Appendix C: Using Fathom

Importing Data from the Web into Fathom

There are several ways you can import data into Fathom depending on how the data is displayed on a website.

The Easy Method

1. Open the webpage with the data you want to import.
2. Open a blank Fathom document.
3. Drag the icon to the left of the url in your web browser onto the Fathom document.
4. Fathom will create a new collection with the data from the webpage.
5. To see if the data imported properly, highlight the collection and drag a table icon from the shelf to the workspace. Depending on how the data is formatted on the webpage, you may need to delete some cases (rows) or attributes (columns) or rename some attributes (column headings).
6. If there is more than one data table on the webpage, Fathom will choose one of them to import. If Fathom doesn't pick the one you want or the data didn't import properly, try to the method below.

The Not-as-Easy Method

1. Highlight the data on the webpage, including the variable names. Do not highlight anything else, including rows with totals, unless you want them imported as well.
2. Copy the data by right-clicking and choosing Copy or pressing ctrl-C on the keyboard.
3. In Fathom, drag a blank collection (not a table) from the shelf to the workspace.
4. Highlight the blank collection and paste the data by right-clicking and choosing Paste or pressing ctrl-V on the keyboard.
5. To see if the data imported properly, highlight the collection and drag a table icon to the workspace. Depending on how the data is formatted on the webpage, you may need to delete some cases (rows) or attributes (columns) or rename some attributes (column headings).

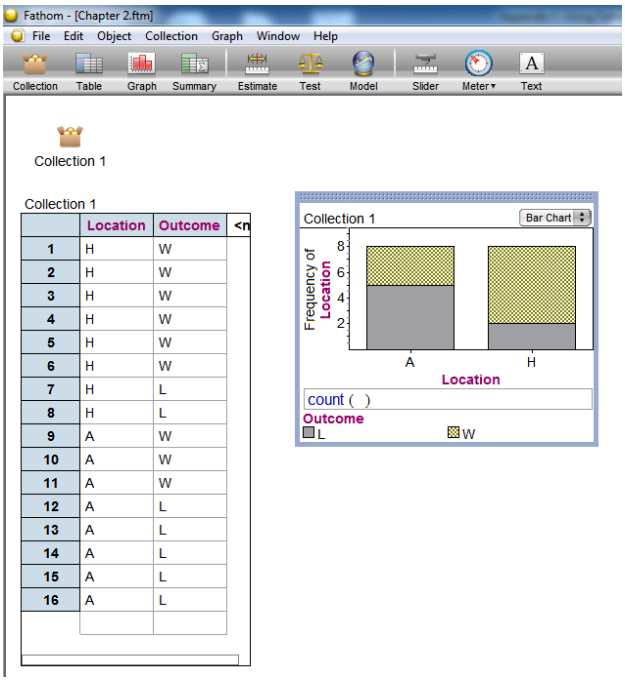
Note: Some websites, such as www.baseball-reference.com and its siblings, allow you to choose how to display the data. The Easy method works well if there is only one table on the webpage. If the Easy method doesn't work, click on the "pre" link above the data table and follow the instructions for the Not-as-Easy Method.

Chapter 1: Graphing Categorical Data

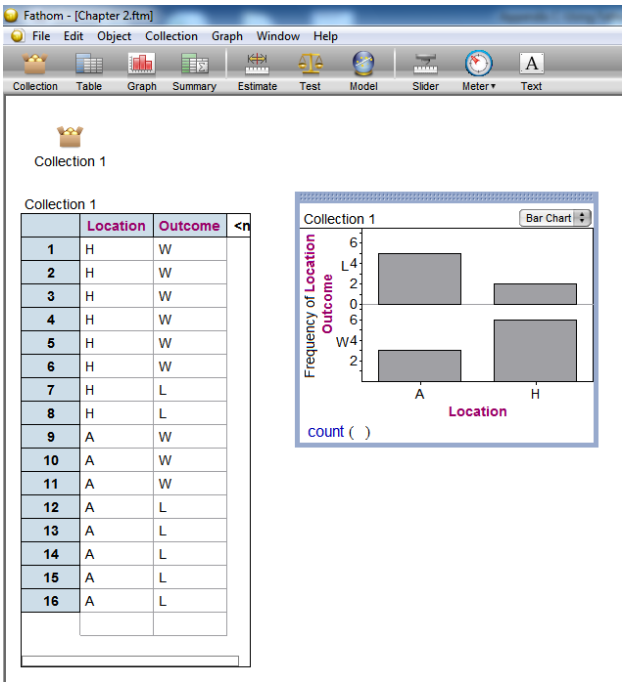
Note: To make graphs of categorical data in Fathom, you first must enter each outcome individually in a table. Unlike Excel, you cannot simply enter the counts from a two-way table. Also, you cannot make pie charts in Fathom.

Follow these steps to make a graph comparing the *PERFORMANCES* of the 2008 Arizona Cardinals at home (6 wins and 2 losses) with their *PERFORMANCES* on the road (3 wins and 5 losses):

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Location" and the second column "Outcome." Fill in the data as shown in the screen shot.
2. Drag a graph icon from the shelf onto the workspace. Then, drag the column heading "Location" to the horizontal axis of the graph. This will create a bar chart showing that there were 8 games at home and 8 games on the road.
3. To create a segmented bar chart, drag the column heading "Outcome" into the middle of the graph. The results should look like the screen shot below.



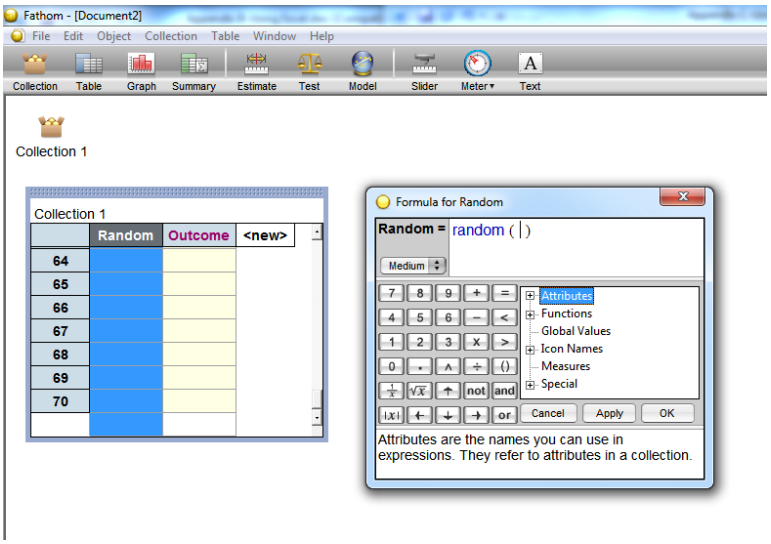
4. To make a comparative bar chart, drag a graph icon from the shelf onto the workspace. Then, drag the column heading "Location" to the horizontal axis of the graph. This will create a bar chart showing that there were 8 games at home and 8 games on the road. Finally, drag the column heading "Outcome" to the vertical axis. The results should look like the screen shot below.


Fathom - [Chapter 2.ftm]
File Edit Object Collection Graph Window Help
Collection Table Graph Summary Estimate Test Model Slider Meter Text
Collection 1
Collection 1
Location Outcome
1 H W
2 H W
3 H W
4 H W
5 H W
6 H W
7 H L
8 H L
9 A W
10 A W
11 A W
12 A L
13 A L
14 A L
15 A L
16 A L
Collection 1
Bar Chart
Frequency of Location Outcome
6
L4
2
0
6
W4
2
A
H
Location
count ( )

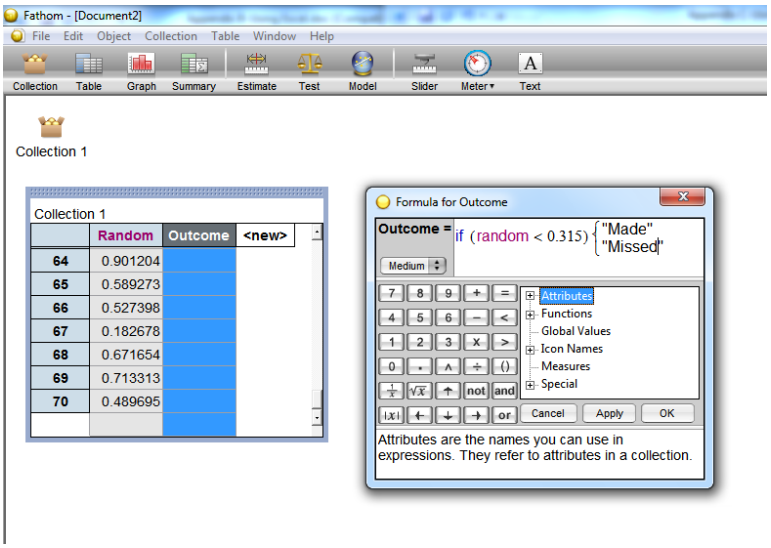
Chapter 1: Simulating Athletic Performance

Follow these steps for simulating LeBron James' three-point shooting *PERFORMANCE* in 70 attempts assuming that his *ABILITY* to make a three-point shot is 31.5%.

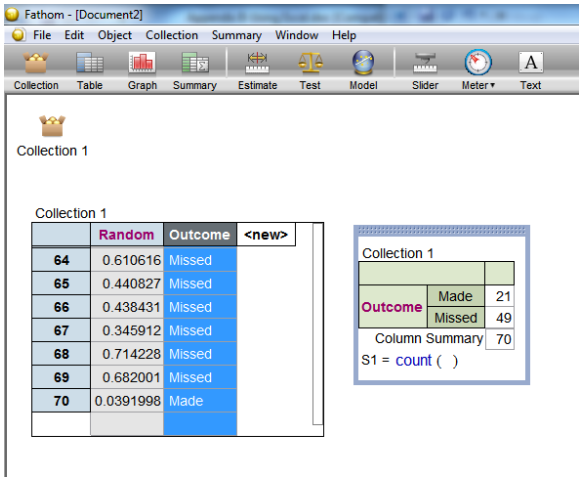
1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Random" and the second column "Outcome."
2. Add 70 cases (rows) by right-clicking in the table and choosing "New Cases…". Then, enter 70 into the dialog box and press OK.
3. To populate the first column with random numbers from 0 to 1, right-click the heading of the "Random" column and choose "Edit Formula." Then, type the following formula: random(. When Fathom recognizes this function, the word random will turn blue and Fathom will put the end parenthesis as shown in the screen shot below. Press OK. *Note: It is possible to choose the random function from a menu when in the formula editor by clicking on "Functions" and then "Random Numbers." You can also find the formula for generating random integers and other types of random numbers in this menu.*



4. To populate the second column with the simulated outcomes of LeBron's shot attempts, right-click the "Outcome" column and choose "Edit Formula." Then, type in the following formula: if(. When Fathom recognizes this function, the word if will turn blue and Fathom will create places to put a conditional statement and two different outcomes. Enter random < 0.315 in the parentheses, "Made" in the top space of the bracket and "Missed" in the bottom space of the bracket as shown in the screen shot below. Press OK. For cases (rows) where the value in the random column is less than 0.315, Fathom will put the word "Made" in the outcome column. Likewise, if the value in the random column is not less than 0.315, Fathom will put the word "Missed" in the outcome column.

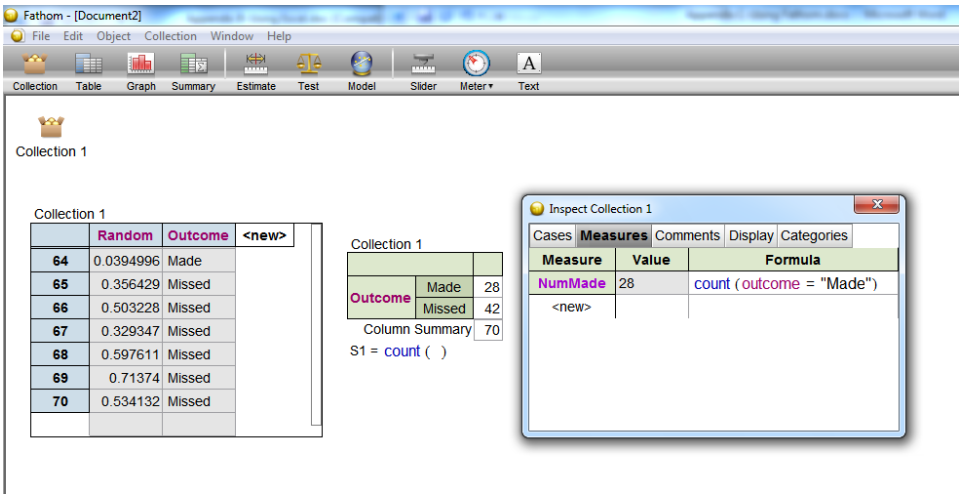


5. To have Fathom count how many simulated shots that LeBron made, drag a summary icon from the shelf onto the workspace. Then, drag the "Outcome" heading to the lower-left cell of the summary box, as shown in the screen shot below. In this simulated set of 70 shots, LeBron made 21 and missed 49.

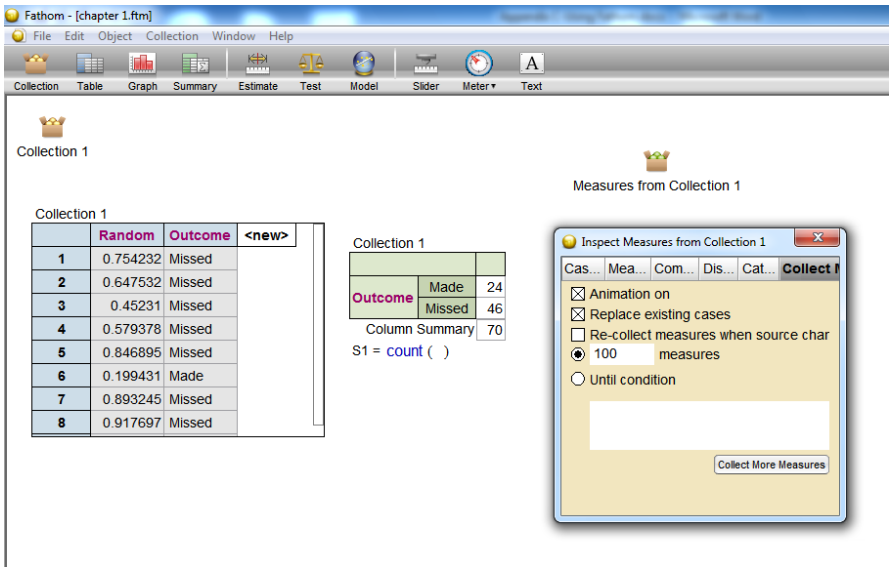


6. To have Fathom generate more simulated sets of 70 shots, simply press ctrl-Y over and over again.

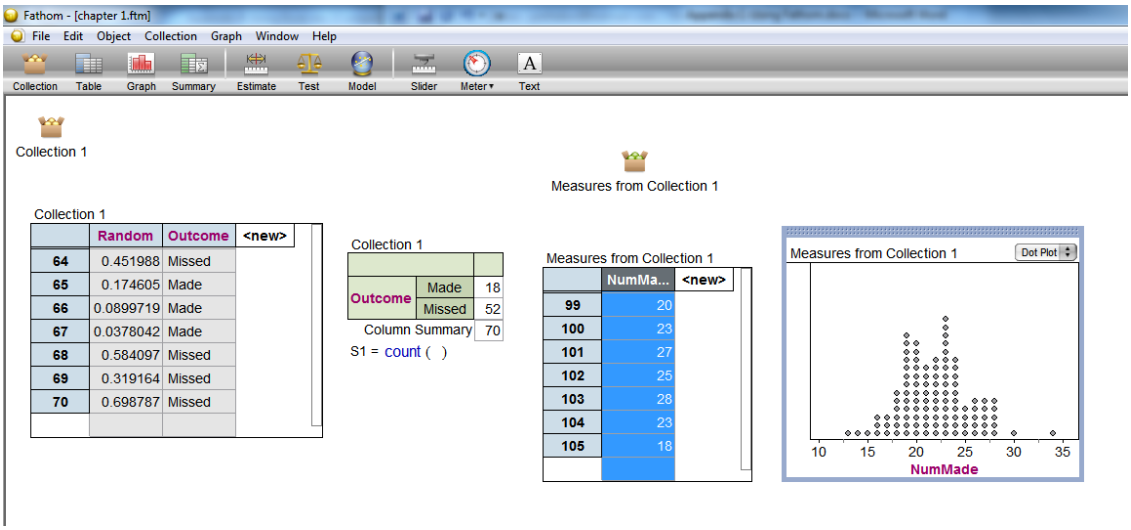
7. To have Fathom collect the results of multiple sets of 70 shots so that you can display the results in a dotplot, double-click on the box called "Collection 1" and choose the Measures tab. In the Measure column, create a measure called "NumMade." Then, double click in the formula box and enter the following function: count(outcome = "Made") as shown in the screen shot below.



8. Next, right-click on the box called “Collection 1” and choose “Collect Measures.” In the inspector, select “Replace existing cases” and enter 100 instead of 5 for the number of measures. Then, press “Collect More Measures.” This will have Fathom simulate 100 sets of 70 shots and record the number of shots made in each of the 100 sets of 70. To increase the number of measures, uncheck the “Replace existing cases” box, enter the number of additional measures desired, and press “Collect more measures.” The simulation will go much faster if you uncheck the “Animation on” box.



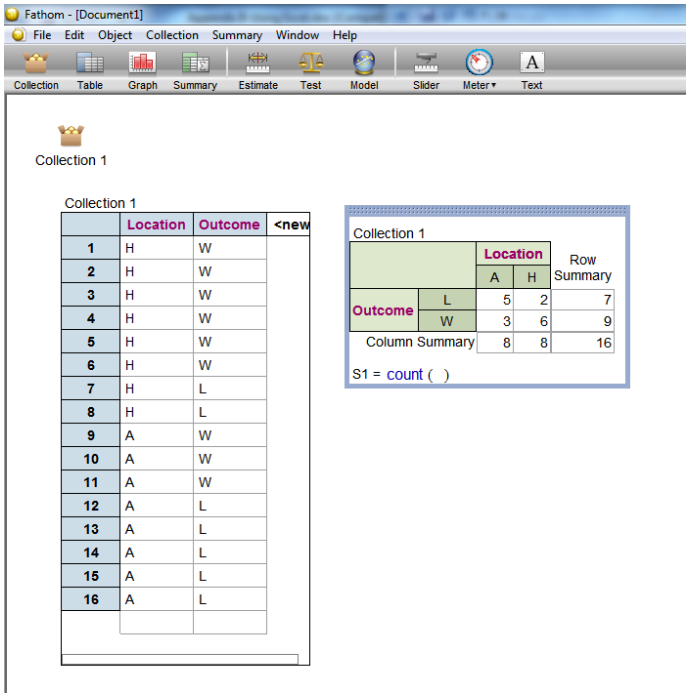
9. Finally, to create a dotplot of the results, drag a graph icon to the workspace. Then, drag the “NumMade” heading to the horizontal axis of the graph, as shown in the screen shot below. If you need to rescale the axis, you can either double click on the axis and change the settings or just drag the values on the horizontal axis to the left or right.


Fathom - [chapter 1.ftm]
File Edit Object Collection Graph Window Help
Collection Table Graph Summary Estimate Test Model Slider Meter Text
Collection 1
Measures from Collection 1
Collection 1
Random Outcome <new>
64 0.451988 Missed
65 0.174605 Made
66 0.0899719 Made
67 0.0378042 Made
68 0.584097 Missed
69 0.319164 Missed
70 0.698787 Missed
Collection 1
Outcome Made 18
Missed 52
Column Summary 70
S1 = count ( )
Measures from Collection 1
NumMa... <new>
99 20
100 23
101 27
102 25
103 28
104 23
105 18
Measures from Collection 1
Dot Plot
10 15 20 25 30 35
NumMade

Chapter 2: Creating a Two-Way Table

Follow these steps to make a two-way table that summarizes the *PERFORMANCES* of the 2008 Arizona Cardinals at home (6 wins and 2 losses) with their *PERFORMANCES* on the road (3 wins and 5 losses):

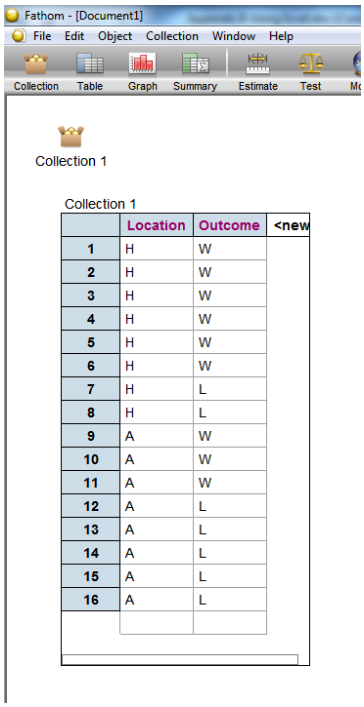
1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Location" and the second column "Outcome." Fill in the data as shown in the screen shot below.
2. Drag a summary icon from the shelf onto the workspace. Drag the "Outcome" heading to the lower-left cell in the summary box and the "Location" heading to the upper-right cell in the summary box.



Chapter 2: Simulating a Difference in Proportions

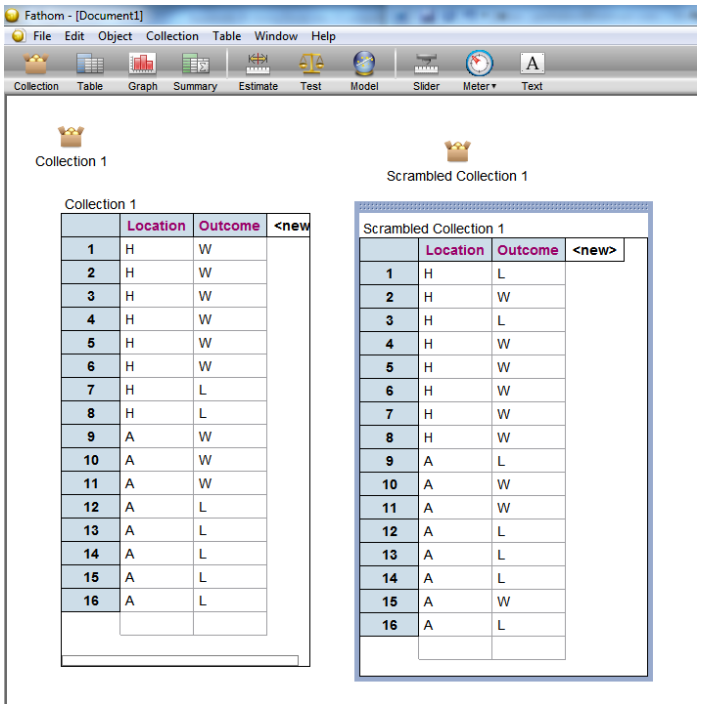
Follow these steps to simulate the difference in the *PERFORMANCE* of the 2008 Arizona Cardinals at home (6 wins and 2 losses) and on the road (3 wins and 5 losses):

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Location" and the second column "Outcome." Fill in the data as shown in the screen shot.



	Location	Outcome	<new
1	H	W	
2	H	W	
3	H	W	
4	H	W	
5	H	W	
6	H	W	
7	H	L	
8	H	L	
9	A	W	
10	A	W	
11	A	W	
12	A	L	
13	A	L	
14	A	L	
15	A	L	
16	A	L	

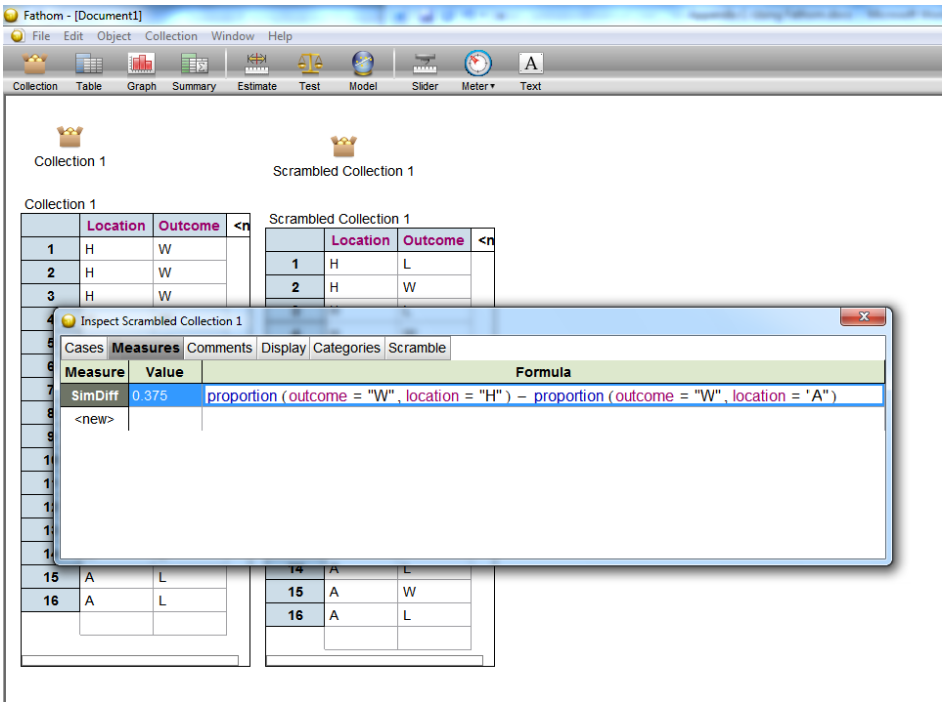
2. To shuffle the outcomes, right-click the box called "Collection 1" and choose "Scramble Attribute Values." In the inspector, change the attribute from Location to Outcome and press Scramble Attribute Values again. This shuffles up the W's and L's in the Outcome column. To see the results, highlight the box called "Scrambled Collection 1" and drag a table icon onto the workspace.



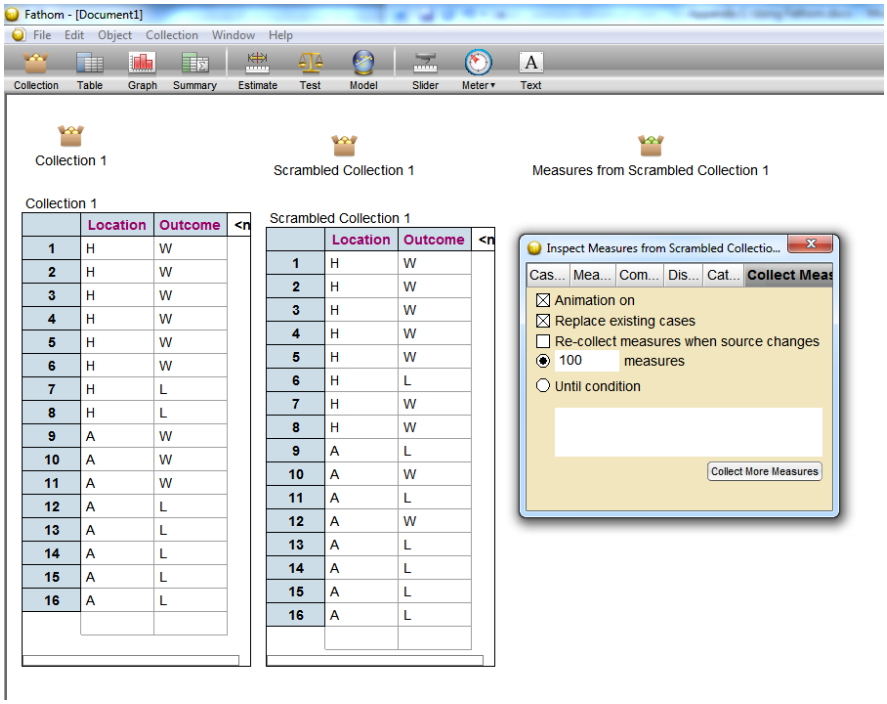
3. To have Fathom repeatedly shuffle the outcomes, calculate the difference in proportion of wins, and record these values, double-click on the box called "Scrambled Collection 1." Then, click on the "Measures" tab in the Scrambled Collection 1 Inspector. Under the column heading measure, type "SimDiff." Now, double click on the cell under the heading formula and type:

   proportion(outcome="W",location="H") – proportion(outcome="W",location="A")

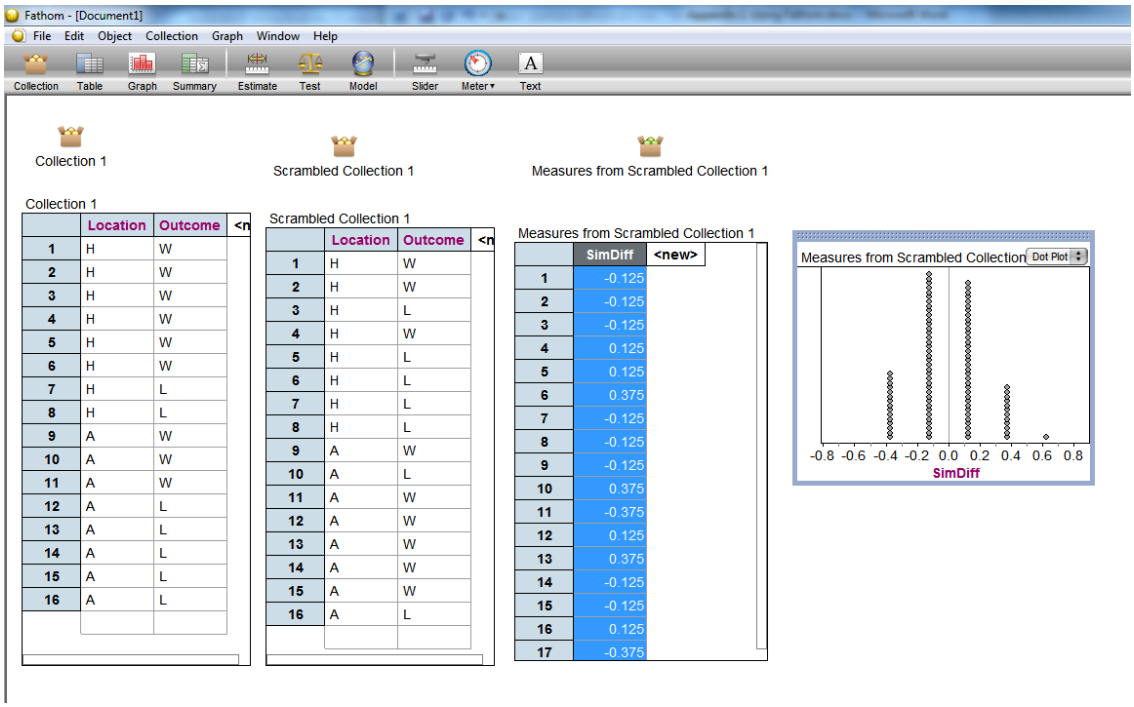
   Note: In the formula above, the commands after the comma indicate that these are conditional formulas. For example, the formula proportion(outcome="W",location="H") calculates the proportion of outcomes that were wins, but only for the games were the location was at home.



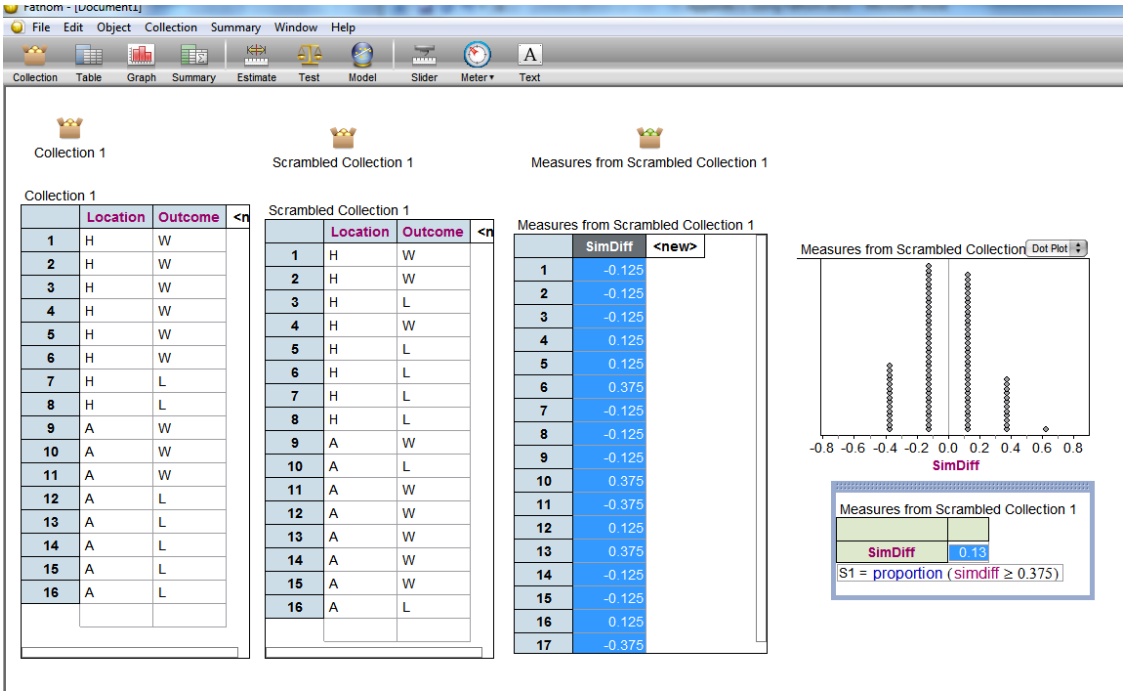
4. Next, right-click the box called “Scrambled Collection 1” and choose Collect Measures. In the inspector, select “Replace existing cases” and enter 100 instead of 5 for the number of measures. Then, press Collect More Measures. *Note:* To increase the number of measures, uncheck the “Replace existing cases” box, enter the number of additional measures desired, and press “Collect more measures.” The simulation will go much faster if you uncheck the “Animation on” box.



5. To see the results of the 100 trials, highlight the box called “Measures from Scrambled Collection 1” and drag a table icon from the shelf onto the workspace. Then, drag a graph icon from the shelf onto the workspace. Finally, drag the heading “SimDiff” from the table to the horizontal axis of the graph. If you need to rescale the axis, you can either double click on the axis and change the settings or just drag the values on the horizontal axis to the left or right.



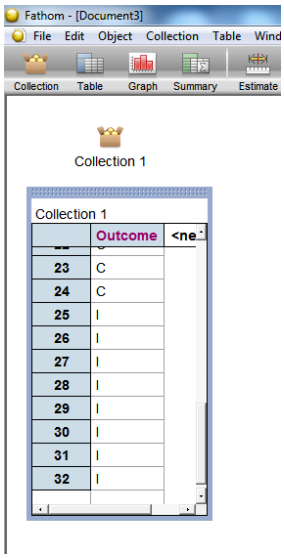
6. To have Fathom calculate the proportion of dots that are 0.375 or beyond, drag a summary icon from the shelf onto the workspace. Then, drag the SimDiff heading from the table into the lower-left corner of the summary box, double click on the formula S1=mean() and enter S1=proportion(SimDiff≥0.375). To find the ≥ symbol, hold the ctrl button and the > symbol will turn into a ≥ symbol in the dialog box.


File Edit Object Collection Summary Window Help
Collection Table Graph Summary Estimate Test Model Slider Meter Text
Collection 1
Scrambled Collection 1
Measures from Scrambled Collection 1
Collection 1
Location Outcome
Scrambled Collection 1
Location Outcome
Measures from Scrambled Collection 1
SimDiff <new>
Measures from Scrambled Collection
Dot Plot
-0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8
SimDiff
Measures from Scrambled Collection 1
SimDiff 0.13
S1 = proportion (simdiff ≥ 0.375)

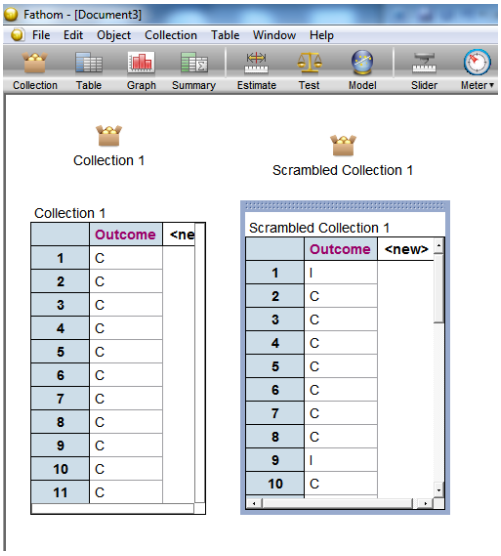
Chapter 3: Simulating the Longest Streak

To simulate the length of the longest streak using Tony Romo's data from Chapter 3, follow the following steps:

1. In a new Fathom document, drag a blank table onto the workspace and Label the first column Outcome. Then, enter 24 "C"s for Tony Romo's 24 complete passes and 8 "I"s for Romo's 8 incomplete passes.

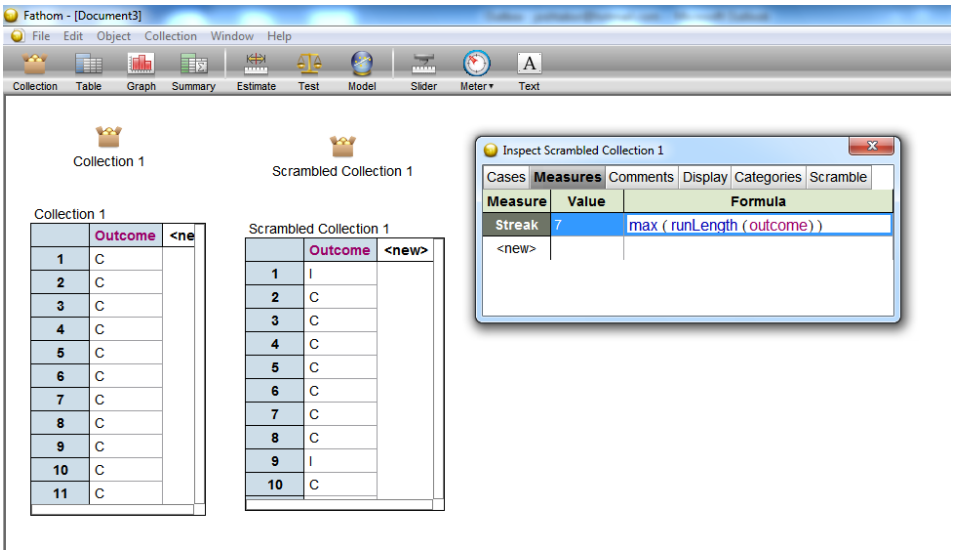


2. Right-click on the box called "Collection 1" and choose Scramble Attribute Values. This will shuffle the 32 outcomes in a random order. To see the results, highlight Scrambled Collection 1 with your mouse and drag a blank table onto the workspace.

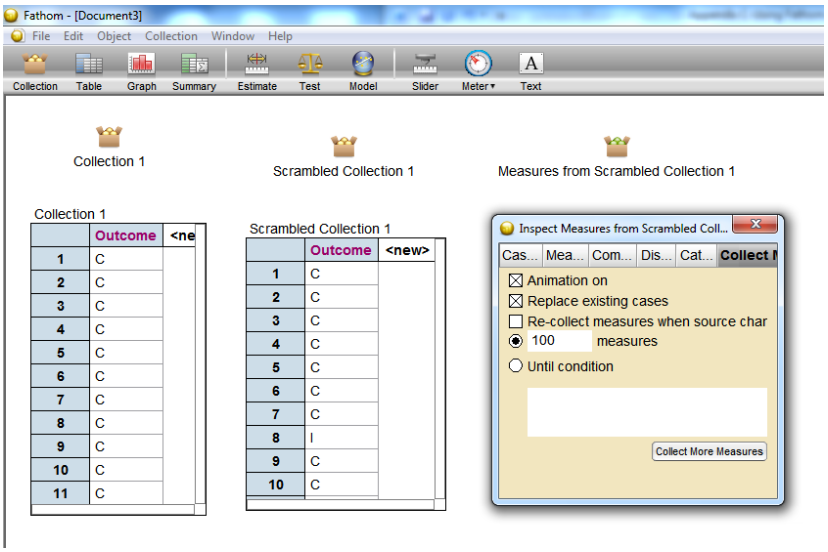


3. Double-click on the box called "Scrambled Collection 1" to open the inspector. In the "Measures" tab, enter the name "Streak" and the formula: max(runLength(outcome)). Press

Enter. This will display the longest streak of consecutive complete passes or incomplete passes.

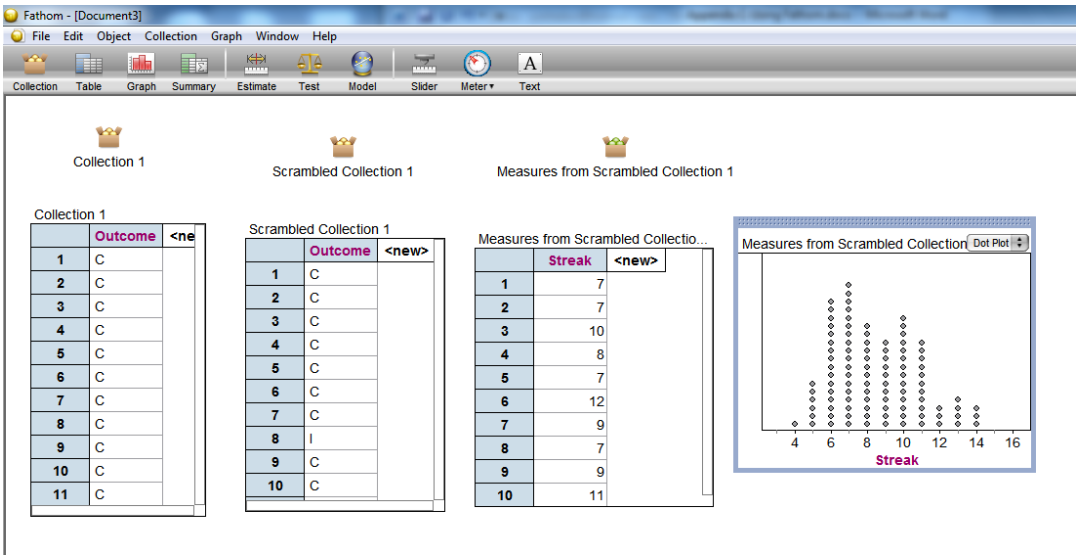


4. Right-click on the box called “Scrambled Collection 1” and choose Collect Measures. In the inspector, select “Replace existing cases” and enter 100 instead of 5 for the number of measures. Then, press Collect More Measures. *Note:* To increase the number of measures, uncheck the “Replace existing cases” box, enter the number of additional measures desired, and press “Collect more measures.” The simulation will go much faster if you uncheck the “Animation on” box.

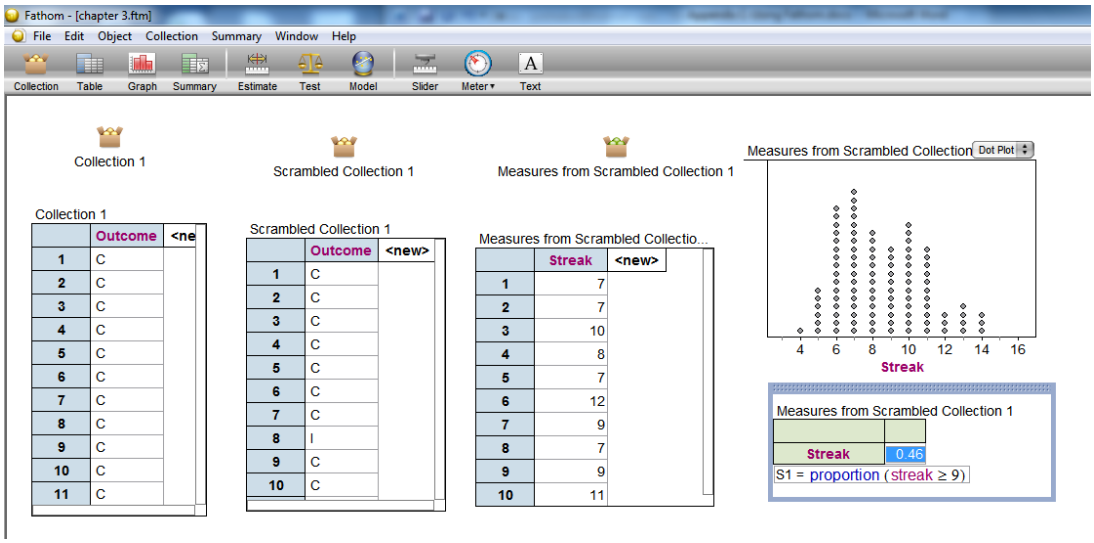


5. To see the results of the 100 trials, highlight the box called “Measures from Scrambled Collection 1” and drag a table icon from the shelf onto the workspace. Then, drag a graph icon from the shelf onto the workspace. Finally, drag the heading “Streak” from the table to the horizontal axis of the graph. If you need to rescale the axis, you can either double click

on the axis and change the settings or just drag the values on the horizontal axis to the left or right.



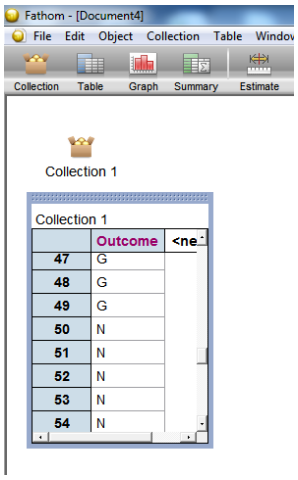
6. To have Fathom calculate the proportion of dots that are 9 or beyond, drag a summary icon from the shelf onto the workspace. Then, drag the Streak heading from the table into the lower-left corner of the summary box, double click on the formula S1=mean() and enter S1=proportion(Streak ≥ 9). To find the ≥ symbol, hold the ctrl button and the > symbol will turn into a ≥ symbol in the dialog box.



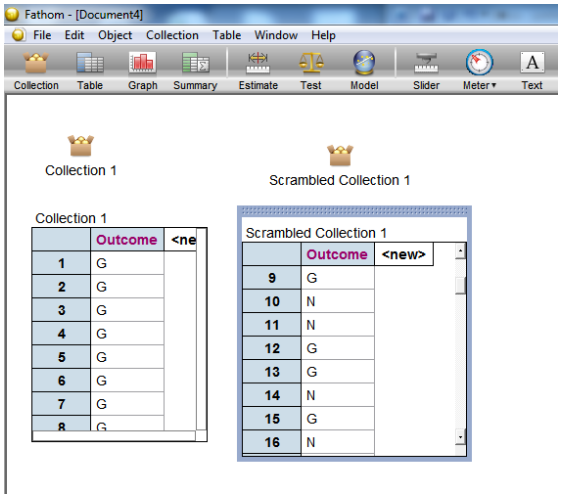
Chapter 3: Simulating the Number of Streaks

To simulate the number of streaks of 4 in a row using Mario Lemieux's data from Chapter 3, follow the steps below:

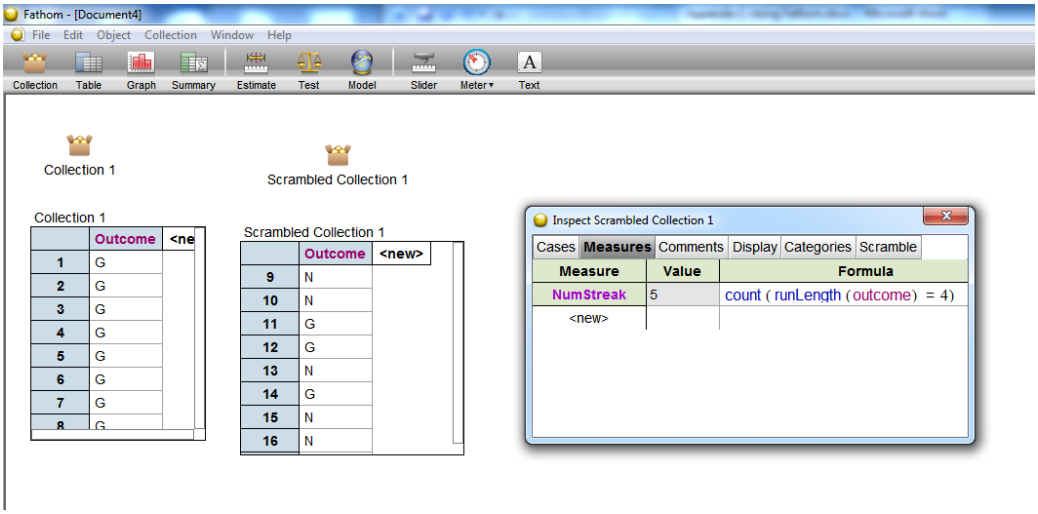
1. In a new Fathom document, drag a blank table onto the workspace and Label the first column Outcome. Then, enter 49 "G"s for the 49 games where Lemieux had a goal and 27 "N"s for the 27 games where he did not have a goal.



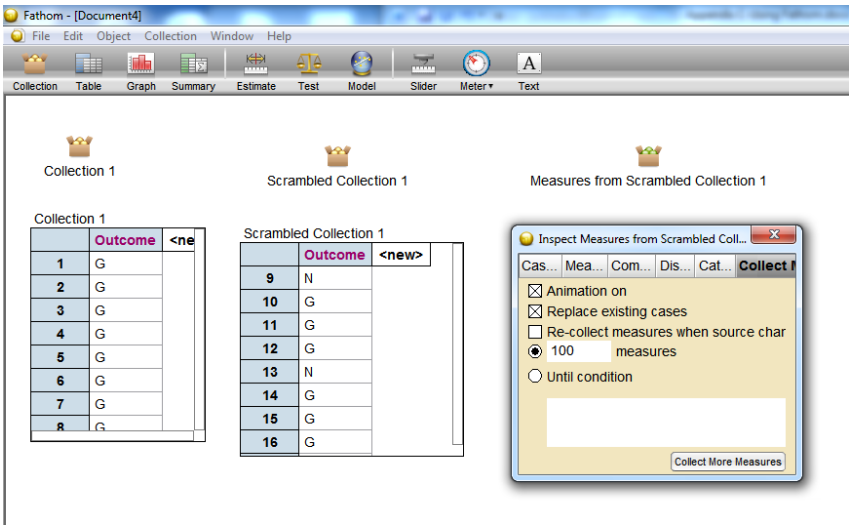
2. Right-click on the box called "Collection 1" and choose Scramble Attribute Values. This will shuffle the 76 outcomes in a random order. To see the results, highlight Scrambled Collection 1 with your mouse and drag a blank table onto the workspace.



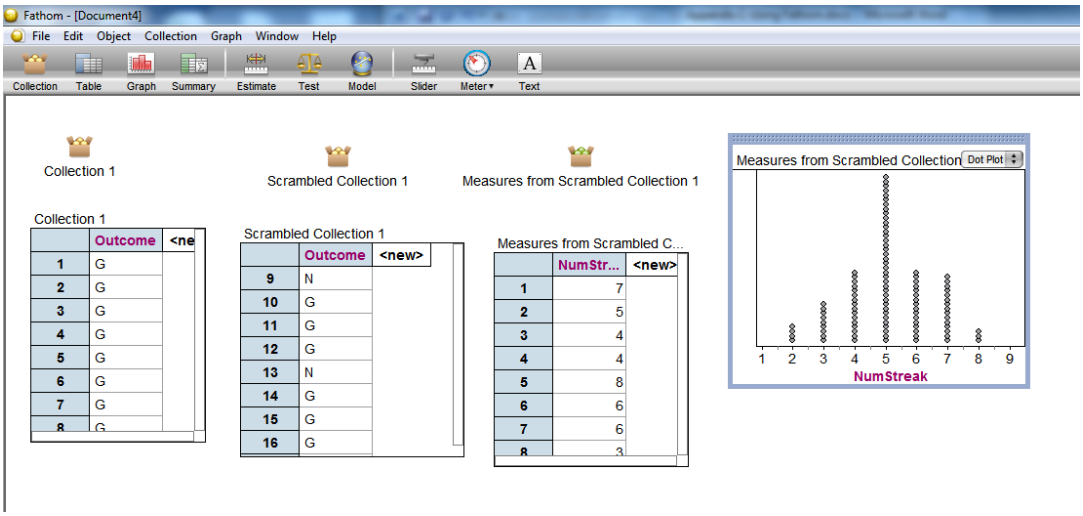
3. Double-click on the box called "Scrambled Collection 1" to open the inspector. In the "Measures" tab, enter the name "NumStreak" and the formula: count(runLength(outcome)=4). Press Enter. This will display the number of streaks of 4 or more in a row. To change the streak length, replace the "4" in the formula to the desired number.



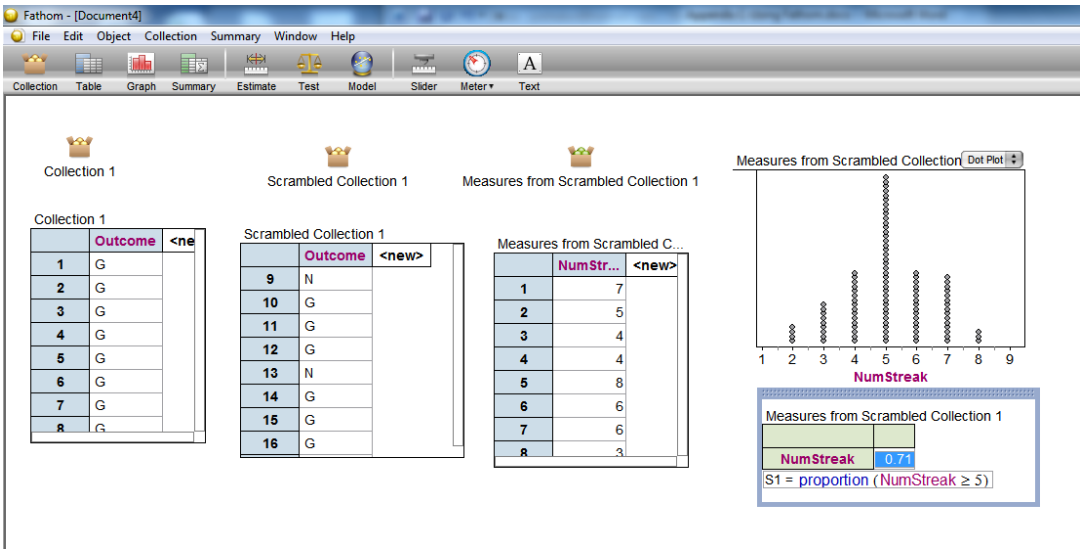
4. Right-click on the box called "Scrambled Collection 1" and choose Collect Measures. In the inspector, select "Replace existing cases" and enter 100 instead of 5 for the number of measures. Then, press Collect More Measures. *Note:* To increase the number of measures, uncheck the "Replace existing cases" box, enter the number of additional measures desired, and press "Collect more measures." The simulation will go much faster if you uncheck the "Animation on" box.



5. To see the results of the 100 trials, highlight the box called "Measures from Scrambled Collection 1" and drag a table icon from the shelf onto the workspace. Then, drag a graph icon from the shelf onto the workspace. Finally, drag the heading "NumStreak" from the table to the horizontal axis of the graph. If you need to rescale the axis, you can either double click on the axis and change the settings or just drag the values on the horizontal axis to the left or right.



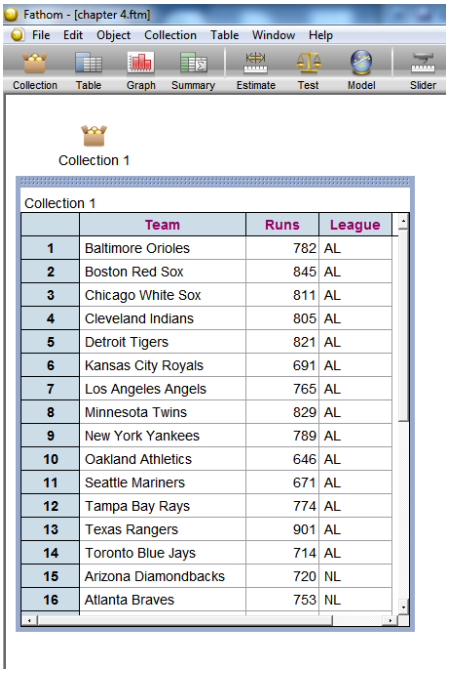
6. To have Fathom calculate the proportion of dots that are 5 or beyond, drag a summary icon from the shelf onto the workspace. Then, drag the Streak heading from the table into the lower-left corner of the summary box, double click on the formula S1=mean() and enter S1=proportion(NumStreak ≥ 5). To find the ≥ symbol, hold the ctrl button and the > symbol will turn into a ≥ symbol in the dialog box.



Chapter 4: Graphs and Summary Statistics for Numerical Data

To make graphs and calculate summary statistics for the 2008 Major League Baseball runs scored data, follow these steps:

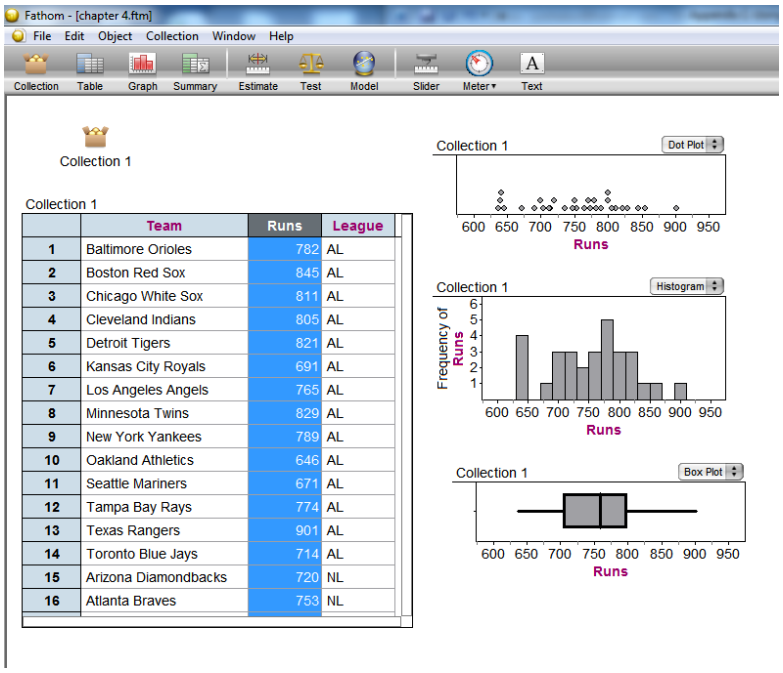
1. To enter the data, drag a blank table from the shelf to the workspace. In the heading of one column type "Runs" and in another column heading type "League". A column for team names is optional. Then, fill in the data from Chapter 4.



Collection 1

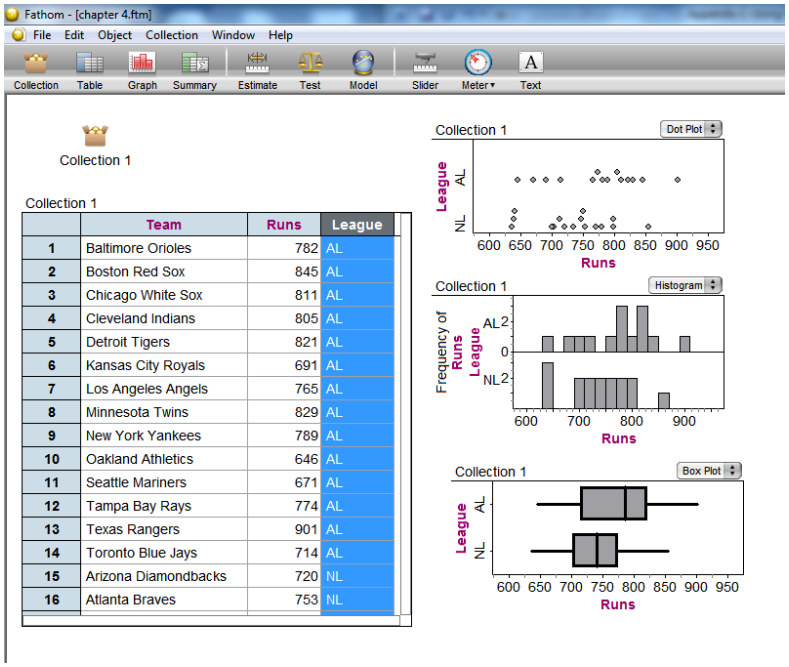
	Team	Runs	League
1	Baltimore Orioles	782	AL
2	Boston Red Sox	845	AL
3	Chicago White Sox	811	AL
4	Cleveland Indians	805	AL
5	Detroit Tigers	821	AL
6	Kansas City Royals	691	AL
7	Los Angeles Angels	765	AL
8	Minnesota Twins	829	AL
9	New York Yankees	789	AL
10	Oakland Athletics	646	AL
11	Seattle Mariners	671	AL
12	Tampa Bay Rays	774	AL
13	Texas Rangers	901	AL
14	Toronto Blue Jays	714	AL
15	Arizona Diamondbacks	720	NL
16	Atlanta Braves	753	NL

2. To graph the data, drag a graph icon from the shelf to the workspace. Then, drag the "Runs" heading to the horizontal axis. The default graph for numerical data is a dotplot, but you can use the menu in the upper-right corner to choose other types of graphs, including boxplots and histograms. To adjust the scale, double click on the axis and change the settings or drag a value on the axis to the left or right. To change the width of the bars on a histogram, move your mouse over the boundary between two bars so that it turns into a double-sided arrow. Then, click and drag to the right or left. To make the histogram into a relative frequency (or relative percentage) histogram, right-click in the histogram, choose "Scale" and then choose the desired scaling for the vertical axis.



	Team	Runs	League
1	Baltimore Orioles	782	AL
2	Boston Red Sox	845	AL
3	Chicago White Sox	811	AL
4	Cleveland Indians	805	AL
5	Detroit Tigers	821	AL
6	Kansas City Royals	691	AL
7	Los Angeles Angels	765	AL
8	Minnesota Twins	829	AL
9	New York Yankees	789	AL
10	Oakland Athletics	646	AL
11	Seattle Mariners	671	AL
12	Tampa Bay Rays	774	AL
13	Texas Rangers	901	AL
14	Toronto Blue Jays	714	AL
15	Arizona Diamondbacks	720	NL
16	Atlanta Braves	753	NL

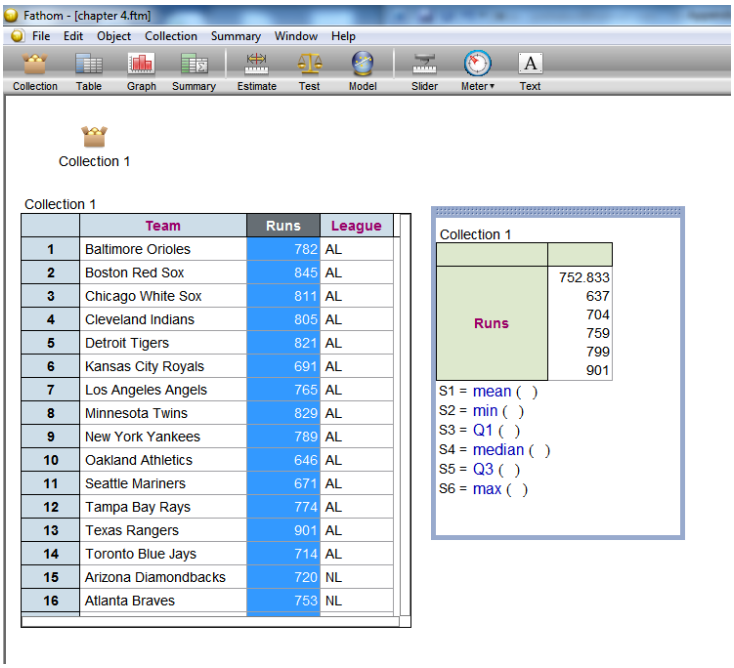
3. To compare the distributions for AL teams and NL teams, drag the "League" heading to the vertical axis of each type of graph. Fathom recognizes that the "League" variable is categorical, so it knows to split the data into groups.



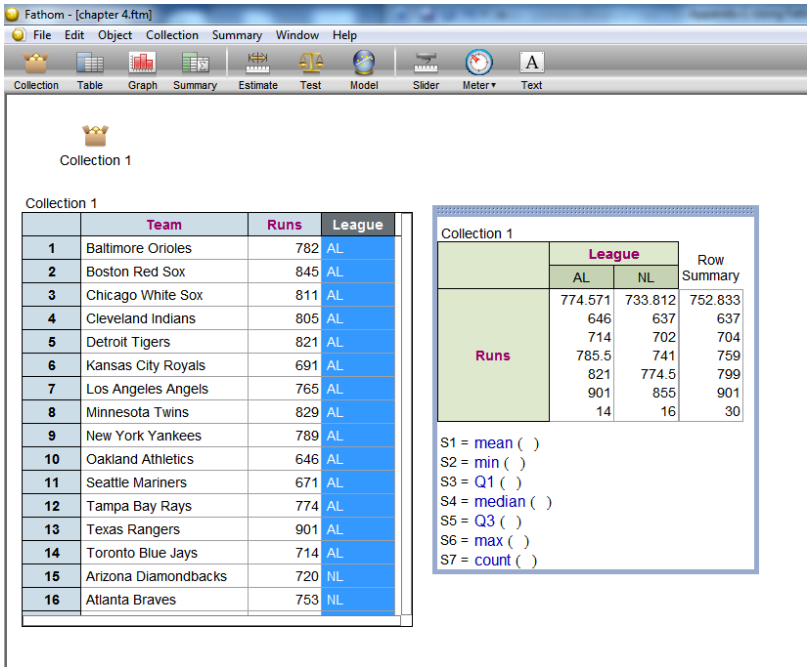
	Team	Runs	League
1	Baltimore Orioles	782	AL
2	Boston Red Sox	845	AL
3	Chicago White Sox	811	AL
4	Cleveland Indians	805	AL
5	Detroit Tigers	821	AL
6	Kansas City Royals	691	AL
7	Los Angeles Angels	765	AL
8	Minnesota Twins	829	AL
9	New York Yankees	789	AL
10	Oakland Athletics	646	AL
11	Seattle Mariners	671	AL
12	Tampa Bay Rays	774	AL
13	Texas Rangers	901	AL
14	Toronto Blue Jays	714	AL
15	Arizona Diamondbacks	720	NL
16	Atlanta Braves	753	NL

4. To calculate summary statistics for the data, drag a summary icon to the workspace. Then, drag the "Runs" heading to the lower left cell of the summary box. The default statistic for a single numerical variable is the mean, but you can right-click and choose options such as adding the five-number summary or many other statistics. You can also double-click on the mean( ) formula to choose other functions from the menu. Choose

Functions, then Statistical, then One Attribute for a list of what functions are available and how to use them.



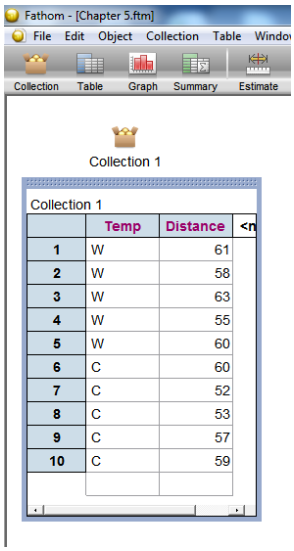
5. To compare the distributions for AL teams and NL teams, drag the "League" heading to the upper-right cell of the summary box. Again, Fathom recognizes that "League" is categorical and knows to split the teams into two groups.



Chapter 5: Simulating the Difference in Means (or Medians)

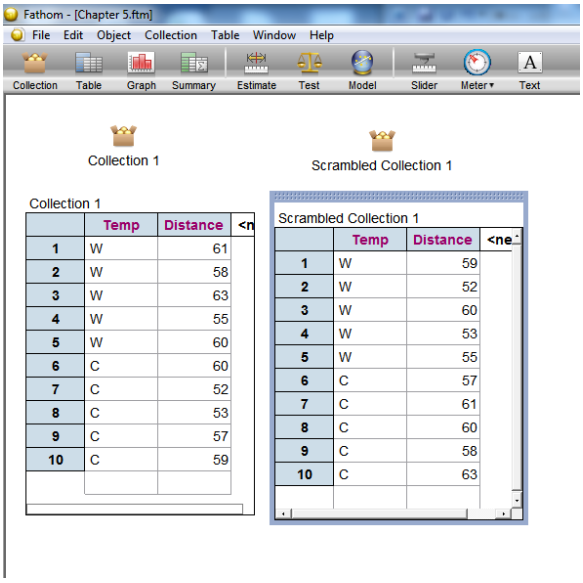
Follow these steps to simulate the difference in the *PERFORMANCE* of the kicker in the warm football experiment from Chapter 5:

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Temp" and the second column "Distance." Fill in the data as shown in the screen shot.



	Temp	Distance	<n
1	W	61	
2	W	58	
3	W	63	
4	W	55	
5	W	60	
6	C	60	
7	C	52	
8	C	53	
9	C	57	
10	C	59	

2. To shuffle the outcomes, right-click the box called "Collection 1" and choose "Scramble Attribute Values." In the inspector, change the attribute from Temp to Distance and press Scramble Attribute Values again. This shuffles up the distances in the Distance column. To see the results, highlight the box called "Scrambled Collection 1" and drag a table icon onto the workspace.



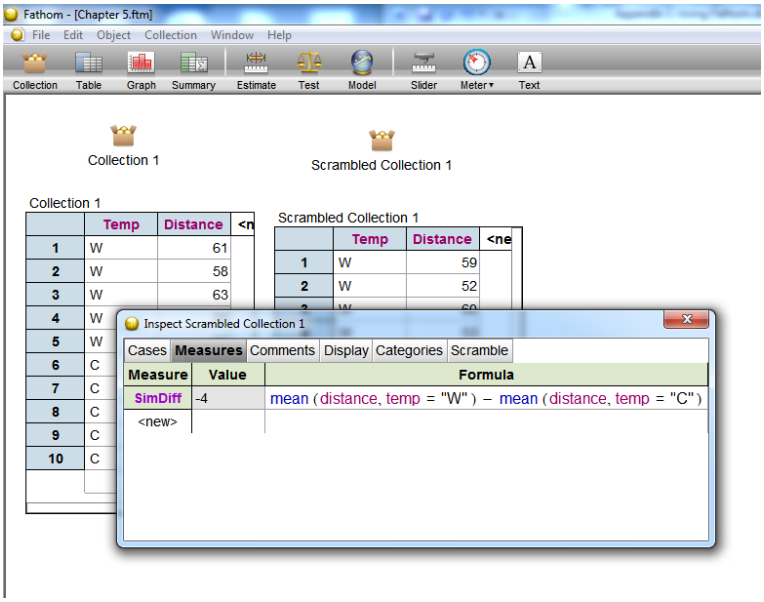
	Temp	Distance	<n
1	W	61	
2	W	58	
3	W	63	
4	W	55	
5	W	60	
6	C	60	
7	C	52	
8	C	53	
9	C	57	
10	C	59	

	Temp	Distance	<ne
1	W	59	
2	W	52	
3	W	60	
4	W	53	
5	W	55	
6	C	57	
7	C	61	
8	C	60	
9	C	58	
10	C	63	

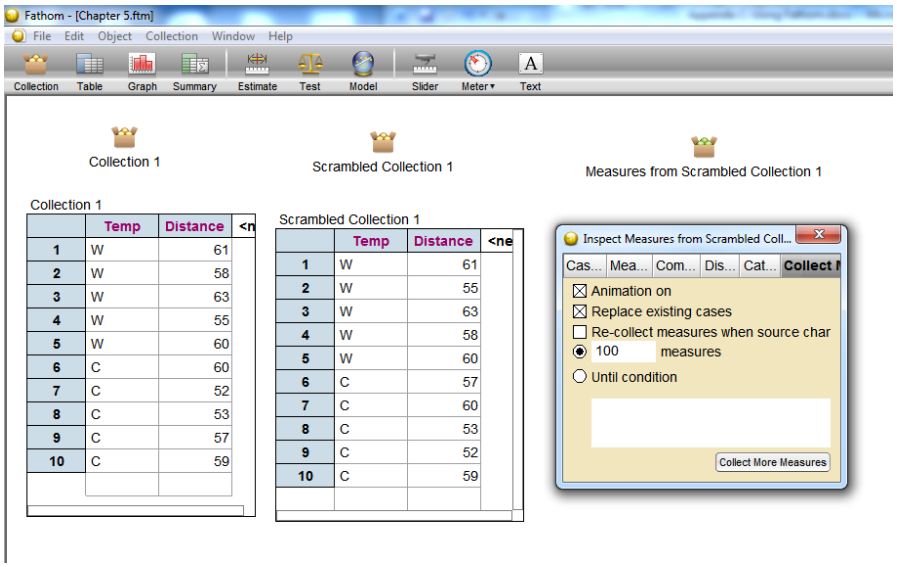
3. To have Fathom repeatedly shuffle the outcomes, calculate the difference in mean distances, and record these values, double-click on the box called “Scrambled Collection 1.” Then, click on the “Measures” tab in the Scrambled Collection 1 Inspector. Under the column heading measure, type “SimDiff.” Now, double click in the empty formula box and type:

   mean(distance,temp=”W”) – mean(distance,temp=”C”)

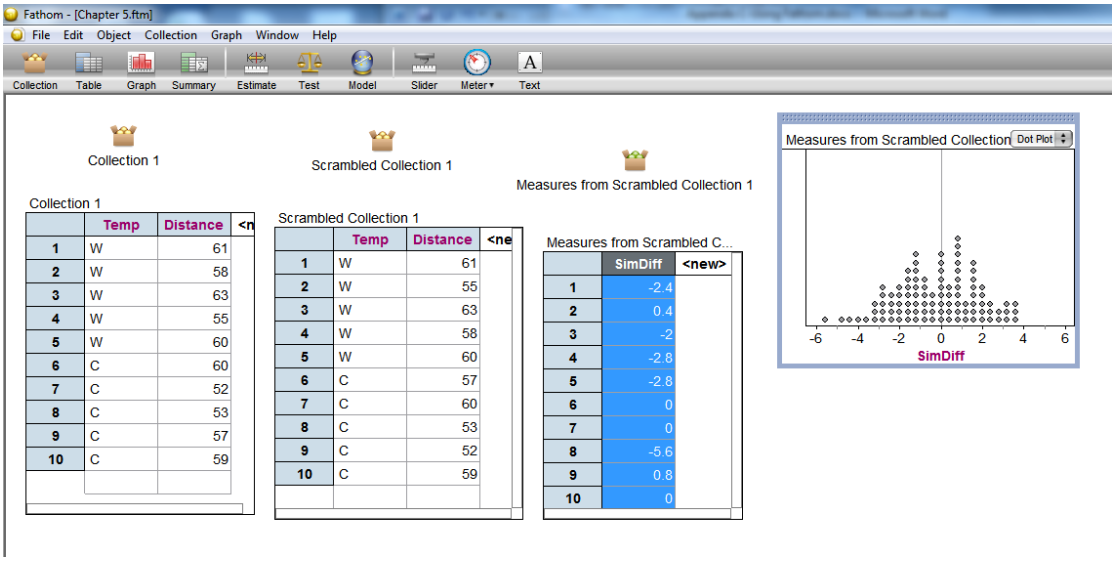
   This formula calculates the difference in mean distance when the temperature variable is warm and the mean distance when the temperature is cold.



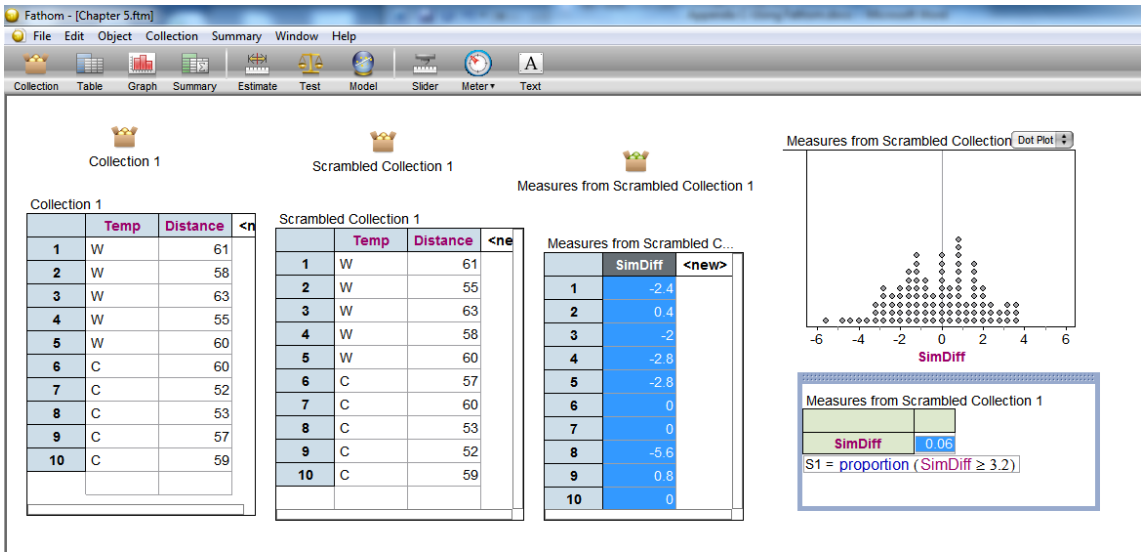
4. Next, right-click the box called “Scrambled Collection 1” and choose Collect Measures. In the inspector, select “Replace existing cases” and enter 100 instead of 5 for the number of measures. Then, press Collect More Measures. *Note:* To increase the number of measures, uncheck the “Replace existing cases” box, enter the number of additional measures desired, and press “Collect more measures.” The simulation will go much faster if you uncheck the “Animation on” box.



5. To see the results of the 100 trials, highlight the box called "Measures from Scrambled Collection 1" and drag a table icon from the shelf onto the workspace. Then, drag a graph icon from the shelf onto the workspace. Finally, drag the heading "SimDiff" from the table to the horizontal axis of the graph. If you need to rescale the axis, you can either double click on the axis and change the settings or just drag the values on the horizontal axis to the left or right.



6. To have Fathom calculate the proportion of dots that are 3.2 or beyond, drag a summary icon from the shelf onto the workspace. Then, drag the SimDiff heading from the table into the lower-left corner of the summary box, double click on the formula S1=mean() and enter S1=proportion(SimDiff $\geq$ 3.2). To find the $\geq$ symbol, hold the ctrl button and the > symbol will turn into a $\geq$ symbol in the dialog box.

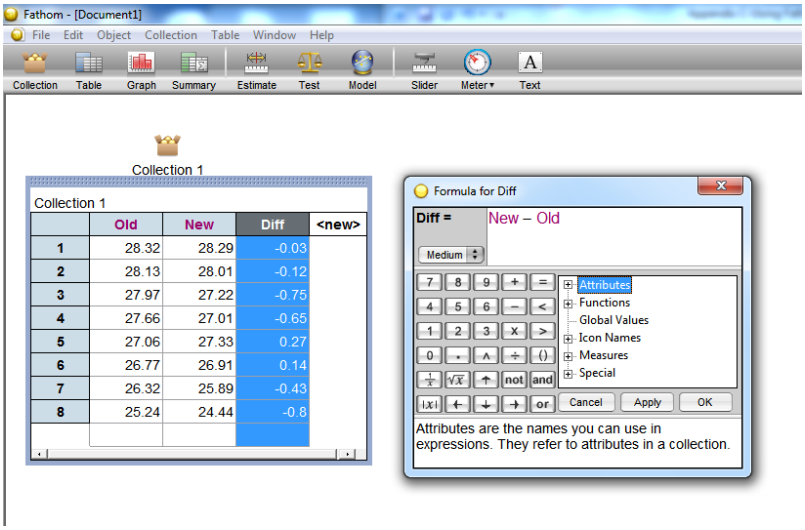


7. To simulate a difference in medians instead of a difference in means, simply replace the word “mean” with the word “median” in step 4.

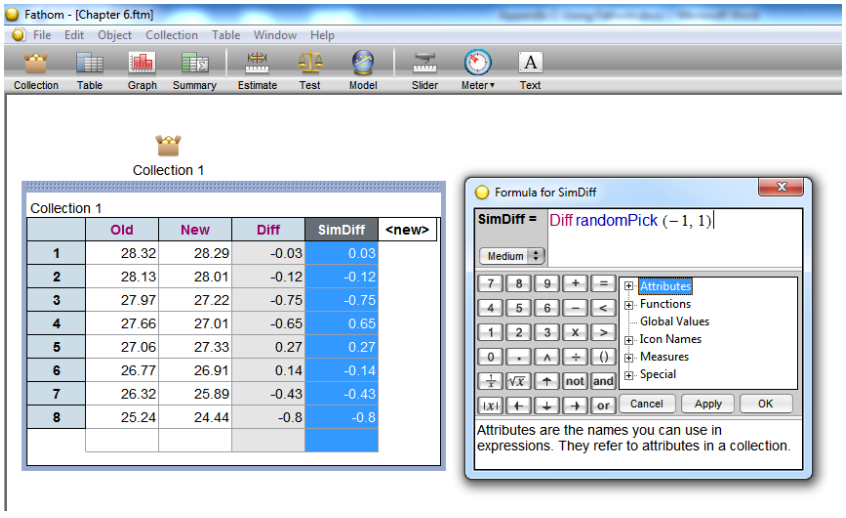
Chapter 6: Simulating Mean Differences

To simulate the mean difference for the swimsuit example in Chapter 6, follow the steps below:

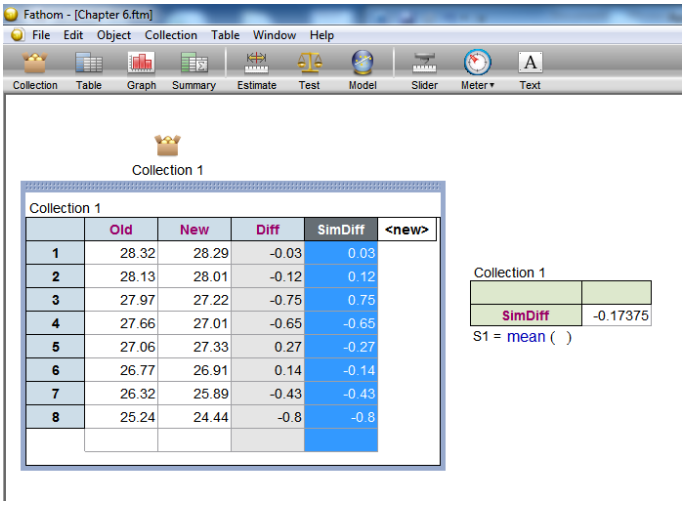
1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Old" and the second column "New." Fill in the data as shown in the screen shot. In the third column, type "Diff" in the heading. Then, right click on the heading, choose Edit Formula and enter the following formula: New – Old .



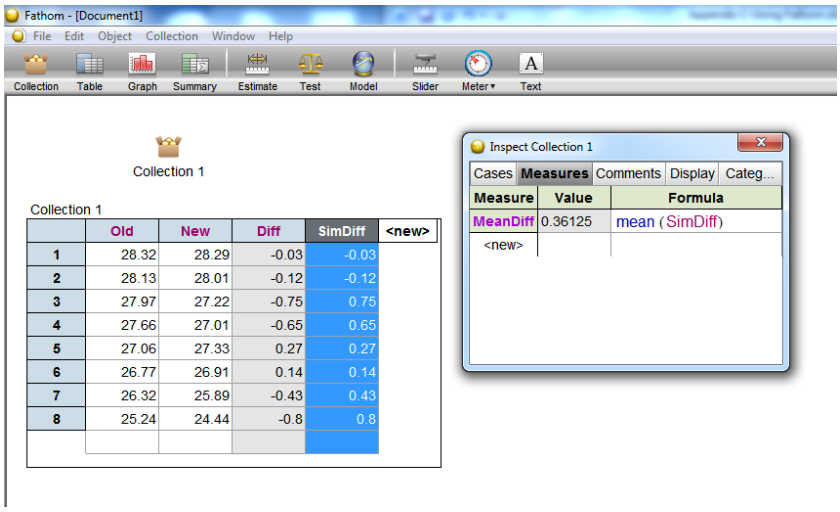
2. To simulate the distribution of the mean difference, type "SimDiff" in the heading of the fourth column. Then, right click the heading, choose Edit Formula and enter the formula: Diff · randomPick(-1,1) . This will make each simulated difference either positive or negative, with equal likelihood.



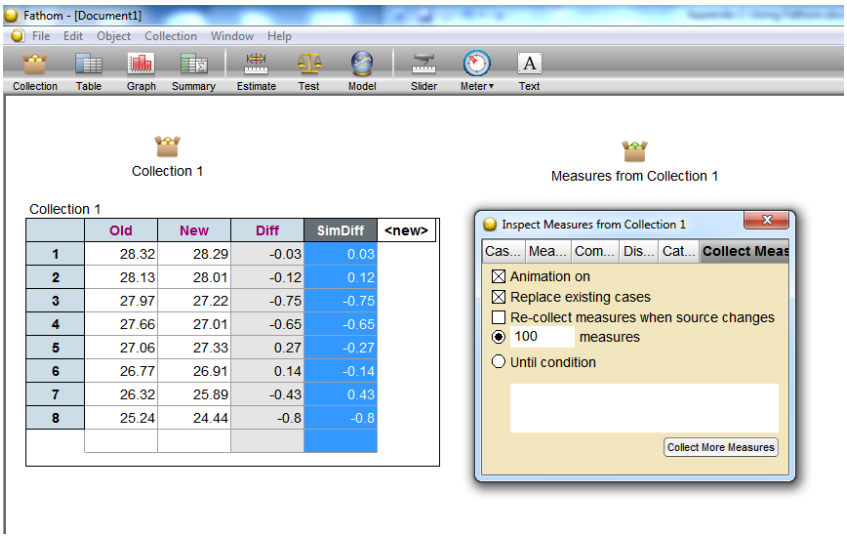
3. To see the simulated mean difference, drag a summary icon from the shelf to the workspace. Then, drag the "SimDiff" heading to the lower-left cell of the summary box. Press ctrl-Y to re-randomize.



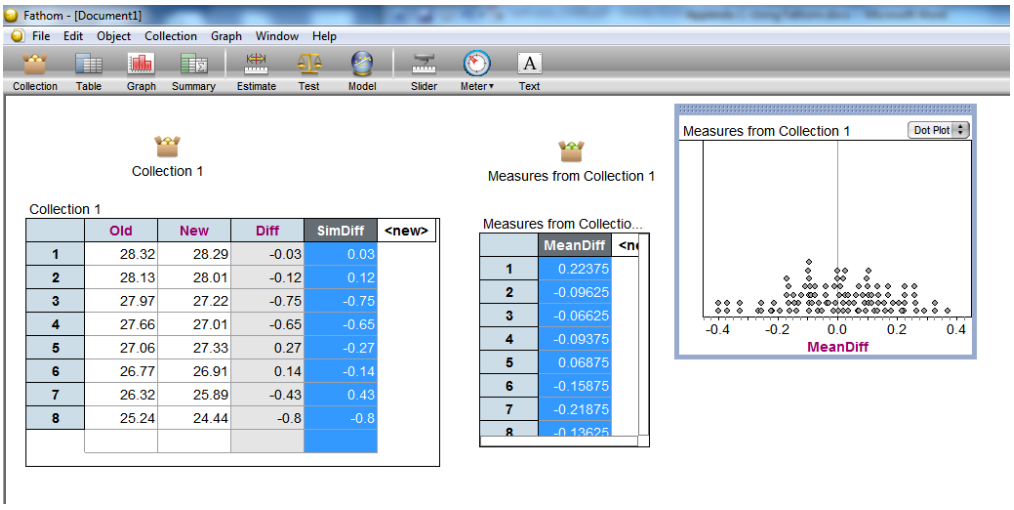
4. To have Fathom repeatedly re-randomize, calculate the simulated mean difference, and collect these values, double-click on Collection 1 and choose the Measures tab. In the Measures column, type SimMean. Double click in the formula cell and enter the function: mean(SimDiff) . Press OK.



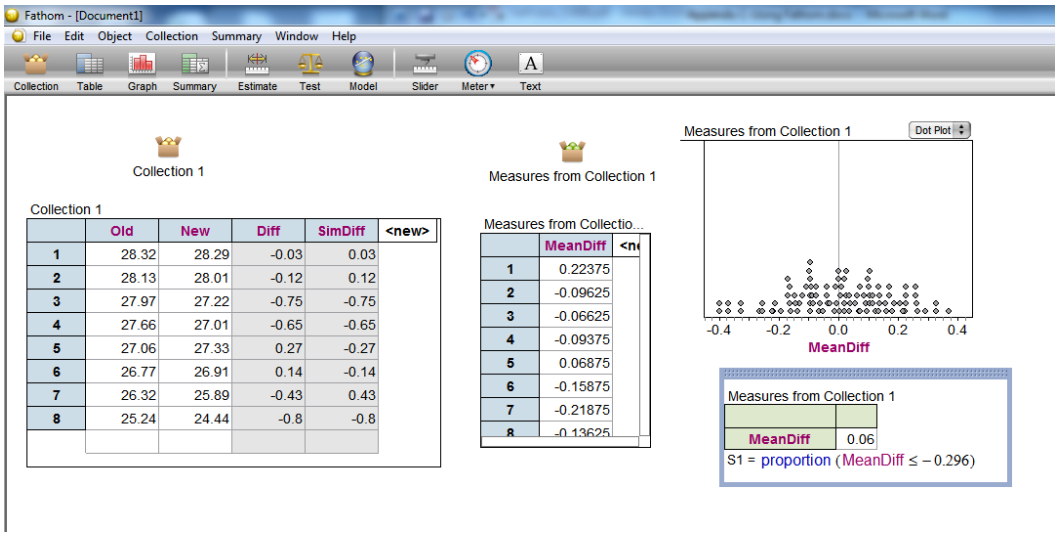
5. Next, right-click the box called “Scrambled Collection 1” and choose Collect Measures. In the inspector, select “Replace existing cases” and enter 100 instead of 5 for the number of measures. Then, press Collect More Measures. *Note:* To increase the number of measures, uncheck the “Replace existing cases” box, enter the number of additional measures desired, and press “Collect more measures.” The simulation will go much faster if you uncheck the “Animation on” box.



6. To see the results of the 100 trials, highlight the box called "Measures from Scrambled Collection 1" and drag a table icon from the shelf onto the workspace. Then, drag a graph icon from the shelf onto the workspace. Finally, drag the heading "MeanDiff" from the table to the horizontal axis of the graph. If you need to rescale the axis, you can either double click on the axis and change the settings or just drag the values on the horizontal axis to the left or right.



7. To have Fathom calculate the proportion of dots that are –0.296 or below, drag a summary icon from the shelf onto the workspace. Then, drag the MeanDiff heading from the table into the lower-left corner of the summary box, double click on the formula S1=mean() and enter S1=proportion(MeanDiff $\leq$ –0.296). To find the $\leq$ symbol, hold the ctrl button and the < symbol will turn into a $\leq$ symbol in the dialog box.



Collection 1

	Old	New	Diff	SimDiff	<new>
1	28.32	28.29	-0.03	0.03	
2	28.13	28.01	-0.12	0.12	
3	27.97	27.22	-0.75	-0.75	
4	27.66	27.01	-0.65	-0.65	
5	27.06	27.33	0.27	-0.27	
6	26.77	26.91	0.14	-0.14	
7	26.32	25.89	-0.43	0.43	
8	25.24	24.44	-0.8	-0.8	

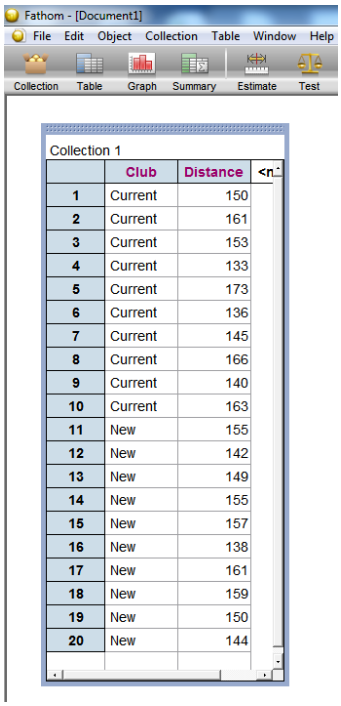
Measures from Collectio...

	MeanDiff	<n
1	0.22375	
2	-0.09625	
3	-0.06625	
4	-0.09375	
5	0.06875	
6	-0.15875	
7	-0.21875	
8	-0.13625	

Chapter 7: Calculating the Standard Deviation

To calculate the standard deviation for the two drivers in the golf example from Chapter 7, follow these steps:

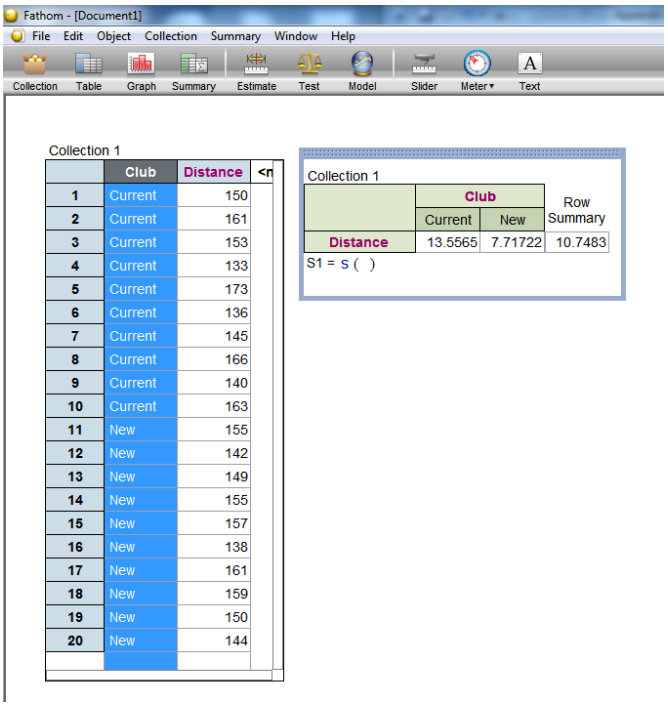
1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Club" and the second column "Distance." Fill in the data as shown in the screen shot.



Collection 1

	Club	Distance
1	Current	150
2	Current	161
3	Current	153
4	Current	133
5	Current	173
6	Current	136
7	Current	145
8	Current	166
9	Current	140
10	Current	163
11	New	155
12	New	142
13	New	149
14	New	155
15	New	157
16	New	138
17	New	161
18	New	159
19	New	150
20	New	144

2. To calculate the standard deviation for each, drag a summary icon to the workspace. Then, drag the "Distance" heading to the lower left cell. The default statistic for a single numerical variable is the mean. To change this to the standard deviation, double click on the mean formula and change the formula to: `s( )`. To compare the standard deviations, drag the "Club" attribute to the upper right cell.


Fathom - [Document1]
File Edit Object Collection Summary Window Help
Collection Table Graph Summary Estimate Test Model Slider Meter Text
Collection 1
Club Distance
1 Current 150
2 Current 161
3 Current 153
4 Current 133
5 Current 173
6 Current 136
7 Current 145
8 Current 166
9 Current 140
10 Current 163
11 New 155
12 New 142
13 New 149
14 New 155
15 New 157
16 New 138
17 New 161
18 New 159
19 New 150
20 New 144
Collection 1
Club
Current New
Row Summary
Distance 13.5565 7.71722 10.7483
S1 = S ( )

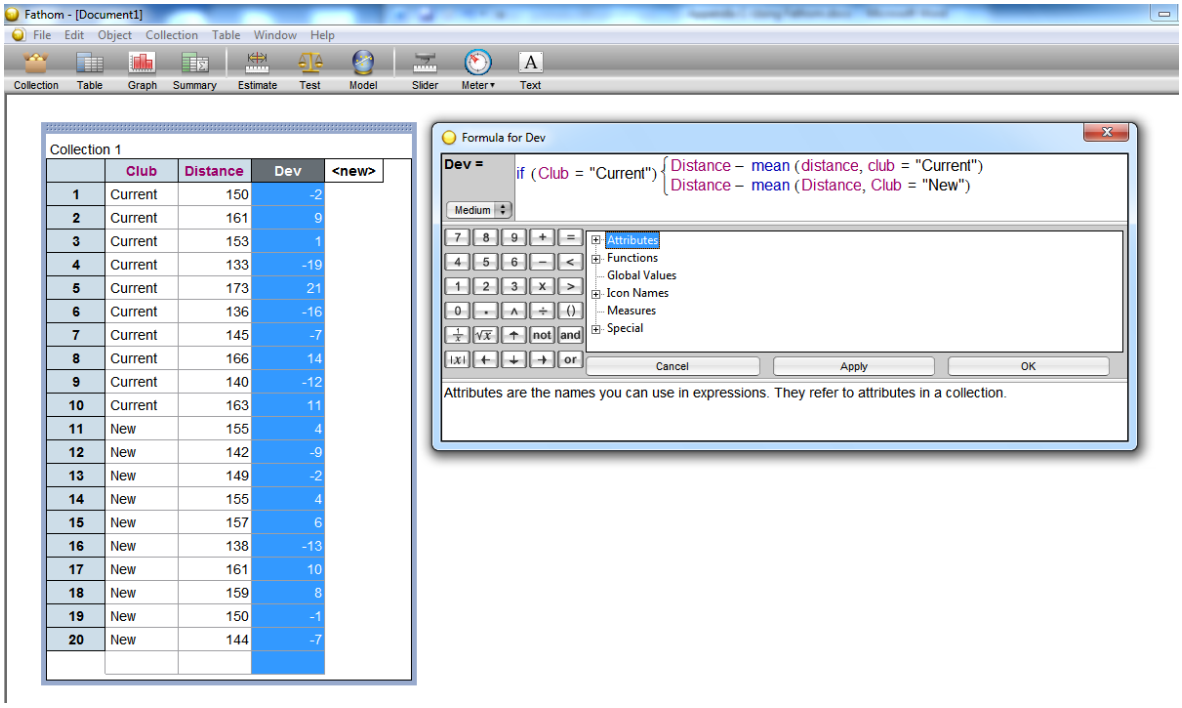
Chapter 7: Simulating a Difference in Standard Deviations

To simulate the difference in standard deviations for the golf club example in Chapter 7, follow the steps below:

1. Enter the golf club data as described in step #1 above.
2. To calculate the deviations, at the top of the third column enter a new attribute called "Dev". Right click on the heading and choose Edit Formula. Enter the following:

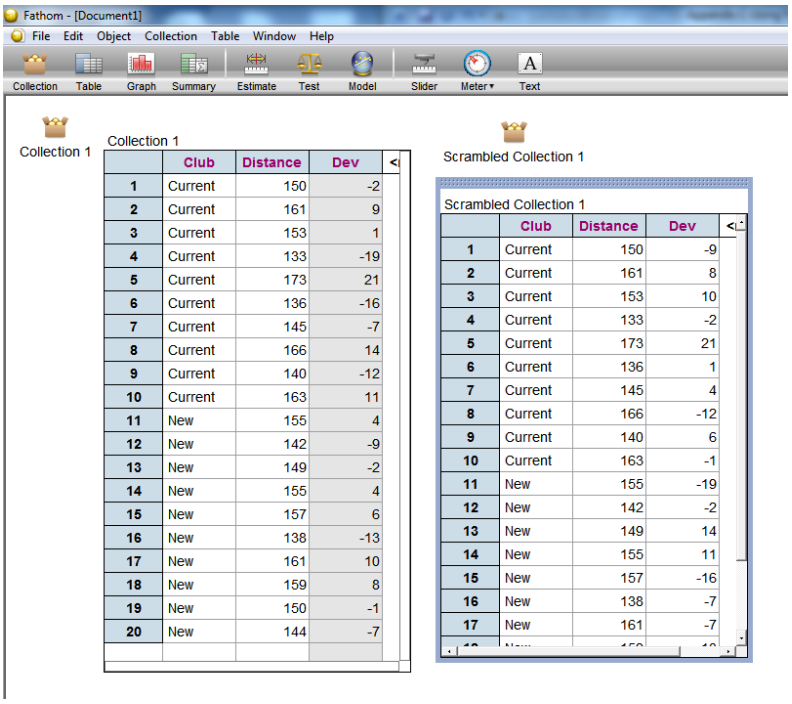
$$\text{if(Club="Current")}\begin{cases}\text{distance} - \text{mean(distance,club="Current")} \\ \text{distance} - \text{mean(distance,club="New")}\end{cases}$$

This will calculate the deviations from the appropriate mean. More specifically, if the club is the current club, it will find the mean distance for all the current clubs and subtract the mean from the distance of the shot. If the club is not the current club, then it will find the mean distance for all the new clubs and subtract the mean from the distance of the shot.



	Club	Distance	Dev	<new>
1	Current	150	-2	
2	Current	161	9	
3	Current	153	1	
4	Current	133	-19	
5	Current	173	21	
6	Current	136	-16	
7	Current	145	-7	
8	Current	166	14	
9	Current	140	-12	
10	Current	163	11	
11	New	155	4	
12	New	142	-9	
13	New	149	-2	
14	New	155	4	
15	New	157	6	
16	New	138	-13	
17	New	161	10	
18	New	159	8	
19	New	150	-1	
20	New	144	-7	

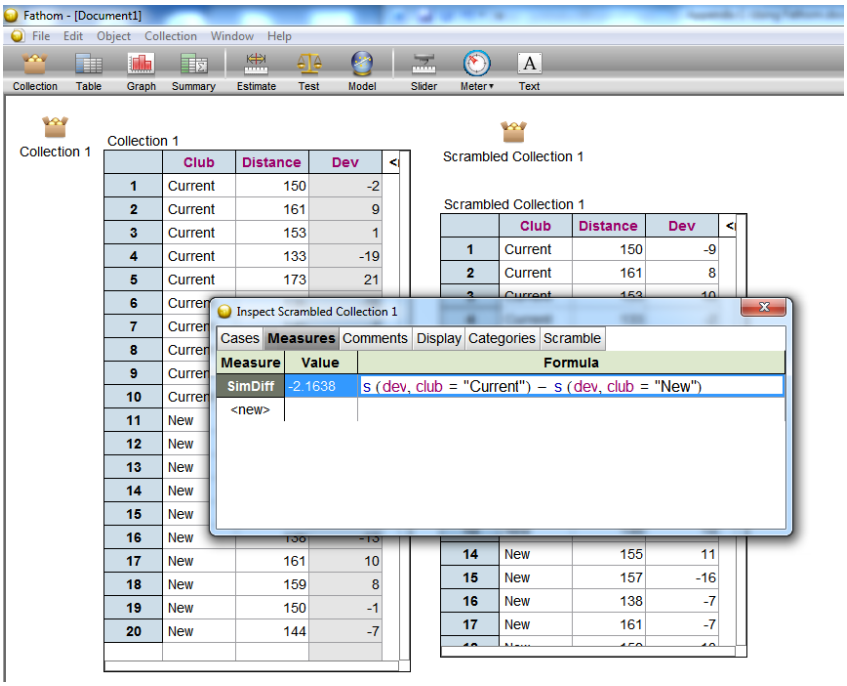
3. To shuffle the deviations, right-click the box called "Collection 1" and choose "Scramble Attribute Values." In the inspector, change the attribute from Club to Dev and press Scramble Attribute Values again. This shuffles up the deviations in the Dev column. To see the results, highlight the box called "Scrambled Collection 1" and drag a table icon onto the workspace.



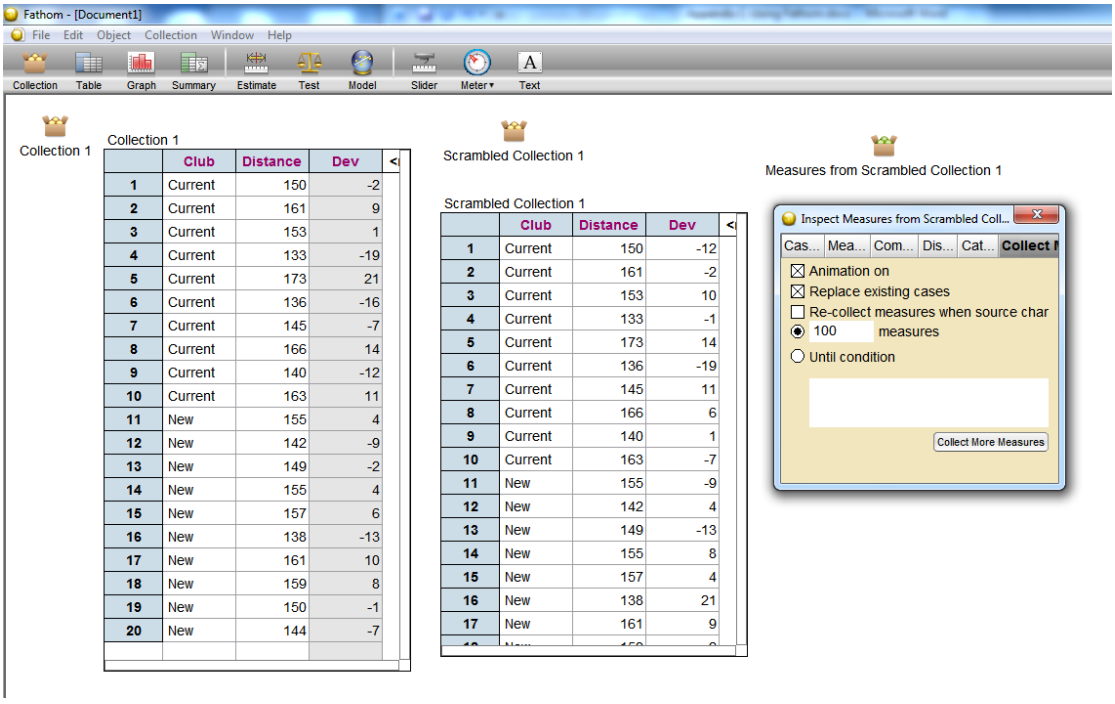
	Club	Distance	Dev
1	Current	150	-2
2	Current	161	9
3	Current	153	1
4	Current	133	-19
5	Current	173	21
6	Current	136	-16
7	Current	145	-7
8	Current	166	14
9	Current	140	-12
10	Current	163	11
11	New	155	4
12	New	142	-9
13	New	149	-2
14	New	155	4
15	New	157	6
16	New	138	-13
17	New	161	10
18	New	159	8
19	New	150	-1
20	New	144	-7

	Club	Distance	Dev
1	Current	150	-9
2	Current	161	8
3	Current	153	10
4	Current	133	-2
5	Current	173	21
6	Current	136	1
7	Current	145	4
8	Current	166	-12
9	Current	140	6
10	Current	163	-1
11	New	155	-19
12	New	142	-2
13	New	149	14
14	New	155	11
15	New	157	-16
16	New	138	-7
17	New	161	-7

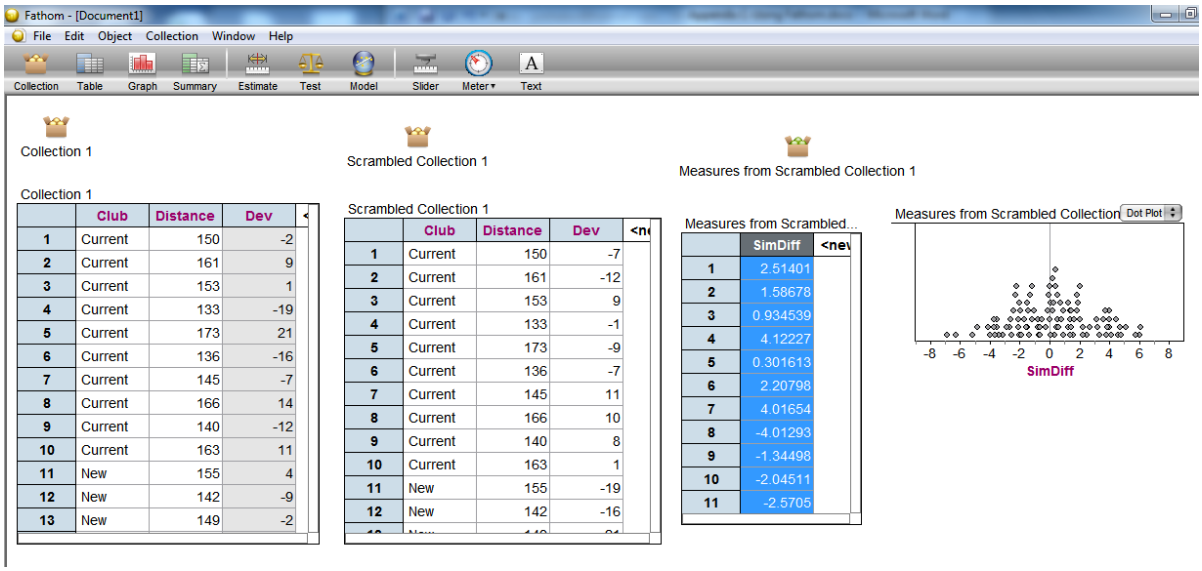
4. To have Fathom repeatedly shuffle the deviations, calculate the difference in standard deviations, and record these values, double-click on the box called “Scrambled Collection 1.” Then, click on the “Measures” tab in the Scrambled Collection 1 Inspector. Under the column heading measure, type “SimDiff.” Now, double click on the cell under the heading formula and type: s(dev,club=”Current”) – s(dev,club=”New”). This will calculate the standard deviation of the deviations when the club is new and subtract that value from the standard deviation of the deviations when the club is current.



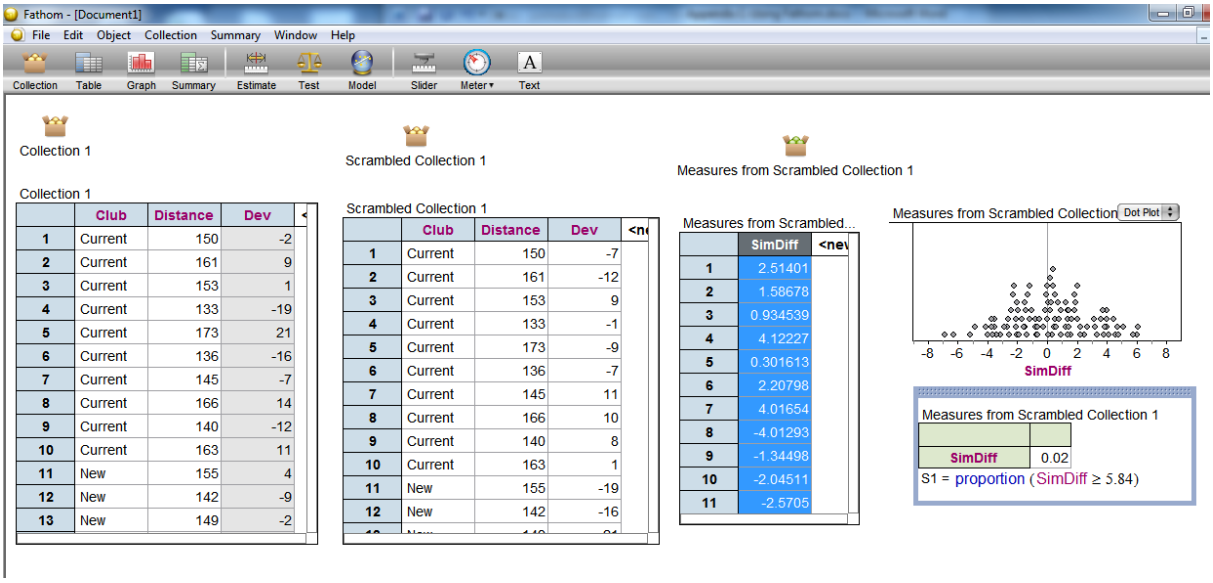
5. Next, right-click the box called “Scrambled Collection 1” and choose Collect Measures. In the inspector, select “Replace existing cases” and enter 100 instead of 5 for the number of measures. Then, press Collect More Measures. *Note:* To increase the number of measures, uncheck the “Replace existing cases” box, enter the number of additional measures desired, and press “Collect more measures.” The simulation will go much faster if you uncheck the “Animation on” box.



6. To see the results of the 100 trials, highlight the box called “Measures from Scrambled Collection 1” and drag a table icon from the shelf onto the workspace. Then, drag a graph icon from the shelf onto the workspace. Finally, drag the heading “SimDiff” from the table to the horizontal axis of the graph. If you need to rescale the axis, you can either double click on the axis and change the settings or just drag the values on the horizontal axis to the left or right.



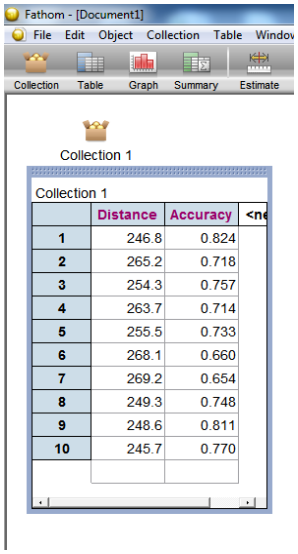
7. To have Fathom calculate the proportion of dots that are 5.84 or beyond, drag a summary icon from the shelf onto the workspace. Then, drag the SimDiff heading from the table into the lower-left corner of the summary box, double click on the formula S1=mean() and enter S1=proportion(SimDiff ≥ 5.84). To find the ≥ symbol, hold the ctrl button and the > symbol will turn into a ≥ symbol in the dialog box.



Chapter 10: Making Scatterplots and Calculating the Correlation

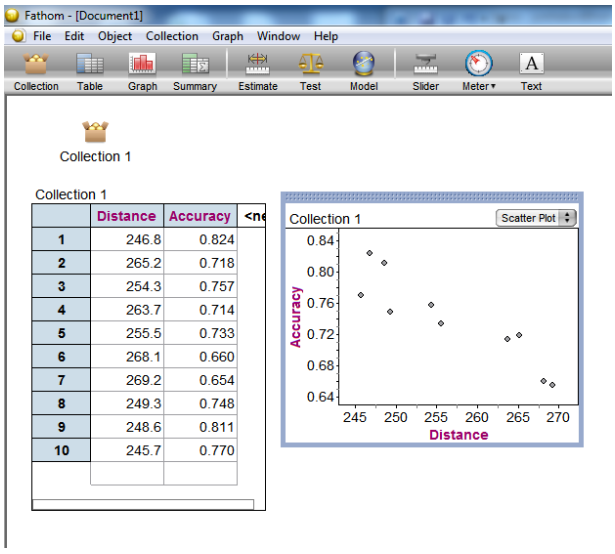
To create a scatterplot and calculate the correlation for the average driving distance and driving accuracy of the top-10 LPGA golfers in 2009, follow the steps below:

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Distance" and the second column "Accuracy." Fill in the data as shown in the screen shot.



	Distance	Accuracy	<ne
1	246.8	0.824	
2	265.2	0.718	
3	254.3	0.757	
4	263.7	0.714	
5	255.5	0.733	
6	268.1	0.660	
7	269.2	0.654	
8	249.3	0.748	
9	248.6	0.811	
10	245.7	0.770	

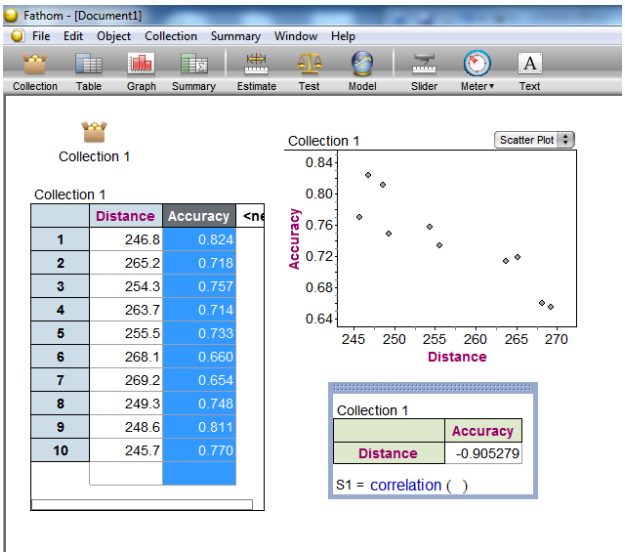
2. To create a scatterplot, drag a graph icon to the workspace. Then, drag the "Distance" heading to the horizontal axis and the "Accuracy" heading to the vertical axis as shown in the screen shot below. As with other graphs, you can double-click on axis to change the scale or grab a number on the scale and drag it to the left or right as needed.



	Distance	Accuracy	<ne
1	246.8	0.824	
2	265.2	0.718	
3	254.3	0.757	
4	263.7	0.714	
5	255.5	0.733	
6	268.1	0.660	
7	269.2	0.654	
8	249.3	0.748	
9	248.6	0.811	
10	245.7	0.770	

3. To calculate the correlation, drag a summary icon to the workspace. Then, drag the "Distance" heading to the lower-left cell of the summary box and the "Accuracy"

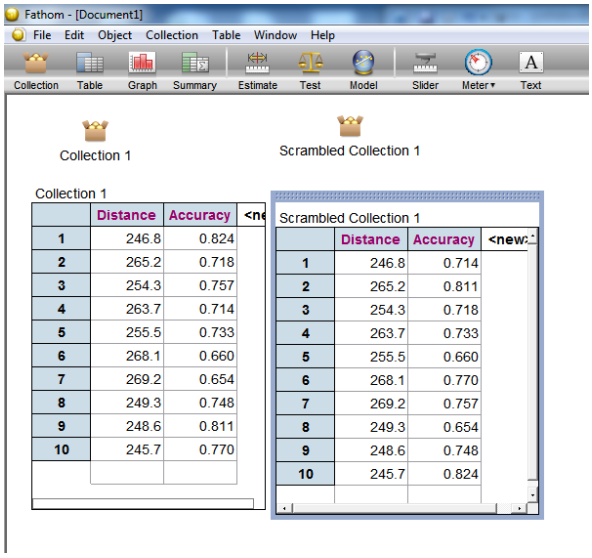
heading to the upper-right cell. The default statistic for the relationship between two numerical variable is the correlation.



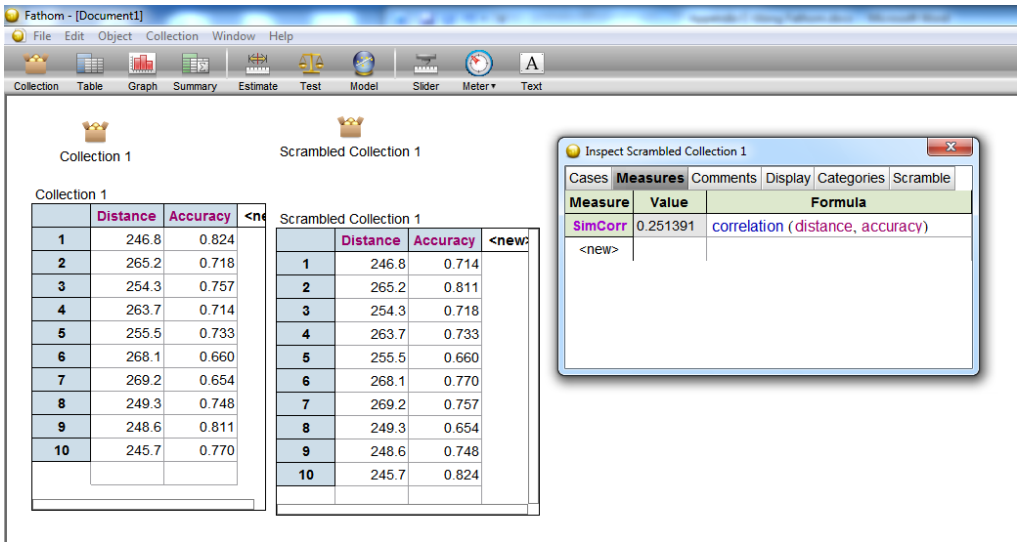
Chapter 10: Simulating the Correlation

Follow these steps to simulate the correlation between the average driving distance and driving accuracy for the top-10 LPGA golfers in 2009:

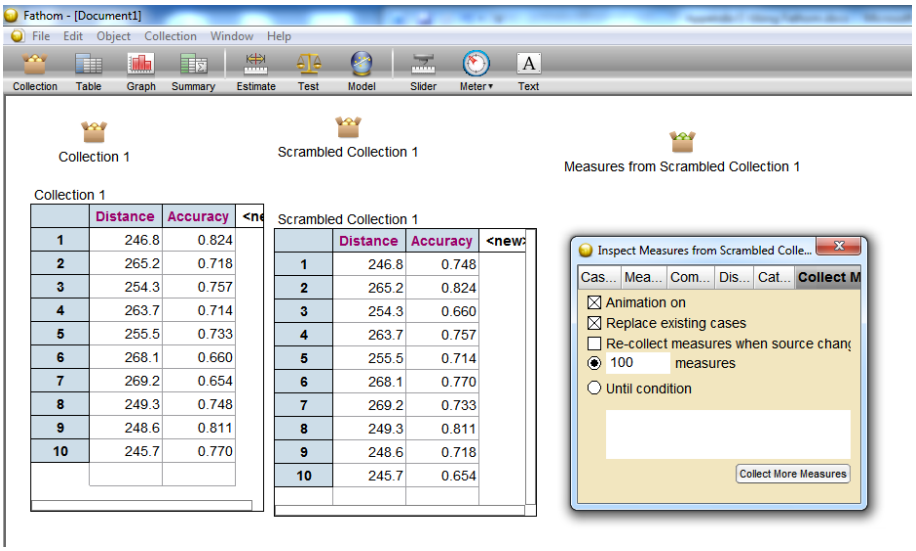
1. Enter the data as described in step 1 above.
2. To shuffle the accuracy values, right-click the box called "Collection 1" and choose "Scramble Attribute Values." In the inspector, change the attribute from Distance to Accuracy and press Scramble Attribute Values again. This shuffles the proportions in the Accuracy column. To see the results, highlight the box called "Scrambled Collection 1" and drag a table icon onto the workspace.



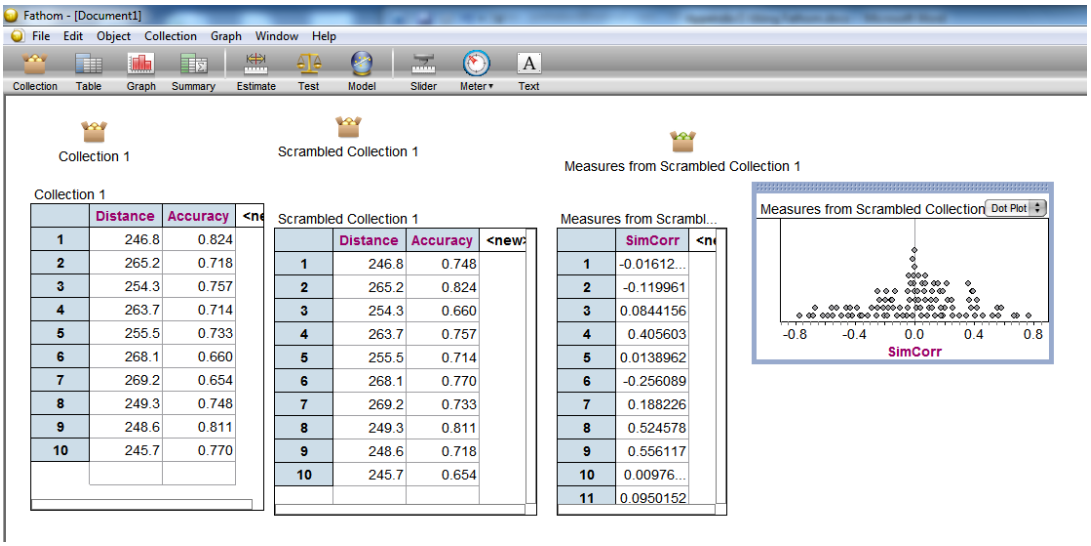
3. To have Fathom repeatedly shuffle the accuracy values, calculate the correlation, and record these values, double-click on the box called "Scrambled Collection 1." Then, click on the "Measures" tab in the Scrambled Collection 1 Inspector. Under the column heading measure, type "SimCorr." Now, double click on the cell under the heading formula and type: Correlation(distance,accuracy)



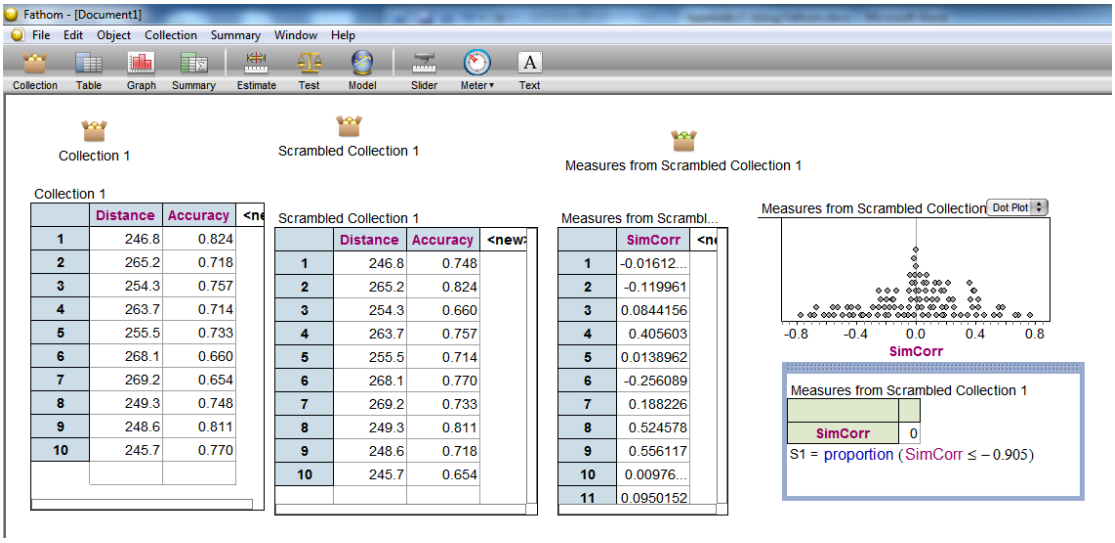
4. Next, right-click the box called “Scrambled Collection 1” and choose Collect Measures. In the inspector, select “Replace existing cases” and enter 100 instead of 5 for the number of measures. Then, press Collect More Measures. *Note:* To increase the number of measures, uncheck the “Replace existing cases” box, enter the number of additional measures desired, and press “Collect more measures.” The simulation will go much faster if you uncheck the “Animation on” box.



5. To see the results of the 100 trials, highlight the box called “Measures from Scrambled Collection 1” and drag a table icon from the shelf onto the workspace. Then, drag a graph icon from the shelf onto the workspace. Finally, drag the heading “SimCorr” from the table to the horizontal axis of the graph. If you need to rescale the axis, you can either double click on the axis and change the settings or just drag the values on the horizontal axis to the left or right.



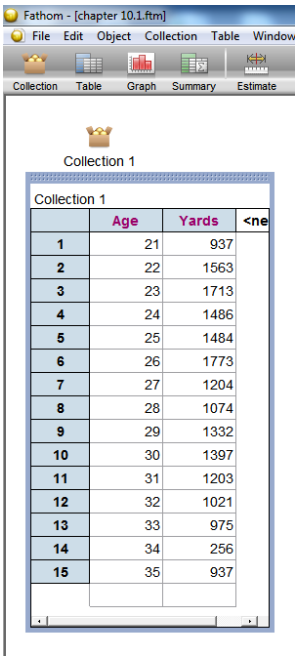
6. To have Fathom calculate the proportion of dots that are –0.905 or lower, drag a summary icon from the shelf onto the workspace. Then, drag the SimCorr heading from the table into the lower-left corner of the summary box, double click on the formula S1=mean() and enter S1=proportion(SimCorr ≤–0.905). To find the ≤ symbol, hold the ctrl button and the < symbol will turn into a ≤ symbol in the dialog box.



Chapter 10: Making Timeplots

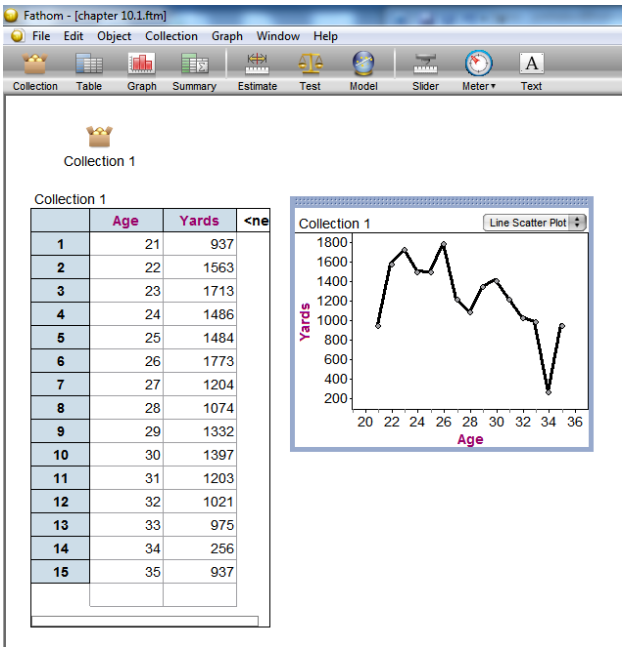
To create a timeplot showing Emmitt Smith's yearly rushing *PERFORMANCES* during his career, follow the steps below:

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Age" and the second column "Yards." Fill in the data as shown in the screen shot.

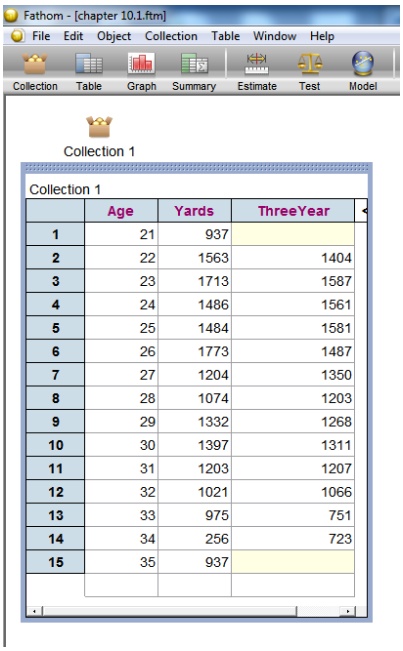


	Age	Yards	<ne
1	21	937	
2	22	1563	
3	23	1713	
4	24	1486	
5	25	1484	
6	26	1773	
7	27	1204	
8	28	1074	
9	29	1332	
10	30	1397	
11	31	1203	
12	32	1021	
13	33	975	
14	34	256	
15	35	937	

2. Drag a graph icon from the shelf onto the workspace. Then, drag the "Age" heading to the horizontal axis and the "Yards" heading to the vertical axis. Finally, in the drop-down menu in the upper-right corner of the graph, choose "Line Scatter Plot." If you want to adjust the scales, double-click on the axis or grab one of the numbers on the scale and drag it to the left or right.

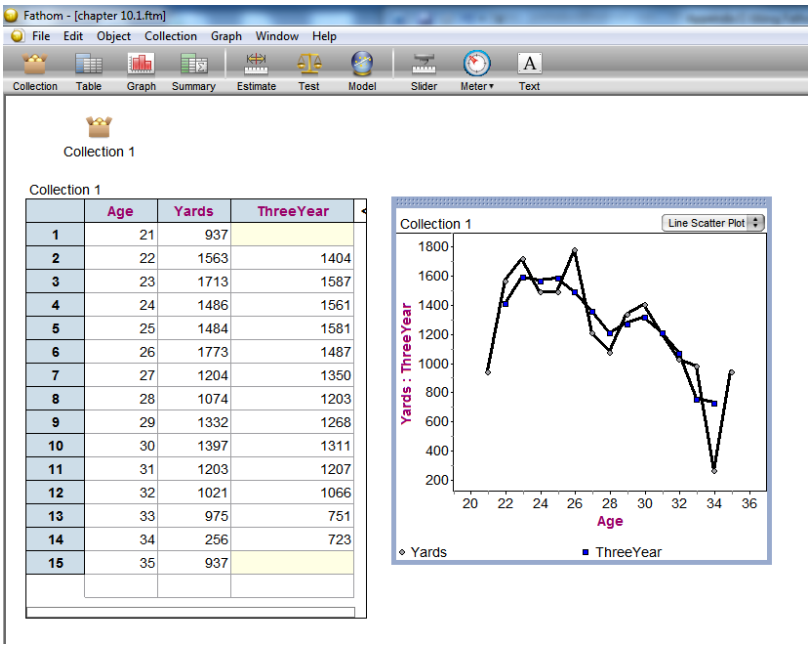


3. If you want to plot two different timeplots on the same axis, such as a 3-year moving average, enter the values as shown in the screen shot below. Unfortunately, there is no way to have Fathom calculate the three-year averages for you.



4. To make a comparative timeplot with the yearly totals and the three-year average, drag a graph icon to the workspace. Then, drag the "Age" heading to the horizontal axis, drag the "Yards" heading to the vertical axis, and choose Line Scatter Plot from the drop down menu. To add the graph of the three-year average, drag the "ThreeYear" heading to the vertical axis, hold it there *without releasing the mouse* until a plus sign appears at the top

of the axis, and drop the heading on the plus. This will add the three-year variable to the yards variable.

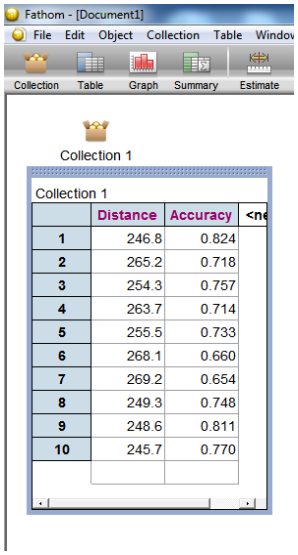


	Age	Yards	ThreeYear
1	21	937	
2	22	1563	1404
3	23	1713	1587
4	24	1486	1561
5	25	1484	1581
6	26	1773	1487
7	27	1204	1350
8	28	1074	1203
9	29	1332	1268
10	30	1397	1311
11	31	1203	1207
12	32	1021	1066
13	33	975	751
14	34	256	723
15	35	937	

Chapter 11: Calculating the Equation of a Least-Squares Regression Line and Displaying the Line on a Scatterplot

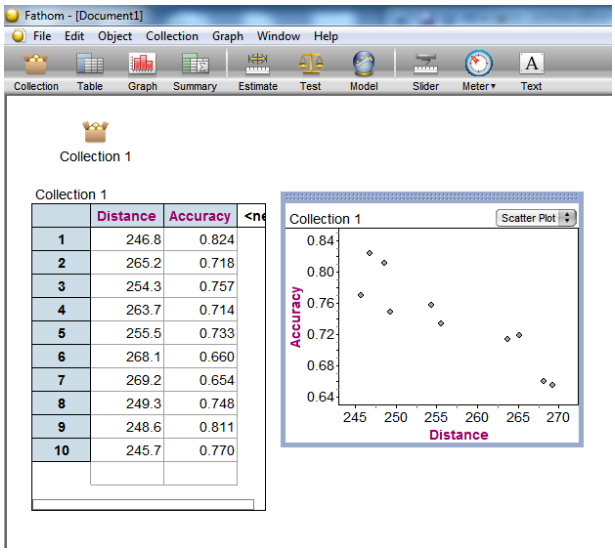
To calculate the equation of the least-squares regression line and display it on a scatterplot of average driving distance and driving accuracy for the top-10 LPGA golfers in 2009, follow the steps below:

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Distance" and the second column "Accuracy." Fill in the data as shown in the screen shot.

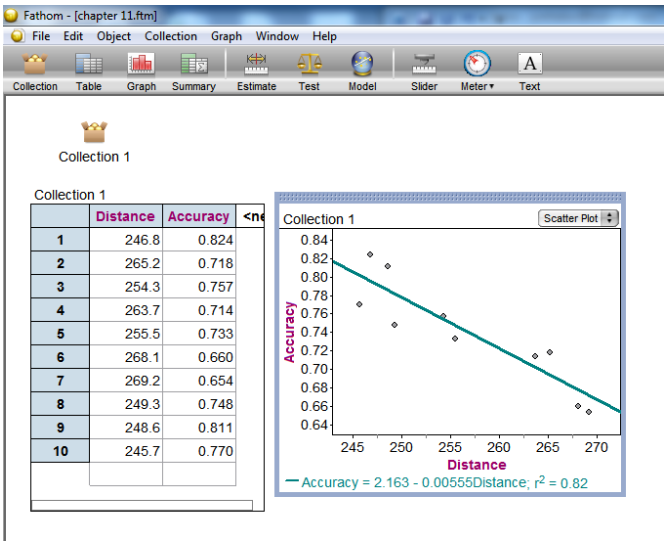


	Distance	Accuracy	<ne
1	246.8	0.824	
2	265.2	0.718	
3	254.3	0.757	
4	263.7	0.714	
5	255.5	0.733	
6	268.1	0.660	
7	269.2	0.654	
8	249.3	0.748	
9	248.6	0.811	
10	245.7	0.770	

2. To create a scatterplot, drag a graph icon to the workspace. Then, drag the "Distance" heading to the horizontal axis and the "Accuracy" heading to the vertical axis as shown in the screen shot below. As with other graphs, you can double-click on axis to change the scale or grab a number on the scale and drag it to the left or right as needed.

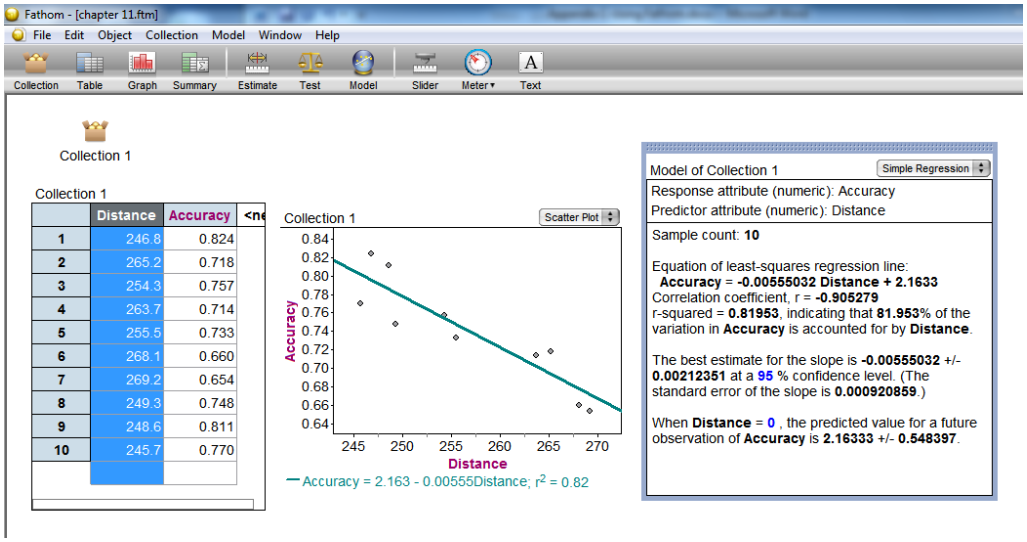


3. To calculate the equation of the least-squares regression line and display its graph on the scatterplot, right-click in the middle of the scatterplot and choose "Least-Squares Line." This will add the line to the scatterplot and display the equation below the plot. You can choose to report the equation of the line in $y = ax + b$ form or $y = a + bx$ form by choosing Preferences from the Edit menu. The screen shot below shows the graph along with the equation: Accuracy = 2.163 – 0.00555Distance. It also shows $r^2 = 0.82$, which is the square of the correlation.



4. An alternate way to calculate the equation of the least-squares regression line is to create a model. To do this, drag a model icon from the shelf to the workspace and choose "Simple Regression" from the drop down menu. Then, drag the "Accuracy" heading to the response attribute and drag the "Distance" heading to the Predictor attribute. The resulting output includes the equation of the least-squares regression line, the correlation, the value and interpretation of r^2, a 95% confidence interval for the slope, and a 95%

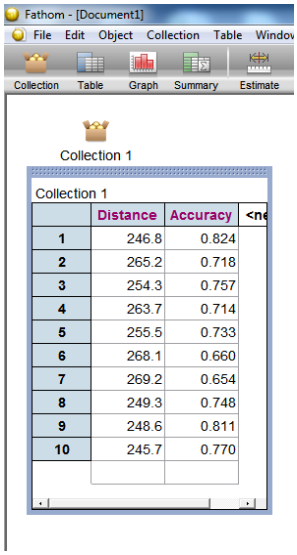
prediction interval for the accuracy when distance = 0 (not a very plausible situation in this context!)



Chapter 11: Calculating the Standard Deviation of the Residuals

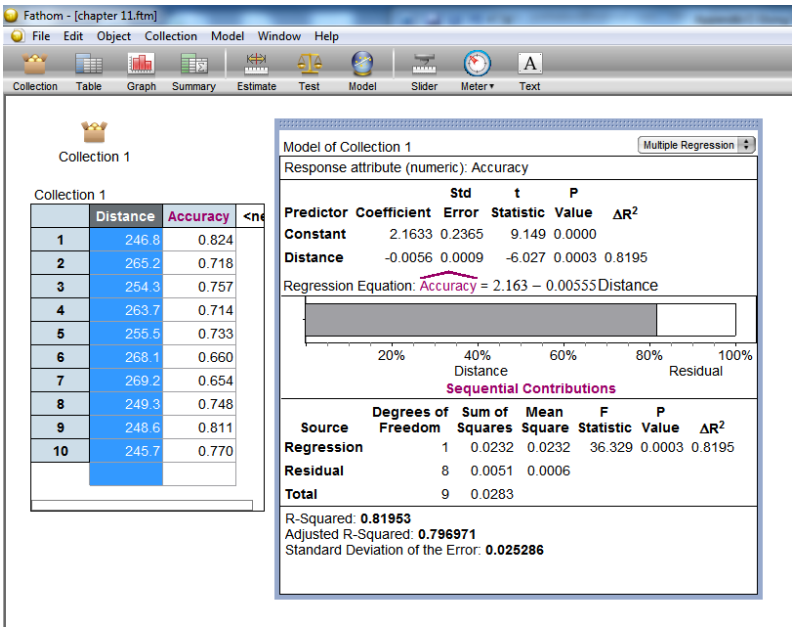
To calculate the standard deviation of the residuals for the relationship between average driving distance and driving accuracy for the top-10 LPGA golfers in 2009, follow the steps below. *Note*: to calculate and display the residuals, see Chapter 13: Making Residual Plots.

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Distance" and the second column "Accuracy." Fill in the data as shown in the screen shot.



	Distance	Accuracy
1	246.8	0.824
2	265.2	0.718
3	254.3	0.757
4	263.7	0.714
5	255.5	0.733
6	268.1	0.660
7	269.2	0.654
8	249.3	0.748
9	248.6	0.811
10	245.7	0.770

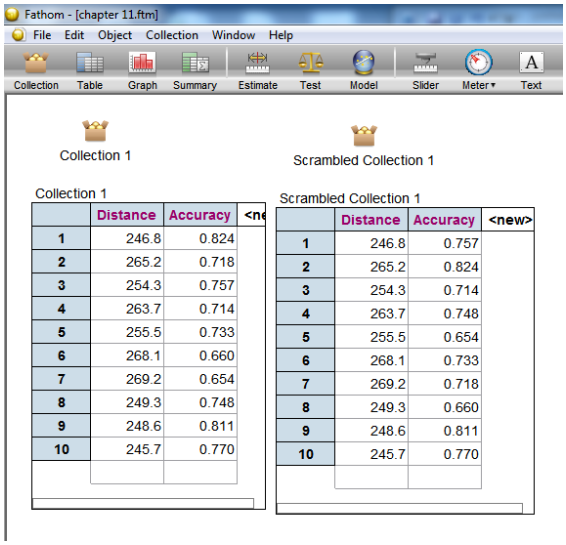
2. Drag a model icon from the shelf to the workspace. Choose "Multiple Regression" from the drop-down menu in the upper-right corner. Then, drag the "Accuracy" heading to the Response attribute. Then, drag the "Distance" heading to the line just below "Constant" that says "Drop attributes here to add predictors to the model." Expand the window so that it looks like the screen shot below. The standard deviation of the residuals is at the very bottom and is called the "standard deviation of the error." In this case, s = 0.025286. The least-squares regression line is displayed just under the location where you dropped the "Distance" attribute.


Fathom - [chapter 11.ftm]
File Edit Object Collection Model Window Help
Collection Table Graph Summary Estimate Test Model Slider Meter Text
Collection 1
Collection 1
Distance Accuracy
1 246.8 0.824
2 265.2 0.718
3 254.3 0.757
4 263.7 0.714
5 255.5 0.733
6 268.1 0.660
7 269.2 0.654
8 249.3 0.748
9 248.6 0.811
10 245.7 0.770
Model of Collection 1
Multiple Regression
Response attribute (numeric): Accuracy
Predictor Coefficient Std Error t Statistic P Value ΔR²
Constant 2.1633 0.2365 9.149 0.0000
Distance -0.0056 0.0009 -6.027 0.0003 0.8195
Regression Equation: Accuracy = 2.163 – 0.00555 Distance
20% 40% 60% 80% 100%
Distance
Residual
Sequential Contributions
Source Degrees of Freedom Sum of Squares Mean Square F Statistic P Value ΔR²
Regression 1 0.0232 0.0232 36.329 0.0003 0.8195
Residual 8 0.0051 0.0006
Total 9 0.0283
R-Squared: 0.81953
Adjusted R-Squared: 0.796971
Standard Deviation of the Error: 0.025286

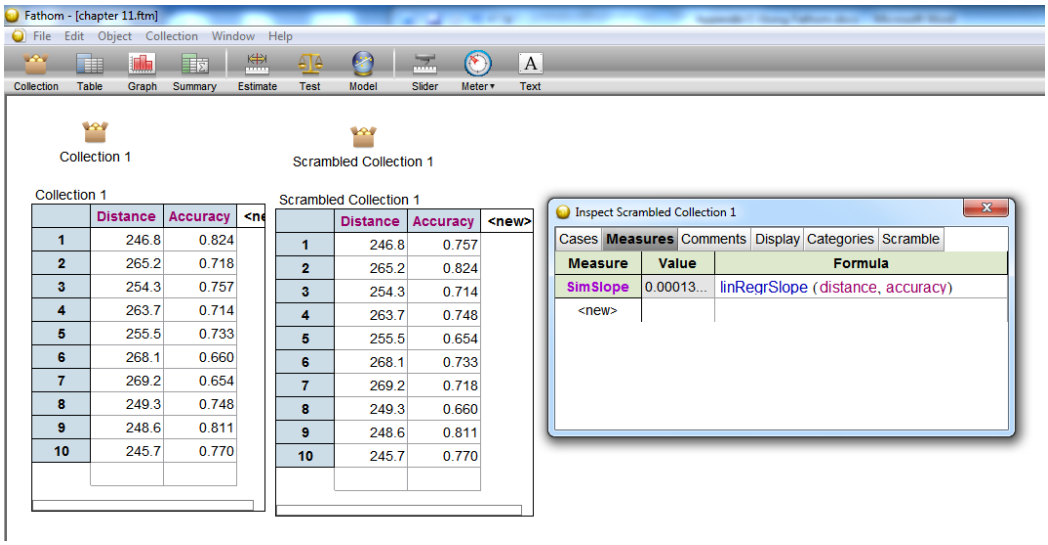
Chapter 11: Simulating the Slope

Follow these steps to simulate the slope of the least-squares regression line for the relationship between the average driving distance and driving accuracy for the top-10 LPGA golfers in 2009:

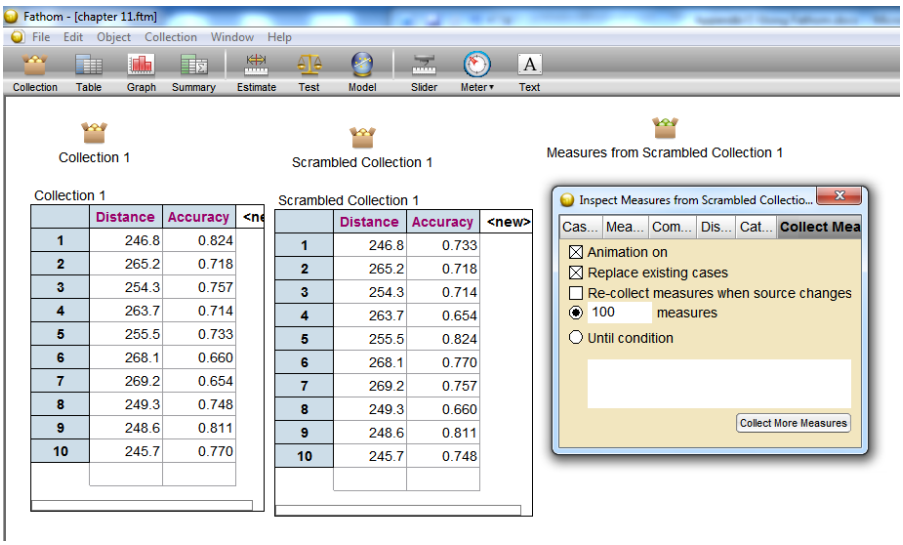
1. Enter the data as described in step 1 above.
2. To shuffle the accuracy values, right-click the box called "Collection 1" and choose "Scramble Attribute Values." In the inspector, change the attribute from Distance to Accuracy and press Scramble Attribute Values again. This shuffles the proportions in the Accuracy column. To see the results, highlight the box called "Scrambled Collection 1" and drag a table icon onto the workspace.



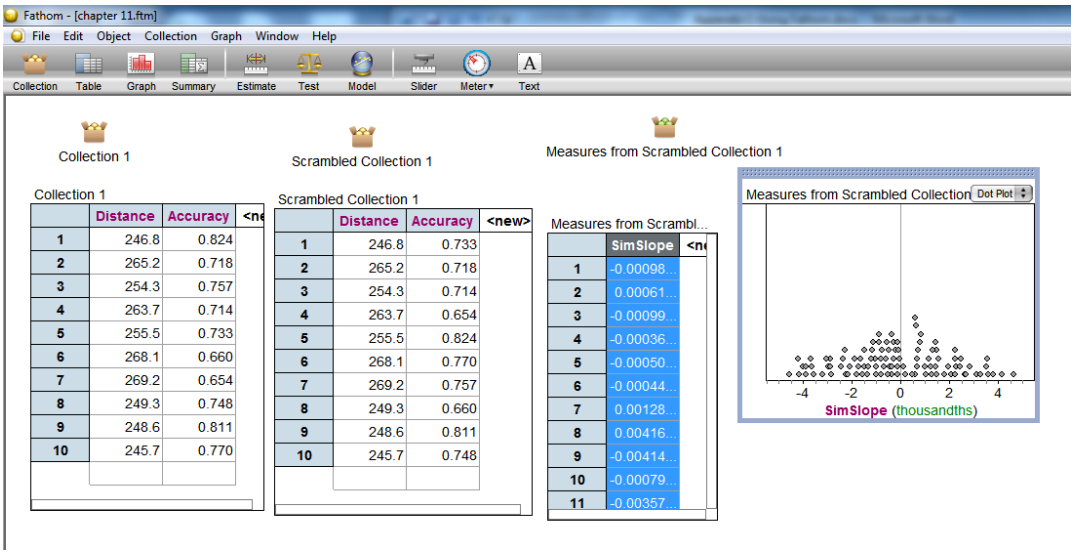
3. To have Fathom repeatedly shuffle the accuracy values, calculate the slope, and record these values, double-click on the box called "Scrambled Collection 1." Then, click on the "Measures" tab in the Scrambled Collection 1 Inspector. Under the column heading measure, type "SimSlope." Now, double click on the cell under the heading formula and type: LinRegrSlope(distance,accuracy)



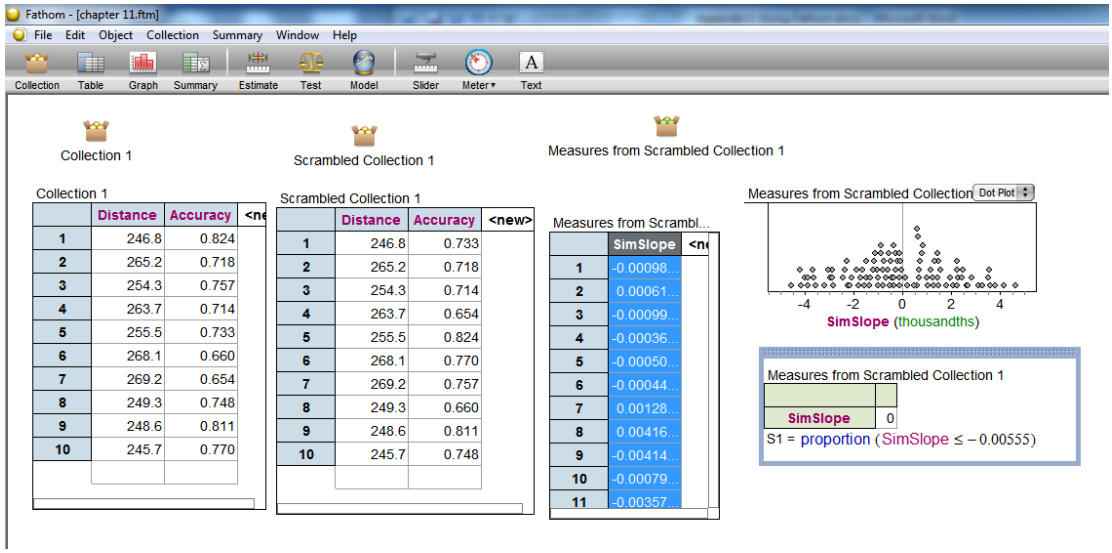
4. Next, right-click the box called “Scrambled Collection 1” and choose Collect Measures. In the inspector, select “Replace existing cases” and enter 100 instead of 5 for the number of measures. Then, press Collect More Measures. *Note:* To increase the number of measures, uncheck the “Replace existing cases” box, enter the number of additional measures desired, and press “Collect more measures.” The simulation will go much faster if you uncheck the “Animation on” box.



5. To see the results of the 100 trials, highlight the box called “Measures from Scrambled Collection 1” and drag a table icon from the shelf onto the workspace. Then, drag a graph icon from the shelf onto the workspace. Finally, drag the heading “SimSlope” from the table to the horizontal axis of the graph. If you need to rescale the axis, you can either double click on the axis and change the settings or just drag the values on the horizontal axis to the left or right.



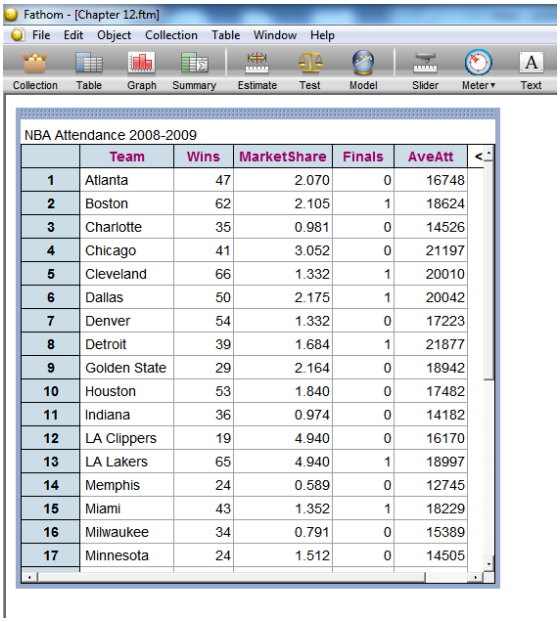
6. To have Fathom calculate the proportion of dots that are –0.00555 or lower, drag a summary icon from the shelf onto the workspace. Then, drag the SimSlope heading from the table into the lower-left corner of the summary box, double click on the formula S1=mean() and enter S1=proportion(SimSlope ≤–0.00555). To find the ≤ symbol, hold the ctrl button and the < symbol will turn into a ≤ symbol in the dialog box.



Chapter 12: Creating Multiple Regression Models

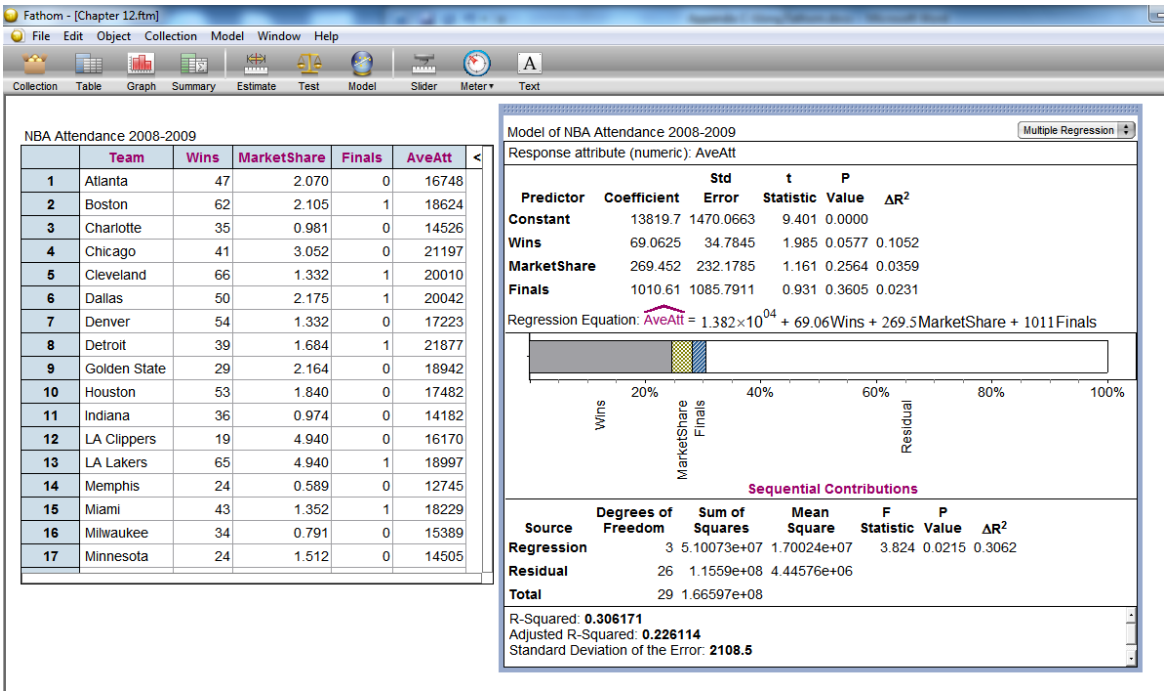
To create a multiple regression model for NBA attendance in 2008–2009 using wins, market share, and finals, follow the steps below:

1. Enter the data for all 30 NBA teams from the example in Chapter 12, as shown in the partial screen shot below.



	Team	Wins	MarketShare	Finals	AveAtt
1	Atlanta	47	2.070	0	16748
2	Boston	62	2.105	1	18624
3	Charlotte	35	0.981	0	14526
4	Chicago	41	3.052	0	21197
5	Cleveland	66	1.332	1	20010
6	Dallas	50	2.175	1	20042
7	Denver	54	1.332	0	17223
8	Detroit	39	1.684	1	21877
9	Golden State	29	2.164	0	18942
10	Houston	53	1.840	0	17482
11	Indiana	36	0.974	0	14182
12	LA Clippers	19	4.940	0	16170
13	LA Lakers	65	4.940	1	18997
14	Memphis	24	0.589	0	12745
15	Miami	43	1.352	1	18229
16	Milwaukee	34	0.791	0	15389
17	Minnesota	24	1.512	0	14505

2. Drag a Model icon from the shelf to the workspace. In the drop-down menu, choose "Multiple Regression." Then, drag the "AveAtt" heading to the Response attribute and drag the "Wins," "MarketShare," and "Finals" headings to the line that says "Drop attributes here to add predictors to the model."



	Team	Wins	MarketShare	Finals	AveAtt
1	Atlanta	47	2.070	0	16748
2	Boston	62	2.105	1	18624
3	Charlotte	35	0.981	0	14526
4	Chicago	41	3.052	0	21197
5	Cleveland	66	1.332	1	20010
6	Dallas	50	2.175	1	20042
7	Denver	54	1.332	0	17223
8	Detroit	39	1.684	1	21877
9	Golden State	29	2.164	0	18942
10	Houston	53	1.840	0	17482
11	Indiana	36	0.974	0	14182
12	LA Clippers	19	4.940	0	16170
13	LA Lakers	65	4.940	1	18997
14	Memphis	24	0.589	0	12745
15	Miami	43	1.352	1	18229
16	Milwaukee	34	0.791	0	15389
17	Minnesota	24	1.512	0	14505

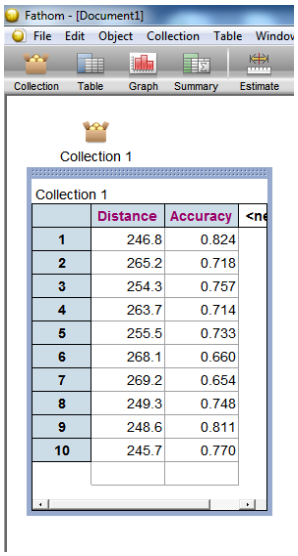
3. The equation of the multiple regression model will be displayed just below the place where you dragged the predictor attributes. The standard deviation of the residuals is listed at the very bottom of the output and is called the standard deviation of the error.

4. To investigate the effect each of the predictor variables has on the model, you can remove variables by right-clicking on a predictor variable and choosing "Remove Attribute." To add the variable back into the model, just drag its heading back to the list of predictor variables in the model.

Chapter 13: Making Residual Plots

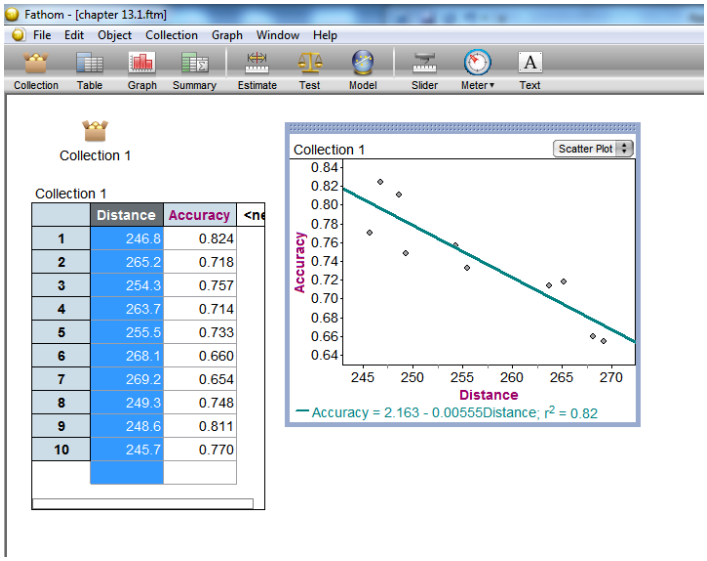
To create a residual plot for the relationship between average driving distance and driving accuracy for the top-10 LPGA golfers in 2009, follow the steps below.

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Distance" and the second column "Accuracy." Fill in the data as shown in the screen shot below.



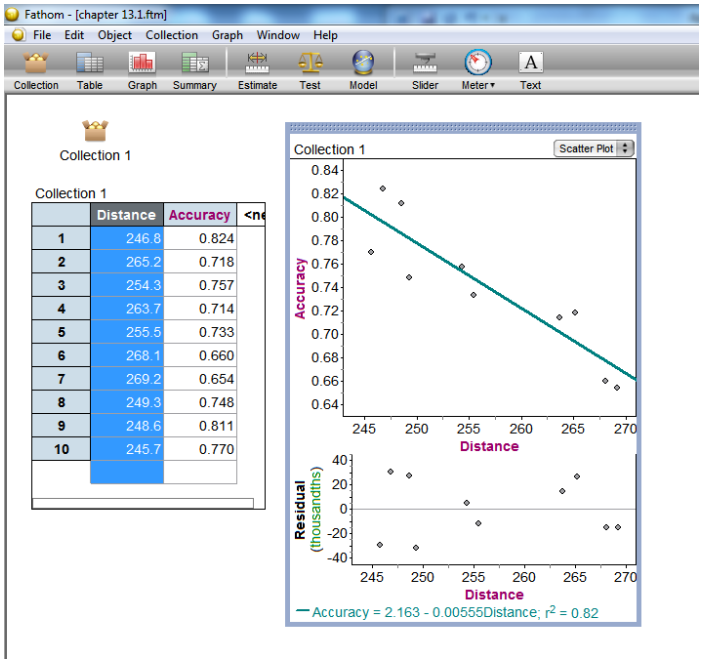
	Distance	Accuracy
1	246.8	0.824
2	265.2	0.718
3	254.3	0.757
4	263.7	0.714
5	255.5	0.733
6	268.1	0.660
7	269.2	0.654
8	249.3	0.748
9	248.6	0.811
10	245.7	0.770

2. Make a scatterplot of the data by dragging a graph icon to the workspace. Then, drag the "Distance" heading to the horizontal axis and the "Accuracy" heading to the vertical axis as shown in the screen shot below. Then, add the least-squares regression line by right-clicking in the middle of the scatterplot and choosing "Least Squares Line." As with other graphs, you can double-click on axis to change the scale or grab a number on the scale and drag it to the left or right as needed.

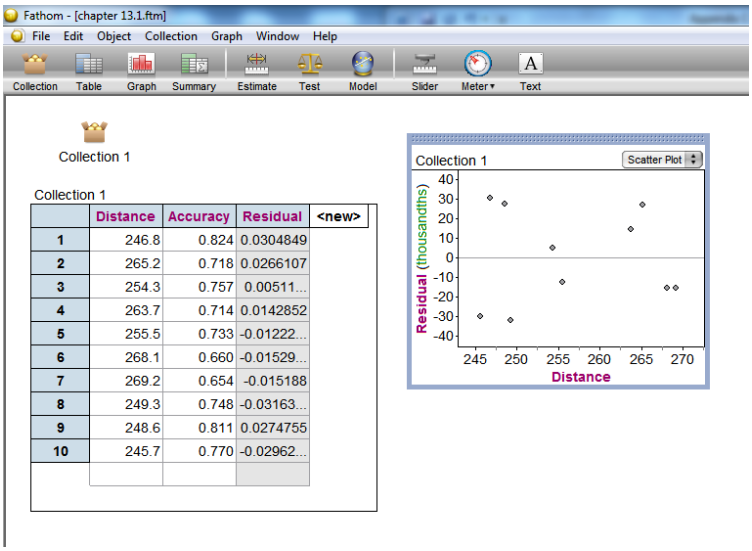


	Distance	Accuracy
1	246.8	0.824
2	265.2	0.718
3	254.3	0.757
4	263.7	0.714
5	255.5	0.733
6	268.1	0.660
7	269.2	0.654
8	249.3	0.748
9	248.6	0.811
10	245.7	0.770

3. To make a residual plot the easy way, right-click in the middle of the scatterplot and choose “Make Residual Plot.” This option only works after graphing a model on the scatterplots (e.g. the least-squares regression line). To make the residual plot larger, grab the bottom of the graph with your mouse and drag it down.



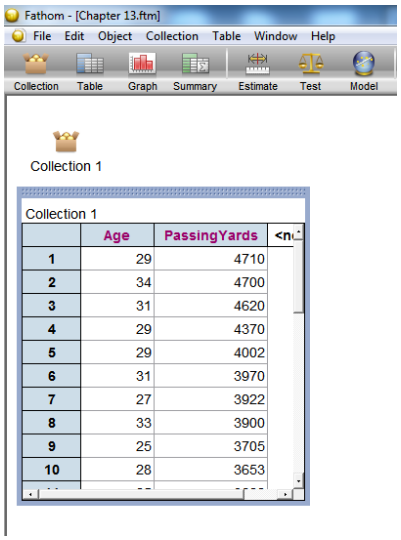
4. You can also make a residual plot by calculating the residuals in a separate column and then making a new graph. To do this, type the attribute “Residual” into the heading of the third column. Then, right-click on the heading and choose “Edit formula.” Then, enter the formula: accuracy – (2.163 – 0.00555distance). This will subtract the predicted accuracy from the actual accuracy for each player. If you are only making residual plots for least-squares regression lines, you can also enter the formula: LinRegrResidual(distance,accuracy). To make the residual plot, drag a graph icon from the shelf, drag the “Distance” heading to the horizontal axis and the “Residual” heading to the vertical axis.


Fathom - [chapter 13.1.ftm]
File Edit Object Collection Graph Window Help
Collection Table Graph Summary Estimate Test Model Slider Meter Text
Collection 1
Collection 1
Distance Accuracy Residual <new>
1 246.8 0.824 0.0304849
2 265.2 0.718 0.0266107
3 254.3 0.757 0.00511...
4 263.7 0.714 0.0142852
5 255.5 0.733 -0.01222...
6 268.1 0.660 -0.01529...
7 269.2 0.654 -0.015188
8 249.3 0.748 -0.03163...
9 248.6 0.811 0.0274755
10 245.7 0.770 -0.02962...
Collection 1
Scatter Plot
Residual (thousandths)
40 30 20 10 0 -10 -20 -30 -40
245 250 255 260 265 270
Distance

Chapter 13: Calculating Quadratic Models

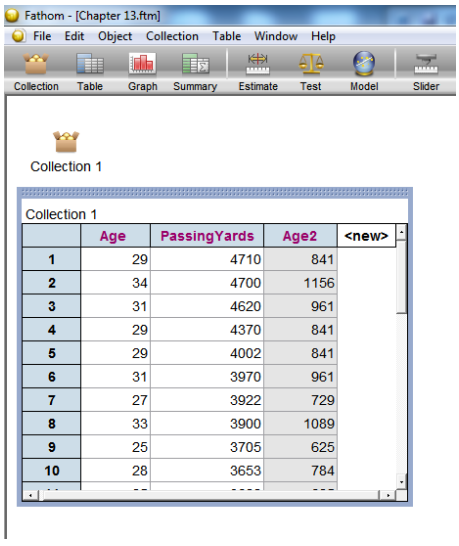
To calculate the equation of a quadratic model for the quarterback age and passing yards data from Chapter 13, you need to create a multiple regression model with two explanatory variables: age and age^2. To do this, follow the steps below:

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Age" and the second column "PassingYards." Fill in the data from Chapter 13 as shown in the partial screen shot below.



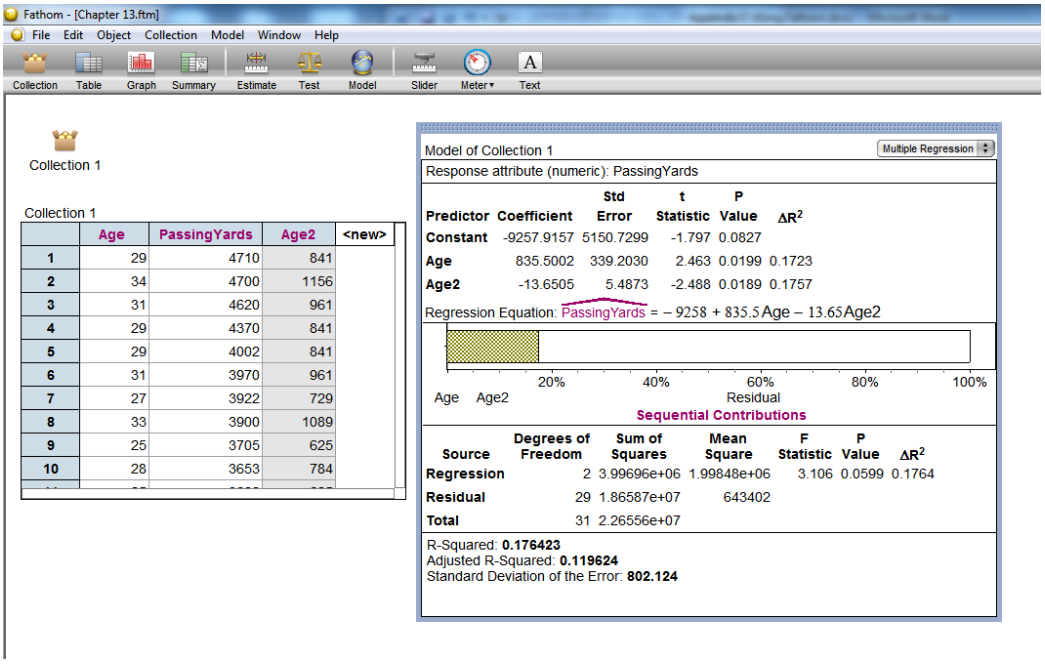
	Age	PassingYards
1	29	4710
2	34	4700
3	31	4620
4	29	4370
5	29	4002
6	31	3970
7	27	3922
8	33	3900
9	25	3705
10	28	3653

2. Create a third attribute called "Age2," right-click on the heading and choose "Edit Formula." Then, enter the formula: Age^2. This will square the age values and allow us to use multiple regression to create a quadratic model.



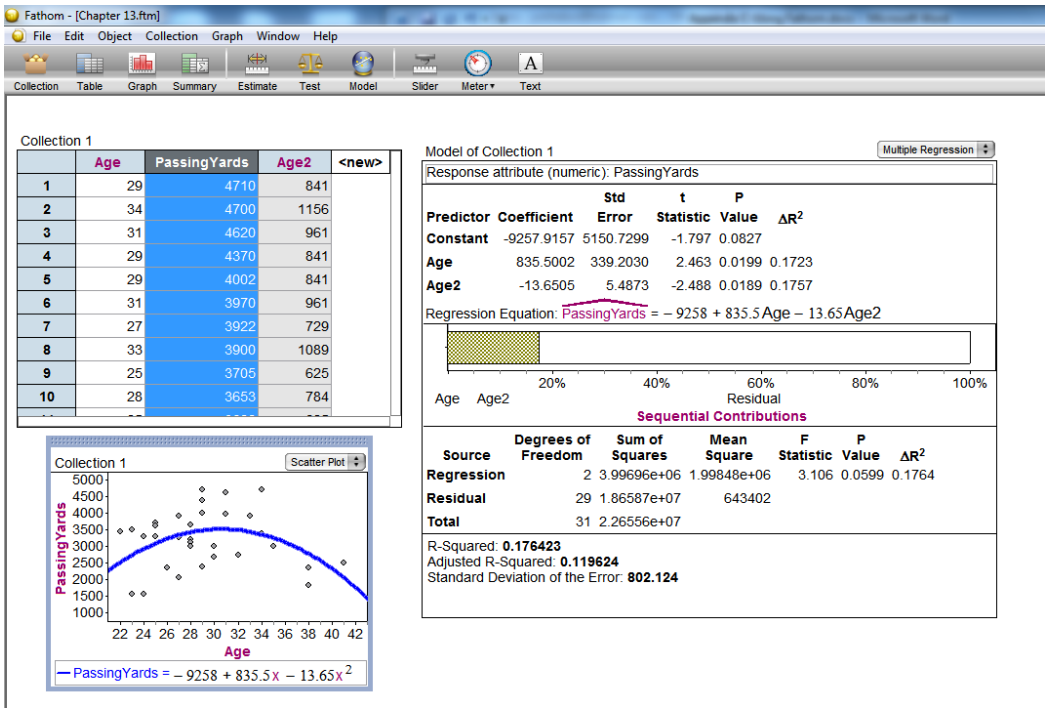
	Age	PassingYards	Age2	<new>
1	29	4710	841	
2	34	4700	1156	
3	31	4620	961	
4	29	4370	841	
5	29	4002	841	
6	31	3970	961	
7	27	3922	729	
8	33	3900	1089	
9	25	3705	625	
10	28	3653	784	

3. Drag a Model icon from the shelf to the workspace. In the drop-down menu, choose "Multiple Regression." Then, drag the "PassingYards" heading to the Response attribute and drag the "Age" and "Age2" headings to the line that says "Drop attributes here to add predictors to the model." The equation of the model will be directly below where you dropped the predictor attributes.



	Age	PassingYards	Age2	<new>
1	29	4710	841	
2	34	4700	1156	
3	31	4620	961	
4	29	4370	841	
5	29	4002	841	
6	31	3970	961	
7	27	3922	729	
8	33	3900	1089	
9	25	3705	625	
10	28	3653	784	

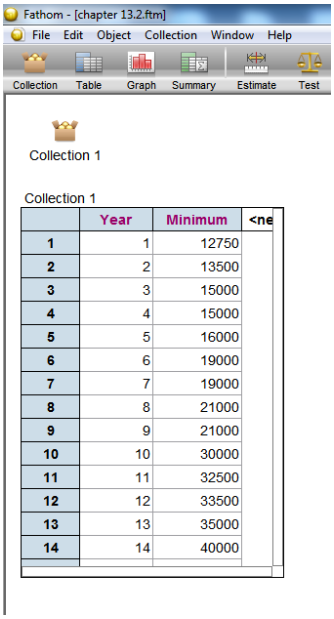
4. To graph the quadratic model on the scatterplot, drag a graph icon from the shelf onto the workspace. Then, drag the "Age" heading to the horizontal axis and the "PassingYards" heading to the vertical axis. Finally, right click in the middle of the graph and choose "Plot Function." Then, enter the formula: –9258 + 835.5x – 13.65*x*^2.


Fathom - [Chapter 13.ftm]
File Edit Object Collection Graph Window Help
Collection Table Graph Summary Estimate Test Model Slider Meter Text
Collection 1
Age PassingYards Age2 <new>
1 29 4710 841
2 34 4700 1156
3 31 4620 961
4 29 4370 841
5 29 4002 841
6 31 3970 961
7 27 3922 729
8 33 3900 1089
9 25 3705 625
10 28 3653 784
Collection 1
Scatter Plot
PassingYards
5000 4500 4000 3500 3000 2500 2000 1500 1000
22 24 26 28 30 32 34 36 38 40 42
Age
PassingYards = – 9258 + 835.5x – 13.65x2
Model of Collection 1
Multiple Regression
Response attribute (numeric): PassingYards
Predictor Coefficient Std Error t Statistic P Value ΔR2
Constant -9257.9157 5150.7299 -1.797 0.0827
Age 835.5002 339.2030 2.463 0.0199 0.1723
Age2 -13.6505 5.4873 -2.488 0.0189 0.1757
Regression Equation: PassingYards = – 9258 + 835.5 Age – 13.65 Age2
20% 40% 60% 80% 100%
Age Age2 Residual
Sequential Contributions
Source Degrees of Freedom Sum of Squares Mean Square F Statistic P Value ΔR2
Regression 2 3.99696e+06 1.99848e+06 3.106 0.0599 0.1764
Residual 29 1.86587e+07 643402
Total 31 2.26556e+07
R-Squared: 0.176423
Adjusted R-Squared: 0.119624
Standard Deviation of the Error: 802.124

Chapter 13: Calculating Exponential Models

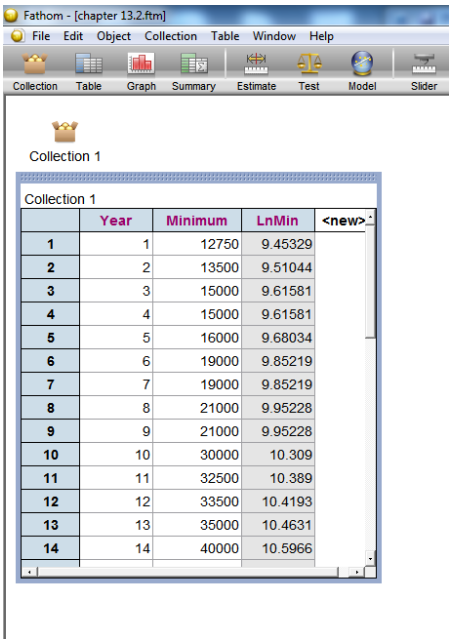
To calculate the equation of a exponential model for the Major League Baseball minimum salary data from Chapter 13, you need to use logarithms to transform the salary data. To do this, follow the steps below:

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Year" and the second column "Minimum." Fill in the data from Chapter 13 as shown in the partial screen shot below.



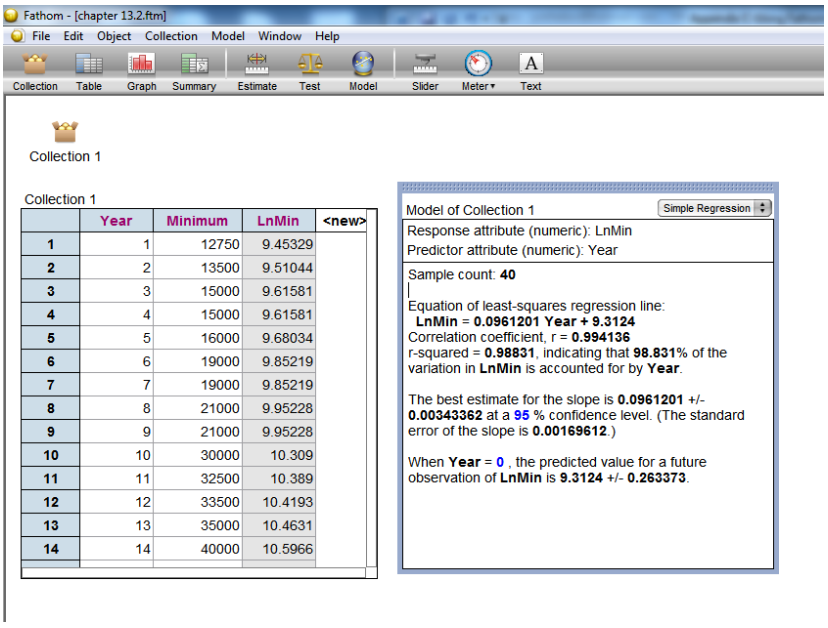
	Year	Minimum	<ne
1	1	12750	
2	2	13500	
3	3	15000	
4	4	15000	
5	5	16000	
6	6	19000	
7	7	19000	
8	8	21000	
9	9	21000	
10	10	30000	
11	11	32500	
12	12	33500	
13	13	35000	
14	14	40000	

2. Create a third attribute called "LnMin," right-click on the heading and choose "Edit Formula." Then, enter the formula: ln(minimum). This will take the natural logarithm of the minimum salaries. *Note:* You can use logarithms of any base.



	Year	Minimum	LnMin	<new>
1	1	12750	9.45329	
2	2	13500	9.51044	
3	3	15000	9.61581	
4	4	15000	9.61581	
5	5	16000	9.68034	
6	6	19000	9.85219	
7	7	19000	9.85219	
8	8	21000	9.95228	
9	9	21000	9.95228	
10	10	30000	10.309	
11	11	32500	10.389	
12	12	33500	10.4193	
13	13	35000	10.4631	
14	14	40000	10.5966	

3. Drag a Model icon from the shelf to the workspace. In the drop-down menu, choose "Simple Regression." Then, drag the "LnMin" heading to the Response attribute and drag the "Year" heading to the predictor attribute.



	Year	Minimum	LnMin	<new>
1	1	12750	9.45329	
2	2	13500	9.51044	
3	3	15000	9.61581	
4	4	15000	9.61581	
5	5	16000	9.68034	
6	6	19000	9.85219	
7	7	19000	9.85219	
8	8	21000	9.95228	
9	9	21000	9.95228	
10	10	30000	10.309	
11	11	32500	10.389	
12	12	33500	10.4193	
13	13	35000	10.4631	
14	14	40000	10.5966	

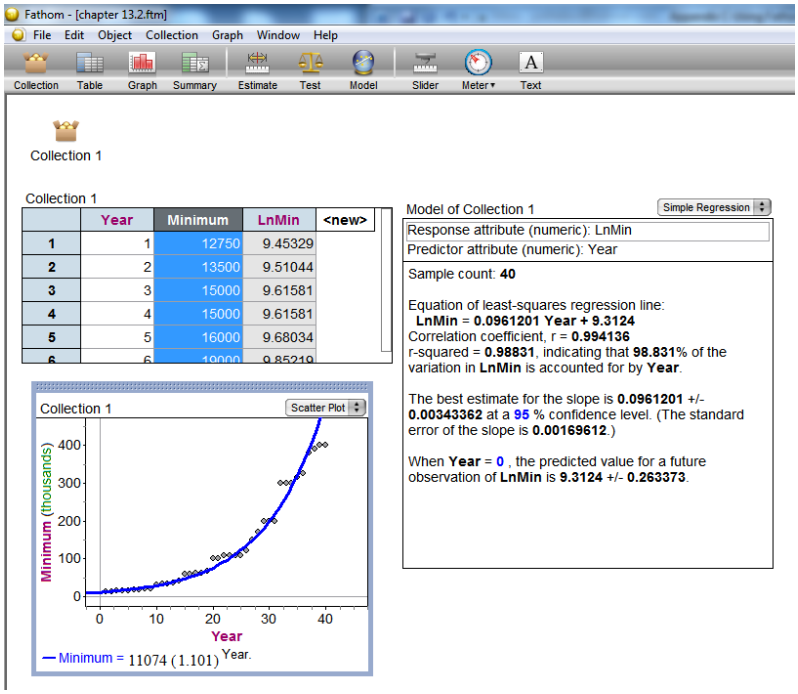
4. The least-squares regression line is LnMin = 0.0961201Year + 9.3124. To create an exponential model, we need to back-transform the equation to get rid of the logarithm:

$$e^{LnMin} = e^{0.0961202Year + 9.3124}$$

$$Min = \left(e^{0.0961202}\right)^{Year}\left(e^{9.3124}\right)$$

$$Min = 11074\left(1.101\right)^{Year}$$

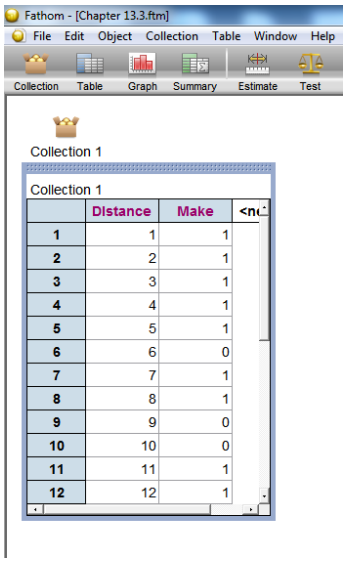
5. To graph the exponential model on the scatterplot, drag a graph icon from the shelf onto the workspace. Then, drag the "Year" heading to the horizontal axis and the "Minimum" heading to the vertical axis. Finally, right click in the middle of the graph and choose "Plot Function." Then, enter the formula: 11074(1.101)^Year.



Chapter 13: Graphing Logistic Models

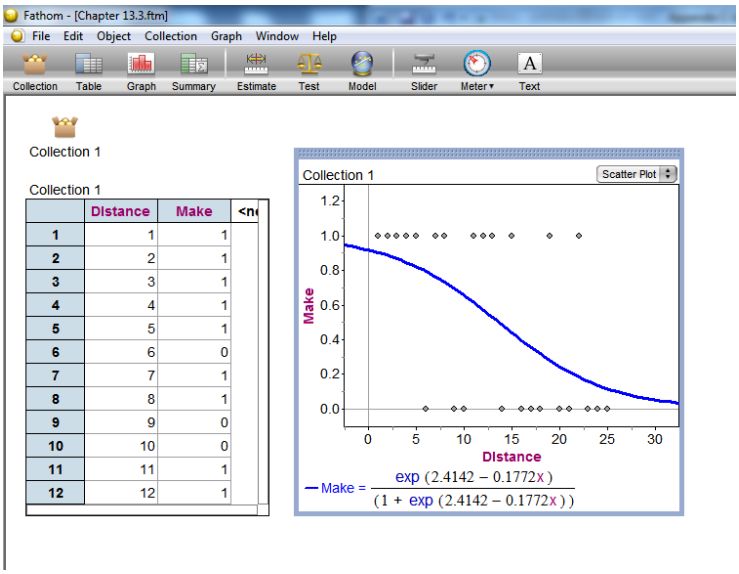
Although Fathom will not calculate a logistic model, it is possible to use Fathom to graph a logistic model. To graph a logistic model for the basketball shooting data on page 521, follow the steps below.

1. Drag a table icon from the shelf onto the workspace. Then, create two attributes by naming the first column "Distance" and the second column "Make." Fill in the data from Chapter 13, page 521, as shown in the partial screen shot below.



	Distance	Make
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1
6	6	0
7	7	1
8	8	1
9	9	0
10	10	0
11	11	1
12	12	1

2. Create a scatterplot by dragging a graph icon from the shelf to the workspace. Then, drag the "Distance" heading to the horizontal axis and the "Make" heading to the vertical axis. Finally, right-click in the middle of the scatterplot, choose "Plot Function" and add the formula: $\exp(2.4142 - 0.1772x)/(1+\exp(2.4142 - 0.1772x))$.



	Distance	Make
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1
6	6	0
7	7	1
8	8	1
9	9	0
10	10	0
11	11	1
12	12	1