Fathom™ Tutorial

Dynamic Statistical Software

for teachers of Ontario grades 7-10 mathematics courses

by Roland W. Meisel
Version 2.1 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 wishes to use Fathom™ in teaching the concepts and meeting the technology expectations in the Ontario grades 7-10 mathematics courses. Fathom™ is an attractive alternative to using spreadsheets or graphing calculators to meet the technology expectations in these courses. The software facilitates the import of data, and its "drag and drop" approach to graphing makes graphical display a fast and easy process.

By working through this tutorial, the user will gain a solid knowledge of Fathom™ as applied to these courses, 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 scatter plots and lines of best fit, and calculating correlation coefficients
* performing linear and non-linear regressions
* performing dynamic curve-fitting using sliders

About Fathom™

Fathom™ is a powerful dynamic statistical software package published by

Key Curriculum Press®
Innovators in Mathematics Education

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

Introduction to Fathom™

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.

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.

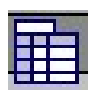
Fathom™ considers a collection as a set of **cases**. Each case in a collection can have a number of **attributes**. For example, the **cases** in a **collection** of medical records could have **attributes** such as the patient's name, age, sex, height, weight, blood pressure, and so on. There are two 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.

Entering, Displaying, Sorting, and Filtering Data

Example 1 Tables and Graphs

a) Set up a Fathom™ document for a CD collection.
b) Graph the Title and Time attributes.

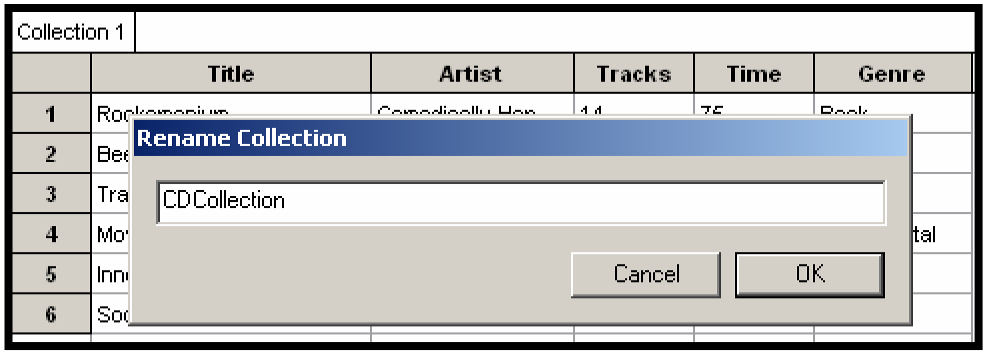
Solution

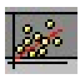
Launch Fathom™ and drag the **case table**  icon from the shelf to the workspace. Click on the attribute **<new>**, type the heading **Title**, and press **Enter**. Similarly, create attribute columns for **Artist, Tracks, Time,** and **Genre**. Enter the data into each attribute column. When you are finished, your **case table** will look like the screen shot shown below.

Collection 1

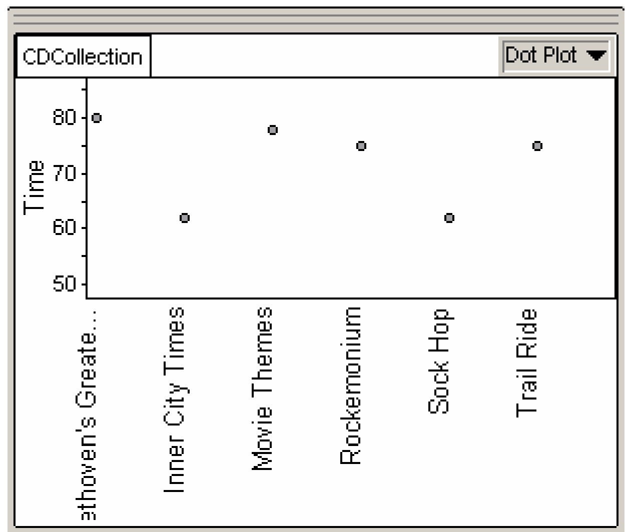
	Title	Artist	Tracks	Time	Genre
1	Rockemonium	Comedically Hep	14	75	Rock
2	Beethoven's Greatest Hits	Eastern Symphony	10	80	Classical
3	Trail Ride	Cowboy Bob	12	75	Country
4	Movie Themes	James First	18	78	Instrumental
5	Inner City Times	Ice Box	11	62	Rap
6	Sock Hop	Various	16	62	Oldies

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 **CDCollection**.



Drag the graph icon  to the workspace.

Drag the **Title** attribute from the **case table** to the horizontal axis of the graph, and the **Time** 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 CDs in Example 1 by decreasing time first, then by increasing tracks if two CDs have the same time.
b) List only those CDs with a time of less than 76 minutes.
c) Set up a separate table showing only the time and tracks data for each title. and rank the CDs by track length.

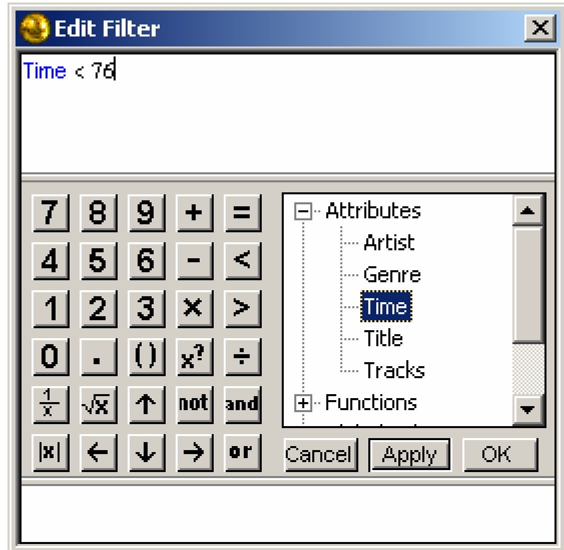
Solution

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

CDCollection

	Title	Artist	Tracks	Time	Genre
1	Beethoven's Greatest Hits	Eastern Symphony	10	80	Classical
2	Inner City Times	Ice Box	11	62	Rap
3	Trail Ride	Cowboy Bob	12	75	Country
4	Rockemonium	Comedically Hep	14	75	Rock
5	Sock Hop	Various	16	62	Oldies
6	Movie Themes	James First	18	78	Instrumental

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 **Time** attribute, choose the less-than button, and type 76. Click the **Apply** button, and then **OK**.



The results should look like the screen shot below:

CDCollection

	Title	Artist	Tracks	Time	Genre
1	Inner City Times	Ice Box	11	62	Rap
2	Trail Ride	Cowboy Bob	12	75	Country
3	Rockemonium	Comedically Hep	14	75	Rock
4	Sock Hop	Various	16	62	Oldies

Time < 76

The **Filter** is listed at the bottom as **Time** < 76.

c) Click on the **CDCollection** collection box, and drag a new **case table** onto the workspace. Click on the **Genre** attribute. Choose **Hide Attribute** from the **Display** menu. Use the same method to hide the **Artist** attribute. Right-click on the **Tracks** attribute, and use **Sort Descending** to rank the teams. Your result will look like the screen shot at the right.

CDCollection

	Title	Tracks	Time
1	Movie Themes	18	78
2	Sock Hop	16	62
3	Rockemonium	14	75
4	Trail Ride	12	75
5	Inner City Times	11	62
6	Beethoven's Greatest Hits	10	80

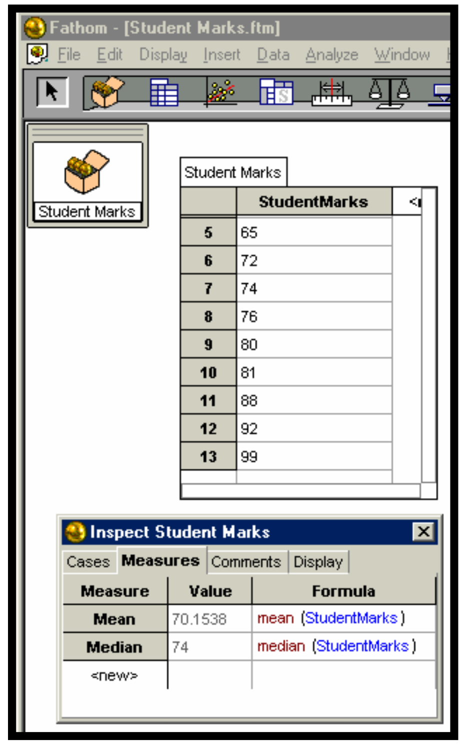
Practice Question

1. 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) 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.

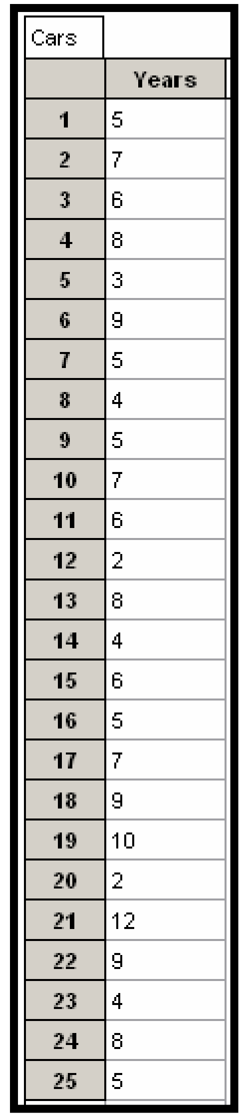
Special Topic: Measures of Spread

Example 1 Using Technology to Calculate Standard Deviations

An automotive engineer has collected data on the number of years a particular model of car spends on the road before it wears out, for 25 cars. Determine the mean, standard deviation, and the variance for these data.

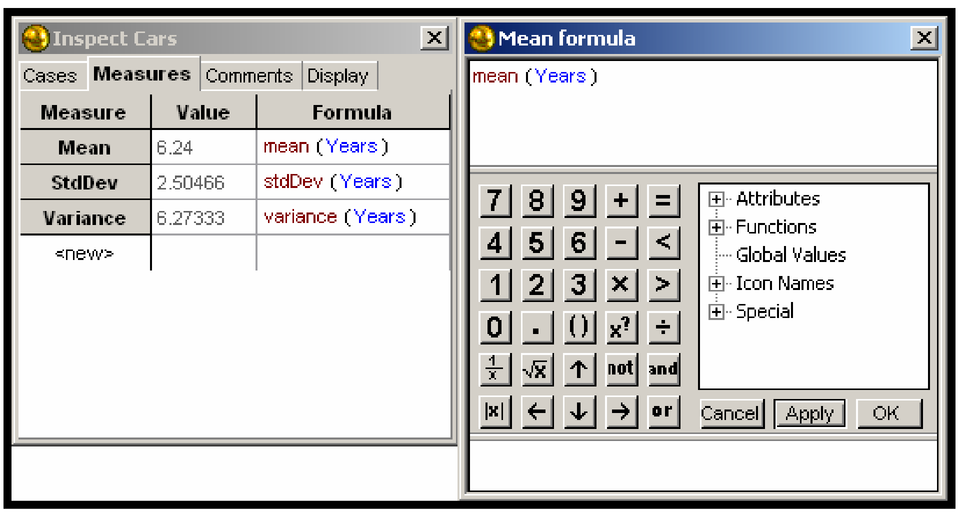
Solution 1 Using Fathom™

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


Cars

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

Double-click on the **Cars** 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 **Years**. Click on **Apply** and then **OK**. The mean of 6.24 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, 2.50466 and 6.27333, as shown in the screen shot below.



Example 2 Determining Quartiles and Interquartile Ranges

A random survey of people attending a motorcycle show asked them how much money they had spent on accessories over the previous year, to the nearest $50. 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.

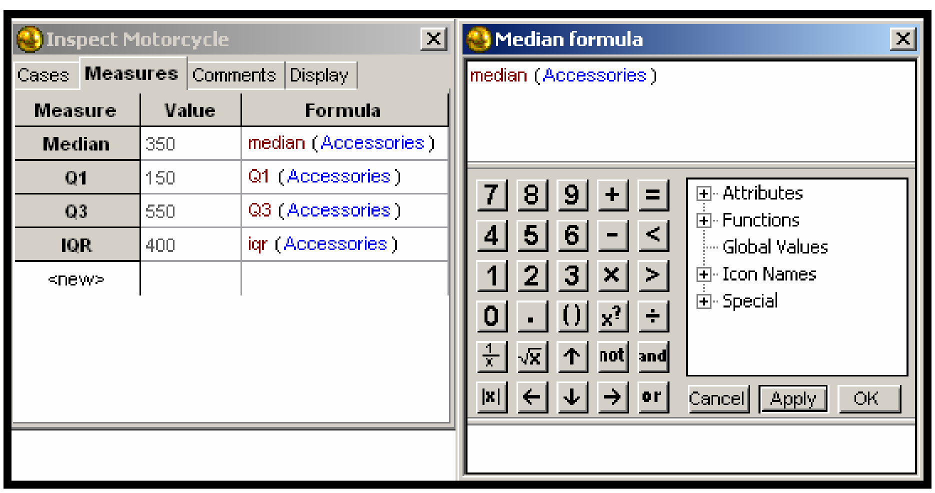
Motorcycle

	Accessories
1	200
2	150
3	350
4	500
5	100
6	50
7	450
8	300
9	200
10	150
11	650
12	700
13	1200
14	850
15	150
16	100
17	350
18	600
19	450
20	500

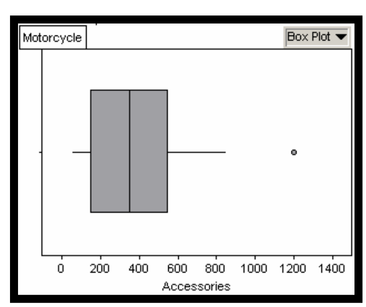
Solution 2 Using Fathom™

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

Double-click on the **Motorcycle** 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 **Accessories**. Click on **Apply** and then **OK**. The median of 350 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, 150, 550 and 400, as shown in the screen shot below. The semi-interquartile range is half of the interquartile range, or 200.



b) Drag a **graph icon** to the workspace. Drag the **Accessories** 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.



Scatter Plots and Linear Correlation

Example 1 Using Technology to Determine Correlation Coefficients

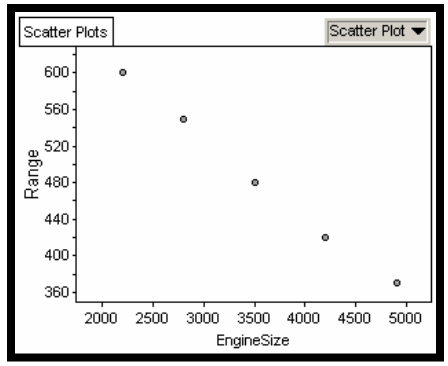
A model of truck has a choice of five engines according to their displacement in cubic centimeters. Each engine results in a different range using the fuel in the 60 litre tank. Determine whether there is a linear correlation between displacement and range for these five engines by creating a scatter plot and calculating the correlation coefficient.

Scatter Plots

	EngineSize	Range
1	2200	600
2	2800	550
3	3500	480
4	4200	420
5	4900	370

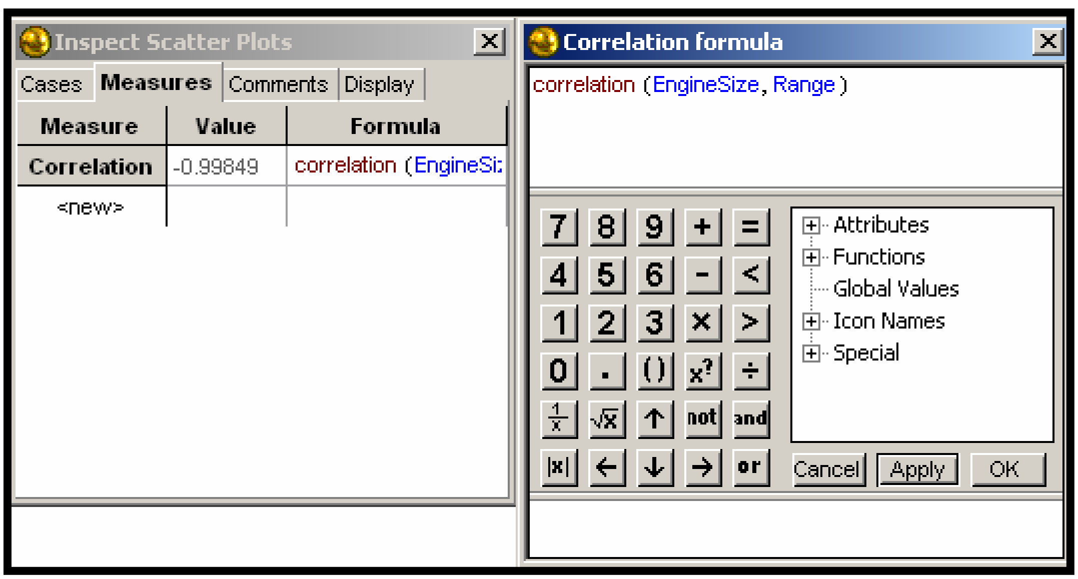
Solution 1   Using Fathom™

Launch Fathom™ and open a new document if necessary. Drag a new collection box to the workspace and rename it **Scatter Plots**. Drag a **case table** to the workspace. Rename the <**new**> column **EngineSize.** Also create an attribute column named **Range**. Enter the data shown in the screen shot.



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

Double-click on the **Scatter Plots** collection box to open the **inspector**. Choose the **Measures** tab. Double-click on **<new>** and rename it **Correlation**. 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 **EngineSize**, type a comma, and double-click on **Range**. Click on **Apply** and then **OK**. The correlation coefficient of -0.99849 will appear in the **Value** column.



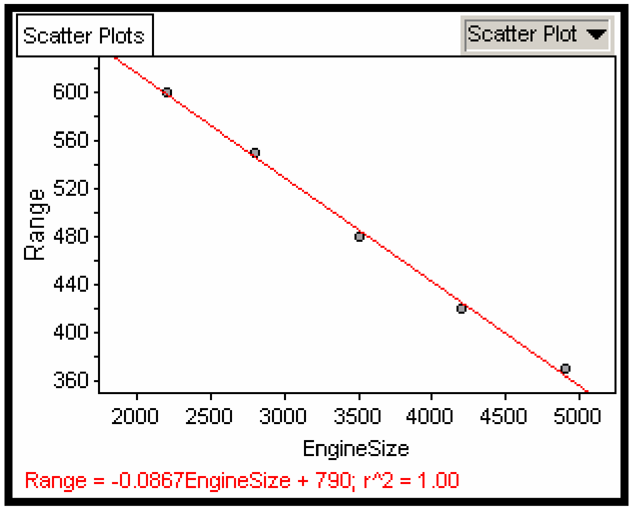
Line of Best Fit

Example 1 Linear Regression Using Technology

Using the automobile data from the previous example, graph the data and the line of best fit.

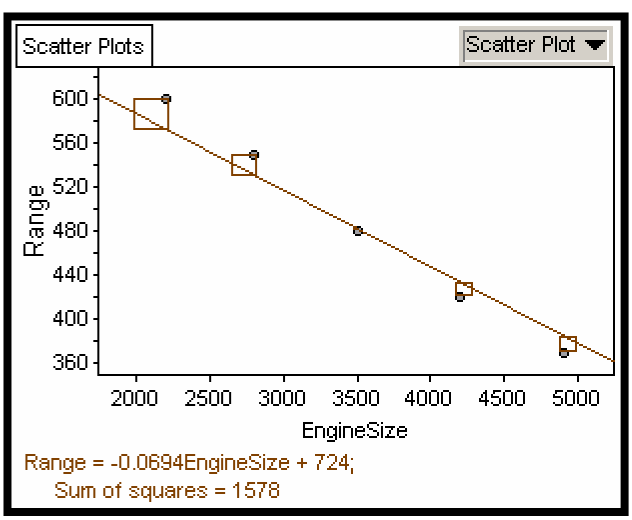
Solution 1 Using Fathom™

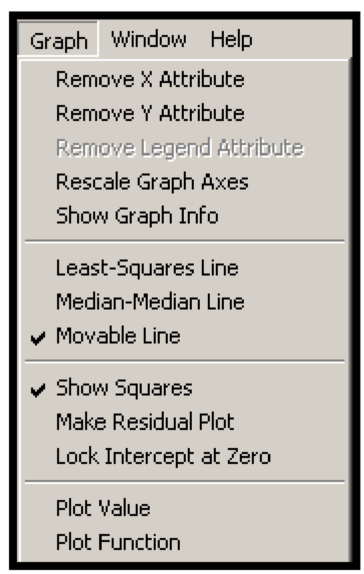
Launch Fathom™ and enter the automobile data from the previous example, if necessary. Construct a scatter plot with engine size on the horizontal axis, and range on the vertical axis. Ensure that the graph is selected. Click on the drop-down **Graph** menu, and choose **Least-Squares Line**. A line of 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, which is the square of the correlation coefficient.



Special Topic: Movable Lines and Least Squares

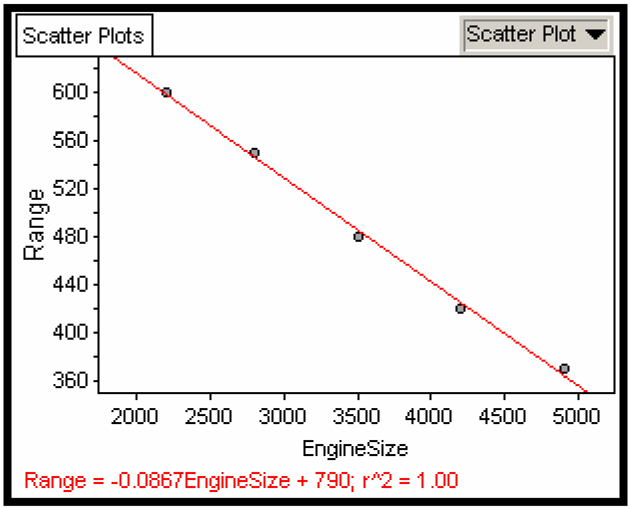
As a learning tool, you can precede the automated calculation of the line of best fit with a **Movable Line**. Turn off the **Least-Squares** line, and choose **Movable Line** from the menu shown below. Also turn on **Show Squares**. You can both translate and rotate the line. As you do so, you can see 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.

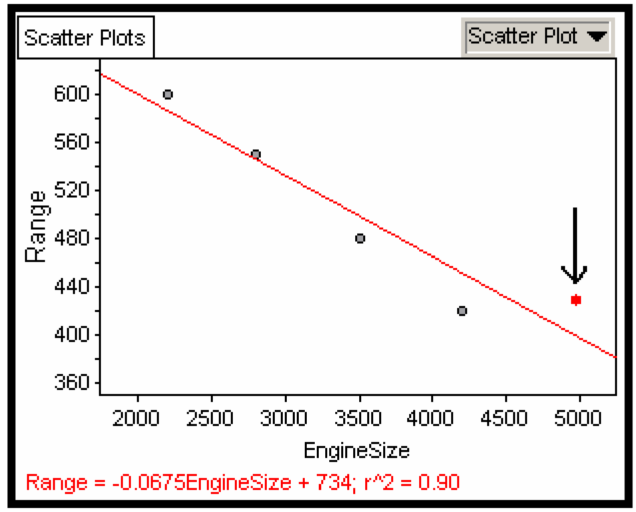




Special Topic: The Effect of Outliers

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

Quadratic Curve Fitting

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. Here is an example of quadratic curve-fitting applicable to the grade 10 course.

Example 1 Quadratic Curve-Fitting

The position of a roller coaster in metres at times measured in seconds is shown at the right.
a) Make a graph of position versus time.
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.

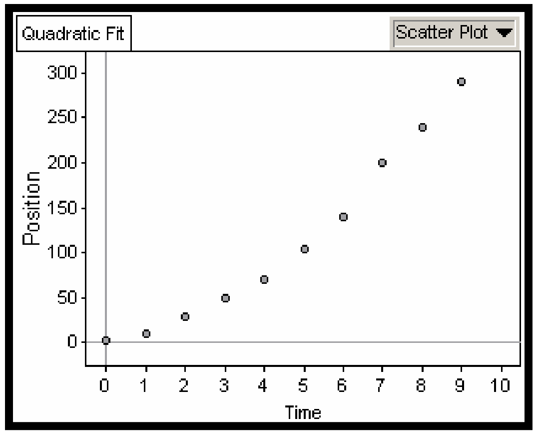
Quadratic Fit

	Position	Time
1	3	0
2	10	1
3	29	2
4	50	3
5	70	4
6	104	5
7	140	6
8	200	7
9	240	8
10	290	9

Solution 1

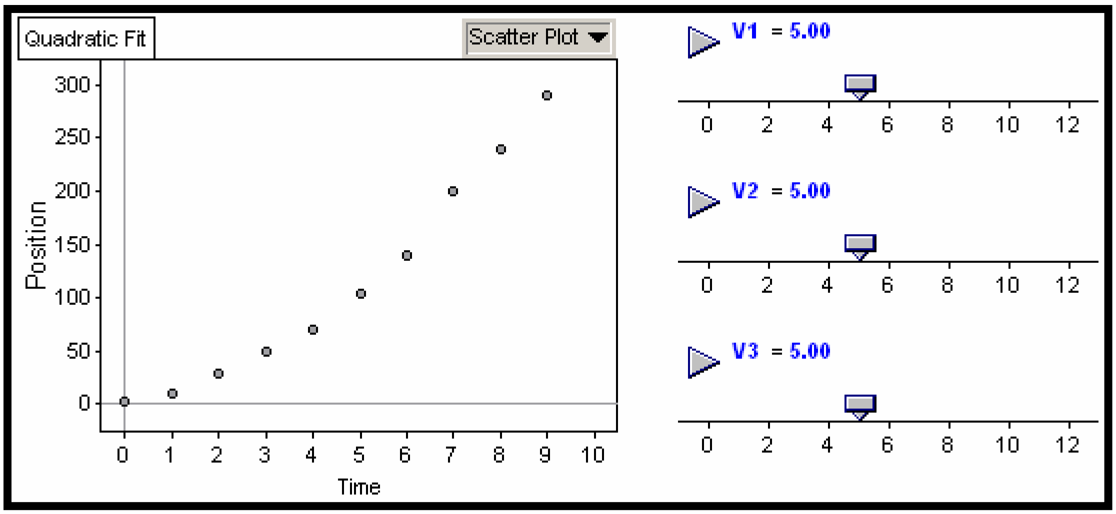
a) Launch Fathom™ and open a new document, if necessary. Drag a new collection box to the workspace and rename it **Quadratic Fit**. Drag a **case table** to the workspace. Rename the <**new**> column **Position**. Also create an attribute column named **Time**. Enter the data shown in the screen shot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

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



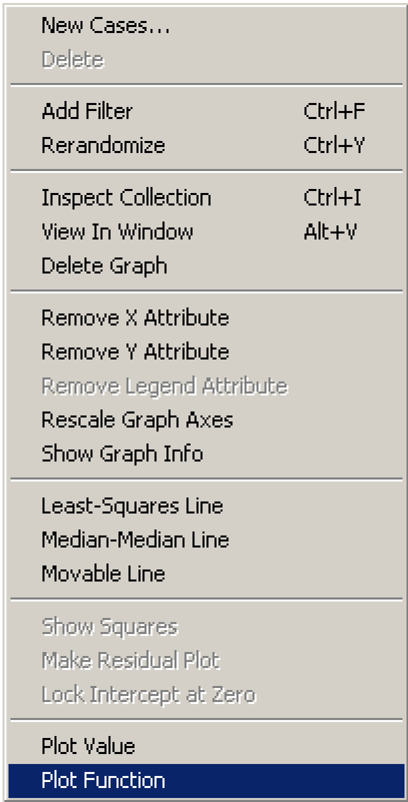
b) The simplest curve that might fit the data appears to be a quadratic.

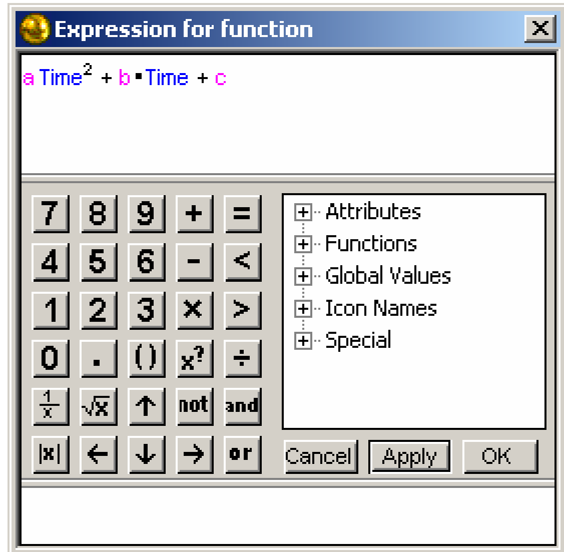
c) The general equation of a quadratic is $y = ax^2 + bx + c$. Therefore, you will need three sliders to plot this curve such that it can be adjusted dynamically to fit the data. Drag three sliders from the tool shelf to the workspace, as shown below.



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.

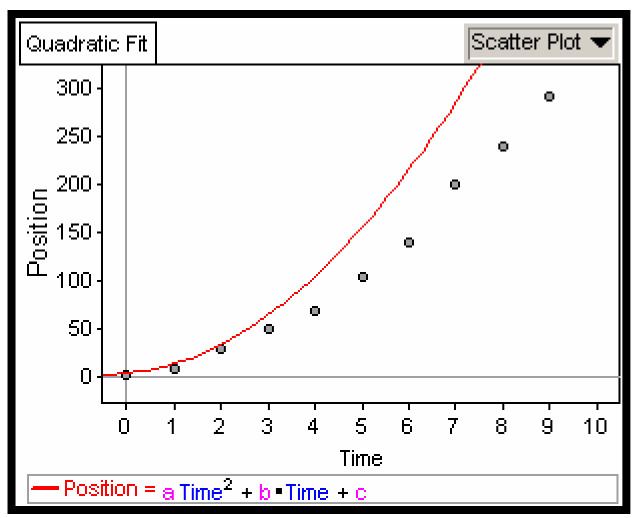




Enter the formula:

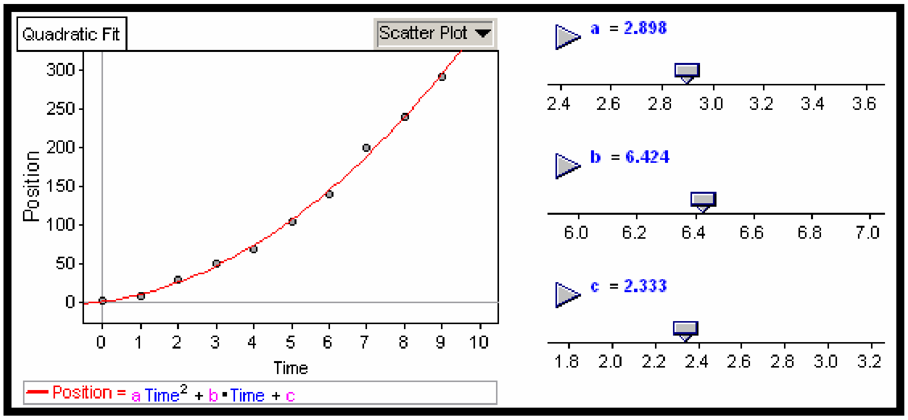
a*Time^2 + b*Time + c

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.

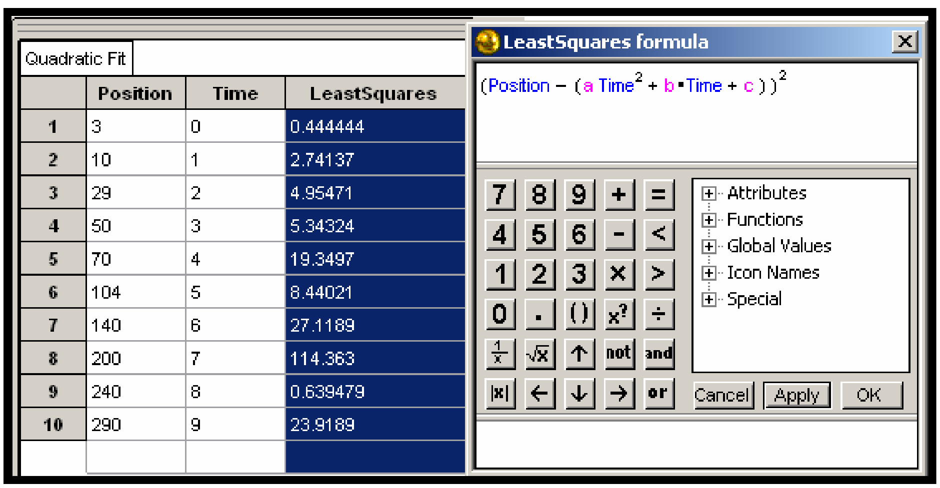


Special Topic: Quadratic Regression

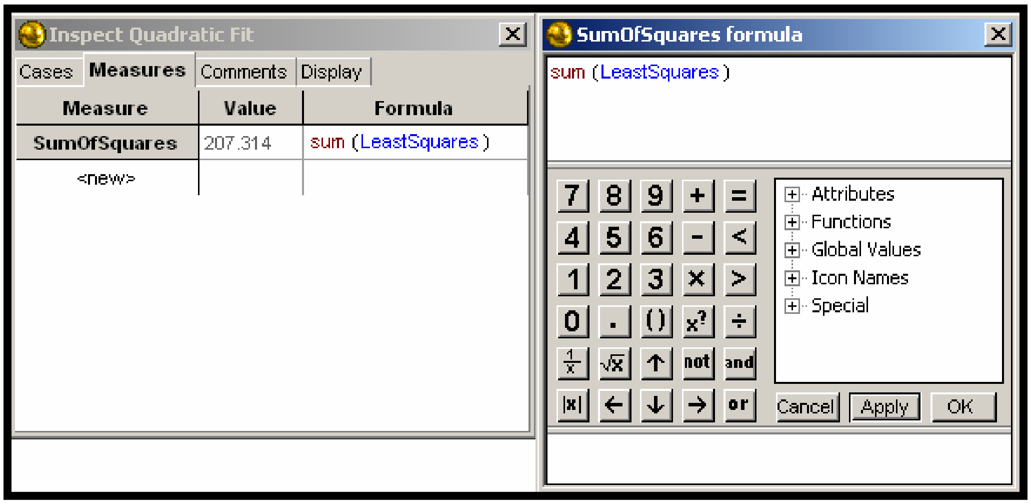
d) At this point, you have done the curve fit by "eyeball". To get a better measure of the accuracy of the fit, add a "least squares" column to your **case table**. Right-click on the **LeastSquares** attribute, and edit the formula to:

(Position – (a*Time^2 + b*Time + 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 **Quadratic Fit** 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.

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