M1 Maths

A5-7 Model Functions

- finding model functions for data using technology

Summary Learn Solve Revise Answers

Summary

If a relation shows a pattern, we can express it as a formula. Sometimes a relation shows an approximation to a pattern, though with some deviation from the actual pattern. In this case we can write a formula for the relation that it approximates. This is called a model relation. As such relations are generally functions, we more often call them model functions.

If we enter data as a set of ordered pairs into a graphics calculator or spreadsheet, it can find a formula for the model function of the type we specify. We can specify a linear function, a quadratic function, an exponential function or a function of various other types.

Along with the formula, it will give a value for r, the correlation coefficient, or r^2. These indicate how close the data is to the model function.

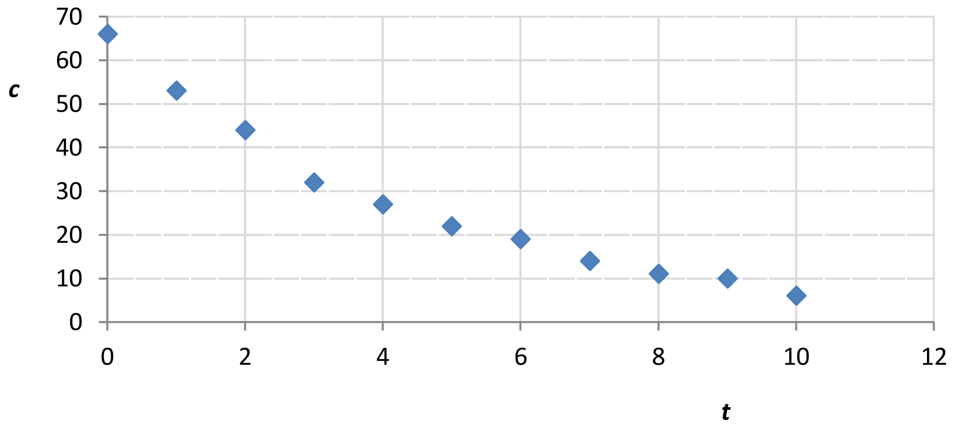
To choose the most appropriate function type (linear, power or whatever), for each type, we consider three criteria: the value for r^2; the likelihood that the situation the data came from would produce that type of relation; and whether the model function would produce appropriate values of the dependent variable for values of the independent variable higher or lower than in our data.

Learn

When you make a cup of coffee, it's hot. If you let it sit for a while, it gets cooler, eventually approaching room temperature. If we recorded the temperature difference between the coffee and the room every minute, we might get a relation something like the one in the table below. (c is the difference between the temperature of the coffee and room temperature, and t is the time since making the coffee.)

t (minutes)	0	1	2	3	4	5	6	7	8	9	10
c (degrees)	66	53	44	32	27	21	19	14	11	10	6

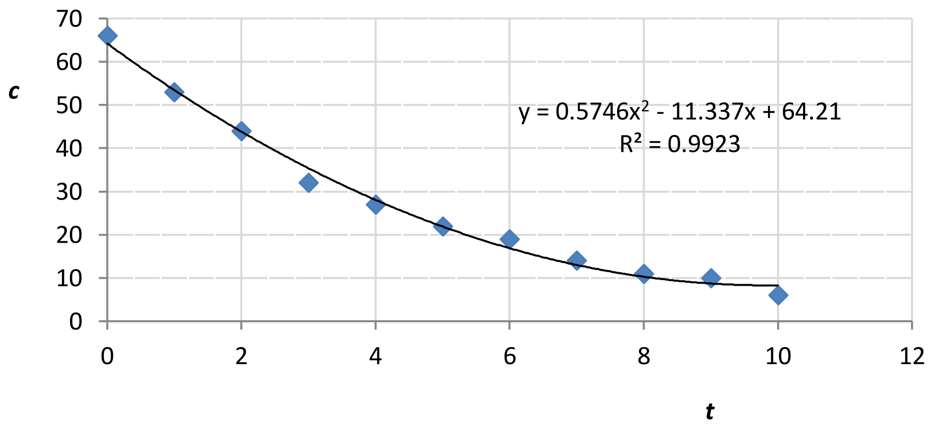
As a graph, the relation would look like this:



The graph seems to show a pattern, though with a little bit of random variation superimposed – possibly from inaccuracies in the measurements taken. If it showed an exact pattern, we could find a formula to represent it. But it is also possible to find a formula to represent the pattern even when the data doesn't fit it exactly. We find a formula for a function whose graph passes close to the data points. Such a formula is called a model function because it isn't exactly the same function as the data, but it is similar, does have its most important features and could be used to make predictions about further value pairs.

Graphics calculators and spreadsheets can produce model functions for data. They can present the model functions both as formulae and as graphs superimposed on the data graph. You choose the type of function (linear, quadratic, cubic, power, exponential etc.), then they adjust the parameters of the chosen function type to give the best fit to the data.

If we choose a quadratic function in Excel we get $y = 0.5746x^2 - 11.337x + 64.21$. This function superimposed on the data graph is shown below.



The R^2 value indicates how close the data is to the values given by the model function. In other words it shows how well the model function fits the data. A value of 1 means a perfect fit; the lower the value, the worse the fit. 0.9923 is quite a good fit. R (or *r*) is the correlation coefficient you met with linear functions in Module S5-1.

To get a model function with Excel, make a table of the data, then select it and insert a scatter graph without lines. Then right click on a data point on the graph and select 'Add Trendline…'

You will then get a dialogue box. Choose the type of function you want, tick the 'Display Equation on chart' and the 'Display R^2 value on chart' boxes, then select 'Close'.

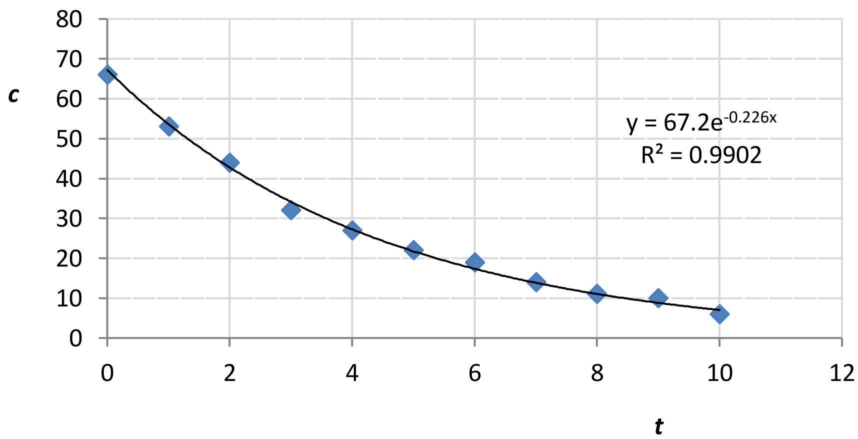
To get a model function with a Casio CFX-9850GB PLUS, go into STAT mode (2 on the main menu). Then put the times (values for t) in List 1 and the corresponding temperature differences (values for c) in List 2.

Then press GRPH (F1). Then press SET (F6) and check that Graph 1 is set to Scatter with XList being List1 and YList being List2. Frequency should be 1. Then press EXIT. Then press GRPH1 and you should see a scatter graph of your data.

Along the bottom of the screen will be a number of model functions types. Select X^2, which stands for quadratic. If you press DRAW (F6), the calculator will draw the model function superimposed on the scatter plot of your data. This allows you to see how good the fit is.

Choosing the Type of Function

We chose a quadratic function for the coffee and it gave quite a good fit with an R^2 of 0.9923. But there are other function types which might have given a good fit, for instance a higher-degree polynomial like a cubic or an exponential function. If we chose the exponential function, we would have got the following:



[Note the use of e for the base. You should remember from Module A5-4 on exponential functions that e is approximately 2.718281828.]

The exponential function is also a good fit. How do we decide which type is the most appropriate. There are three things to consider.

The first is which gives the best R^2 value. The exponential function gave a slightly lower R^2 than the quadratic, so based on this, the quadratic is better.

But the second thing to consider is which function type is more likely to represent the situation. We might know something about the physics of cooling that would make one function type more likely than another. In fact, the physics of the situation tells us that the cooling would be exponential (the temperature difference is multiplied by the same fraction each minute or the rate of decrease of c is proportional to c).

The third thing to consider is how the different function families would behave if we extrapolated to greater values of the independent variable, i.e. to greater values of t (like if we kept the observations going to say $t = 30$). The quadratic function, being a parabola, would start to increase again as time went on. This would not be expected to happen in reality, making the quadratic a somewhat dubious choice. On the other hand, the exponential function graph would continue to get closer and closer to $c = 0$ as t increased, exactly as we would expect for the cooling coffee.

Taking these three criteria into consideration, on balance, the exponential choice would seem to be the best one. So we would conclude that the best model function is $y = 67.2e^{-0.226x}$, or, using c and t, $c = 67.2e^{-0.226t}$.

Practice

Q1 Use a spreadsheet to find model functions for the following data. Try polynomials of degree 1, 2, 3 and 4, a power function and an exponential function. For each of these six types, give the function and the R^2 value. Then decide which is the best fit.

x	1	2	3	4	5	6	7
temperature (*T*)	48	38	35	29	24	22	18

Q2 Repeat Q1 using a graphics calculator, making note of any differences in the results.

Q3 The following table is measurements of the difference between the temperature of a dead cat and room temperature over a period of 3.5 hours. Using a spreadsheet or a calculator, find the best linear, quadratic, cubic, power and exponential functions to represent the data, give each function along with its R^2 value, then explain, using the three criteria, which you would consider the most appropriate.

time (*t*) in hours	0	0.5	1	1.5	2	2.5	3	3.5
temperature difference, c	28	23	18	14	11	8	6	5

Q4 The table below shows, for different time periods, the record amount of rain recorded within that time interval anywhere on Earth. By trying different function types and using the three criteria, decide on the most appropriate model of the relation between time interval and rain amount.

Period (*t*) (min)	1	5	10	30	60	120	240	720	1440
rainfall (*r*) (mm)	38	82	121	188	305	445	703	1144	1825

Q5 The table below shows the height of a cannon ball over the 9 seconds that it is in the air. Find the best model function for the data.

Time (s)	0	1	2	3	4	5	6	7	8	9
Height (m)	0	41	71	91	100	99	89	70	42	3

Q6 The table below shows the height of a rocket over the 20 seconds of its flight. Find the best model function for the data.

Time (s)	1	3	5	7	9	11	13	15	17	19	21
Height (m)	3	7	17	28	45	61	70	71	62	44	13

Note that the process of finding a model function for data is sometimes called performing a regression on the data or doing a regression analysis.

Solve

Q51 The relation between the mass (m) in milligrams of a bacterial colony and time (t) in hours can best be modelled by the function $m = 21e^{0.32t}$. The R^2 value is 0.92. Draw a possible graph for the relation between $t = 0$ and $t = 10$ at 1 hour intervals.

Q52 Pick 6 value pairs, then do a regression on them using polynomials of degree2, 3, 4, 5 and 6, noting the R^2 value each time. What do you notice?

Revise

Revision Set 1

Q61 The following table is measurements of the rate of heat loss from a bottle of hot water and the temperature difference between the bottle and the surrounding air.

temperature difference (t)	10	20	30	40	50	60	70	80
rate of heat loss (h)	4.5	10.5	17.9	25.0	32.9	43.1	51.0	60.6

Using a spreadsheet or a calculator, find the best linear, quadratic, cubic, power and exponential functions to represent the data, give each function along with its R^2 value, then explain, using the three criteria, which you would consider the most appropriate.

Revision Set 2

Q71 The following table shows the takings per week on a toll road for various toll charges. Note that if the toll charge is very low, not much money will be taken. On the other hand, if the toll charge is very high, few people will use the road and, again, not much will be taken.

toll charge (c) (\$)	0.1	1	2	3	4	5	6	7
weekly takings (t) (\$1000s)	4	12	20	26	27	22	14	5

Using a spreadsheet or a calculator, find the best linear, quadratic, cubic, power and exponential functions to represent the data, give each function along with its R^2 value, then explain, using the three criteria, which you would consider the most appropriate.

Revision Set 3

Q81 The following table shows the value of a house for different numbers of years after it was built.

time (t) in years	0	2	5	6	10	15	18	24	28
value (v) in \$1000s	132	145	167	172	213	273	315	410	498

Using a spreadsheet or a calculator, find the best linear, quadratic, cubic, power and exponential functions to represent the data, give each function along with its R^2 value, then explain, using the three criteria, which you would consider the most appropriate.

Answers

Q1 Linear y = y = -4.75x + 49.571 $R^2 = 0.9635$
Quadratic $y = 0.4405x^2 - 8.2738x + 54.857$ $R^2 = 0.9883$
Cubic $y = -0.0833x^3 + 1.4405x^2 - 11.69x + 57.857$ $R^2 = 0.9906$
Quartic $y = 0.0417x^4 - 0.75x^3 + 5.0417x^2 - 19.167x + 62.571$ $R^2 = 0.9918$
Power $y = 52.343x^{-0.483}$ $R^2 = 0.9335$
Exponential $y = 54.628e^{-0.158x}$ $R^2 = 0.9909$
On the basis of this, the quartic is the best fit.

Q2 Should be the same as in P2.

Q3 Linear $y = -6.6429x + 25.75$ $R^2 = 0.9595$
Quadratic $y = 1.3571x^2 - 11.393x + 28.125$ $R^2 = 0.9996$
Cubic $y = 0.0606x^3 + 1.039x^2 - 10.976x + 28.045$ $R^2 = 0.9997$
Power You will note that model power functions cannot be obtained if one of the value pairs has 0 as the independent variable. Can you think why?
Exponential $y = 29.321e^{-0.511x}$ $R^2 = 0.9967$
The cubic gives the best R^2, but an exponential function would be more expected. Also, the cubic will start to rise to infinity after more time, something that is not likely to happen to the cat. So the exponential is the best choice.

Q4 If time is measured in minutes,
Linear $y = 1.2006x + 188.7$ $R^2 = 0.9561$
Quadratic $y = -0.0005x^2 + 1.948x + 133$ $R^2 = 0.9809$
Cubic $y = 2E\text{-}06x^3 - 0.0042x^2 + 3.6059x + 71.376$ $R^2 = 0.9989$
Power $y = 35.085x^{0.5343}$ $R^2 = 0.9972$
Exponential $y = 159.42e^{0.002x}$ $R^2 = 0.6133$
The cubic gives the best R^2, but, with a cubic, the pattern of rising with a decreasing gradient won't continue; in fact for long periods of time, the function would predict negative rainfalls. The power function has a similarly high R^2, but will continue to rise with decreasing gradient. This should therefore be the chosen function.

Q5 Linear $y = 4.9833x + 47.067$ $R^2 = 0.1658$
Quadratic $y = -4.9318x^2 + 44.438x + 1.0364$ $R^2 = 0.9993$
Cubic $y = 0.0665x^3 - 5.7298x^2 + 46.845x - 0.0808$ $R^2 = 1$
Power Not possible
Exponential Cannot be obtained because the height is 0 at $t = 0$.
Of those obtained, the cubic gives the best R^2, but will predict that the cannon ball will rise again at a later time. The quadratic is therefore the best choice.

Q6 Linear $y = 2.0182x + 16.073$ $R^2 = 0.2693$
Quadratic $y = -0.5163x^2 + 13.377x - 25.749$ $R^2 = 0.8193$
Cubic $y = -0.0547x^3 + 1.2874x^2 - 2.5718x + 4.1922$ $R^2 = 0.9968$
Power $y = 3.7589x^{0.9231}$ $R^2 = 0.6527$
Exponential $y = 9.3906e^{0.0949x}$ $R^2 = 0.3569$
The cubic function has by far the best R^2 and is OK with the other two criteria.

Q61 Linear $y = 0.808x - 5.6714$ $R^2 = 0.995$
Quadratic $y = 0.0027x^2 + 0.5685x - 1.6804$ $R^2 = 0.9994$
Cubic $y = -3E\text{-}05x^3 + 0.006x^2 + 0.4398x - 0.4429$ $R^2 = 0.9995$
Power $y = 0.2493x^{1.2529}$ $R^2 = 0.9997$
Exponential $y = 4.9638e^{0.0345x}$ $R^2 = 0.9215$
The power function has the best R^2, though not by much. It is ok with criteria 2 and 3.

Q71 Linear $y = 0.2608x + 15.334$ $R^2 = 0.005$
Quadratic $y = -1.8567x^2 + 13.395x + 1.6945$ $R^2 = 0.9853$
Cubic y = y = $-0.0251x^3 - 1.5889x^2 + 12.67x + 2.0172$ $R^2 = 0.9859$

Power $y = 10.839x^{0.2724}$ $R^2 = 0.2678$
Exponential $y = 12.09e^{0.0292x}$ $R^2 = 0.0093$
The cubic and quadratic have very similar R^2. Neither would be appropriate outside the domain of the data, so we should go for the simper function – the quadratic.

Q81 Linear $y = 12.771x + 105.08$ $R^2 = 0.9722$
Quadratic $y = 0.2819x^2 + 5.0021x + 133.35$ $R^2 = 0.9996$
Cubic $y = 0.0026x^3 + 0.1723x^2 + 6.1391x + 131.44$ $R^2 = 0.9997$
Power Not possible
Exponential $y = 131.68e^{0.0477x}$ $R^2 = 0.9995$
The quadratic, cubic and exponential functions all have very high R^2. But in this situation, one would expect an exponential function. So the exponential is the most appropriate.