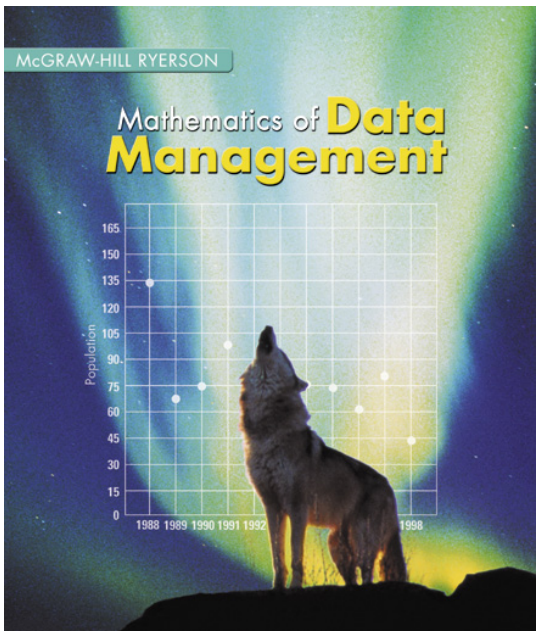
Fathom™ Tutorial

Dynamic Statistical Software

designed for teachers using

McGraw-Hill Ryerson
Mathematics of Data Management



by Roland W. Meisel
Version 1.1 2004-07-27

Please Return This Guide to the Presenter at the End of the Workshop

if you would like an electronic copy of this guide, please send email to:

rollym@vaxxine.com

Thank you!

Foreword

This tutorial is designed for the teacher who will be using Fathom™ to teach the Grade 12 Mathematics of Data Management MDM4U course. It is keyed to the *McGraw-Hill Ryerson Mathematics of Data Management* text, © 2002, McGraw-Hill Ryerson Limited. All worked examples and technical extensions that specifically make use of Fathom™ are included in this tutorial. Step-by-step keystroke instructions and liberal use of screen shots will ease the novice along the learning curve for this powerful new technology. The user will find it helpful to follow through the text as he or she works through this tutorial. By working through the tutorial, the user will gain a solid knowledge of Fathom™ as applied to the MDM4U course, including topics such as:

* entering, displaying, sorting, and filtering data

* calculating values such as maximum, minimum, and mean

* calculating measures of central tendency and measures of spread

* creating histograms

* creating and automating simulations of processes involving random numbers

* creating scatter plots and lines of best fit, and calculating correlation coefficients

* performing linear and non-linear regressions

* performing dynamic curve-fitting using sliders

* detecting a hidden variable

* analysing distributions

* creating normal probability plots

About Fathom™

Fathom™ is a powerful dynamic statistical software package published by

An excellent web site, with Fathom™ resources and links to other web sites, is at **www.keypress.com**.

Fathom™ is a statistics software package that offers a variety of powerful data analysis tools in an easy-to-use format. This section introduces basic features of Fathom™ such as entering, displaying, sorting, and filtering data. A complete guide is available on the Fathom™ CD. The real power of this software will be demonstrated in later chapters with examples that apply its sophisticated tools to statistical analysis and simulations.

When you enter data into Fathom™, it creates a **collection**, an object that contains the data. Fathom™ can then use the data from the collection to produce other objects, such as a **graph**, **table**, or **statistical test**. These secondary objects display and analyse the data from the collection, but they do not actually contain the data themselves. If you delete a graph, table, or statistical test, the data still remains in the collection.

Fathom™ considers a collection as a set of **cases**. Each case in a collection can have a number of **attributes**. For example, the **cases** in a **collection** of medical records could have **attributes** such as the patient's name, age, sex, height, weight, blood pressure, and so on. There are two basic types of attributes, **categorical** (such as male/female) and **continuous** (such as height or weight). The **case table** feature displays the cases in a collection in a format similar to a spreadsheet, with a row for each case and a column for each attribute. You can add, modify, and delete cases using a case table.

Example 1 Tables and Graphs

a) Set up a collection for the hockey league standings from Example 3 on page 17 of the Student text.
b) Graph the Team and Points attributes.

***Note*:** The Interactive Student e-book CD-ROM contains a Microsoft® PowerPoint presentation showing the solutions for this section. Appendix B of the Student text contains detailed information on using Fathom™.

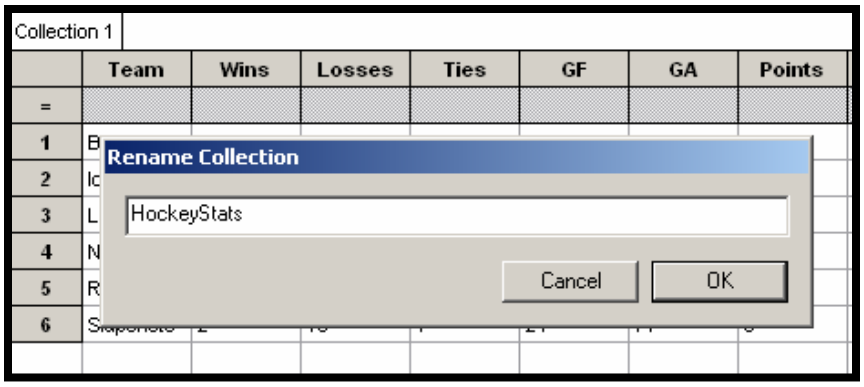
Solution

Launch Fathom™ and drag the **case table**  icon from the shelf to the workspace. Click on the attribute **<new>**, type the heading **Team**, and press **Enter**. Similarly, create attribute columns for **Wins**, **Losses**, **Ties**, **GF**, **GA**, and **Points**. Enter the data into each attribute column. When you are finished, your **case table** will look like the screen shot shown below.

Collection 1

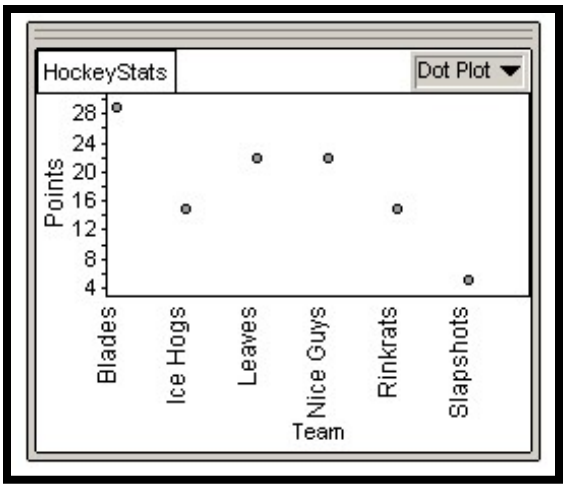
	Team	Wins	Losses	Ties	GF	GA	Points
=							
1	Blades	12	3	5	70	56	29
2	Ice Hogs	6	8	3	46	46	15
3	Leaves	7	5	8	55	51	22
4	Nice Guys	10	5	2	64	42	22
5	Rinkrats	6	7	3	58	63	15
6	Slapshots	2	15	1	24	71	5

Note that your collection has been given the generic name **Collection 1**. You can double-click on the collection box and change the name to something more descriptive, like **HockeyStats**.



Drag the graph icon  to the workspace.

Drag the **Team** attribute from the **case table** to the horizontal axis of the graph, and the **Points** attribute to the vertical axis of the graph. Your graph will look like the screen shot at the right.



Example 2 Sorting and Filtering

a) Rank the hockey teams in Example 1 by points first, then by wins if two teams have the same number of points, and finally by losses if two teams have the same number of points and wins.
b) List only those teams with fewer than 16 points.
c) Set up a separate table showing only the goals for (GF) and goals against (GA) data for the teams and rank the teams by their goals scored.

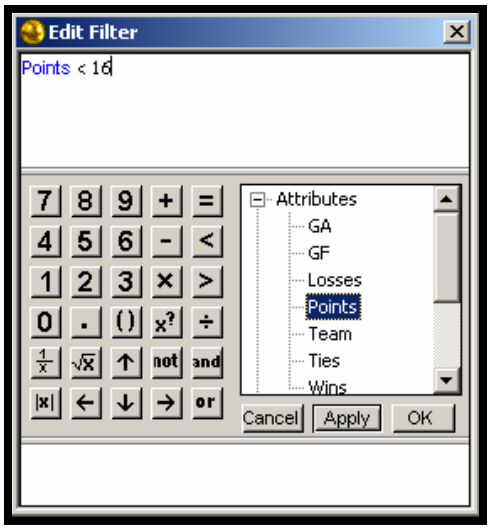
Solution

a) To sort the data, right-click on the **Points** attribute, and choose **Sort Descending**. This will set the primary sort to use the **Points** attribute, and sort the **case table** in descending order of points. Set the secondary sort by right-clicking on the **Wins** attribute, and choosing **Sort Descending**. Set the tertiary sort by right-clicking on the **Losses** attribute, and choosing **Sort Ascending**. Your final sort will result in a **case table** that looks like the following screen shot.

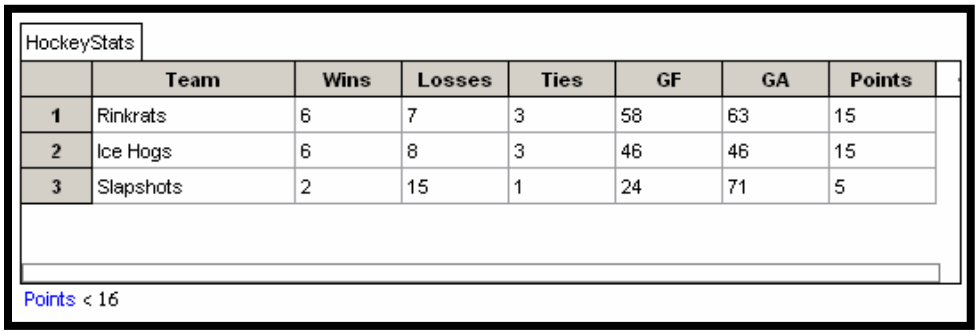
HockeyStats

	Team	Wins	Losses	Ties	GF	GA	Points
1	Blades	12	3	5	70	56	29
2	Nice Guys	10	5	2	64	42	22
3	Leaves	7	5	8	55	51	22
4	Rinkrats	6	7	3	58	63	15
5	Ice Hogs	6	8	3	46	46	15
6	Slapshots	2	15	1	24	71	5

b) To filter the data, choose **Add Filter** from the **Data** menu. To expand the **Attributes** list, click on the plus sign to the left of **Attributes**. The result is shown in the screen shot at the right. Double-click on the **Points** attribute, choose the less-than button, and type 16. Click the **Apply** button, and then **OK**.



The results should look like the screen shot below:

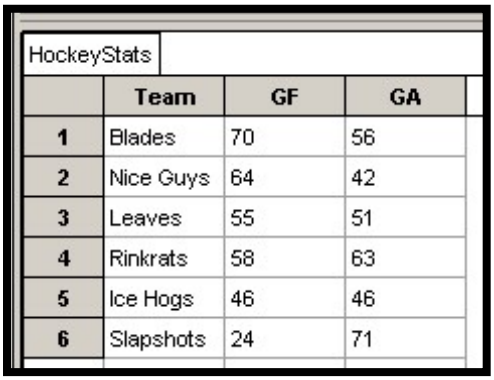

HockeyStats

	Team	Wins	Losses	Ties	GF	GA	Points
1	Rinkrats	6	7	3	58	63	15
2	Ice Hogs	6	8	3	46	46	15
3	Slapshots	2	15	1	24	71	5

Points < 16

The **Filter** is listed at the bottom as **Points** < 16.

c) Click on the **HockeyStats** collection box, and drag a new **case table** onto the workspace. Note that the filter you applied in part b) is not active on this case table. Click on the **Wins** attribute. Choose **Hide Attribute** from the **Display** menu. Use the same method to hide the **Losses**, **Ties**, and **Points** attributes. Right-click on the **GF** attribute, and use **Sort Descending** to rank the teams. Your result will look like the screen shot at the right.


HockeyStats

	Team	GF	GA
1	Blades	70	56
2	Nice Guys	64	42
3	Leaves	55	51
4	Rinkrats	58	63
5	Ice Hogs	46	46
6	Slapshots	24	71

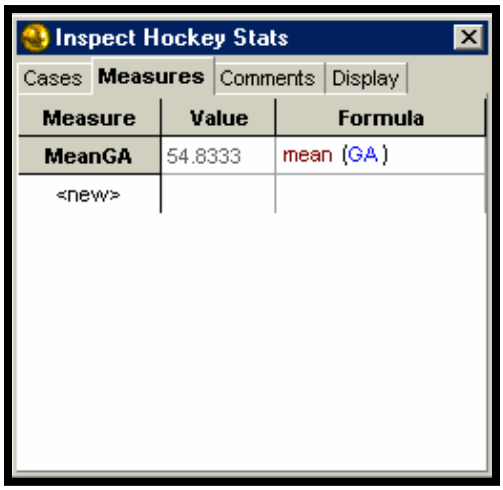
Questions

1. Enter the data from Example 1 into Fathom™. Use the built-in functions in Fathom™ to find the following. (For details on functions in Fathom™, see the Fathom™ section of Appendix B or consult the Fathom™ Help screen or manual).
a) the mean of goals against (GA)
b) the largest value of goals for (GF)
c) the smallest value of GF
d) the sum of GA
e) the sum of GA and GF for each case

2. a) Set up a new collection with the following student marks:
65, 88, 56, 76, 74, 99, 43, 56, 72, 81, 80, 30, 92
b) Sort the marks from lowest to highest.
c) Calculate the mean mark.
d) Determine the median (middle) mark.

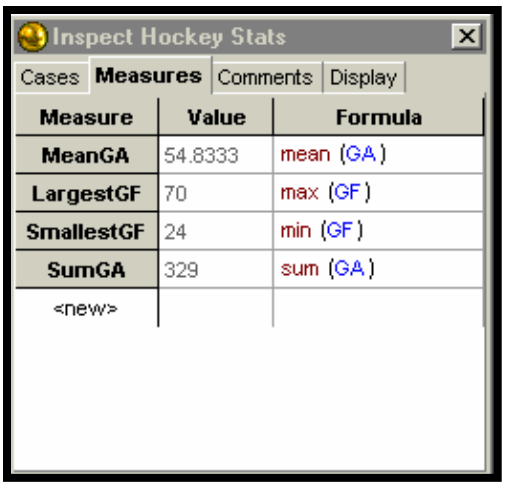
Solution for Question 1

a) Double-click on the **HockeyStats** collection box to open the **inspector**. Choose the **Measures** tab. Double-click on **<new>** and rename it **MeanGA**. Right-click on the box under **Formula**, and choose **Edit Formula**. Expand **Functions/Statistical/One Attribute**. Double-click on **Mean**. Expand **Attributes**, and double-click on **GA**. Your final formula will look like the screen shot at the left below.



Click on **Apply** and then **OK**. The **inspector** will look like the screen shot at the right above. Note that the mean of goals against has been calculated and appears in the **Value** column.

b) In a manner similar to part a), add the largest value of goals for **GF** to the **inspector** using the **max** function under **Functions/Statistical/One Attribute**.



c) Add the smallest value of goals for **GF** to the **inspector** using the **min** function under **Functions/Statistical/One Attribute**.

d) Add the sum of goals against **GA** to the **inspector** using the **sum** function under **Functions/Statistical/One Attribute**. The final result will look like the screen shot at the right.

e) Add an attribute called **GAplusGF**. Right-click on the title, and select **Edit** formula. Edit the formula to GA + GF, as shown in the screen shot below. You can add the formula line by selecting **Show Formulas** under the **Edit** menu.

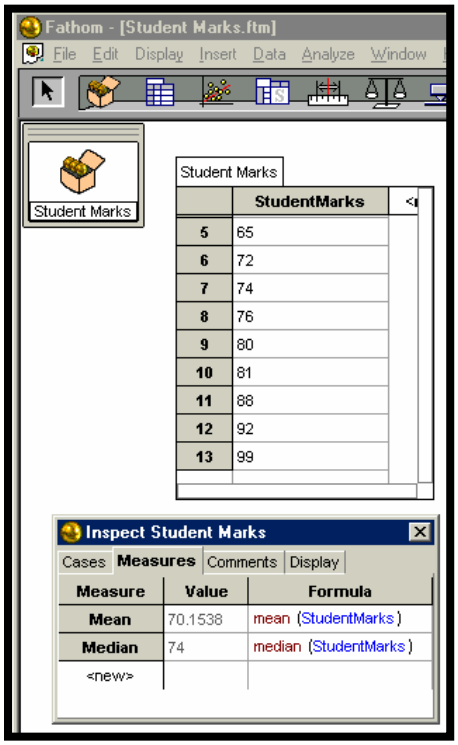
HockeyStats

	Team	Wins	Losses	Ties	GF	GA	Points	GAplusGF	<ne
=								GA + GF	
1	Blades	12	3	5	70	56	29	126	
2	Nice Guys	10	5	2	64	42	22	106	
3	Leaves	7	5	8	55	51	22	106	
4	Rinkrats	6	7	3	58	63	15	121	
5	Ice Hogs	6	8	3	46	46	15	92	
6	Slapshots	2	15	1	24	71	5	95	

Solution for Question 2

a) Choose **New** from the **File** menu, drag a **case table** to the workspace, rename the **<new>** attribute **StudentMarks**, and enter the data.

b) Right-click on the **StudentMarks** attribute, and choose **Sort Ascending**. Your result will look like the screen shot at the right.



c) Double-click on the **Student Marks** collection box to open the **inspector**. Choose the **Measures** tab. Double-click on **<new>** and rename it **Mean**. Right-click on the box under **Formula**, and choose **Edit Formula**. Expand **Functions/Statistical/One Attribute**. Double-click on **Mean**. Move up to **Attributes**, and double-click on **StudentMarks**. Click on **Apply** and then **OK**. The mean of 70.1538 will appear in the **Value** column.

d) In a similar manner to part c), add the median to the **inspector** using the **median** function under **Functions/Statistical/One Attribute**. The value is 74, as shown in the screen shot at the right.

1.4 Simulations

Page 33 MHR MDM

A simulation is an experiment, model, or activity that imitates real or hypothetical conditions.

Example 1 Simulating a Multiple-Choice Test

When writing a multiple-choice test, you may have wondered "What are my chances of passing just by guessing?" Suppose that you make random guesses on a test with 20 questions, each having a choice of 5 answers. The expected mark based on pure guessing is 4 out of 20, since there is a 1 in 5 chance of guessing right on each question. However, it is possible that you could get any number of the questions right—anywhere from zero to a perfect score.

a) Devise a simulation for making guesses on the multiple-choice test.
b) Run the simulation 100 times and use the results to estimate the mark you are likely to get.
c) Would it be practical to run your simulation 1000 times or more?

Solution 4 Using Fathom™

a) Fathom™ has built-in functions to generate random numbers and count the scores in the simulations.

MCTest	Guess
1	2
2	1
3	3
4	5
5	2
6	3
7	3
8	5
9	3
10	5
11	4
12	5
13	5
14	1
15	4
16	5
17	5
18	5
19	4
20	4

Launch Fathom™ and open a new document, if necessary. Drag a new collection box to the workspace and rename it **MCTest**. Right-click on the box and create 20 new cases.

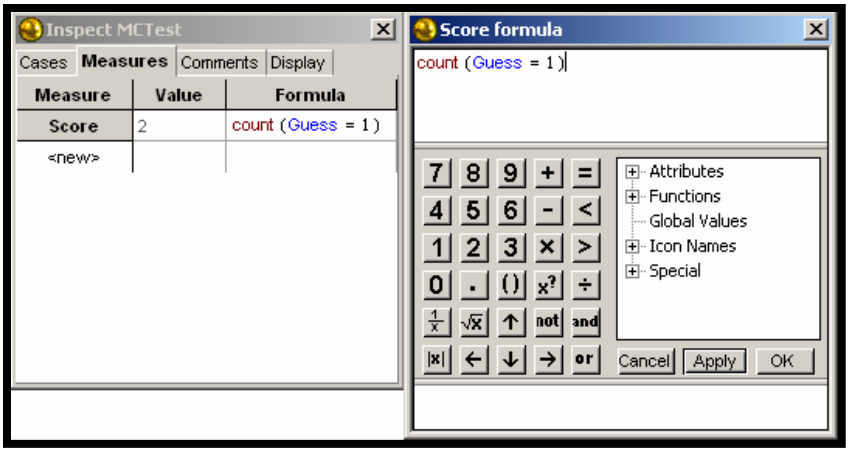
Drag a **case table** to the workspace. You should see your 20 cases listed. If you cannot see all the cases on the screen, expand the table by dragging the edges.

Rename the <**new**> column **Guess**. Right-click on **Guess** and choose **Edit Formula**. Double-click on **Functions**, then **Random Numbers**. Double-click the **randomInteger()** function. Enter 1,5 between the brackets of the **randomInteger()** function, and click **OK.** This will fill the **Guess** column with random integers between 1 and 5. For the purposes of this simulation, assume that the correct answer for each question is 1. Scroll down the column to see how many correct guesses there are in this simulation.

***Note*:** Since these numbers are random, your numbers are unlikely to be the same as those shown in the screen shot at the right.

b) You can run a new simulation by pressing **Ctrl-Y**, which will fill the **Guess** column with a new set of random numbers. Try this a couple of times.

Better still, you can program Fathom™ to repeat the simulation 100 times automatically and keep track of the number of correct guesses.

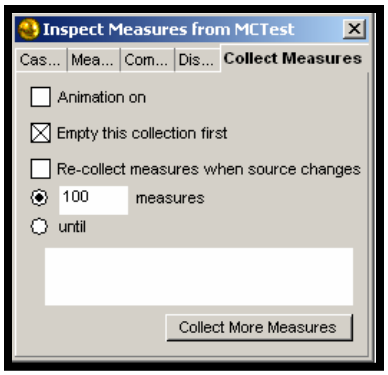


First, set up the counting function. Double-click on the collection box to open the inspector. Choose the **Measures** tab and rename the <**new**> column **Score**. Then, right-click the column below **Formula** and choose **Edit Formula**. Expand **Functions**/**Statistical/One Attribute**. Choose **count**, enter **Guess = 1** between the brackets, and click **OK** to count the number of correct guesses in your **case table**.

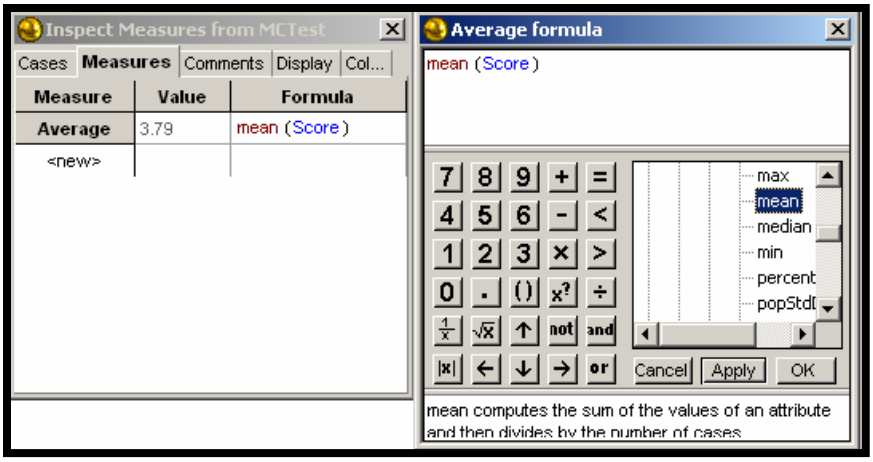
Click on the **MCTest** collection box. Now, choose **Analyse/Collect Measures** from the main menu bar. This will create a new collection called **Measures from MCTest**. Click on this new collection box, and drag a new **case table** to the workspace. Fathom™ will automatically run five simulations of the multiple-choice test and show the results in this **case table**. Note: your results will likely differ from those shown.

Measures from MCTest	Score	<new>
1	4	
2	4	
3	3	
4	2	
5	7	

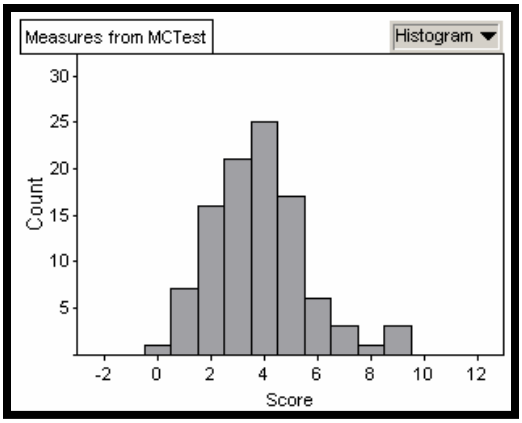
To simulate 100 tests, double-click on the **Measures from MCTest** collection box to open the inspector. Turn off the animation in order to speed up the simulation. Change the number of measures to 100. Then, click on the **Collect More Measures** button. You should now have 100 measures in the **case table** for **Measures** from **MCTest**. Expand the **case table**, and scroll down to see the 100 measures.



Now, you will calculate the mean or average score for these 100 simulations. Go back to the **Inspect Measures from MCTest** box, click on the **Measures** tab, and change the column heading **<new>** to **Average**. Right-click the column below **Formula** and choose **Edit Formula**. Expand **Functions**/**Statistical/One Attribute**. Choose **mean**, enter **Score** between the brackets, and click **OK** to display the mean mark on the 100 tests.



Finally, plot a **histogram** of the scores from the simulations. Drag the graph icon onto the workspace. Then, drag the **Score** attribute from the **Measures from the MCTest** case table to the horizontal axis of the graph. Fathom™ automatically produces a dot plot of your data. To display a histogram instead, simply click the menu in the upper right hand corner of the graph and choose **Histogram**.



c) Fathom™ can easily run this simulation 1000 times or more. To do this, change the number of measures desired to 1000, and press the **Collect More Measures** button, as described above.

Example 2 **Using Technology to Calculate Standard Deviations**

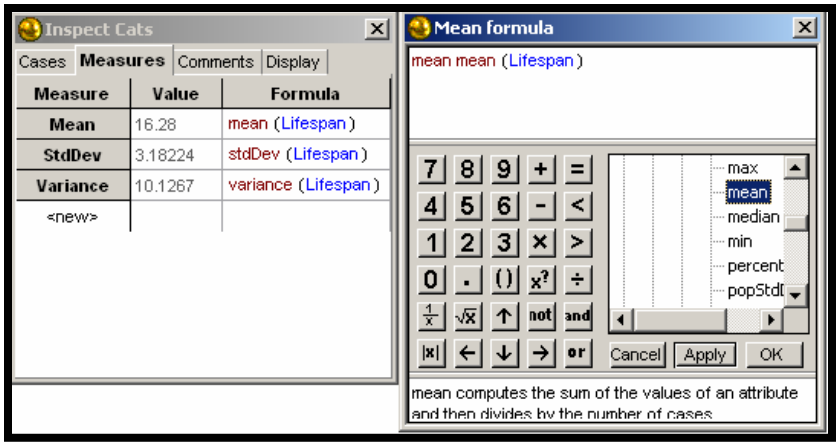
A veterinarian has collected data on the life spans, in years, of a rare breed of cat. Determine the mean, standard deviation, and the variance for these data.

Cats	
	Lifespan
1	16
2	18
3	19
4	12
5	11
6	15
7	20
8	21
9	18
10	15
11	16
12	13
13	16
14	22
15	18
16	19
17	17
18	14
19	9
20	14
21	15
22	19
23	20
24	15
25	15

Solution **Using Fathom™**

Launch Fathom™ and open a new document, if necessary. Drag a new collection box to the workspace and rename it **Cats**. Drag a **case table** to the document. Rename the <**new**> column **Lifespan**. Enter the 25 data shown in the screen shot at the right.

Double-click on the **Cats** collection box to open the **inspector**. Choose the **Measures** tab. Double-click on **<new>** and rename it **Mean**. Right-click on the box under **Formula**, and choose **Edit Formula**. Expand **Functions/Statistical/One Attribute**. Double-click on **Mean**. Move up to **Attributes**, and double-click on **Lifespan**. Click on **Apply** and then **OK**. The mean of 16.28 will appear in the **Value** column. In a similar manner, add the standard deviation and variance to the **inspector** using the **stdDev** and **variance** functions under **Functions/Statistical/One Attribute**. The values are, respectively, 3.18224 and 10.1267, as shown in the screen shot below.



Example 3 Determining Quartiles and Interquartile Ranges

A random survey of people at a science-fiction convention asked them how many times they had seen *Star Wars*. The results are shown at the right.
a) Determine the median, the first and third quartiles, and the interquartile and semi-interquartile ranges. What information do these measures provide?
b) Prepare a suitable box plot of the data.

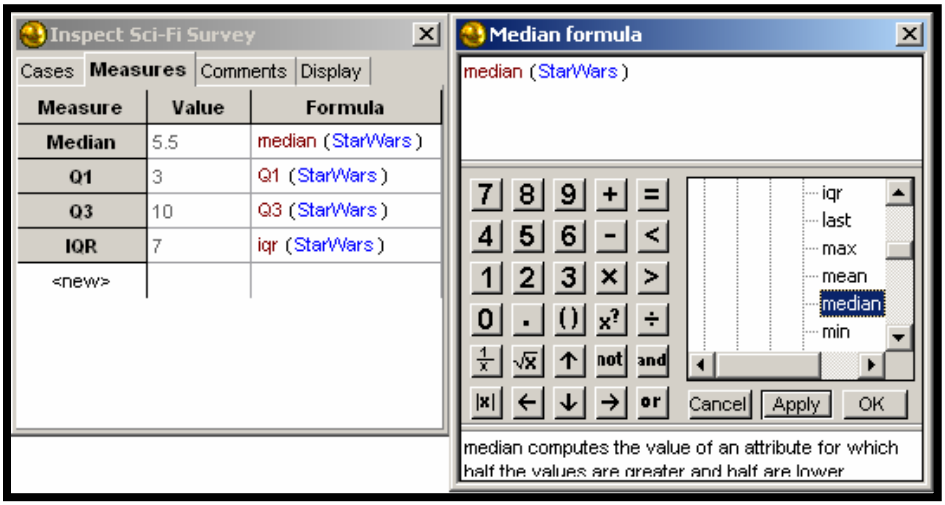
Sci-Fi Survey

	StarWars
1	3
2	4
3	2
4	8
5	10
6	5
7	1
8	15
9	5
10	16
11	6
12	3
13	4
14	9
15	12
16	3
17	30
18	2
19	10
20	7

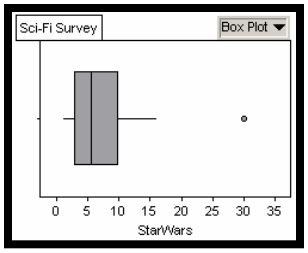
Solution 3 Using Fathom™

a) Launch Fathom™ and open a new document if necessary. Drag a new collection box to the workspace and rename it **Sci-Fi Survey**. Drag a **case table** to the workspace. Rename the <**new**> column **StarWars**. Enter the 20 data shown in the screen shot at the right.

Double-click on the **Sci-Fi Survey** collection box to open the **inspector**. Choose the **Measures** tab. Double-click on **<new>** and rename it **Median**. Right-click on the box under **Formula**, and choose **Edit Formula**. Expand **Functions/Statistical/One Attribute**. Double-click on **Median**. Move up to **Attributes**, and double-click on **StarWars**. Click on **Apply** and then **OK**. The median of 5.5 will appear in the **Value** column. In a similar manner, add the first quartile, the third quartile, and the interquartile range to the **inspector** using the **Q1, Q3** and **iqr** functions under **Functions/Statistical/One Attribute**. The values are, respectively, 3, 10 and 7, as shown in the screen shot below. The semi-interquartile range is half of the interquartile range, or 3.5.



b) Drag a **graph icon** to the workspace. Drag the **StarWars** attribute to the horizontal axis of the graph. Change the graph to a **Box Plot** using the drop-down menu in the upper right corner of the graph.



Example 3 Using Technology to Determine Correlation Coefficients

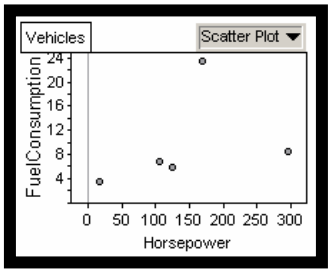
Determine whether there is a linear correlation between horsepower and fuel consumption for these five vehicles by creating a scatter plot and calculating the correlation coefficient.

	Vehicle	Horsepower	FuelConsumption
1	midsize sedan...	105	6.7
2	minivan	170	23.5
3	small SUV	124	5.9
4	motorcycle	17	3.4
5	sports car	296	8.4
6			

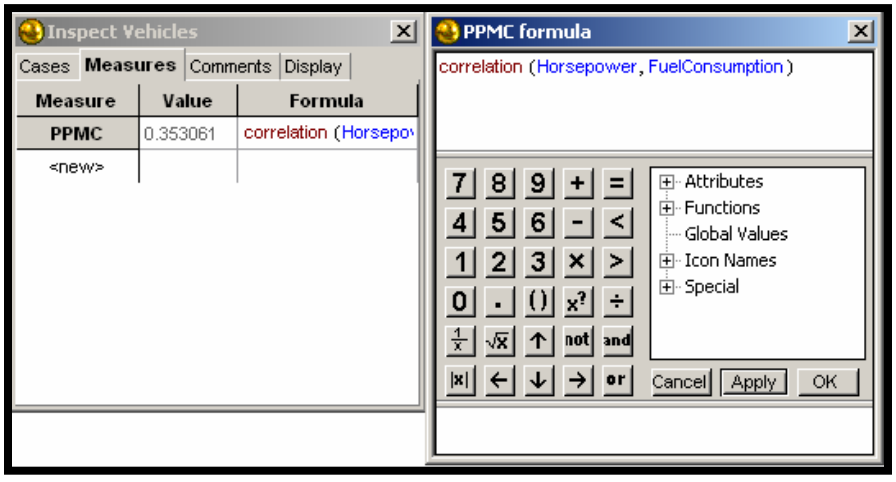
Solution 3 Using Fathom™

Launch Fathom™ and open a new document if necessary. Drag a new collection box to the workspace and rename it **Vehicles**. Drag a **case table** to the workspace. Rename the <**new**> column **Vehicle**. Also create attribute columns named **Horsepower** and **FuelConsumption**. Enter the 5 data shown in the screen shot.

To create a scatter plot, drag a **graph icon** to the workspace. Drag the **Horsepower** attribute to the horizontal axis of the graph and the **FuelConsumption** attribute to the vertical axis of the graph, as shown in the screen shot at the right.



Double-click on the **Vehicle** collection box to open the **inspector**. Choose the **Measures** tab. Double-click on **<new>** and rename it **PPMC**. Right-click on the box under **Formula**, and choose **Edit Formula**. Expand **Functions/Statistical/Two Attributes**. Double-click on **correlation**. Move up to **Attributes**, and double-click on **Horsepower**, type a comma, and double-click on **FuelConsumption**. Click on **Apply** and then **OK**. The correlation coefficient of 0.353061 will appear in the **Value** column.



Example 2 Linear Regression Using Technology

Researchers monitoring the numbers of wolves and rabbits in a wildlife reserve hypothesize that the wolf population depends on the rabbit population, since wolves prey on rabbits. Over the years, the researchers collected the data shown at the right.

a) Determine the line of best fit and the correlation coefficient for these data.

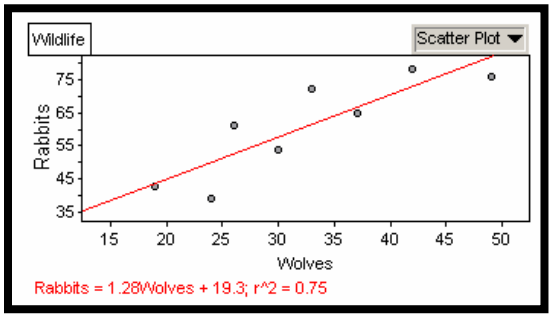
b) Graph the data and the line of best fit.

	Year	Rabbits	Wolves
1	1994	61	26
2	1995	72	33
3	1996	78	42
4	1997	76	49
5	1998	65	37
6	1999	54	30
7	2000	39	24
8	2001	43	19

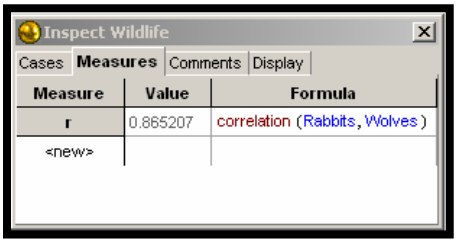
Solution 3 Using Fathom™

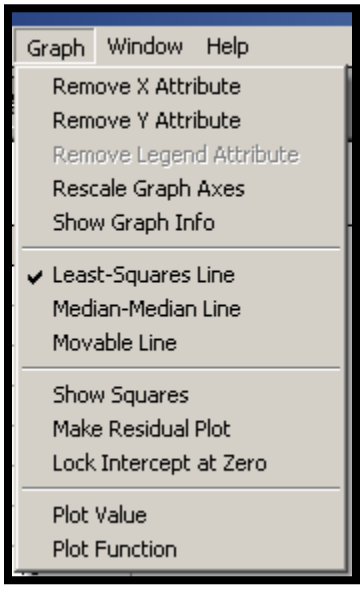
Launch Fathom™ and open a new document if necessary. Drag a new collection box to the workspace and rename it **Wildlife**. Drag a **case table** to the workspace. Rename the <**new**> column **Year**. Also create attribute columns named **Rabbits** and **Wolves**. Enter the 8 data shown in the screen shot above.

Drag a graph icon to the workspace. Drag the **Wolves** attribute to the horizontal axis of the graph and the **Rabbits** attribute to the vertical axis of the graph, as shown in the screen shot at the right. Ensure that the graph is selected. Click on the drop-down **Graph** menu, and choose **Least-Squares Line**. A line of best fit has been added to your graph. Note the equation for the line appears below the graph, as well as the coefficient of determination.



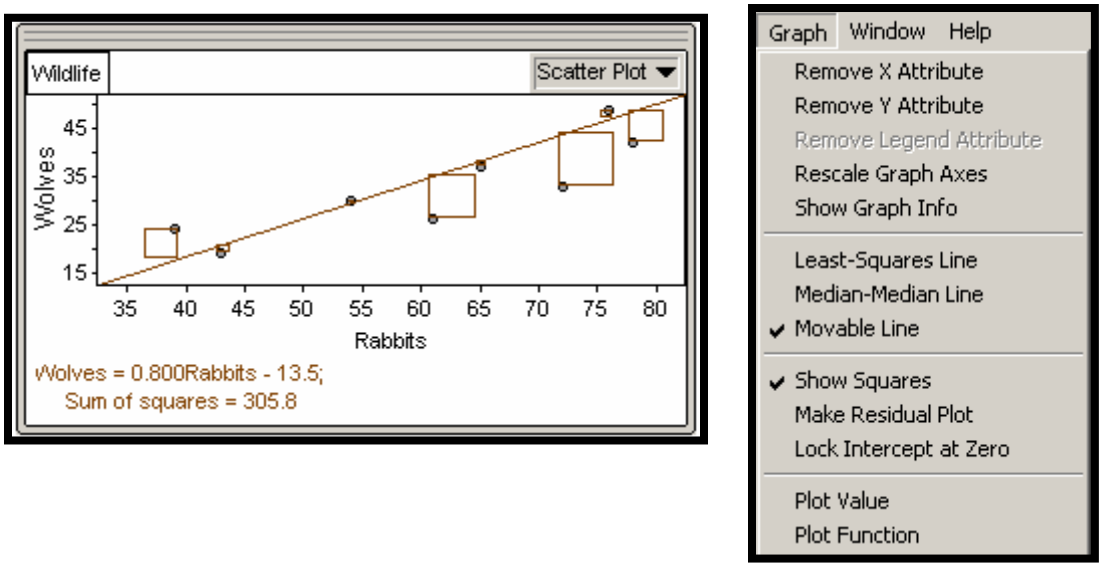
Double-click on the **Wildlife** collection box to open the **inspector**. Choose the **Measures** tab. Double-click on **<new>** and rename it **r**. Right-click on the box under **Formula**, and choose **Edit Formula**. Expand **Functions/Statistical/Two Attributes**. Double-click on **correlation**. Move up to **Attributes**, and double-click on **Rabbits**, type a comma, and double-click on **Wolves**. Click on **Apply** and then **OK**. The correlation coefficient of 0.865207 will appear in the **Value** column. Note that squaring the correlation coefficient will result in the coefficient of determination.





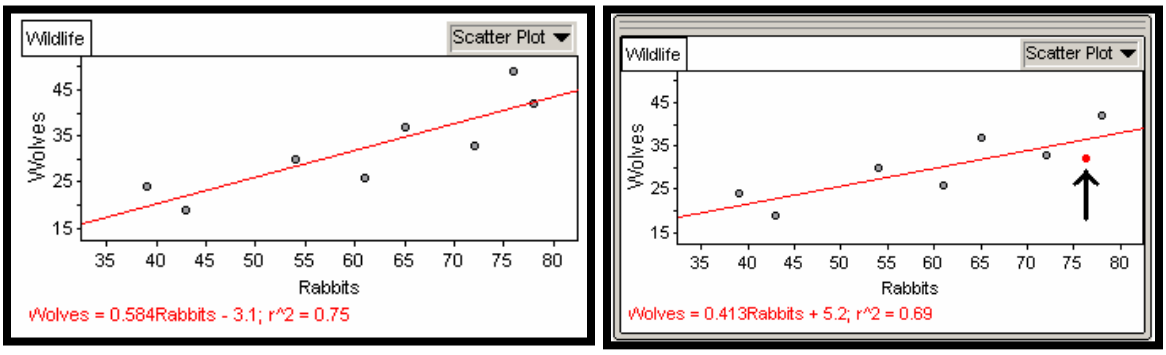
Special Topic: Movable Lines and Least Squares

As a learning tool, you can precede the automated calculation of the line of best fit with a **Movable Line**. Turn off the **Least-Squares** line, and choose **Movable Line** from the menu shown below. Also turn on **Show Squares**. You can both translate and rotate the line. As you do so, you can see the effect of line placement on the least-squares fit. The **Sum of squares** value that appears below the graph will change dynamically as you move the line, as will the equation of the line of best fit.



Special Topic: The Effect of Outliers

Fathom™ lets you drag points directly on your graph, changing the coordinates dynamically in the related **case table** as you do so. Select the second right-most point on the **Wildlife** plot, and drag it downwards, as shown below.



Notice the effect on the location and equation of the line of best fit. Note how the coordinates of this point also change in the related **case table**.

Special Topic: Non-linear Regression

Fathom™, in its current incarnation, will only perform linear regressions. However, the powerful curve-plotting and dynamic display capabilities of Fathom™ make it an excellent learning and analysis tool. With a little effort, you can set up Fathom™ to perform any kind of curve-fitting that you can define as an expression. This makes it much more powerful than most spreadsheets.

Example Quadratic Curve-Fitting

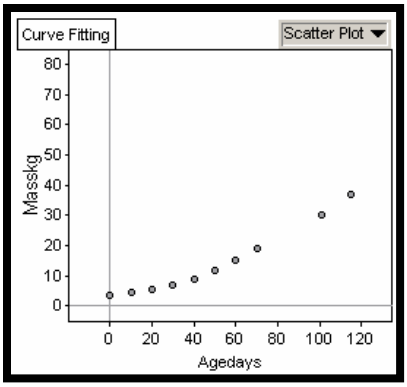
The mass in kilograms of a puppy versus its age in days is shown in the table at the right.
a) Make a graph of mass versus age.
b) Hypothesize the type of curve that might fit these data.
c) Use sliders to fit this curve.
d) Add a least-squares column to your data table, and find the sum of the squares.
e) Adjust the sliders to find the best fit according to the sum of the squares.

Curve Fitting

	Agedays	Masskg
1	0	3.25
2	10	4.25
3	20	5.5
4	30	7
5	40	9
6	50	11.5
7	60	15
8	70	19
9	101	30
10	115	37

Solution

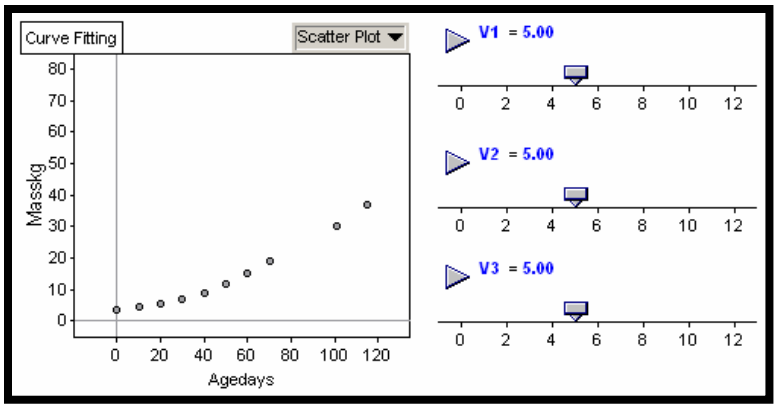
a) Launch Fathom™ and open a new document, if necessary. Drag a new collection box to the workspace and rename it **Curve Fitting**. Drag a **case table** to the workspace. Rename the <**new**> column **Agedays**. Also create an attribute column named **Masskg**. Enter the 10 data shown in the screen shot above.



Drag a **graph icon** to the workspace. Drag the **Agedays** attribute to the horizontal axis of the graph and the **Masskg** attribute to the vertical axis of the graph, as shown in the screen shot at the right.

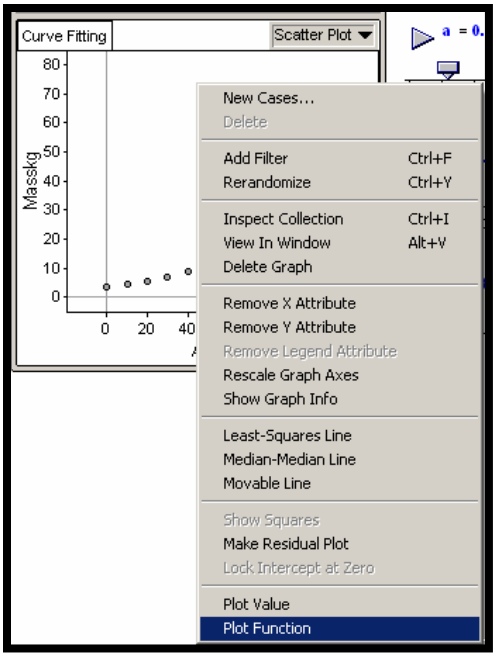
b) The simplest curve that might fit the data appears to be a quadratic.

c) The general equation of a quadratic is $y = ax^2 + bx + c$. Therefore, you will need three sliders to plot this curve such that it can be adjusted dynamically to fit the data. Drag three sliders from the tool shelf to the workspace, as shown at the right.



Note that Fathom™ will assign each slider a generic "V" name. You can double-click on each of these names in turn, and change them to *a*, *b*, and *c*.

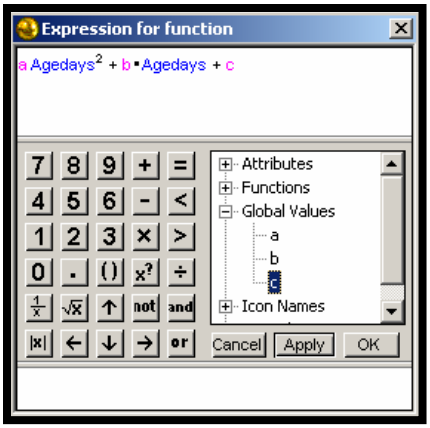
Right-click on the graph, and choose **Plot Function**. The expression dialogue box will appear.



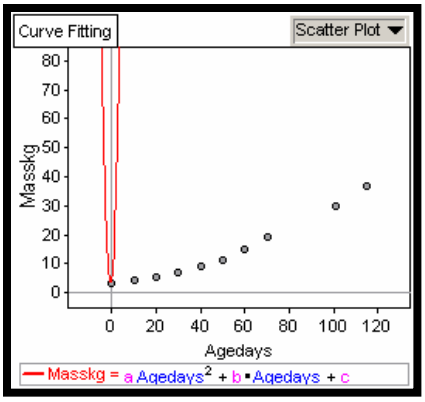
Enter the formula:

a*Agedays^2 + b*Agedays + c

You will find the sliders under **Global Values**. Press the **Apply** button and then **OK**.

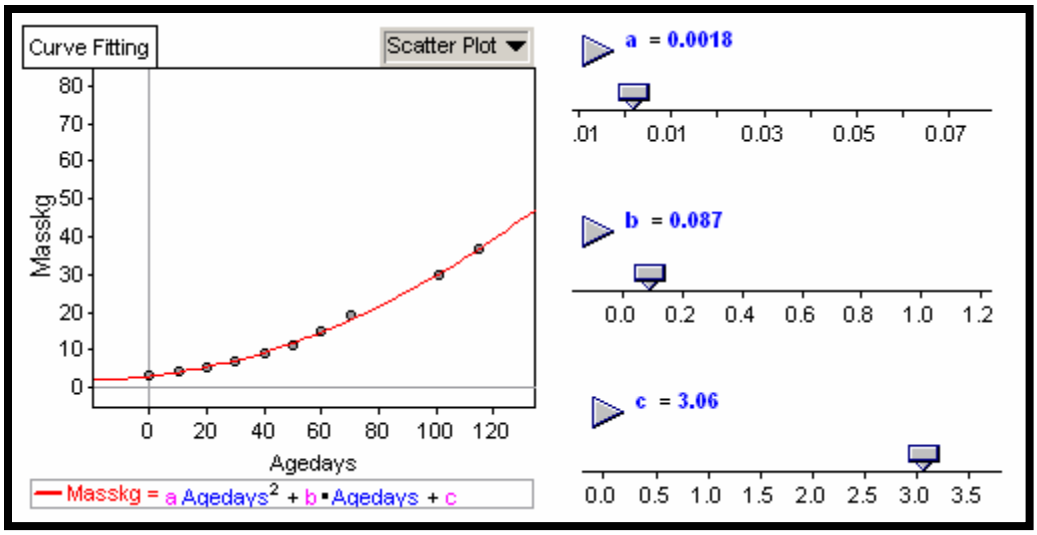


The quadratic curve will be plotted as shown in at the right. Note that it is nowhere near the scatter plot. You must now adjust the sliders in order to "fit" the curve.



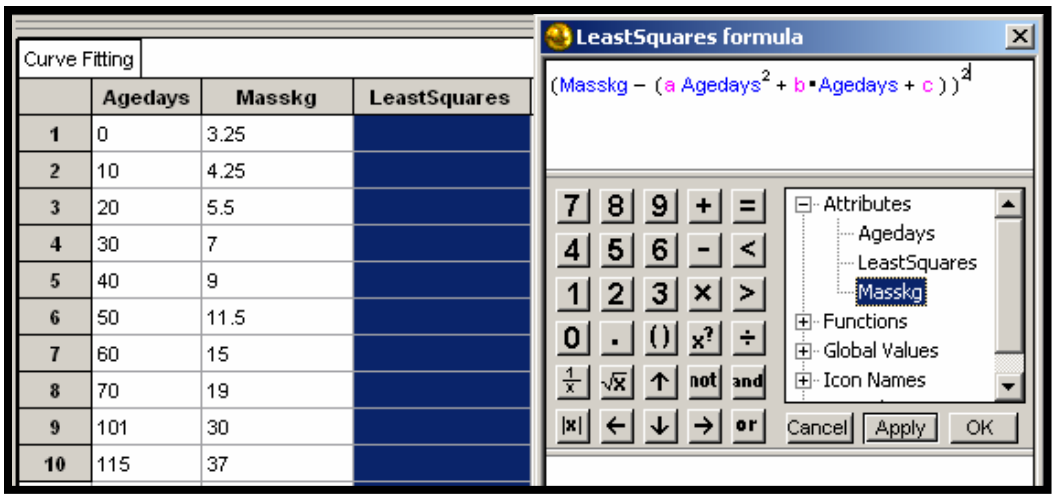
Adjust the sliders until the curve makes its best fit to the scatter plot. In most cases, it will not fit perfectly. You can drag the scale on each slider to obtain a wider or narrower range of values. When you are finished, your screen should look much like the one shown below.

Note: Take your time with this step. Adjusting the sliders and slider scales properly takes some practice. When you have the correct values, you can make small adjustments to each slider, and observe the smooth dynamic effects on the graph.



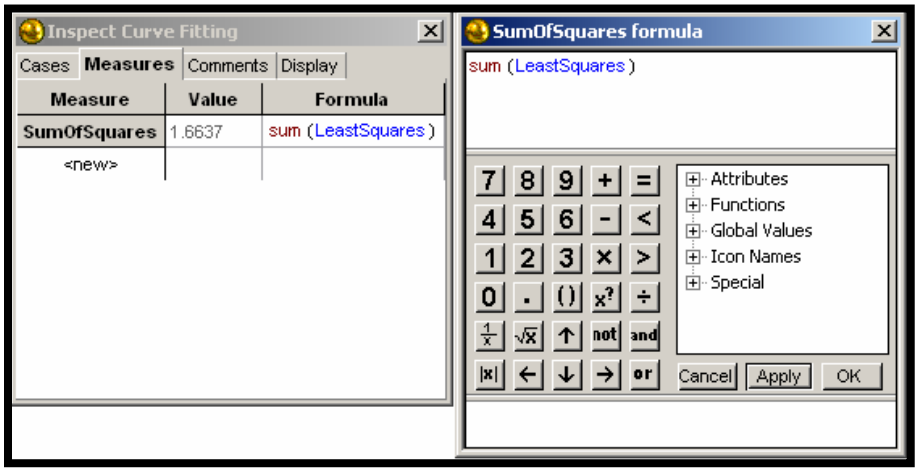
d) At this point, you have done the curve fit by "eyeball". To get a better measure of the accuracy of the fit, add a "least squares" column to your **case table**. Right-click on the **LeastSquares** attribute, and edit the formula to:

(Masskg – (a*Agedays^2 + b*Agedays + c))^2



This will calculate the square of the difference between the scatter plot data and the value predicted by your quadratic fit.

Next, calculate the sum of the squares. Double-click on the **Curve Fitting** collection box to open the **inspector**. Choose the **Measures** tab. Double-click on **<new>** and rename it **SumOfSquares**. Right-click on the box under **Formula**, and choose **Edit Formula**. Expand **Functions/Statistical/One Attribute**. Double-click on **sum**. Move up to **Attributes**, and double-click on **LeastSquares**. Click on **Apply** and then **OK**. The sum of the squares will appear in the **Value** column.



You can now move your sliders, and watch the sum of the squares change as you change your quadratic fit formula. With some practice, you can dynamically find a good approximation of the least squares fit.

Note that almost any spreadsheet will automatically generate the proper regression formula for you. On the other hand, using Fathom™ in this way offers two advantages:

1) Fathom™ is an excellent learning tool, and allows the student to see the effects of changing the parameters of the curve fit formula in real time, rather than just generating a "black box" answer.

2) Fathom™ can fit any curve that can be defined. Most spreadsheets have a limited number of curves that can be used for regression purposes.

Example 3 Detecting a Hidden Variable

	Year	Orchestras	Professionals
1	1991	10	16
2	1992	11	18
3	1993	12	20
4	1994	12	23
5	1995	14	26
6	1996	14	32
7	1997	16	13
8	1998	16	16
9	1999	18	20
10	2000	20	26

An arts council is considering whether to fund the start-up of a local youth orchestra. The council has a limited budget and knows that the number of youth orchestras in the province has been increasing. The council needs to know whether starting another youth orchestra will help the development of young musicians. One measure of the success of such programs is the number of youth-orchestra players who go on to professional orchestras. The council has collected the data shown at the right.

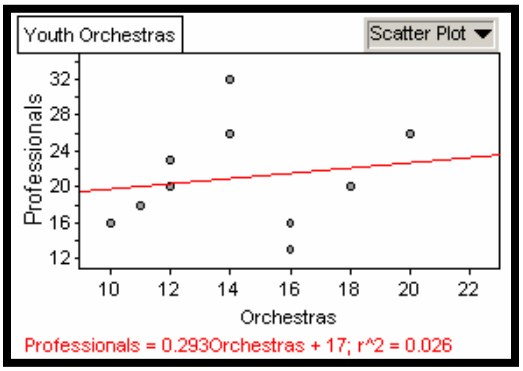
a) Does a linear regression allow you to determine whether the council should fund a new youth orchestra? Can you draw any conclusions from other analysis?

b) Suppose you discover that one of the country's professional orchestras went bankrupt in 1997. How does this information affect your analysis?

Solution Using Fathom™

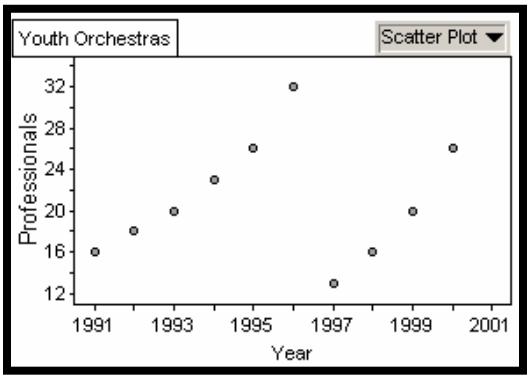
a) Launch Fathom™ and open a new document if necessary. Drag a new collection box to the workspace and rename it **Youth Orchestras**. Drag a **case table** to the workspace. Rename the <**new**> column **Year**. Also create attribute columns named **Orchestras** and **Professionals**. Enter the 10 data shown in the screen shot above.

Drag a **graph icon** to the workspace. Drag the **Orchestras** attribute to the horizontal axis of the graph and the **Professionals** attribute to the vertical axis of the graph, as shown in the screen shot at the right. Ensure that the graph is selected. Click on the drop-down **Graph** menu, and choose **Least-Squares Line**. The correlation appears weak. However, note the unusual pattern of the scatter plot. This suggests the presence of a hidden variable.



To make a time-series graph, drag a graph icon to the workspace. Drag the **Year** attribute to the horizontal axis of the graph and the **Professionals** attribute to the vertical axis of the graph, as shown in the screen shot at the right. This pattern suggests that something unusual happened in 1997.

b) The collapse of a major orchestra means both that there is one less orchestra hiring young musicians and that a large number of experienced musicians are now available for hiring. The two distinct regions of the scatter plot should be analysed separately for correlation.



Example 1 Jury Selection

a) Determine the probability distribution for the number of women on a six-person civil-court jury selected from a pool of 8 men and 10 women.

b) What is the expected number of women on the jury?

Solution 4 Using Fathom™

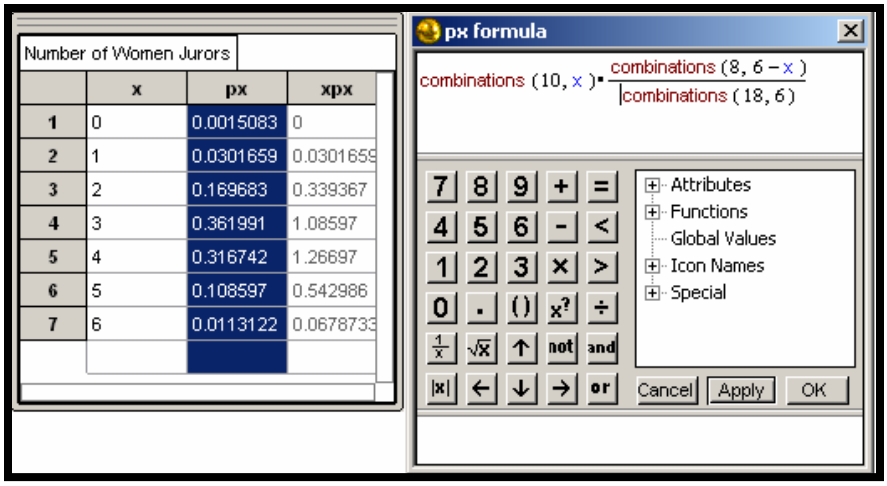
a) Launch Fathom™ and open a new document, if necessary. Drag a new collection box to the workspace and rename it **Number of Women Jurors**. Drag a **case table** to the workspace. Rename the <**new**> column **x**. Also create attribute columns named **px** and **xpx**. Since there can be from 0 to 6 women on the jury, enter integers from 0 to 6 in the **x** column.

To calculate the probability for each value of **x**, right-click on the **px** attribute, choose **Edit Formula**, and enter the following formula:

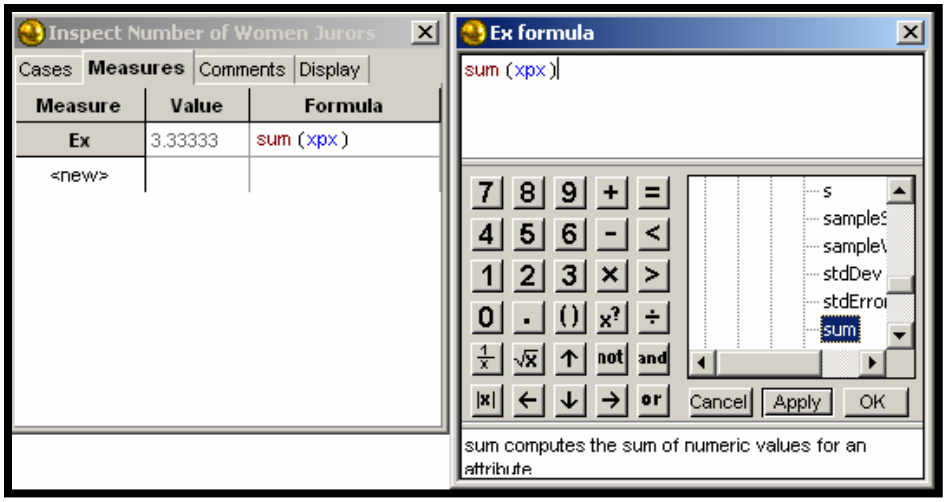
combinations(10,x)*combinations(8,6–x)/combinations(18,6)

Right-click on the **xpx** attribute, choose **Edit Formula**, and enter the formula:

x*px



b) Double-click on the **Number of Women Jurors** collection box to open the **inspector**. Choose the **Measures** tab. Double-click on **<new>** and rename it **Ex**. Right-click on the box under **Formula**, and select **Edit Formula**. Expand **Functions/Statistical/One Attribute**. Double-click on **sum**. Move up to **Attributes**, and double-click on **xpx**. Click on **Apply** and then **OK**. The expected number of women jurors of 3.33333 will appear in the **Value** column.



Technology Extension: Normal Probability Plots Page 442 MHR MDM

A toy tricycle comes with this label: "Easy-To-Assemble. An adult can complete this assembly in 20 min or less." Thirty-six adults were asked to complete the assembly of a tricycle, and record their times. Here are the results:

16	10	20	22	19	14	30	22	12	24	28	11
17	13	18	19	17	21	29	22	16	28	21	15
26	23	24	20	8	17	21	32	18	25	22	20

a) Make a normal probability plot of the data. Are the assembly times normally distributed?

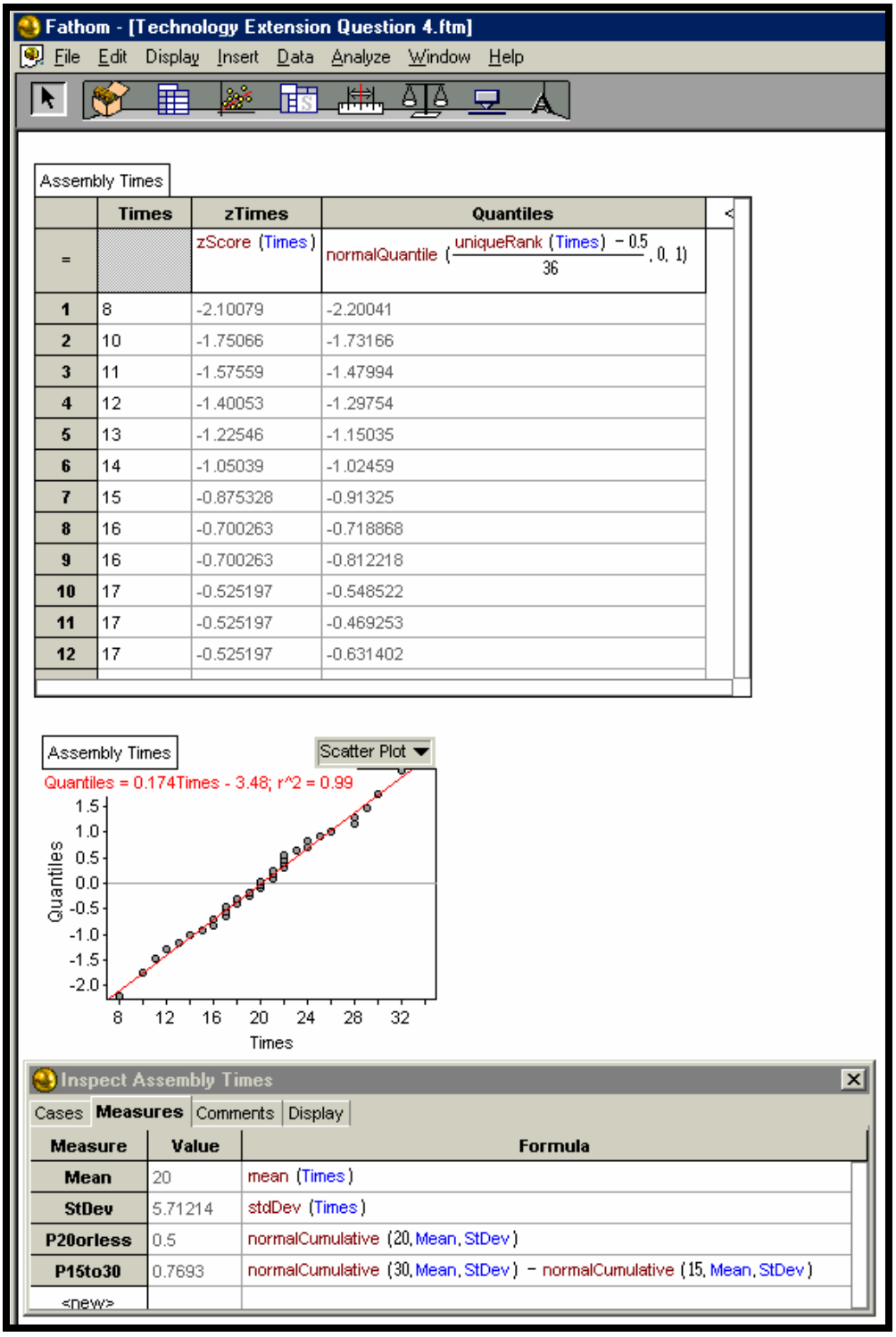
b) Find the mean and standard deviation of the data. What is the probability that an adult can complete this assembly in 20 min or less? What proportion of adults should complete this assembly within 15 to 30 min?

Solution Using Fathom™

Launch Fathom™, and open a new document, if necessary. Drag a new **collection** box to the workspace. Rename the collection **Assembly Times**, and create 36 new cases. Drag a new **case table** to the workspace. Name the first column **Times**, the second column **zTimes**, and the third column **Quantiles**.

Enter the time data in the first column. **Sort** it in ascending order. Edit the formula in the second column to **zScore(Times)**. This will calculate the *z*-scores for the data.

Edit the formula in the third column to **normalQuantile((uniqueRank(Times)**-0.5**)**/36,0,1. The 0.5 in the formula ensures that the normal quantile is calculated for the midpoint of each of the 36 segments of the distribution. This formula will calculate the expected *z*-score if the data were indeed distributed normally. The **uniqueRank()** function returns the "row number" of the sorted data. Note that most of the quantile *z*-scores in the following screen shot are different from the *z*-scores for the corresponding data. Your **case table** will look like the following screen shot.



a) Drag a new graph to the workspace. Drag the **Times** attribute to the horizontal axis, and the **Quantiles** attribute to the vertical axis to generate a normal probability plot. Choose **Least-Squares Fit** from the **Graph** menu. The linear correlation coefficient for **Times** and **Quantiles** is 0.995, indicating that the data are likely normally distributed.

b) Double-click on the **collection** to open the **inspector**. Choose the **Measures** tab. Create four measures: **Mean**, **StdDev**, **P20orLess**, and **P15to30**. Use the **mean**, **standard deviation**, and **normalCumulative** functions to calculate the mean, the standard deviation, and the answers to question 3, as shown in the previous screen shot.

This concludes this Fathom™ Tutorial. For additional Fathom™ resources, visit the Key Curriculum Press web site at www.keypress.com.