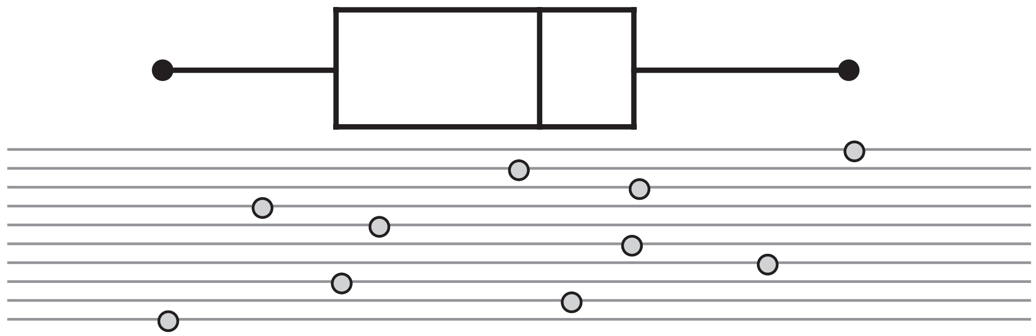
8

Probability and Data



Normal Distribution

Although statistical analysis uses mathematics, it is a science in that it involves the analysis of observations. Recent computer advances have made statistical analysis much easier and more efficient. Statistics is now an indispensable tool in such diverse fields as medicine, biology, economics, sociology, meteorology, and sports.

RANDOM SIMULATION

Suppose you are taking a test, but you are entirely unprepared. In fact, you have no knowledge of the subject matter at all. It's a true/false test with 100 questions, and you need to get at least 60 right to pass. You know that there are 200 students taking that same test, and not one of them knows anything about the subject.

Q1 None of this worries you. Anyone can answer half of the questions correctly simply by guessing, so you only need to be a little bit luckier than most. What would you guess is your probability of getting a passing score? Express your guess as a decimal number between 0 and 1.

1. Open page 1 of **Normal Distribution.gsp.** This sketch will display a random distribution of scores on the test. You can control three parameters:

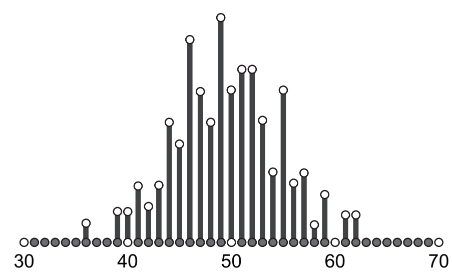
   p, the probability of a correct answer on any one problem

   pass mark, the minimum score needed to pass the test

   n, the number of people taking the test

Double-click on a parameter to edit it.

2. Edit these three parameters to model the true/false test described above. To try a new random sampling, select the red circle interior and press the exclamation point (**!**) key. Do this several times, and as you do so, watch the number of passing scores.



Q2 Based on these observations, about how many people out of 200 are likely to pass the test? What is the approximate probability of passing?

3. Select the red circle interior and press the exclamation point (**!**) key again several times. This time watch the overall shape of the distribution.

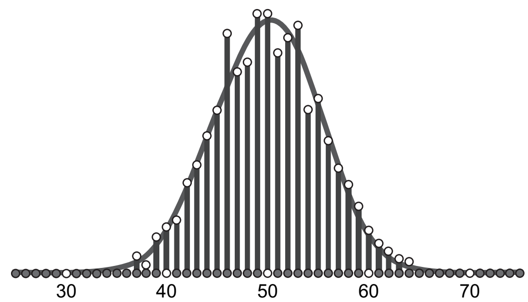
This is something you may have noticed in other data distributions. It has a bell shape, high in the middle and tapering down toward the extremes. Early investigators of probability theory noticed it too. They were successful in deriving a function that would predict the distribution.

4. Open page 2. The curve you see is a prediction of the data distribution. The vertical scale in this model is automatically adjusted so that the curve stays the same when you add more data.

The curve is actually growing taller when you increase *n*. The variable scale makes it appear stable and keeps it from outgrowing the screen.

5. Select the parameter *n* and press the + key several times. This will increase *n* by 100 each time. Run the number up to at least 1000.

Q3 What do you notice about the relationship between the random data and the curve as you increase *n*?



For large values of *n*, this function can give you a reliable approximation of the data distribution before you even see the data.

6. Press *Hide Data, Show Total Region,* and *Show Success Region.* When the regions appear, you will also see measurements representing their areas.

Q4 Based on the way you have seen the data fit the graph, how can you use the two area measurements to approximate the probability of passing the test? Do so now. What is the approximate probability?

THE NORMAL DENSITY CURVE

If you have a reasonable level of confidence in the shape of the distribution, you can estimate the probability of an event by comparing areas, just as you did above. The *normal density curve* is actually a family of curves. It has two parameters: μ (mu), the mean; and σ (sigma), the standard deviation. Here is the function definition:

Perhaps you are not yet familiar with *e*. For now, you only need to know that it is a constant. Its value is about 2.72.

$$f(x) = \frac{e^{-((x-\mu)^2/2\sigma^2)}}{\sqrt{2\pi} \cdot \sigma}, \sigma > 0$$

7. Open page 3. This is a normal density curve. Experiment with changing parameters μ and σ.

Q5 Describe in detail how changing μ and σ affects the shape of the curve. Does it cross the x-axis?

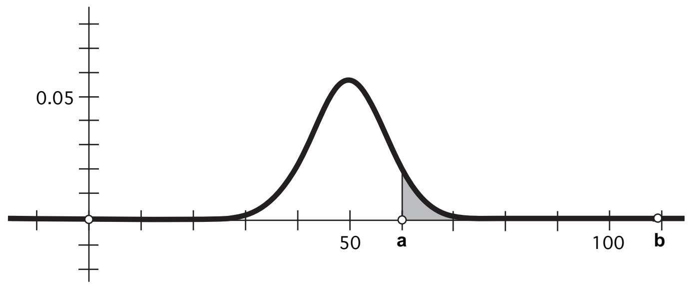
To set the limits more precisely, edit the parameters *set a* and *set b,* and then press *Set Limits a and b.*

8. Press the *Show Limits* button. The points *a* and *b* on the *x*-axis control the limits of a region under the curve. Notice the measurement for the area of this region. This time it has no units because it is based on a coordinate system.

Q6 With the earlier curve the height was based on the size of the data sample. That's not the case here. Drag the region limits to the edge of the screen left and right. What is the approximate total area under the curve? Try this again for several μ and σ settings. How does this simplify the probability calculation?

For the upper end of this interval, simply drag *b* to the edge of the screen. The missing area is negligible.

Q7 The true/false test example has normal parameters $\mu = 50$ and $\sigma = 5$. Enter these parameters and adjust the limits so that the region covers $x > 59.5$. What is your probability estimate?



Q8 If the test contained 50 questions worth 2 points each, the parameters would be $\mu = 50$ and $\sigma \approx 7.07$. In that case, would you be more likely or less likely to pass the test?

EXPLORE MORE

It was possible to calculate μ and σ for these true/false test examples, but random variables are usually much more complex. Take a person's weight as an example. It is influenced by age, diet, health, and a combination of genetic material from thousands of generations of ancestors. No one could possibly gather and process so much information. What you have to do instead is to make measurements of the population or, more likely, a sample of the population. You can then derive μ and σ from that.

Sports is an excellent field for investigating statistics because there is plenty of readily available data on athletes' heights, weights, ages, performances, and even salaries. You will probably have to do your own calculations of the mean and standard deviation.

Q9 The published rosters of six National Basketball Association teams indicate a mean height of 79.4 inches, with a standard deviation of 4.03 inches. Based on this information, what is the probability of a randomly chosen NBA player having a height between 70 and 74 inches?

Objective: Students simulate a random distribution of test scores and discover properties of the normal density curve.

Student Audience: Algebra 2/Precalculus/Statistics

Prerequisites: An understanding of simple probability and the mean is necessary. The standard deviation is used here, but no formula is given. Students can use the given σ values.

Sketchpad Level: Easy. The sketches are complex, but students only need edit parameters and click action buttons.

Activity Time: 30–40 minutes

Setting: Paired/Individual Activity (use **Normal Distribution.gsp**)

RANDOM SIMULATION

(The test score is what's called a *binomial random variable,* and it results in a *binomial distribution.* Students don't need to be familiar with these terms to do the activity.)

Q1 The probability of scoring 60 or more is approximately 0.028. Don't expect students to calculate this; the question is only a prompt for discussion.

Q2 This question does not ask for a precise calculation either, but students should be able to give a refined answer based on the observations. Encourage them to keep track of results of new random samples.

The expected number of passing scores is about 6, which makes the probability about 0.03. Students may find this counterintuitive. Although scoring 6 out of 10 is no great feat, scoring 60 out of 100 is highly unlikely.

Q3 As n increases, the data distribution tends to fit the curve more closely.

Q4 To approximate the probability, calculate the ratio of the areas:

$$\text{probability} \approx \frac{\text{success area}}{\text{total area}} \approx 0.03$$

This method uses a continuous integration although the random variable is discrete. The regions represent not only the integers between 0 and 100 (the only possible scores), but also all real numbers between the integers and even a very small area outside of the range. This may cause some confusion. Remind students that it is only an approximation. There is some justification for changing the minimum score to 59.5, and that would make a slight difference in the result.

THE NORMAL DENSITY CURVE

Q5 Changing μ causes a horizontal translation of the curve, and μ corresponds to the score at the maximum point of the curve. Increasing σ stretches the curve horizontally and compresses it vertically.

The curve is above the x-axis for any real x. That can be confirmed from the function definition. Students may be intimidated by the complexity of the definition. Help them break it into manageable parts. Since σ is a positive number, the denominator must be positive. The numerator is a positive base, e, with an exponent, so it must be greater than zero no matter what the exponent is.

Q6 No matter what the settings are for parameters μ and σ, the total area under the curve is 1. That corresponds to the denominator of the area ratio from Q4. Therefore, the probability is simply the area of the success region.

Q7 This is the same probability again, 0.03. Students should set limit a to 59.5 and drag b far to the right. In theory there should be no upper limit, but that is not possible here, and the missing area is negligible.

This may raise some questions about σ. Here is the formula:

$$\sigma = \sqrt{kp(1 - p)}$$

where p is the probability of success on any one problem, and k is the number of problems on the test.

Q8 You would be more likely to pass. Increasing σ flattens out the curve, forcing more area into the success region. The logical place to set the lower limit is 59 (since only even scores are possible), and the probability of passing is now about 0.10.

EXPLORE MORE

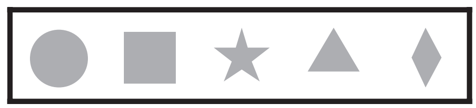
Q9 Based on the data, the probability is about 0.08. Students should set the parameters $\mu = 79.4$, $\sigma = 4.03$, $a = 70$, and $b = 74$.

Permutation and Combination

You probably learned to count before you even started school, but the method you learned (1, 2, 3, . . .) can be limiting. For example, how many seating arrangements are possible for your classroom? Don't answer that now. Just imagine how long it would take to list them all and count. Permutation and combination formulas help us to count things without actually having to see them.

PERMUTATION

Ms. Caba is giving a small prize to each of the top three students in her class. Each prize is different, and she is distributing them randomly among the winners. How many ways are there to distribute three prizes among three students? Each arrangement is a *permutation.*



1. Open **Permutation and Combination.gsp.** The first page, labeled Practice, has a set of icons grouped in a box. Each icon represents a different prize. The buttons allow you to adjust the number of prizes between zero and five.
2. Set the number of prizes to three. Select a prize and drag it out of the box. A copy of it will remain. Drag one of each prize and arrange them in a row.
3. Again, drag one of each prize out of the box and arrange them in another row, but this time in a different order. Continue forming rows of the three prizes until you have formed every possible arrangement.

Q1 How many arrangements were you able to form? In other words, what is the number of permutations for three distinct objects drawn from a set of three?

Q2 Without counting, how many arrangements do you think you could form with four prizes? (*Hint:* How many arrangements are there in which the first prize is a triangle?)

4. Open the Factorial page. Adjust the number of prizes to three, and press the *List Permutations* button. It will answer Q1 and display the possible arrangements, six in all.

Note that the permutations are listed in a logical order. For the first position there are three prizes to choose from. No matter what you choose for the first, there are two prizes available for the second position. That leaves only one remaining for the last position. Hence, the number of permutations is $3 \cdot 2 \cdot 1 = 6$. This number is written 3! (three factorial).

5. This page will list the number of permutations for any set of prizes numbering between zero and five. Try them all.

Q3 As you can see, the number of permutations gets out of hand in a hurry, which is why this demonstration has an upper limit. How many permutations are there for eight distinct objects drawn from a set of eight?

Q4 For any positive integer n you can calculate $n!$ by taking the product of all positive integers less than or equal to n. But what about zero? Why is $0! = 1$?

SUBSETS

In his class Mr. Brownlow has decided to give a prize to anyone who scores 100% on the final exam. He has wrapped four different prizes, but only two students have qualified. In this case you are still counting permutations, but you are not using up all of the available prizes.

Q5 Given a set of four prizes, you must choose two. How many prizes are available for the first student? How many are available for the second? What is the number of permutations for two objects chosen from a set of four?

The number of permutations for r objects chosen from a set of n is written ${}_nP_r$. You can compute it from the following formula:

You may also see this written as P_r^n or $P(n, r)$.

$${}_nP_r = \frac{n!}{(n-r)!}, \quad \text{where } n \text{ and } r \text{ are integers and } n \geq r \geq 0$$

6. Go to the page labeled Permutation. This allows you to change the n and r parameters in a permutation calculation, again with an upper limit of five. List the permutations and check your answer to Q5. Experiment with other settings.

Q6 You may observe the fact that ${}_nP_0 = 1$ for any non-negative integer n. Explain why this is true.

Q7 It is also true that ${}_nP_n = {}_nP_{n-1}$. Explain why.

COMBINATION

Mr. Bozich promised that at the end of the semester he would give a prize to every student who scored 100% on any test. Stacey was the only student who accomplished this feat, and she did it three times. Mr. Bozich has five different prizes wrapped, and he tells Stacey to choose three.

In this case the order of the selections does not matter, because they are all going to the same student. Here you are not counting permutations; you are counting *combinations.*

7. Stay on the Permutation page for now. List the permutations for ${}_5P_3$. There should be 60.

Q8 The first permutation has a circle, a square, and a star. Looking carefully, you can see that six of the permutations have this same combination of prizes. Find them. Explain why there must be six.

You cannot count this same combination six times, and the same goes for all of the other combinations. Therefore, divide the number of permutations by six to get the number of combinations. The expression ${}_nC_r$ represents the number of possible combinations of r objects chosen from a set of n.

The combination may also be written C_r^n, $\binom{n}{r}$, or $C(n, r)$, and is often pronounced "n choose r."

$$ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!}, \quad \text{where } n \text{ and } r \text{ are integers and } n \geq r \geq 0 $$

8. Go to the page labeled Combination. List ${}_5C_3$. Experiment with other combinations.

Q9 What are ${}_nC_0$, ${}_nC_1$, and ${}_nC_n$?

Q10 Explain why ${}_nC_r$ is always the same as ${}_nC_{n-r}$.

CALCULATIONS

9. Open the page labeled Calculations. This page has calculations for factorial, permutation, and combination. There are no graphical representations, but there are also no upper limits for the parameters.

Q11 A basketball team has ten players, and there are five different player positions on the floor. How many different starting lineups are possible?

Q12 Nine apartment tenants all drive, and their parking lot has only nine spaces. How many ways are there to arrange the cars in the spaces?

Q13 At the start of the game of euchre, each player gets five cards from a deck that includes only the cards from nine up to ace (9, 10, J, Q, K, A) in the usual four suits. How many hands are possible in euchre?

Objective: Students manipulate a prepared sketch to model permutations and combinations drawn from a given set of objects.

Student Audience: Algebra 2/Statistics

Prerequisites: This activity could serve as an introduction to permutation and combination, but it would work better as a review. At the least, students should understand factorial notation before doing this.

Sketchpad Level: Easy. Students only need to press buttons on the pre-made sketch.

Activity Time: 40–50 minutes

Setting: Paired/Individual Activity (use **Permutation and Combination.gsp**)

PERMUTATION

Q1 There are six ways to arrange the prizes.

Q2 By adding one more prize, there are now 24 ways: For each prize chosen first, there are 6 ways to arrange the remaining three prizes.

Q3 For eight objects the number is $8! = 40{,}320$.

Q4 We usually say that $0! = 1$ by definition, for convenience, but there is a perfectly logical way of justifying this. Imagine starting with an empty set. If you do nothing at all, then you have one arrangement, and there is no way to change that arrangement. Therefore, there is one way to arrange zero objects.

SUBSETS

Q5 There are four prizes available for the first prize. After choosing that, there are three available for the second. There are 12 permutations.

Q6 Here is a symbolic proof that ${}_nP_0 = 1$:

$${}_nP_0 = \frac{n!}{(n-0)!} = \frac{n!}{n!} = 1$$

Another explanation is that ${}_nP_0$ is the number of ways to draw zero objects from a set of n. There is one way to do this, which is to draw nothing at all.

Q7 Symbolically:

$${}_nP_n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = \frac{n!}{1!} = \frac{n!}{[n-(n-1)]!} = {}_nP_{n-1}$$

Imagine drawing n objects from a set of n. You start by drawing $n - 1$ objects. There are ${}_nP_{n-1}$ ways to do this. In each case there is only one object left to complete the group.

COMBINATION

Q8 There must be six because that is the number of ways there are to arrange three objects ($3! = 6$).

Q9 ${}_nC_0 = 1$ ${}_nC_1 = n$ ${}_nC_n = 1$

Q10 Symbolically:

$${}_nC_r = \frac{n!}{r!(n-r)!} = \frac{n!}{[n-(n-r)]!(n-r)!} = {}_nC_{n-r}$$

When you draw r objects, you could think of that as separating the set into two groups, the r objects that you drew and the $n - r$ objects that you left behind. If ${}_nC_r$ is the number of ways to draw r objects, then it must also be the number of ways to leave $n - r$ objects behind.

CALCULATIONS

Q11 You are drawing five players from a set of ten, and their positions do matter, so use the permutation formula.

$${}_{10}P_5 = 30{,}240$$

Q12 This is simply the number of ways to arrange nine objects.

$$9! = 362{,}880$$

Q13 Each player gets five cards from a set of 24 different cards, and the order of the cards does not matter.

$${}_{24}C_5 = 42{,}504$$

Box and Whiskers

The *box-and-whiskers plot* (sometimes just called a *box plot*) is a recent development in statistical analysis. You cannot derive any detailed information from it, but it gives you a convenient, easily understood graphical representation of the data distribution.

SKETCH AND INVESTIGATE

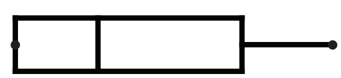
1. Open **Box and Whiskers.gsp.**

The sketch contains ten data values represented as points on parallel lines. Above the points are a box and whiskers. You can change a value by sliding its corresponding point right or left. The data are ordered and displayed on the left, but the actual numerical values are not important for this activity.

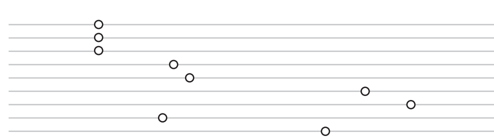
2. Before answering any of the questions, take a minute to experiment with the sketch. Drag the data points and observe the effect.

Each of the following questions suggests a special shape for the box and whiskers. In each case, state whether it is possible. If it is not possible, explain why not. If it is possible, make a rough sketch of the data points that will create that configuration, and suggest a real data set that might make this happen. The first one is done as an example.

Q1 Can one whisker have zero length?



A1 This will occur if the lower one-fourth of the data points all have the same value. This might happen if there is a lower limit to the data range. Test a group of people to see how far they can throw a heavy weight. Those who cannot even lift the weight will all score zero.



Q2 Can the median fall outside of the box?

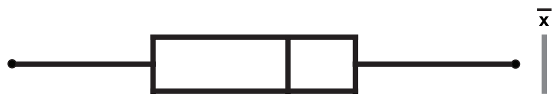
Q3 Can the box have zero width?

Q4 Can both whiskers have zero length?

3. Press the *Show Mean* button. The mean is represented by a green bar. Normally, the mean is not shown on a box-and-whisker plot. It appears here so that you can observe its relationship to the data distribution.

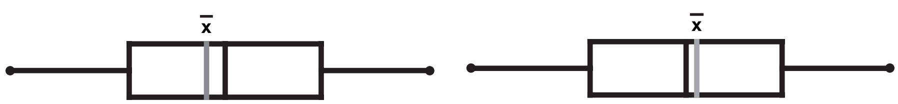
Q5 Can the mean fall outside of the box?

Q6 Can the mean be greater than the maximum?



Questions Q7–Q9 involve moving data points. In doing so, you are actually moving from one data set to another. In real life this could happen when a restaurant manager changes the prices of certain menu items, or when a sports team makes a player trade.

Q7 By moving one or more data points, can you move the mean without changing the box, the whiskers, or the median?



Q8 Can you move a single data point without changing the mean?

Q9 Can you move two data points without changing the mean?

Q10 Can you change the median by moving a single data point?

Objective: Students manipulate data points and observe the effect of their changes on a box-and-whiskers plot. In the process they gain an intuitive understanding of the box-and-whiskers representation of data distribution.

Student Audience: Algebra 1/Algebra 2/Precalculus/Statistics

Prerequisites: Students should already have seen the box-and-whiskers plot, and should understand the meaning of the five-number summary and the mean of a data sample.

Sketchpad Level: Easy. Students drag points on a prepared sketch; they do no constructions.

Activity Time: 40–50 minutes. The actual computer lab work can be performed in 30 minutes or less, but more time will be needed to answer the open-ended questions.

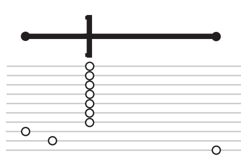
Setting: Paired/Individual Activity (use **Box and Whiskers.gsp**) or Whole-Class Presentation (use **Box and Whiskers Present.gsp**)

SKETCH AND INVESTIGATE

Q1 See A1 in the student section.

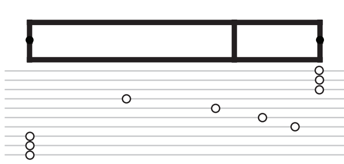
Q2 No, this is not possible. The middle half of the data must fall within the box, and the median must be somewhere in that middle half.

Q3 Yes. It will happen if the middle half of the data have the same value somewhere in the middle of the range.



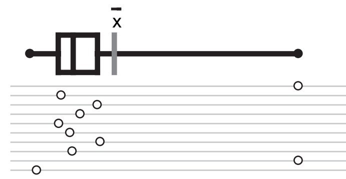
Imagine a math test with only two questions, one very easy and one very difficult. Nearly all of the students would score 1, but a few outliers could score 0 or 2.

Q4 Yes. It is similar to the distribution in Q1, except that there is both a lower limit and an upper limit.



Show a photo of the cast of *Gilligan's Island* to a large group of people. Ask them to identify the characters. Many people have no knowledge of the show, and would score zero. Of the people who have seen the show, most would have no difficulty naming all seven.

Q5 Yes. This can be caused by a fairly close grouping with some extreme outliers.



This situation has been modeled in real life in baseball labor disputes. The mean income of professional baseball players is very high. However, most professional players play in the minor leagues and earn a modest income. Relatively few outliers in the majors earn exorbitant salaries, raising the mean but having little effect on the median or the quartiles.

Q6 No, this is not possible. The mean must be within the range of the sample. You can prove this algebraically.

$$x_{max} \geq x_1, x_{max} \geq x_2, \cdots, x_{max} \geq x_{10}$$

$$10x_{max} \geq x_1 + x_2 + \cdots + x_{10}$$

$$x_{max} \geq \frac{x_1 + x_2 + \cdots + x_{10}}{10}$$

$$x_{max} \geq \bar{x}$$

Q7 Yes. When ten data elements are ordered, the five-number summary is defined by only six of the values. The minimum is x_1, the maximum is x_{10}, the first quartile is x_3, the third quartile is x_7, and the median is the mean of x_5 and x_6. Moving any of the remaining four values will change the mean with no effect on the box, provided they do not pass any of these above-mentioned points.

Q8 No. The mean formula includes every data point, so changing a single value always affects the mean.

Q9 Yes. If you move two data points the same distance in opposite directions, the sum stays the same, and so does the mean.

Q10 Yes. The median is the mean of x_5 and x_6, so it is only necessary to move or replace one of these two points.

1. Open **Box and Whiskers Present.gsp.** The sketch has a set of ten data values controlled by ten points on parallel lines. A box-and-whisker summary appears above the points. The actual ordered data are in a column on the left. However, the scale is arbitrary, so you probably will have no use for the actual data.
2. Drag some of the red points to show the class how this affects the image.

Q1 How can you make the box narrower? (Group the data more closely.)

Q2 How can you have a narrow box but long whiskers? (Group most of the data closely, but have one outlier an each side.)

For Q3–Q8, press *Hide Box and Whiskers* before asking. Have students guide your manipulation of the data. Press *Show Box and Whiskers* after the class has reached something approaching a consensus. For each configuration, try to imagine how it could occur with real data. Some suggestions are in the activity notes.

This presentation can be particularly engaging if you give the controls to a student while you and the class direct him or her.

Q3 How can one whisker have zero length? (Group one end of the data on the same value.)

Q4 How can the median fall outside of the box? (This is not possible.)

Q5 Can the box have zero width? (Group the middle part of the data on the same value.)

Q6 Can both whiskers have zero length? (Put the three lowest on the same value and the three highest on the same value.)

3. Press the *Show Mean* button.

Q7 Can the mean fall outside of the box? (Try closely grouped data, but with one extreme outlier.)

Q8 Can the mean be greater than the maximum? (This is not possible.)

For the remaining questions, do not hide the box and whiskers. You will have to be able to see it as the data points are moving.

Q9 Is it possible to move the mean without changing the box, the whiskers, or the median? (Moving any data point will change the mean, but the box and whiskers are unchanged when you move the 2nd, 4th, 7th, or 9th data points.)

Q10 Is it possible to move one point without moving the mean? (It is not.)

Q11 Is it possible to move two points without moving the mean? (Yes, provided they are moved the same distance in opposite directions.)

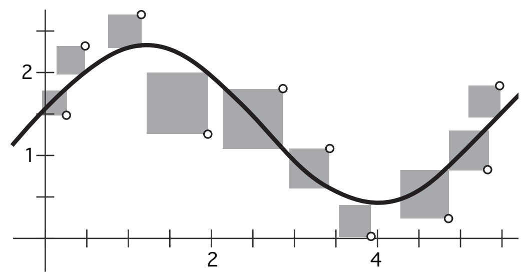
Fitting Functions to Data

The real world doesn't often behave as cleanly as mathematical functions do. The precision and consistency of pure mathematics is one of its attractions, but also one of its challenges: It's not easy to fit the messy data of real life into the more orderly world of mathematics. In this activity you'll use a general technique for fitting functions (which are smooth and well defined) to data (which may not be).

The technique involves starting with a parent function and using translations, stretches, and shrinks to generate a transformed function that more closely fits the data. To get the best fit, you need to be able to measure how close the fit is. You will use the method of *least squares* for this purpose: You will find the vertical distances from the data points to the function, square those distances, and add them up. The smaller the sum, the better the fit. (This is the reason for the method's name.)

MAKE THE SQUARES

1. Open **Fitting Functions.gsp.**



The sketch has ten data points and four sliders. Your job is to transform a parent function using the sliders so that it fits the data points as closely as possible.

Q1 Looking at the pattern of these data points, what parent function might you use?

2. Press *Show Parent Function* to reveal the parent function to use on this page.

To graph $g(x)$, choose **Graph | Plot New Function.** To enter a slider value into the formula, click it in the sketch.

3. Use the parent function and the sliders to graph the transformed function

$$g(x) = a \cdot f\left(\frac{x - h}{b}\right) + k$$

Q2 Drag the sliders to change the transformation. Which slider adjusts the horizontal stretch/shrink? Which one adjusts the vertical translation?

4. Arrange the transformed function plot in a rough approximation of the data.

Now you're ready to measure the distances and find the sum of the squares.

You cannot see the sum yet because it is hidden.

5. Press and hold the **Custom** tools icon and choose **Initialize Function and Sum** from the menu that appears. To start the summation process, click this tool on the transformed function.

Press and hold the **Custom** tools icon to see the menu again.

6. To construct the square for the first data point, choose the **Next Square** custom tool from the Custom Tools menu, and click it on the point.

7. Construct squares for each of the remaining data points.

8. To see the sum of the squares, choose the **Show Result** custom tool. You don't even have to click this tool; it shows the sum as soon as you choose it from the menu.

MAKE THE SQUARES SMALL (LEAST SQUARES)

As you adjust, you'll need to switch back and forth among the different sliders to get the best fit.

Q3 Drag the sliders while you watch the *Sum of Squares* calculation, and try to make this sum as small as possible. When you're satisfied, record the slider values and the *Sum of Squares* that you used.

Q4 Page 2 contains data to fit with a linear function. Change the parent function to $f(x) = x$. Record the slider values and the sum of squares that gives the best fit.

To make very small slider adjustments, select the point on the slider and press the right or left arrow key on the keyboard.

Q5 How many sliders do you really need in order to adjust a linear function? Explain why you don't need all four transformations in this case.

Q6 Page 3 contains data to fit with a quadratic function. The parent function is $f(x) = x^2$. Chris adjusted the sliders to make the sum of the squares 12.10. Can you do better? Record the slider values and the sum of squares that gives the best fit.

Q7 How many sliders do you really need in order to adjust a quadratic function? Explain why you don't need all four transformations in this case.

Q8 Page 4 contains yet more data and an exponential parent function. Fit the transformed function to the data and record your results.

Q9 Why do you think we square the distances from the point to the function before adding them up? Why don't we just find the sum of the distances and use that to measure how good the fit is?

EXPLORE MORE

Q10 Pages 5 and 6 contain additional data, but no functions. On these pages you must create your own parent function, transform it, and fit it to the data points. Choose your function carefully; for some data you may want to try two different parent functions to see which one fits best. (To do a second least-squares calculation in the same sketch, start over with the **Initialize Function and Sum** tool, and then use the **Next Square** tool on each data point.)

9. Collect some data of your own. The data can come from your own measurements, from a science lab, from the Internet, or from some other source. Plot the data in Sketchpad, and then choose an appropriate function and fit it to the data. Present the results to your group or to your entire class.

Fitting Functions to Data

Objective: Students transform functions to fit data, using a least-squares calculation to judge how good the fit is. They use several different parent functions, depending on the shape of the data.

Student Audience: Algebra 2/Precalculus/Statistics

Prerequisites: Students must be familiar with function transformations (translations and stretches/shrinks). It's useful if they also have some experience with fitting linear functions to data.

Sketchpad Level: Intermediate. There's not much construction, but students must know how to create and graph functions. Students use custom tools to perform the least-squares calculations.

Activity Time: 35–45 minutes

Setting: Paired/Individual Activity (use **Fitting Functions.gsp**) or Whole-Class Presentation (use **Fitting Functions Present.gsp**)

MAKE THE SQUARES

Q1 The data appear to be periodic, so the sin x or cos x function may be a good choice for the parent function.

Q2 Slider b adjusts the horizontal stretch/shrink. Slider k adjusts the vertical translation.

5. If students are not familiar with the Custom Tools menu, show them how to press and hold the **Custom** tools icon to make it appear.

MAKE THE SQUARES SMALL (LEAST SQUARES)

Answers will vary for the questions asking students to actually fit functions. Some typical values for Q3, Q4, Q6, Q8, and Q10 are shown in the table below, but there is the possibility for considerable variation. For instance, the sine function used in Q3 will require very different values for the horizontal translation (h) if the vertical stretch (a) is negative instead of positive.

Q5 You need only two of the sliders for a linear function. The values of a and b combine to determine the slope. The values of h and k combine to determine the intercept.

Q7 You need only three of the sliders for a quadratic function. The values of a and b combine to determine the width of the parabola. (The values of h and k determine the vertex.) Similarly, square root and absolute value functions need only three sliders.

Q9 One reason for squaring the values is to make them all positive. If you just add the deviations, a large negative deviation and a large positive one might add up to zero, even though both points are far from the graph.

Q3 (page 1)	$y - 2.05 \sin\left(\frac{x - 5.20}{1.45}\right) + 0.90$
Q4 (page 2)	$y = -0.60x + 2.40$
Q6 (page 3)	$y = -\left(\frac{x - 14.95}{3.20}\right)^2 + 25.25$
Q8 (page 4)	$y = 50.25 \cdot 2^{-(x-0.25)/12.60} - 0.25$
Q10 (page 5)	$y = 1.25\sqrt{x - 7.90} + 0.15$
Q10 (page 6)	$y = \frac{7.5}{x - 5.00} + 1.05$

WHOLE-CLASS PRESENTATION

Use the Presenter Notes and **Fitting Functions Present.gsp** to present this activity to the whole class.

Fitting Functions to Data

Use this presentation to review function transformations and to introduce students to some of the principles of curve fitting.

MAKE THE SQUARES

1. Open **Fitting Functions Present.gsp.** The sketch has ten data points and four sliders. Tell students that their job is to transform a parent function using the sliders so that it fits the data points as closely as possible.

Q1 Looking at the pattern of these data points, what sort of parent function might you use? (The data look periodic, so sin x or cos x might be a good choice.)

2. Press *Show Parent Function* to reveal the parent function for this page.
3. Press *Show Transformed Function* to create and graph the transformed function.

$$g(x) = a \cdot f\left(\frac{x-h}{b}\right) + k$$

It's best to have a student operating the computer and taking direction from you and from other students.

Q2 Ask students how to drag the sliders to experiment with the transformation. You're not trying to fit the data yet; you are familiarizing students with the effects of the sliders. Students should drive this adjustment process. Don't rush them; give them time to discuss, to argue, and to think about what's happening.

Q3 Which slider adjusts the horizontal stretch/shrink? (b)

Q4 Which slider adjusts the vertical translation? (k)

4. Adjust the sliders to arrange the transformed function plot in a rough approximation of the data.

Now you're ready to measure the distances and find the sum of the squares.

5. Use the **Initialize Function and Sum, Next Square,** and **Show Result** custom tools to find the sum of the squares of the deviations. Consult the directions for the student activity for details on using these tools.

To make very small slider adjustments, select the point on the slider and press the right or left arrow key on the keyboard.

Q5 Adjust the sliders to minimize the sum of the squares. Switch back and forth among the various sliders to be sure you've found a minimum result. Record the slider values.

Q6 Why do you think we square the distances from the point to the function before adding them up? Why don't we just find the sum of the distances and use that to measure how good the fit is? (We use squared values to make all the values positive so that positive and negative deviations do not cancel each other out.)

The remaining pages contain a number of interesting data patterns to fit.