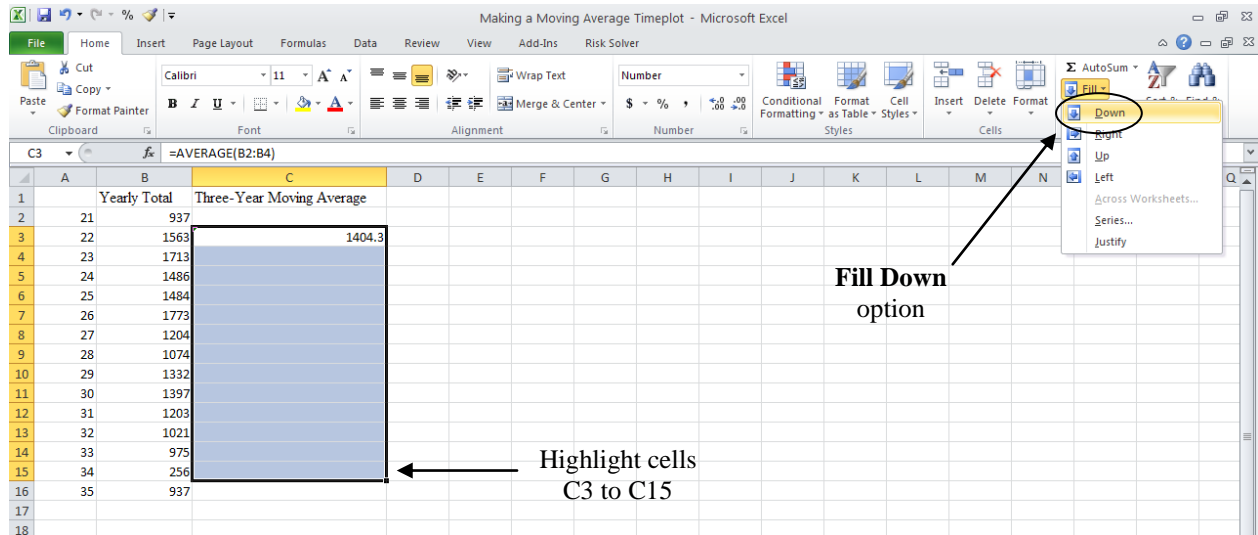

4. To make a moving average timeplot, highlight cells A1 to A16, hold the **Ctrl** button and highlight the moving average data in cells C1 to C16 (including the column heading and the blank cells for his age 21 and age 35 seasons). Open the **Insert** menu, choose the **Scatter** icon, then the **Scatter with Straight Lines and Markers** as indicated in the screen shot below.
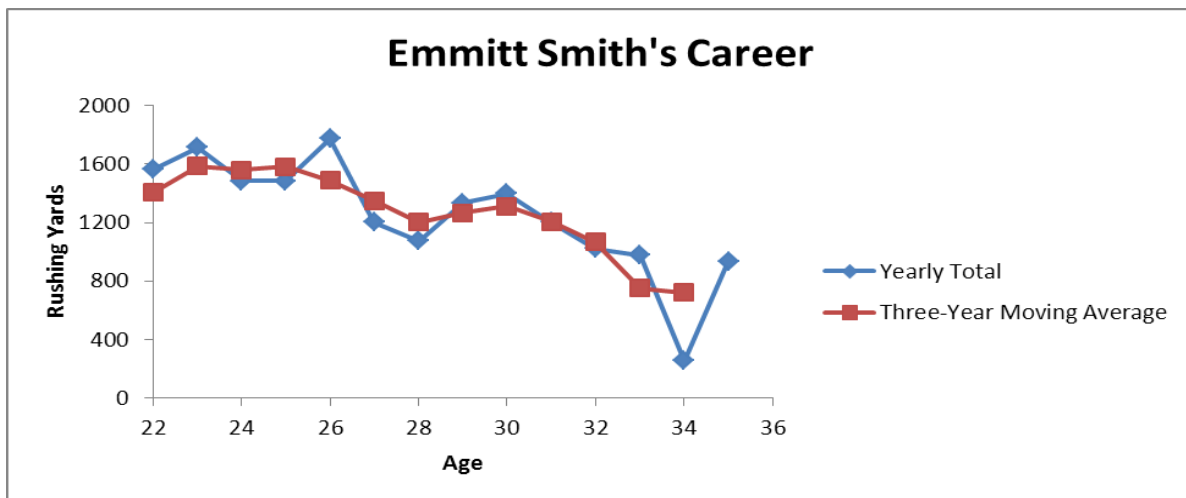
Appendix B: Using Excel

Scatter with Straight Lines and Markers
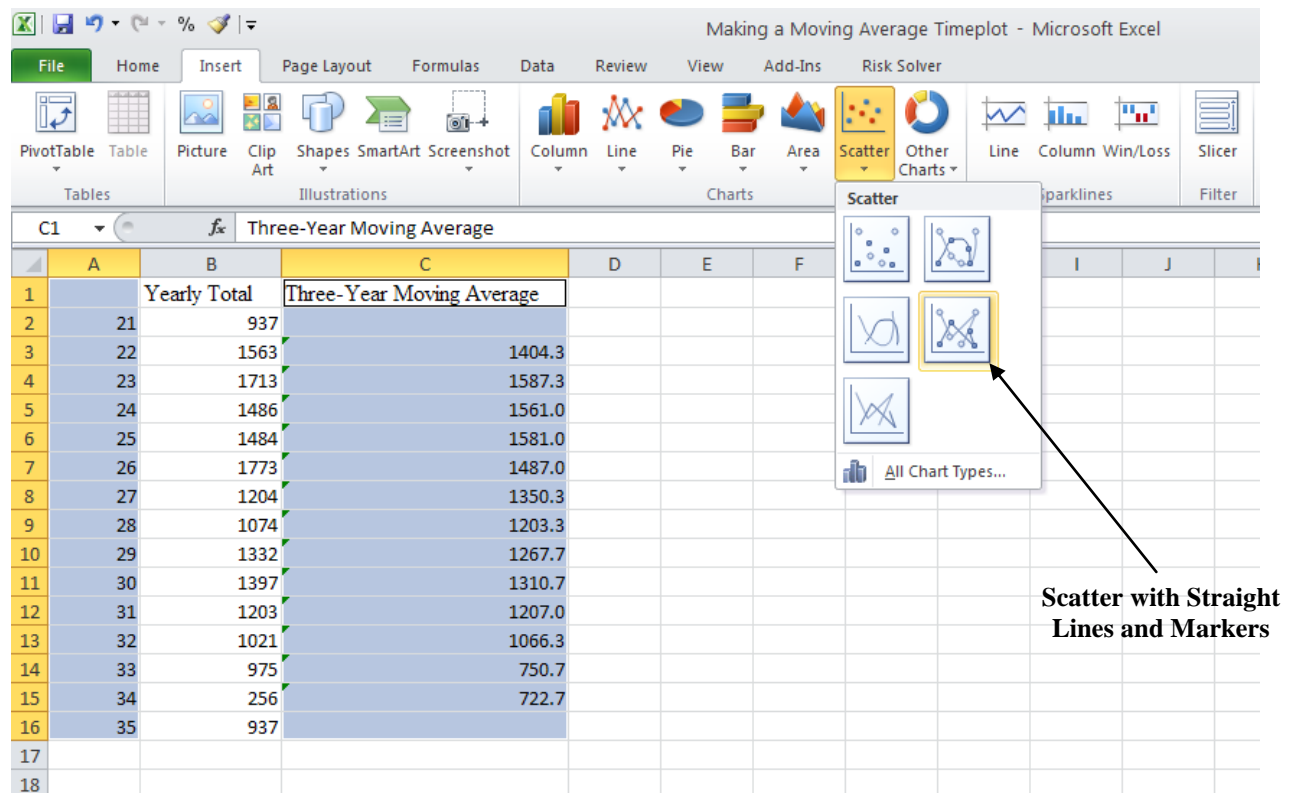
5. To label the title, axes, and adjust the scale, refer to steps 3 and 5 of the previous section. The completed three-year moving average timeplot is shown below.
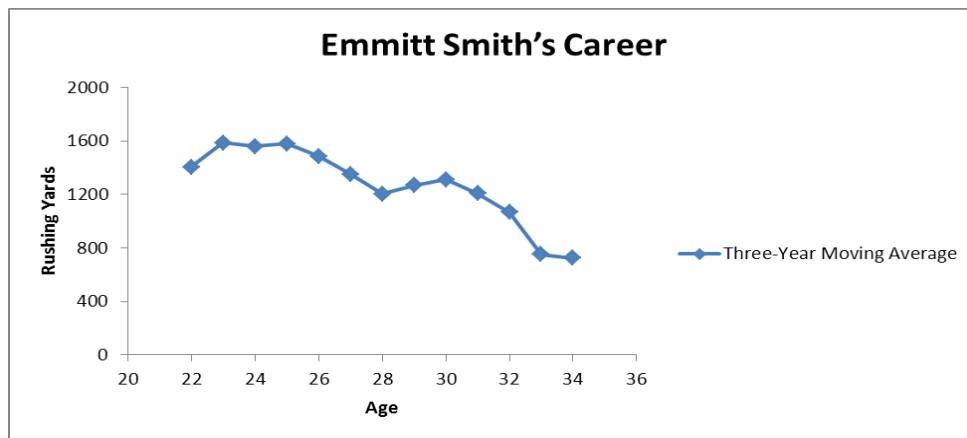
6. To observe both timeplots on the same graph, highlight all three columns of data from A1 to C16, open the **Insert** menu, choose the **Scatter** icon, then the **Scatter with Straight Lines and Markers.** Then follow steps 4 and 5 above. To label the title, axes, and adjust the scale, refer to steps 3 and 5 of the previous section.

7. Here is an example of the two timeplots together.

Appendix B: Using Excel

Emmitt Smith's Career

(Chart showing Rushing Yards vs Age, with series "Yearly Total" and "Three-Year Moving Average")

Chapter 11: Calculating the Equation of a Least-Squares Regression Line

Follow these steps for calculating the equation of a least-squares regression line for the relationship between *x* = average driving distance and *y* = driving accuracy for the top 10 winners on the 2009 LPGA tour. Use the data from the *Chapter 10: Making a Scatterplot* section.

1. Enter the average driving distance data in cells A1 to A10 and the driving accuracy data in cells B1 to B10.

2. To find the slope and the *y* intercept of a least-squares regression line, enter the following labels and formulas using the indicated format. Be sure to use columns F and G for future use.

|	F	G	
1	Slope	=slope(B1:B10,A1:A10)	*
2	y intercept	=intercept(B1:B10,A1:A10)	*

*Note that you <u>must</u> list the *y*-values before the *x*-values when using these functions. The results are shown in the screen shot below.

## Chapter 11: Displaying a Scatterplot with a Least-Squares Regression Line

Follow these steps for displaying the graph of a least-squares regression line on the scatterplot that shows the relationship between x = average driving distance and y = driving accuracy for the top 10 winners on the 2009 LPGA tour. Use the data from the *Chapter 10: Making a Scatterplot* section.

1. Follow steps 1–6 from the *Chapter 10: Making a Scatterplot* section.
2. Right-click on any one of the points in the scatterplot and choose the **Add Trendline** option.
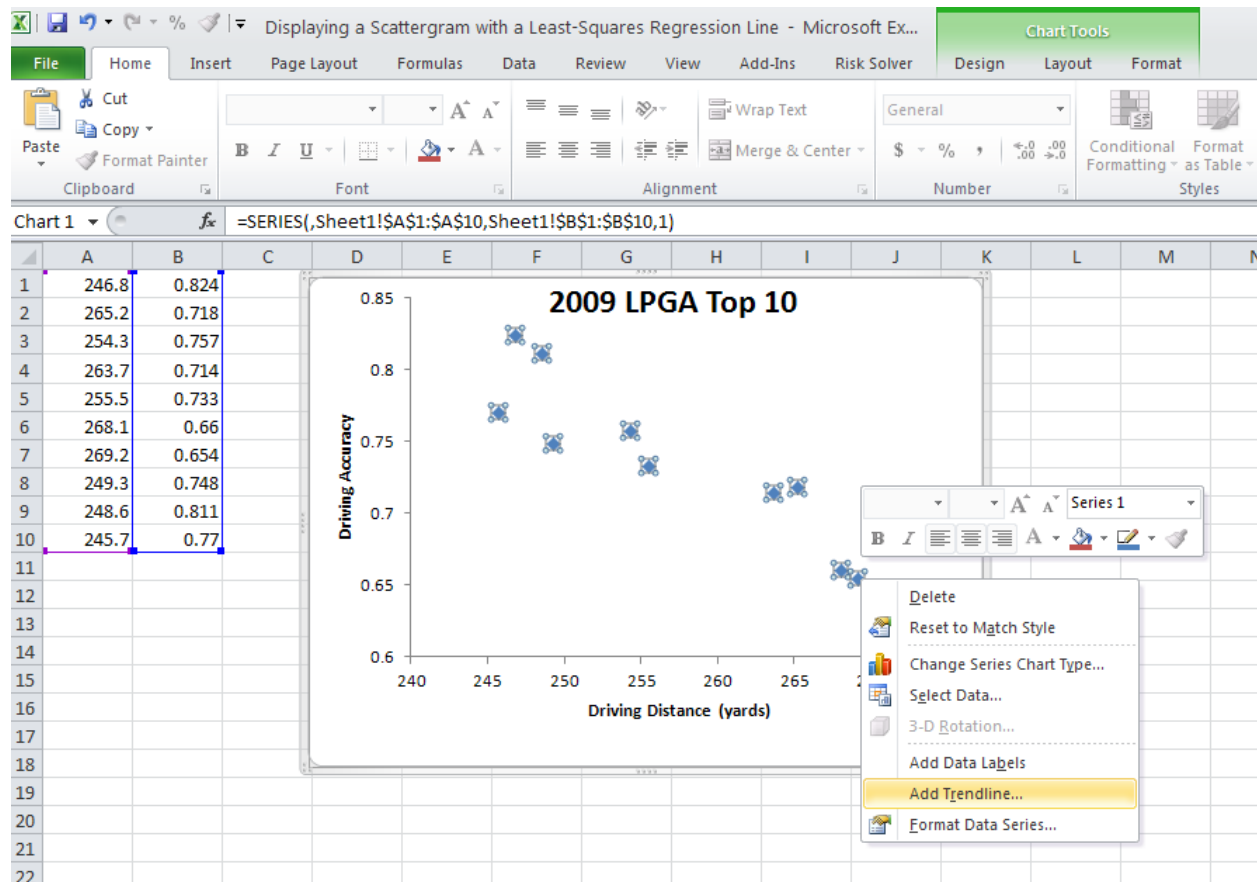

3. Under **Trendline Options**, choose the **Linear** model and check the box to **Display Equation on Chart**.

Appendix B: Using Excel
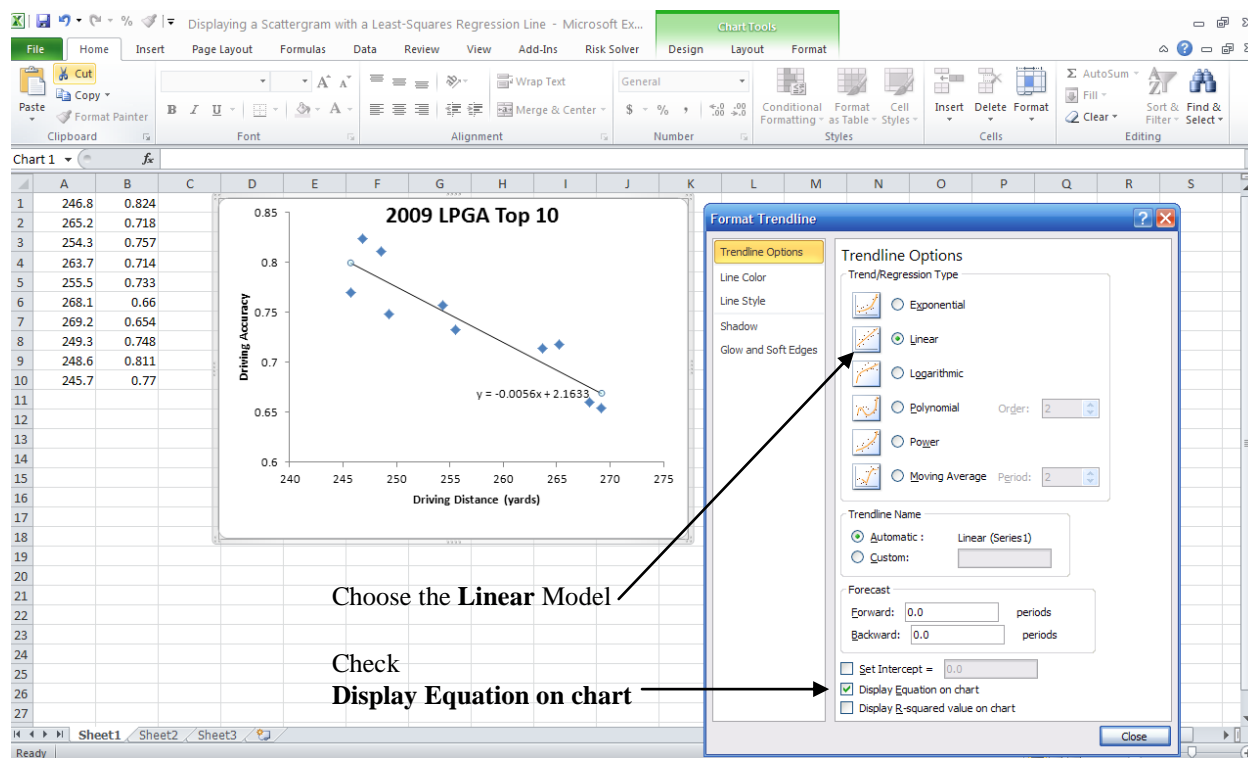
4. If the equation shows up in an awkward place, drag it to a better location.

5. Here is an example of a well-labeled scatterplot with the equation of a least squares regression line displayed. Note that Excel displays the equation in $y = mx+b$ form rather than $\hat{y} = a + bx$ form. These two forms are mathematically equivalent, with the slope displayed as the coefficient of x in both cases.

Appendix B: Using Excel
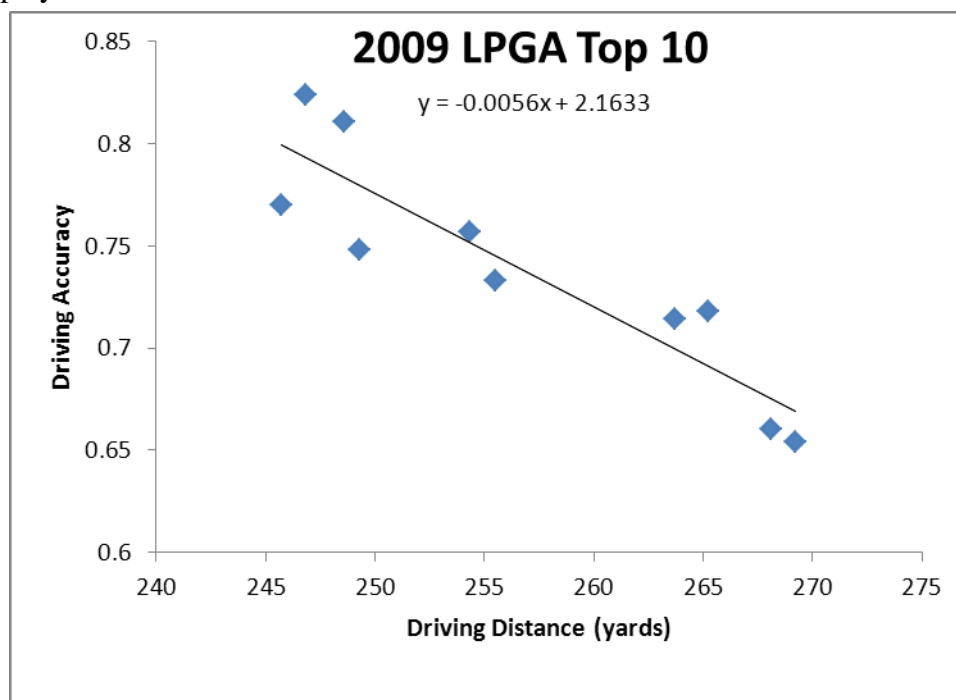
Chapter 11: Calculating the Standard Deviation of the Residuals

Follow these steps for calculating the standard deviation of the residuals for the least-squares regression line summarizing the relationship between x = average driving distance and y = driving accuracy for the top 10 winners on the 2009 LPGA tour. Use the data from the *Chapter 10: Making a Scatterplot* section and the results from the *Chapter 11: Calculating the Equation of a Least-Squares Regression Line* section.
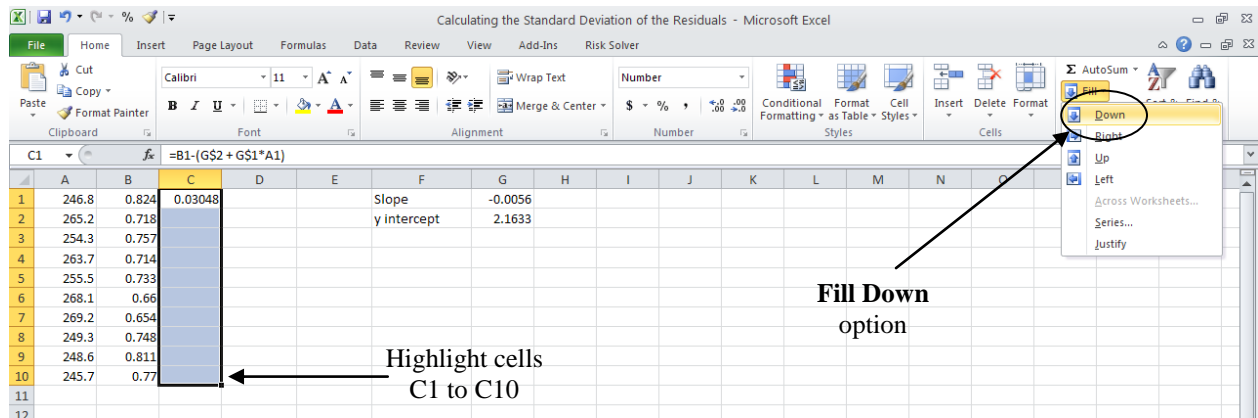
1. Follow steps 1–3 from the *Chapter 11: Calculating the Equation of a Least-Squares Regression Line* section.

2. To calculate the residuals, enter the function =B1–(G$2+G$1*A1) in cell C1. The $ sign locks the row reference so the formula will always refer to cells G2 and G1 when the formula is copied down to other cells.

3. Copy the formula in C1 down to cell C10 because there were a total of 10 golfers. To do this, highlight cell C1, drag your mouse down to cell C10 so that cells C1 to C10 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down**.

Appendix B: Using Excel

4. To calculate the squared residuals, enter the formula $\boxed{\text{=C1^2}}$ in cell D1. Copy this formula down to D10 like in step 3.
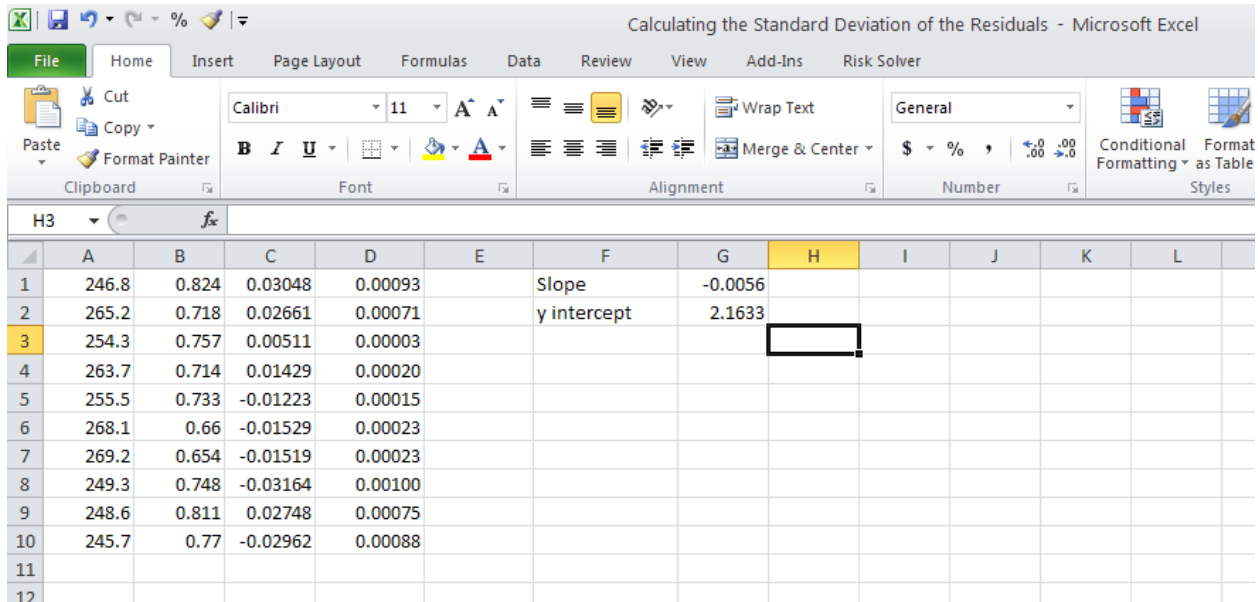
5. To find the standard deviation of the residuals, enter the following label and formula using the indicated format.

	F	G
3	SD of residuals	=sqrt(sum(D1:D10)/(10-2))

The result is shown in cell G3 of the screen shot below.

Appendix B: Using Excel

## Chapter 11: Simulating the Slope of a Least-Squares Regression Line

Follow these steps for simulating the slope of a least-squares regression line for relationship between x = average driving distance and y = driving accuracy for the top 10 winners on the 2009 LPGA tour. Use the data in the *Chapter 10: Making Scatterplots* section.

1. Enter the average driving distance data in cells A1 to A10 and the driving accuracy data in cells B1 to B10.
2. In cell C1, enter the formula =rand() .
3. Copy the formula in C1 down to cell C10 because there were a total of 10 golfers. To do this, highlight cell C1, drag your mouse down to cell C10 so that cells C1 to C10 are highlighted, and fill down by pressing **Ctrl-D** or in the **Home** tab by pressing the **Fill** button and choosing **Down**.
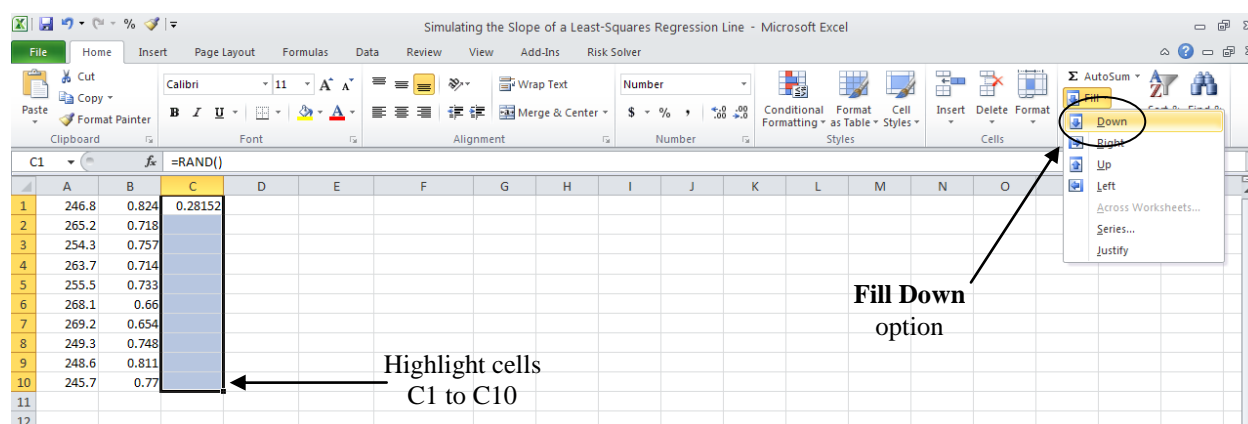
4. To find the slope, enter the following label and formula using the indicated format.

	F	G	
1	Slope	=slope(B1:B10,A1:A10)	*

*Note that you <u>must</u> list the *y*-values before the *x*-values when using these functions. The result is shown in the screen shot below.

Appendix B: Using Excel

5. To simulate the slope, we will sort columns B and C according to the random numbers in column C. This will shuffle the driving accuracy values in column B. To do this, click on cell C1 and highlight cells C1 to B10. Then, select the **Data** tab and choose the **Sort Smallest to Largest** icon $\boxed{\begin{smallmatrix}A\\Z\end{smallmatrix}\downarrow}$. The value in cell G1 is now the *simulated* slope.

6. To generate more simulated slopes, repeatedly press the **Sort Smallest to Largest** icon. *Note*: *Don't be confused when you notice that the random numbers in column C are not in ascending order. Each time Excel sorts the columns, it generates new random numbers for column C.*

Chapter 12: Calculating a Multiple Regression Model

Follow these steps for calculating a multiple regression model for the NBA attendance data. See the screen shot below for the data.

1. Enter the 4 variables (average attendance, wins, market share, and finals) from the NBA attendance data into columns A, B, C, and D.
2. Click on cell G2, open the **Formula** menu, and choose **Insert Function** (far left). Scroll down and select the function **LINEST**. LINEST is an example of a special class of functions in EXCEL called array functions.
3. A Window will pop up asking you for the **Function Arguments** (see the screen shot below).
    - With the cursor in the box for **Known_y's**, highlight the data in cells A2 to A31. Do not highlight the title.
    - Move the cursor to the box of **Known_x's** and highlight the data in cells B2 to D31. Again, do not highlight the titles.
    - In the box labeled **Const**, enter the word *true* or leave the box blank. This will tell Excel to include a constant term in the model.
    - In the box labeled **Stats**, enter the word *true*. This will tell Excel to generate additional statistics such as r^2 and *s*. Press **OK**.

Calculating a Multiple Regression Model - Micro

	A	B	C	D
1	AvgAtt	Wins	MShare	Finals
2	16748	47	2.07	0
3	18624	62	2.105	1
4	14526	35	0.981	0
5	21197	41	3.052	0
6	20010	66	1.332	1
7	20042	50	2.175	1
8	17223	54	1.332	0
9	21877	39	1.684	1
10	18942	29	2.164	0
11	17482	53	1.84	0
12	14182	36	0.974	0
13	16170	19	4.94	0
14	18997	65	4.94	1
15	12745	24	0.589	0
16	18229	43	1.352	1
17	15389	34	0.791	0
18	14505	24	1.512	0
19	15147	34	6.495	0
20	16968	49	0.527	0
21	19287	32	6.495	0
22	18693	23	0.6	0
23	17043	59	1.281	1
24	15802	41	2.578	0
25	18422	46	1.622	0
26	20524	54	1.027	0
27	12571	17	1.223	0
28	18269	54	0.715	1
29	18773	33	4.9	0
30	19903	48	0.803	0
31	16612	19	2.028	0
32				

Function Arguments

LINEST

Known_y's	A2:A31	= {16748;18624;14526;21197;20010;...
Known_x's	B2:D31	= {47,2.07,0;62,2.105,1;35,0.981,0;41,:
Const	true	= TRUE
Stats	true	= TRUE

= {1010.60579414181,269.45155633274

Returns statistics that describe a linear trend matching known data points, by fitting a straight line using the least squares method.

Stats is a logical value: return additional regression statistics = TRUE; return m-coefficients and the constant b = FALSE or omitted.

Formula result = 1010.605794

Help on this function   OK   Cancel

Appendix B: Using Excel

4. After pressing **OK**, there should be a number in cell G2. To get the rest of the model, click on the G2 cell, drag your mouse to create a rectangle 4 columns wide and 3 rows deep (from G2 to J4).   *Note: The number of rows is always 3. The number of columns is based on the number of terms in the model.   Since we will have 3 variables and a constant, we need to make room for 4 coefficients.*

5. Press the F2 button on your keyboard, and then press **Ctrl**-**shift**-**enter**.   The output should look like the table below. The first row lists the coefficients of the variables (in reverse order!) followed by the constant term. The second row lists the standard errors of the estimates. The third row includes the value of r^2 and the standard deviation of the residuals *s*.

Appendix B: Using Excel

6. For ease of reading and interpretation, type the appropriate labels as shown in the screen shot below.

	A	B	C	D	E	F	G	H	I	J	K
1	AvgAtt	Wins	MShare	Finals			Finals	Mshare	Wins	Constant Term	
2	16748	47	2.07	0		Coefficients	1010.606	269.4516	69.06253	13819.70407	
3	18624	62	2.105	1		Standard Errors	1085.791	232.1785	34.78451	1470.066254	
4	14526	35	0.981	0		r-squared and s	0.306171	2108.498	#N/A	#N/A	
5	21197	41	3.052	0							

Chapter 13: Making a Residual Plot

Follow these steps for creating a residual plot for a least-squares regression line summarizing the relationship between x = average driving distance and y = driving accuracy for the top 10 winners on the 2009 LPGA tour. Use the data in the *Chapter 10: Making a Scattterplot* section.

1. Follow steps 1 and 2 from the *Chapter 11: Calculating the Equation of a Least-Squares Regression Line* section to find the slope and y intercept of a least squares regression line.
2. Follow steps 2 and 3 from the *Chapter 11: Calculating the Standard Deviation of the Residuals* section to calculate the residuals. In the screen shot below, the average driving distance (*x*) is listed in column A and the residuals are listed in column C.

|	A	B	C	D	E	F	G	H	I	J	K	L
1	246.8	0.824	0.03048			Slope	-0.0056					
2	265.2	0.718	0.02661			y intercept	2.1633					
3	254.3	0.757	0.00511									
4	263.7	0.714	0.01429									
5	255.5	0.733	-0.01223									
6	268.1	0.66	-0.01529									
7	269.2	0.654	-0.01519									
8	249.3	0.748	-0.03164									
9	248.6	0.811	0.02748									
10	245.7	0.77	-0.02962									
11												

3. To make a residual plot, highlight the average driving distances in cells A1 to A10, hold the **Ctrl** button and highlight the residuals in cells C1 to C10. Open the **Insert** menu, choose the **Scatter** icon, then the **Scatter with only Markers** in the upper left.

4. Follow steps 3 – 6 from the *Chapter 10: Making a Scatterplot* section to format the scatterplot.

5. Here is an example of a well-labeled residual plot.

## Chapter 13: Calculating a Quadratic Model

Follow these steps for creating a quadratic model for the quarterback data on page 499.

1. Enter the age data in cells A1 to A32 and the passing yards data in cells B1 to B32.

2. Highlight the data in cells A1 to B32, open the **Insert** menu, choose the **Scatter** icon, then the **Scatter with only Markers** in the upper left.
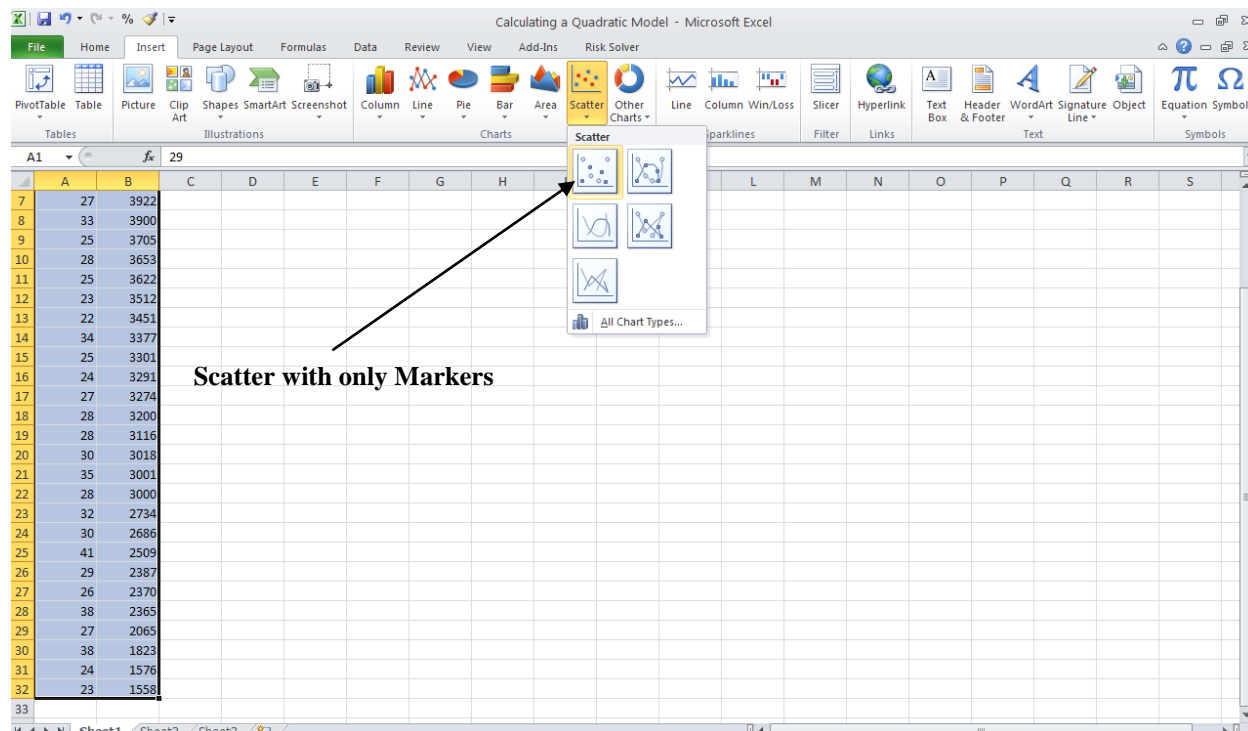
3. Click on the scatterplot and select the **Layout** tab of the **Chart Tools**. Select the **Chart Title** tab and type *2010 NFL Quarterbacks*. Then select the **Axis Titles** tab, label the **Primary Horizontal Axis Title** (*x* axis) *Age* and the **Primary Vertical Axis Title** (*y* axis) *Passing Yards*.
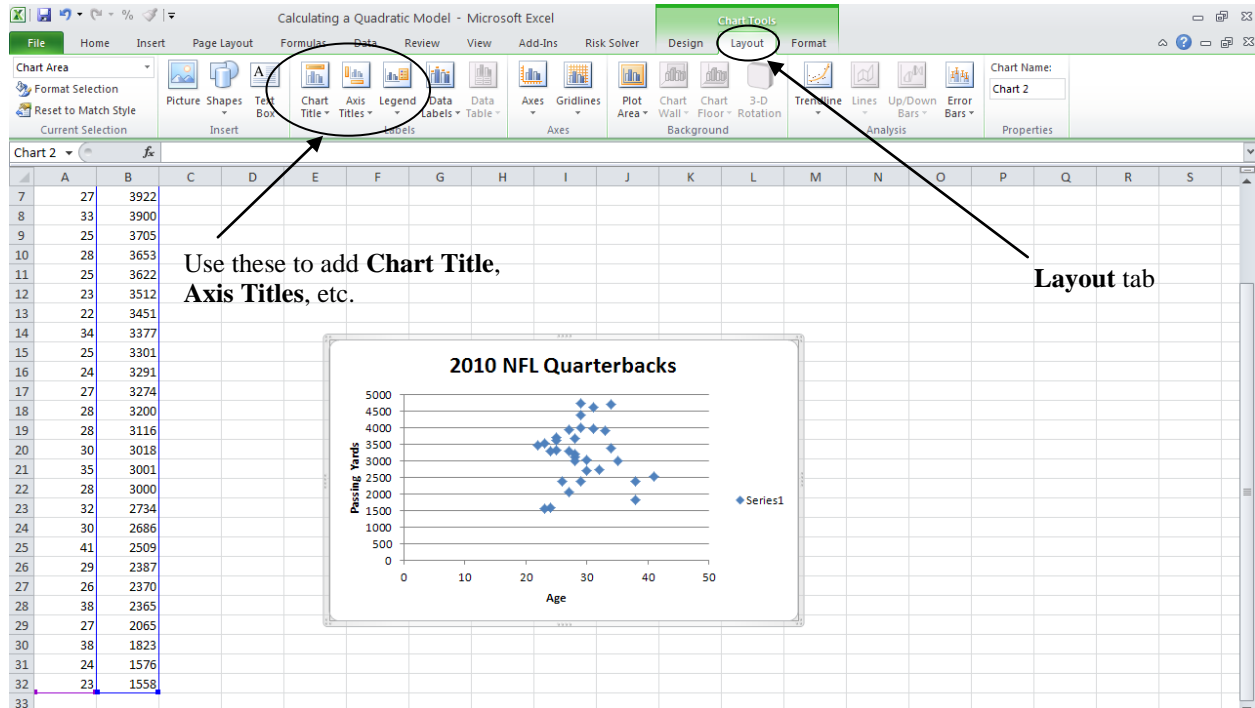

4. To eliminate the legend, select the **Layout** tab of the **Chart Tools**. Select the **Legend** tab and choose **None**. To eliminate the vertical axis major grid lines, right-click on any one of the lines and choose **Delete**.

5. To display a better scale for the horizontal axis, right click anywhere on the horizontal axis, then click on **Format Axis**. Under **Axis Options**, enter 20 for the minimum, 45 for the maximum, and 5 for the major unit. To display a better scale for the vertical axis, right click anywhere on the vertical axis, then click on **Format Axis**. Under **Axis Options**, enter 1000 for the minimum, 5000 for the maximum, and 500 for the major unit.
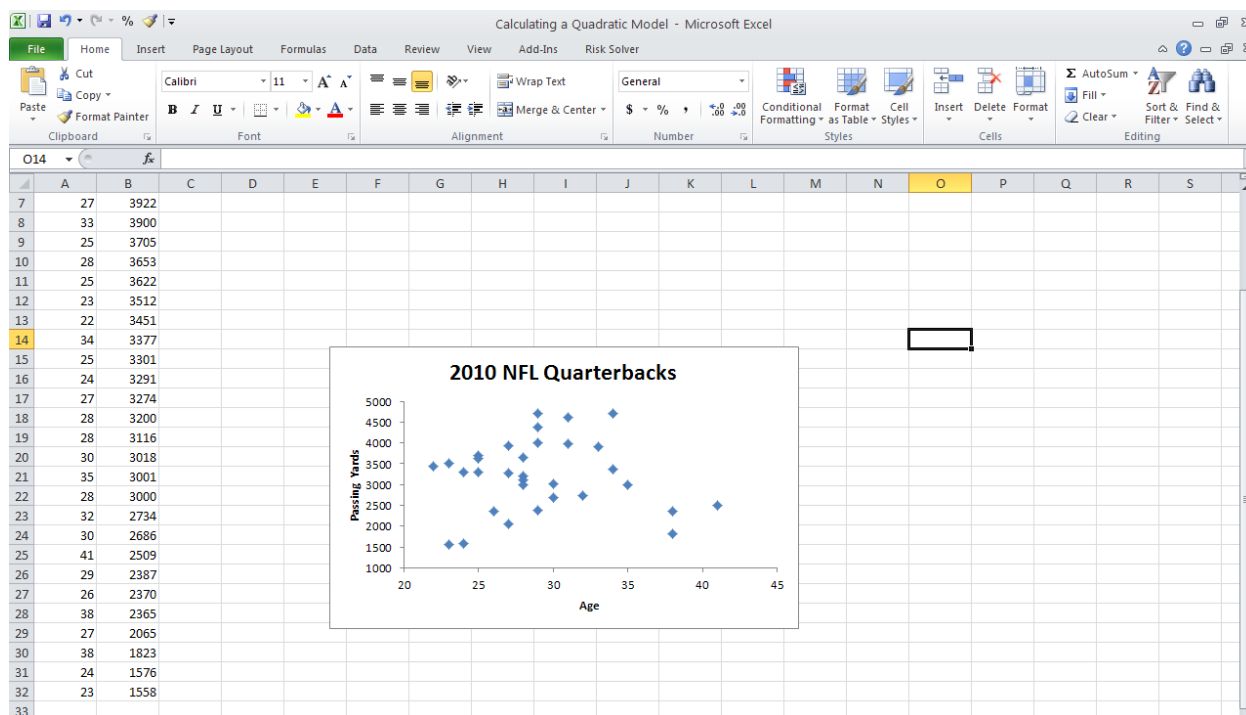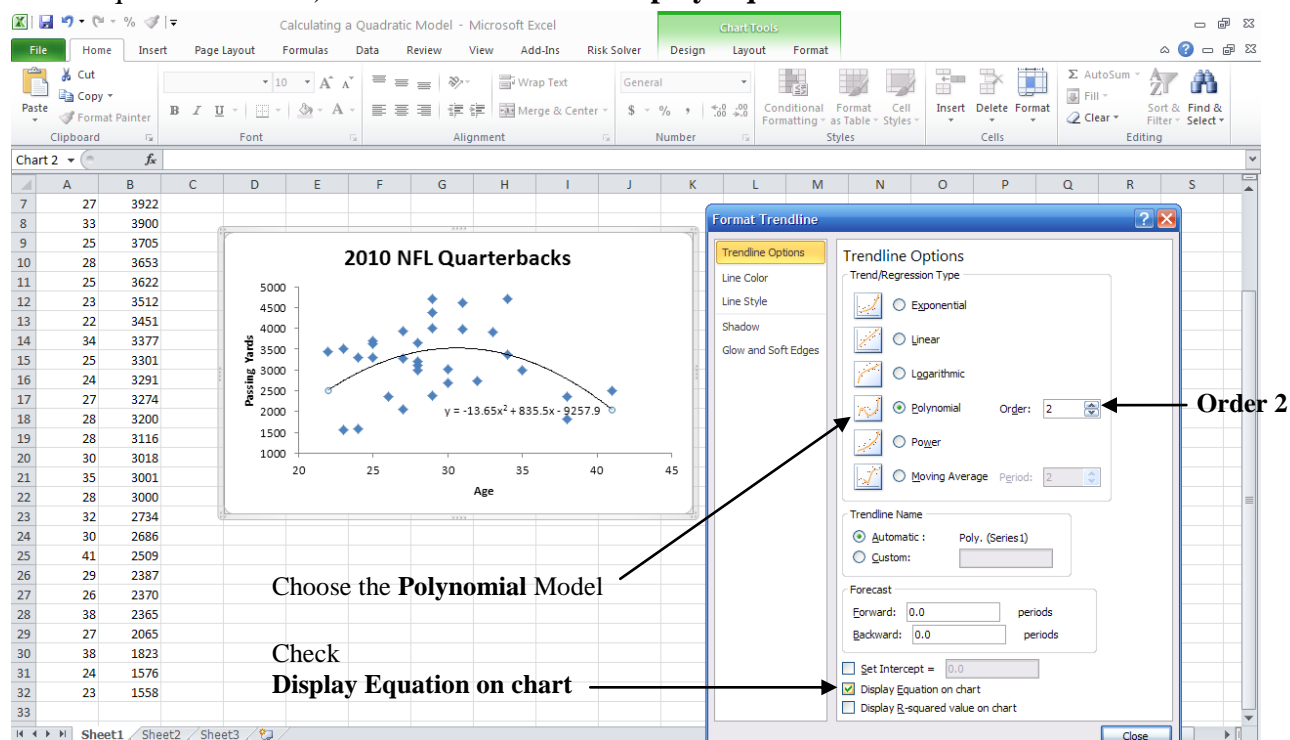
6. Right-click on any one of the points in the scatterplot and choose the **Add Trendline** option.
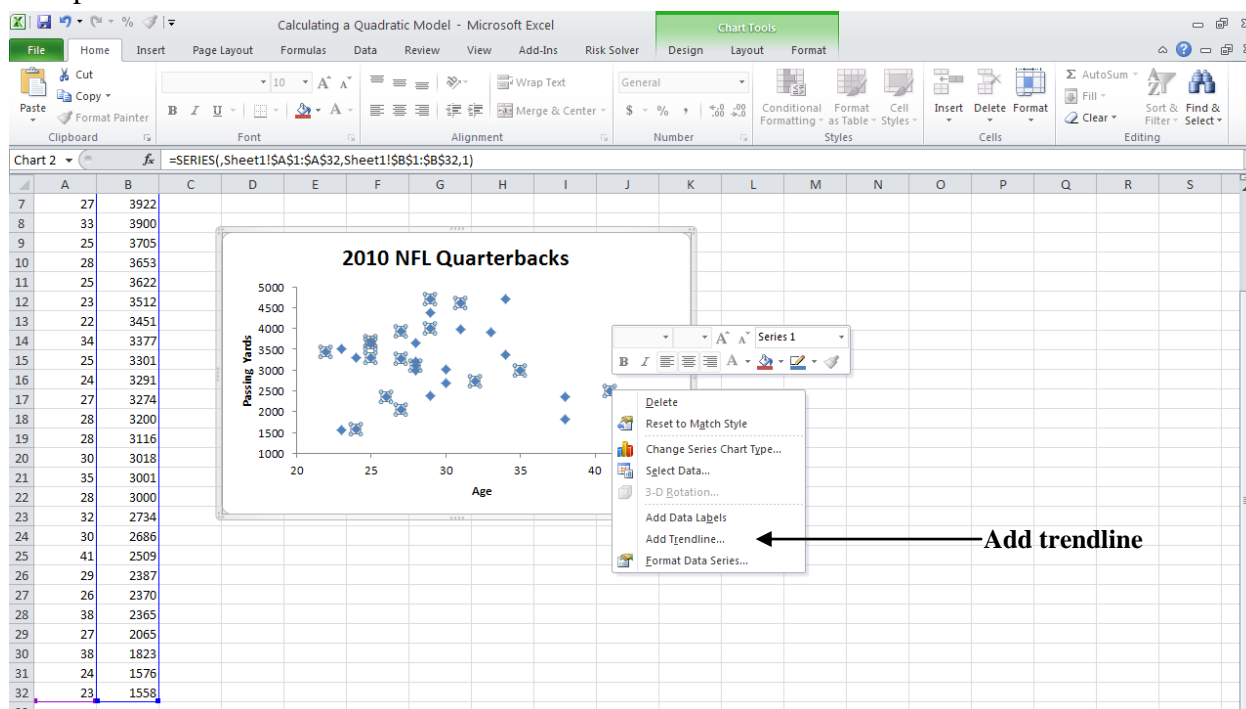
Appendix B: Using Excel

7. Under **Trendline Options**, choose the **Polynomial** model, leave the **Order** as 2 (for a quadratic model) and check the box to **Display Equation on chart**.


8. If the equation shows up in an awkward place, drag it to a better location.
9. If a different background color is desired, right-click on the background and choose **Format Plot Area**. Select **Fill**, then the desired options.
10. Here is an example of a well-labeled scatterplot for a quadratic model with the equation displayed.
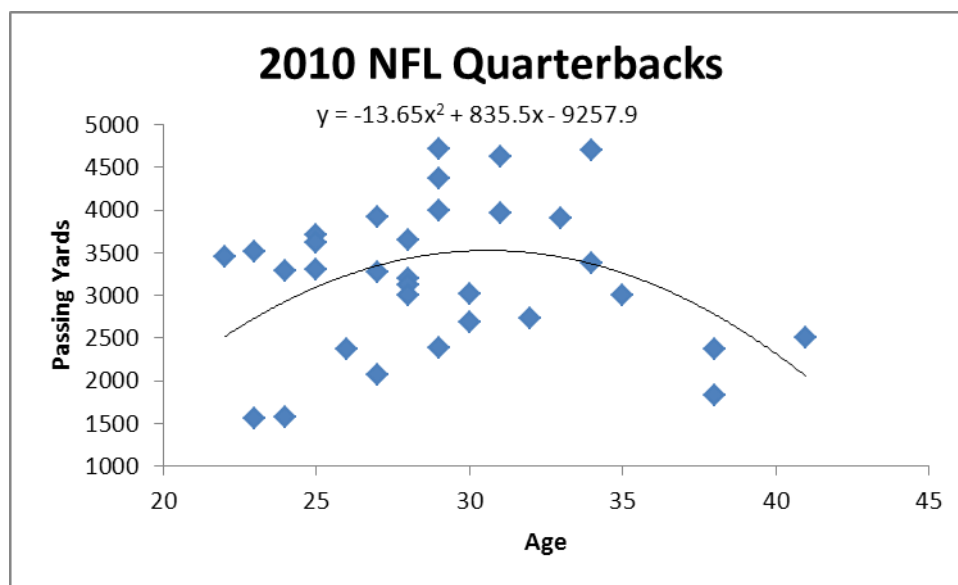
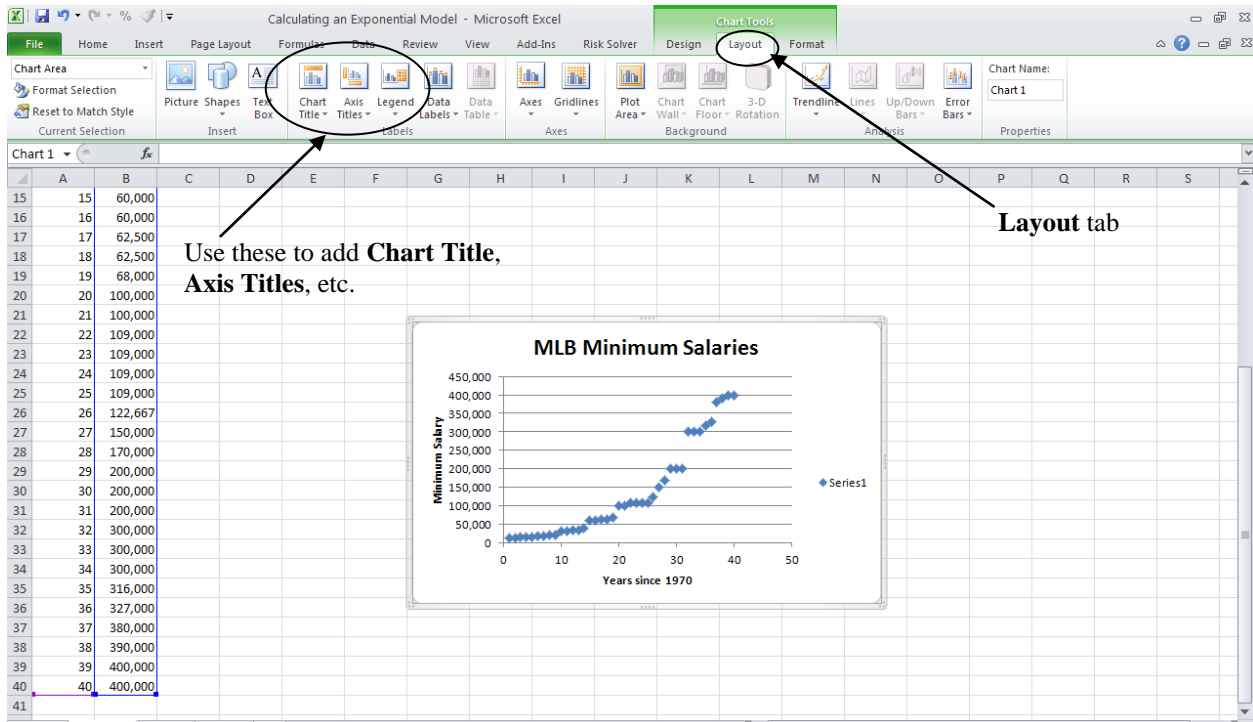## Chapter 13: Calculating an Exponential Model

Follow these steps for creating an exponential model for the baseball salary data on page 506.

1. Enter the year numbers in cells A1 to A40 and the minimum salaries in cells B1 to B40.

2. Highlight the data in cells A1 to B40, open the **Insert** menu, choose the **Scatter** icon, then the **Scatter with only Markers** in the upper left.
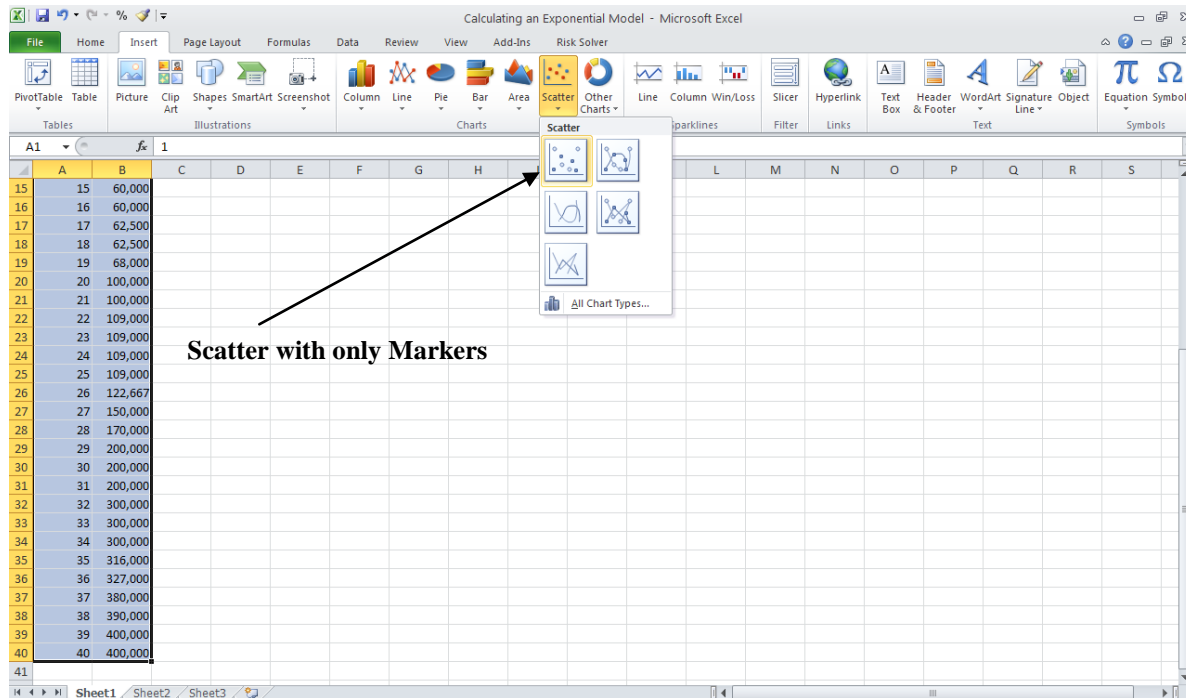

3. Click on the scatterplot and select the **Layout** tab of the **Chart Tools**. Select the **Chart Title** tab and type *MLB Minimum Salaries*. Then select the **Axis Titles** tab, label the **Primary Horizontal Axis Title** (*x* axis) *Years since 1970* and the **Primary Vertical Axis Title** (*y* axis) *Minimum Salary*.

Appendix B: Using Excel

4. To eliminate the legend, select the **Layout** tab of the **Chart Tools**. Select the **Legend** tab and choose **None**. To eliminate the vertical axis major grid lines, right-click on any one of the lines and choose **Delete**.

5. Right-click on any one of the points in the scatterplot and choose the **Add Trendline** option.
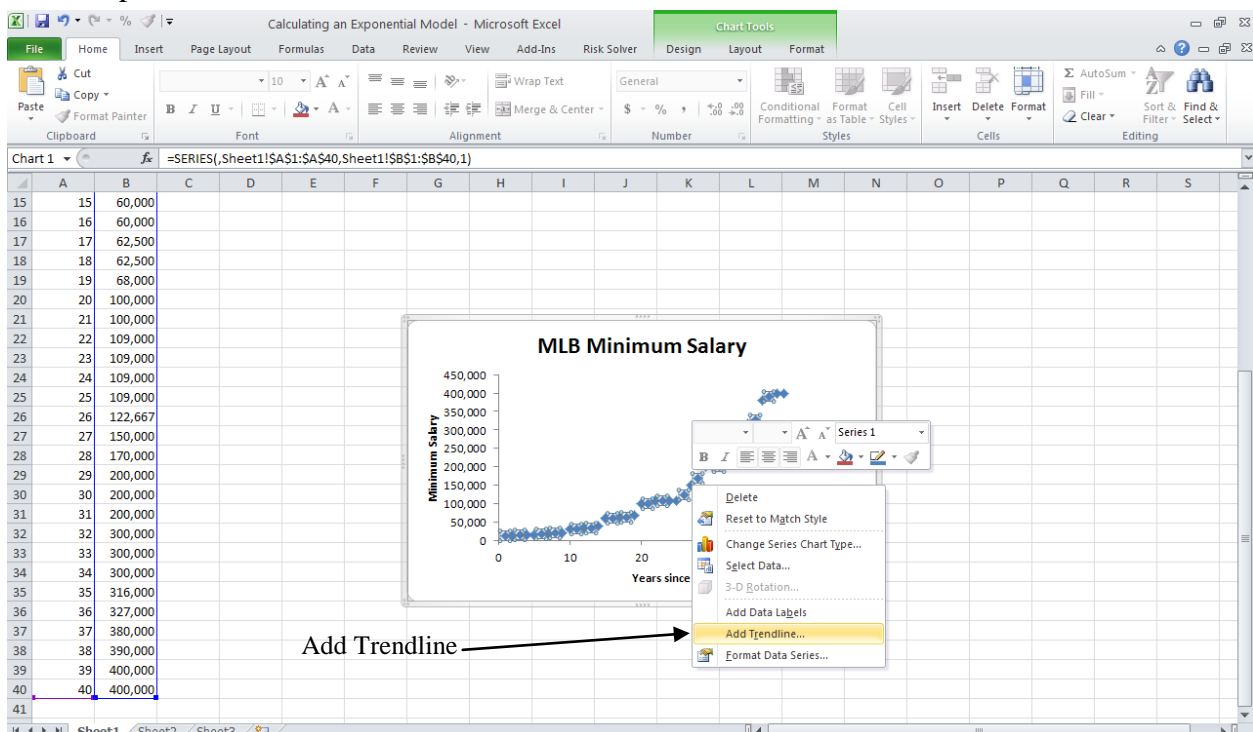
6. Under **Trendline Options**, choose the **Exponential** model and check the box to **Display Equation on chart**. Note that Excel automatically changed the maximum value of the vertical axis (and spacing) in order to accommodate the regression curve. Also, the form Excel uses for exponential models is slightly different than the form that we use in the book. For the baseball data, Excel's exponential model is $\hat{y} = 11075e^{0.0961x}$, where e is the mathematical constant $e \approx 2.718\ldots$. Evaluating the expression $e^{0.0961}$ gives 1.101, so the model is equivalent to $\hat{y} = 11075(1.101)^x$, the model we calculated in the chapter. (To evaluate the expression $e^{0.0961}$ using EXCEL, use the command $\boxed{=\exp(0.0961)}$.)
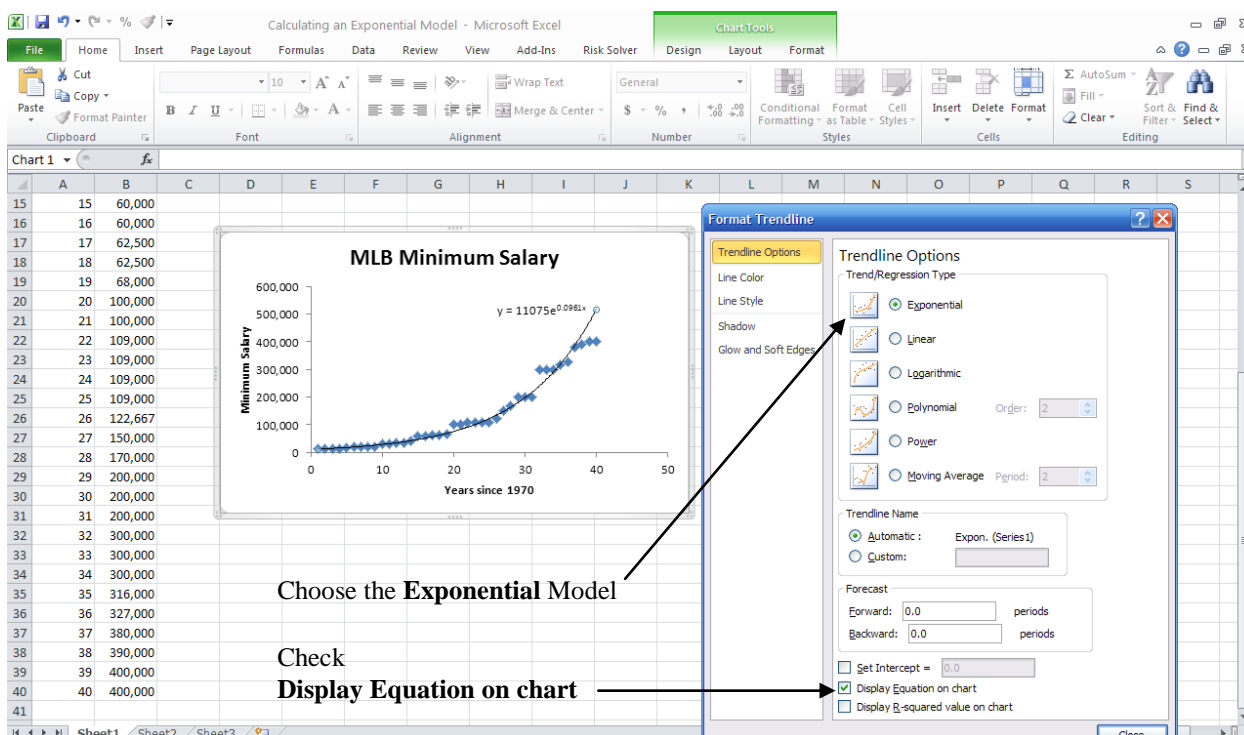
7. If the equation shows up in an awkward place, drag it to a better location.
8. If a different background color is desired, right-click on the background and choose **Format Plot Area**. Select **Fill**, then the desired options.
9. Here is an example of a well-labeled scatterplot for an exponential model with the equation displayed.

Appendix B: Using Excel
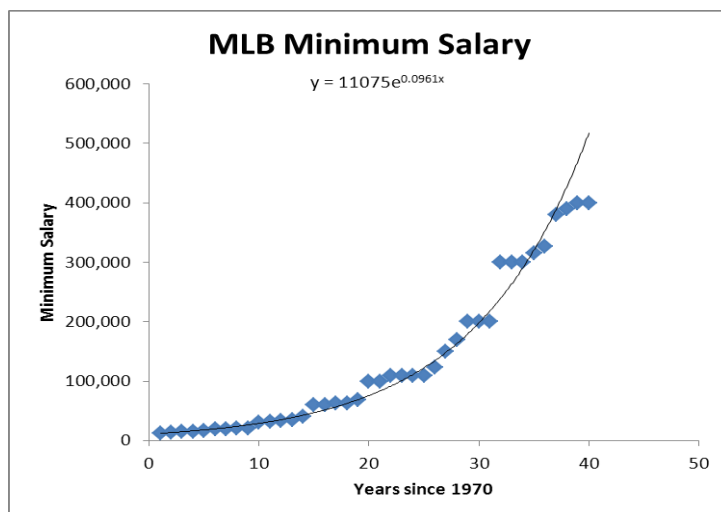
Chapter 13: Calculating Predicted Values for Logistic Models

Follow these steps for calculating a predicted proportion of successes when shooting a basketball from x feet using the logistic model

$$\hat{p} = \frac{e^{3.77-0.226x}}{1+e^{3.77-0.226x}} .$$

1. To predict the proportion of made shots when $x = 7$, enter the value 7 in cell A1.
2. In cell B1, enter the function | =exp(3.77−0.226*A1)/(1+exp(3.77−0.226*A1)) | .
3. To predict the proportion of successes for other values of x, type the value of x into cell A1. You can also enter many different values of x in column A and copy the equation from cell B1 down column B.

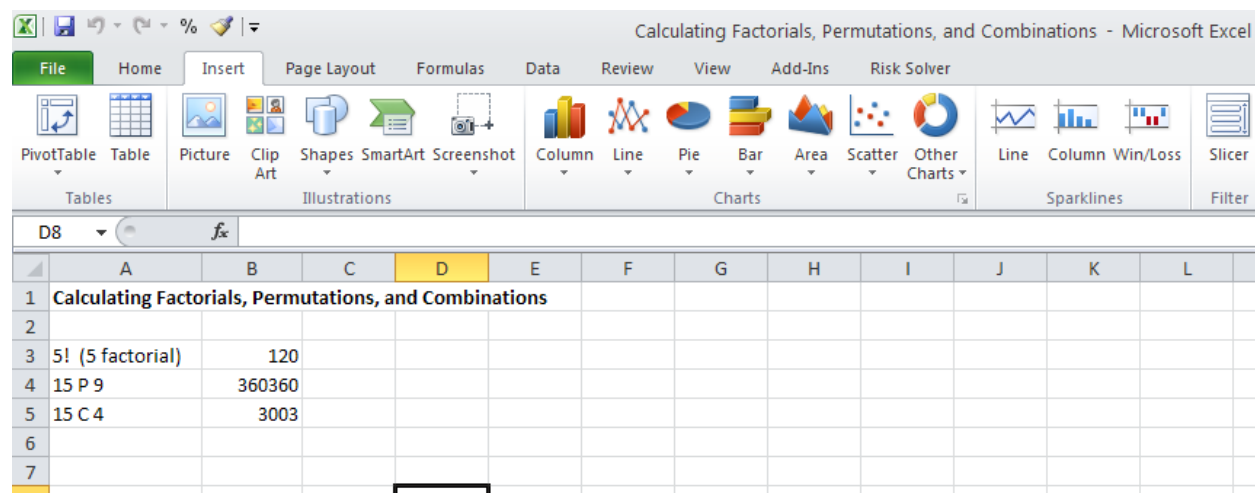## Chapter 14: Calculating Factorials, Permutations, and Combinations

Follow these steps for calculating factorials, permutations, and combinations.
Enter the following labels and formulas using the indicated format.

	A	B
3	5! (5 factorial)	=fact(5)
4	15 P 9	=permut(15,9)
5	6 C 4	=combin(6,4)

The results are shown in the screen shot below.

Appendix B: Using Excel

Chapter 14: Binomial Distribution Calculations

The process for calculating binomial probabilities is to use the function binomdist(x, n, p, cumulative). In this function, x is the number of successes, n is the number of attempts, p is the probability of success, and cumulative is a yes (1) or no (0), depending if you want to calculate the probability of a specific number of successes (cumulative $= 0$) or the probability of less than or equal to a specific number of successes (cumulative $= 1$).

Follow these steps for calculating probabilities for LeBron James assuming that his ABILITY to make a three-point shot is $p = 0.315$ and he takes $n = 70$ attempts.

1. To calculate the probability that he makes exactly 18 shots in 70 attempts, enter the formula =binomdist(18,70,0.315,0) in any cell. Press **Enter** to display the answer 0.0618 as shown in cell B3 of the screen shot below.

2. To calculate the probability that he makes less than or equal to 18 shots in 70 attempts, enter the formula =binomdist(18,70,0.315,1) in any cell. Press **Enter** to display the answer 0.1812 as shown in cell B5 of the screen shot below.

3. To calculate the probability that he makes 18 *or more* shots, we need to find the probability that he makes 17 or fewer and subtract the result from 1. Enter the formula =1-binomdist(17,70,0.315,1) in any cell. Press **Enter** to display the answer 0.8806 as shown in cell B7 of the screen shot below.

Appendix B: Using Excel

Binomial Distribution Calculations - Microsoft Excel

	A	B	C	D	E	F	G	H	I
1	**Binomial Distribution Calculations**								
2									
3	P(Exactly 18 shots in 70 attempts)	0.0618							
4									
5	P(Less than 18 shots in 70 attempts)	0.1812							
6									
7	P(18 or more shots in 70 attempts)	0.8806							
8									
9									