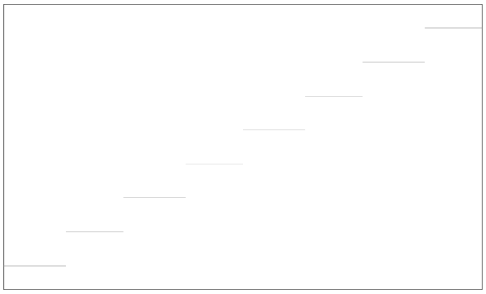
Patterns of Change

When we have information about a relationship between two or more variables, one of the best ways to present that information is with coordinate graphs of (x, y) data pairs. Such graphs may display specific pairs of related numerical values, a line or curve representing the general relationship between the x and y values, or both.

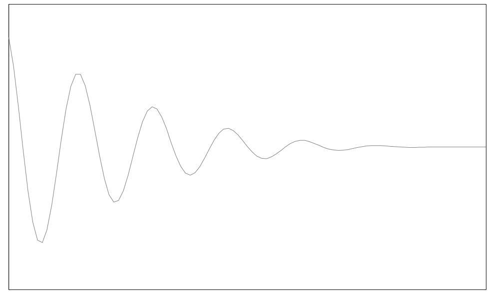
The following statements describe nine different situations in which two variables are (or at least seem to be) related to each other. Match each situation to the graph that you believe is most likely to represent the relation between those variables. Then explain as carefully as you can what the shape of the graph tells about the ways the variables change in relation to each other.

1. When a tennis player hits a high lob shot, its height changes as time passes. What pattern seems likely to relate time and height?
2. The senior class officers at Lincoln High School decided to order and sell souvenir baseball caps with the school insignia, name, and "Class of '95" on them. One supplier said it would charge $100 to make the design and then an additional $4 for each cap made. How would the total cost of the order be related to the number of caps in the order?
3. The population of the world has been increasing for as long as data or estimates have been available. What pattern of population growth has occurred over that time?
4. In planning a bus trip to Florida for spring break, a travel agent worked on the assumption that each bus would hold at most 40 students. How would the number of buses be related to the number of student customers?
5. The depth of water under the U. S. Constellation in Baltimore Harbor changes due to tides as time passes in a day. What pattern would that (time, depth) data fit?
6. When the Lincoln High School class officers decided to order and sell t-shirts with names of everyone in the Class of `95, they checked with a sample of students to see how many would buy at various proposed prices. How would sales be related to price charged?
7. How does the height of a bungee jumper vary as time passes in the jump?
8. In a wildlife experiment, all fish were removed from a lake and the lake was restocked with 1000 new fish. The population of fish then increased over the years as time passed. What pattern would likely describe change in fish population over time?
9. According to *The Old Farmer's Almanac*, if you live near crickets, you can estimate the nighttime outdoor temperature in degrees Fahrenheit by counting the number of cricket chirps in 14 seconds and adding 40 to that number. If you tested that by gathering data and plotting chirps vs. temperature, and the data seemed to support the rule of thumb, what might the graph of observations look like?

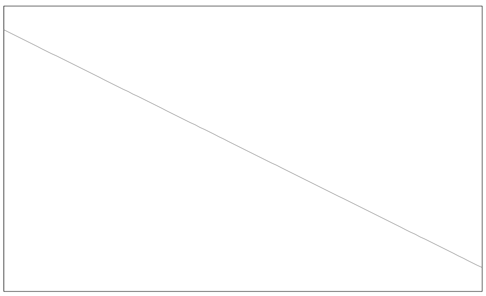
For each of the scenarios above, select the graph from the following page that you think most closely matches the pattern or relationship described. Try not to focus on the units or scales used on the axes of each graph (in fact, none of the graphs have such notations), but on the general nature of the relationships between the quantities described.



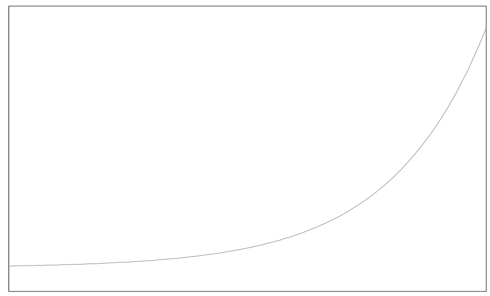
A



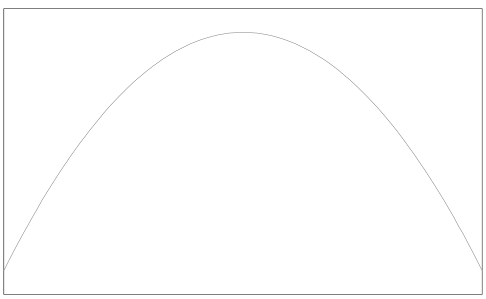
B



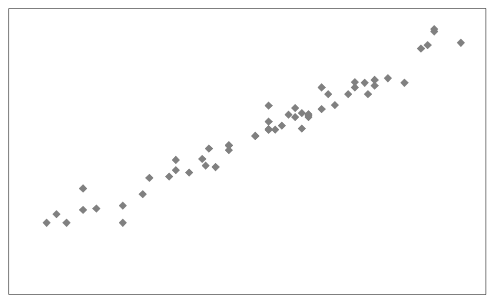
C



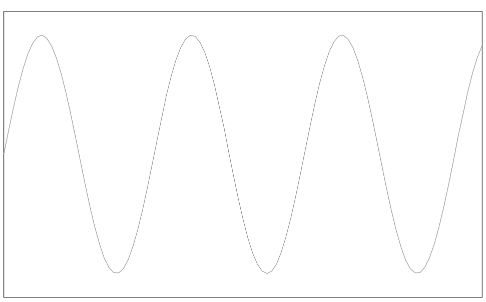
D



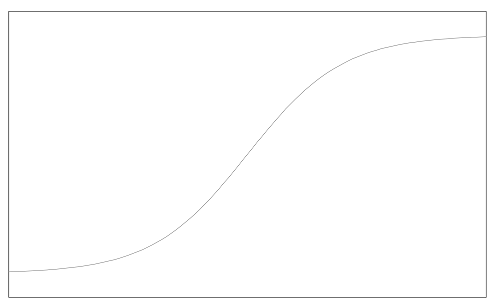
E



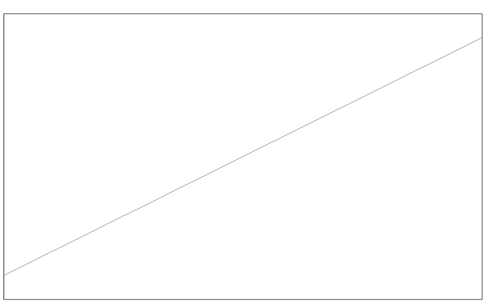
F



G



H



I

Mathematical Models

When the relationships between the quantities of a real-world problem can be expressed in mathematical formulas, we refer to those formulas – along with the "key" that maps the variables and constants in the formulas to real-world quantities and units of measure – as a *mathematical model* of the problem.

For example, while exiting the Lunar Module, Apollo 14 mission commander Alan Shepard rolls a golf ball off the platform at the top of the ladder.[1] The golf ball leaves the platform moving horizontally and falls to the Moon's surface from a height of 3 meters. The Moon's gravitational acceleration close to the surface is 1.63 meters per second per second. (In other words, after 1 second, a body will be falling at 1.63 meters per second; after 2 seconds, it will be falling at 3.26 meters per second; and so on.) Since the Moon's atmosphere is so thin, we can ignore atmospheric drag.

If we want to know how the height above the surface of the golf ball at some time *t* after it leaves the platform, up to the moment it hits the ground, we could take the general mathematical model for a body falling in a vacuum, and adapt it to our problem.

Given

$$\begin{aligned} g &= \text{gravitational acceleration} \\ t &= \text{time} \\ v_0 &= \text{initial vertical speed} \\ h_0 &= \text{initial height} \\ v_t &= \text{vertical speed at time } t \\ h_t &= \text{height at time } t \end{aligned} \tag{1}$$

Then

$$\begin{aligned} v_t &= v_0 + g\,t \\ h_t &= h_0 + v_0 t + \frac{g}{2} t^2 \end{aligned} \tag{2}$$

For our scenario,

$$\begin{aligned} g &= -1.63 \text{ meters/sec}^2 \\ v_0 &= 0 \text{ meters/sec} \\ h_0 &= 3 \text{ meters} \end{aligned} \tag{3}$$

Taken together, (1), (2), and (3) make up a mathematical model that can be used to answer a number of questions about our hypothetical golf ball. We can even set $h_t = 0$ and use the quadratic formula to solve the second equation of (2) for *t*, to find out how long it will take the ball to reach the ground.

---

1 In reality, while Alan Shepard famously used a makeshift golf club to hit two golf balls on the Moon, he didn't roll any golf balls off the LM platform, as far as we know.

Mathematical Formulas as Patterns of Change

By itself, (2) might not be a very useful model, since without (1) it might not be clear what the different variables refer to, and since it doesn't include the information specific to our scenario found in (3). But general formulas like those in (2) are the essential core of a mathematical model: just as a graph visually conveys the relationship between variables, a formula does the same thing symbolically.

See if you can match the equations shown below to the corresponding scenarios or graphs in "Patterns of Change". While none of the formulas is a complete model, some are tied very specifically to the corresponding scenarios. Others fit the scenarios to some degree (at least in the general shape), but might not express the underlying mathematics accurately. Some are written in a general form, where c_0, c_1, etc. represent constant values in the equations.

This is intended to be a challenging exercise, especially when it comes to the last few formulas. Don't worry if you don't understand all of the symbols and notation used; see if you can guess their meaning from their appearance and use. If you can't match all of the formulas to scenarios or graphs, keep in mind that it's possible that some of the formulas don't correspond to any of the scenarios or graphs, and vice versa. On the other hand, it's also possible that one (or more) of the general formulas matches more than one scenario, or that more than one formula can be matched with a single scenario.

i. $y = 4x + 100$

ii. $y = -16(x-2)^2 + 70$

iii. $y = \left\lceil \frac{x}{40} \right\rceil$

iv. $y = x - 40 + \varepsilon$

v. $y = c_0 + c_1 x$

vi. $y = c_0 + c_1 \sin x$

vii. $y = c_0 + c_1 \frac{\cos x}{x+1}$

viii. $y = \frac{c_0}{1 + e^{c_1 - x}}$

ix. $y = c_0 e^{c_1 (x + c_2)}$

Simple Linear Regression

Introduction

In a simple linear model, there's one independent variable and one dependent variable.

$$Y = \alpha + \beta X + \varepsilon \tag{1}$$

where

X and Y are the independent and dependent variables, respectively;

$\alpha, \beta \in \mathbb{R}$ (i.e. the coefficients are real numbers);

ε is the random error or residual, usually assumed to follow a normal distribution.

Least-Squares Regression

If we denote our model estimates for α and β by *a* and *b*, respectively, and the fitted line by

$$\hat{Y} = a + bX, \tag{2}$$

then the sum of the squared residuals (*sum of squared errors*, or SSE) for the model is given by

$$\text{SSE} = \sum \left(y - \hat{y} \right)^2. \tag{3}$$

The values of *a* and *b* which minimize SSE are

$$\begin{aligned} a &= \frac{\sum x^2 \sum y - \sum x \sum x y}{n \sum x^2 - \left(\sum x \right)^2} \\ b &= \frac{n \sum x y - \sum x \sum y}{n \sum x^2 - \left(\sum x \right)^2} \end{aligned} \tag{4}$$

Example

Using the *x* and *y* values shown, fill in the *xy* and x^2 columns, and the bottom row.

x	***y***	***xy***	x^2
0	-0.5		
1	0		
2	1.5		
3	2		
$\sum x =$	$\sum y =$	$\sum x y =$	$\sum x^2 =$

Table 1: Data and calculations for regression example

Use the values from the bottom row of Table 1, along with the number of data points (n), to calculate a and b.

$$a = \frac{\sum x^2 \sum y - \sum x \sum x y}{n \sum x^2 - \left(\sum x\right)^2}$$

$$= \tag{5}$$

$$b = \frac{n \sum x y - \sum x \sum y}{n \sum x^2 - \left(\sum x\right)^2}$$

$$=$$

Use the a and b values from (5) to write the formula of the fitted line, following the form of (2).

$$\hat{Y} = \tag{6}$$

To visualize the fit, plot the data points in Table 1, along with the fitted regression line in (4), on the graph in Figure 1.

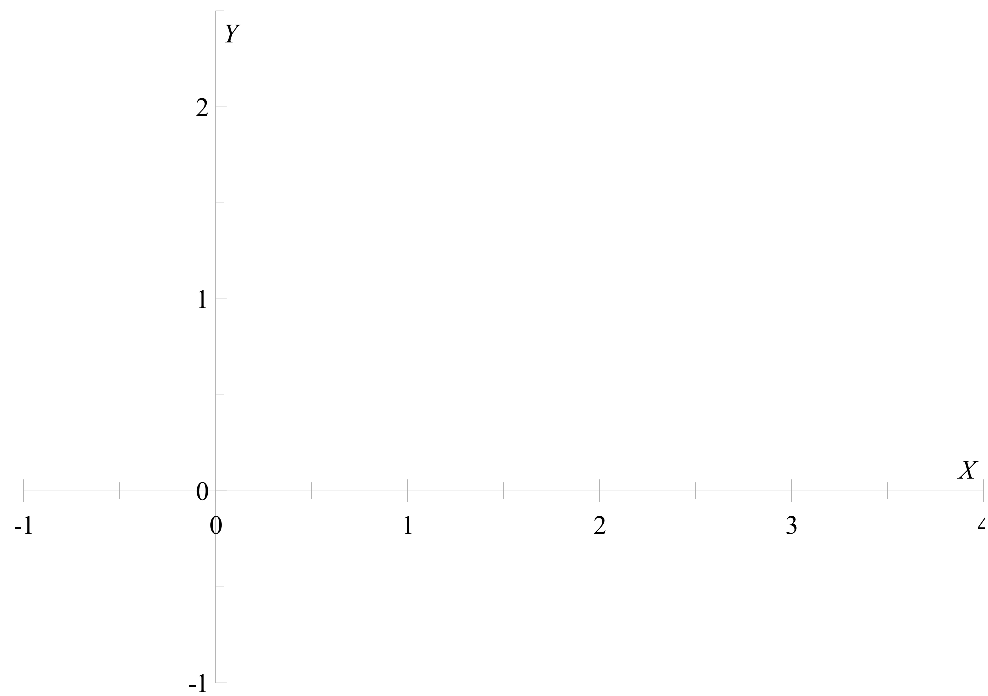


Figure 1: Example data points and fitted regression line

The Coefficient of Determination

To measure the goodness-of-fit of the regression model, we first quantify the total change in the dependent variable, by summing the squared deviations of its observed values from its sample mean. We call this sum the *total sum of squares*, or SST.

$$\begin{aligned} \text{SST} &= \sum \left(y - \bar{y}\right)^2 \\ &= \sum y^2 - \frac{\left(\sum y\right)^2}{n} \end{aligned} \tag{7}$$

SST can also be expressed as the sum of SSE and the *sum of squares of regression* (SSR), i.e.

$$\text{SST} = \text{SSE} + \text{SSR}, \tag{8}$$

where SSE is given by (3), and

$$SSR = \sum \left(\hat{y} - \bar{y}\right)^2. \tag{9}$$

From SST and SSR, we get a fundamental goodness-of-fit measure: the *coefficient of determination*, or R^2.

$$R^2 = \frac{\text{SSR}}{\text{SST}} \tag{10}$$

We can interpret R^2 as the fraction of the variation in the dependent variable that's determined or explained by the model.

Example

To calculate R^2 for the example data set, start by using the original data and the fitted line equation (6) to fill in the additional columns (including the bottom row) in the data table.

x	y	xy	x^2	y^2	$\hat{y}$	$\left(y - \hat{y}\right)^2$
0	-0.5	0	0			
1	0	0	1			
2	1.5	3	4			
3	2	6	9			
$\sum x = 6$	$\sum y = 3$	$\sum xy = 9$	$\sum x^2 = 14$	$\sum y^2 =$		$\sum \left(y - \hat{y}\right)^2 =$

Table 2: Calculations for coefficient of determination in regression example

Now, use formulas (3), (7), (8), and (10), along with the sums computed in Table 2, and the number of data points (n), to find R^2.

$$\text{SSE} = \sum (y - \hat{y})^2$$

$$=$$

$$\text{SST} = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$

$$=$$

$$\text{SSR} = \text{SST} - \text{SSE}$$

$$=$$

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$

$$= \tag{11}$$

What does this value of R^2 allow us to say about the fitted model?

Selected Symbols for Basic Mathematical/Statistical Modeling

These are the key mathematical symbols (other than $+$, $-$, $\cdot$, $/$, $=$, etc.) used in "Mathematical Models & Linear Statistical Models: Basic Concepts & Computations".

Concept	Symbol	Definition	Examples
Floor	$\lfloor \ldots \rfloor$	Rounding down (towards $-\infty$) of a non-integral real number, to the next integer value.	$\lfloor 1.75 \rfloor = 1$ $\lfloor -1.75 \rfloor = -2$ $\lfloor 1 \rfloor = 1$
Ceiling	$\lceil \ldots \rceil$	Rounding up (towards ∞) of a non-integral real number, to the next integer value.	$\lceil 1.75 \rceil = 2$ $\lceil -1.75 \rceil = -1$ $\lceil 1 \rceil = 1$
Exponent	b^n (superscript)	Number of times (not necessary integral) a base b is multiplied by itself in a product.	$x^2 = x \cdot x$ $3^4 = 3 \cdot 3 \cdot 3 \cdot 3 = 81$
Enumeration	s_i (subscript)	Numbered terms of an ordered sequence.	$\boldsymbol{S} = \{s_1, s_2, s_3, \ldots\}$ $\boldsymbol{F} = \{1, 1, 2, 3, 5, \ldots\}$ ($\boldsymbol{F}$ is Fibonacci sequence.)
Sum	$\sum$	Sum of terms in a sequence. $\sum_{i=m}^{n} s_i = s_m + s_{m+1} + \ldots + s_n$ (If the bounds m and n are well understood, they are often omitted from the $\sum$ operator notation.)	$\sum_{i=1}^{4} f_i = f_1 + f_2 + f_3 + f_4$ $= 1 + 1 + 2 + 3$ (Sum of 1st 4 terms of Fibonacci sequence.)
Product	$\prod$	Product of terms in a sequence. $\prod_{i=m}^{n} s_i = s_m \cdot s_{m+1} \cdot \ldots \cdot s_n$	$\prod_{i=3}^{5} \frac{i}{i+1} = \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{5}{6}$ $= \frac{1}{2}$
Factorial	$n!$	$n! = \prod_{i=1}^{n} i$ $= 1 \cdot 2 \cdot \ldots \cdot n$ $0! = 1$	$5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$ $= 120$
Euler's number	e	Base of natural logarithms. $e = \sum_{i=0}^{\infty} \frac{1}{n!}$ $= \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots$ $\approx 2.71828\ldots$	