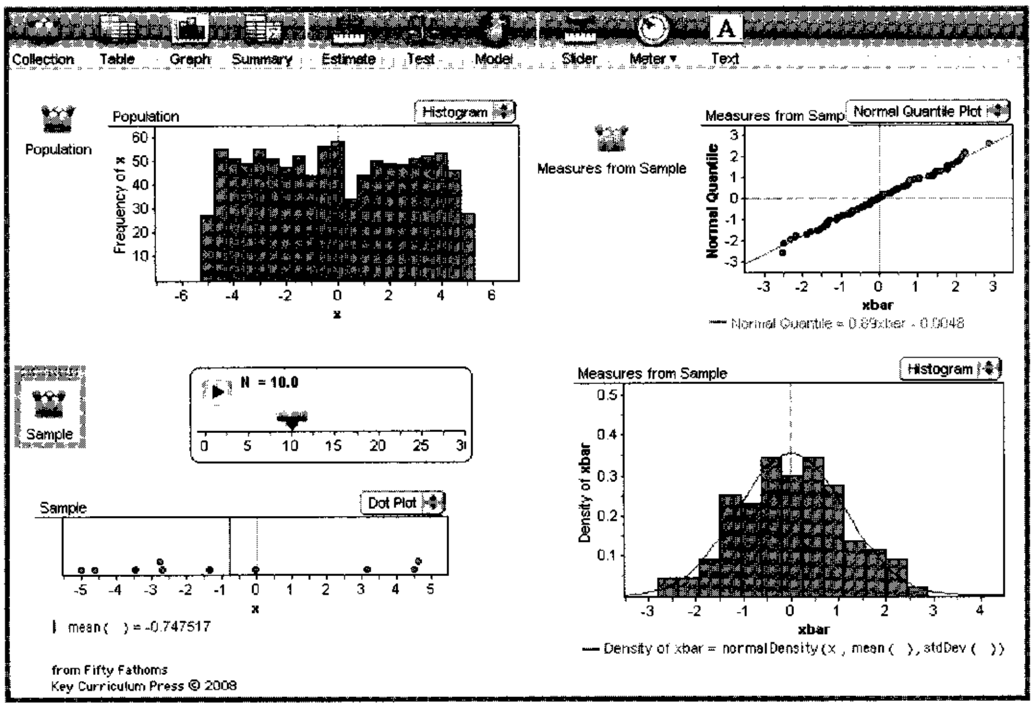
Demo 27: The Central Limit Theorem

A demo of the CLT • How sampling distributions usually look normal • Cases where they do not look normal

In a wide variety of situations, if you take a sample and calculate a statistic—for example, the mean of some quantity—and then repeat the sampling process, thereby collecting a sampling distribution, that distribution will be more or less normal. The bigger the sample, the more normal the distribution will be. This result is true no matter what the shape of the population distribution.

In this demo, you'll get a chance to make any "source" distribution you want, and then make a sampling distribution from that.



What To Do

▹ Open **Central Limit Theorem.ftm.** It will look something like the illustration.

On the left side, you see the **Population** above the **Sample.** The slider controls how many points appear in the sample. On the right, you see the measures collection—which collects 100 means, called **xbar,** from the sample (that is, this is a sampling distribution *of the mean*)—and two graphs of **xbar** itself: a normal quantile plot above, and a histogram below, with a normal density function plotted on it.

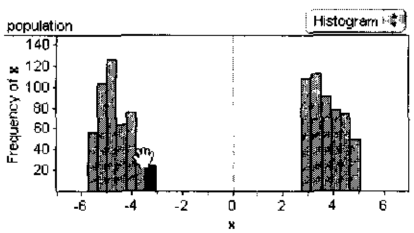
▹ First let's collect a new set of measures. Click once in the measures collection (the box with green balls called **Measures from Sample**) to select it. Then choose **Collect More Measures** from the **Collection** menu. Fathom collects and displays 100 new values for **xbar.**

The shortcut for **Collect More Measures** is **⌘+Y** on the Mac or **Control+Y** in Windows—but you must have the measures collection selected for it to work.

▹ Let's see how different the sampling distribution is if we take a smaller sample size. Set the slider **N** to 2. Notice that now there are only two points in the **sample** graph.

▹ **Collect More Measures** again. (Select the measures collection, then choose **Collect More Measures** from **Collection.**) Chances are the graphs will still look approximately normal (or at least triangular, as in Demo 21, "Adding Uniform Random Variables").

▹ So far, our source data have been roughly uniform. Let's change that. Select bars in the upper-left **population** histogram, and drag them to the two ends (with the *hand,* as shown; if you use "arrows," you'll just change the bin widths). That is, make a sharply bimodal distribution, as shown in the illustration. We have also rescaled the vertical axis.



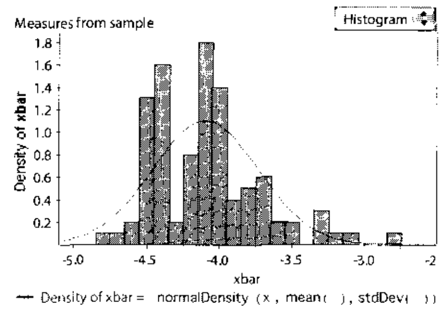
▹ Now, with **N** still equal to 2, collect measures again and see what you get. You should see three spikes in the **xbar** graph. Not normal at all!

This shows that you don't always get a normal distribution; sometimes you need a larger sample size to make the sampling distribution look more normal.

▹ So, make **N** larger (between 25 and 30 will do) and collect measures again. The **xbar** graph looks normal! (More precisely, approaching normal, which is all the Central Limit Theorem guarantees.)

▹ Even so, we can still make it non-normal. Take most of the points from one of your two humps and move them to the other hump. That is, the distribution should be bimodal, but grossly asymmetrical: a lot in one hump, a few in the other. With **N** between 25 and 30, **Collect More Measures** again.

You should see something like the illustration—again, not normal.



▹ Sample size to the rescue! Set **N** to about 100, **Collect More Measures** again, and wait patiently for Fathom to do its work. Probably normality returned (depending on how sharp you made your big spike).

Extension

Make different distributions and sample sizes, and see when the sampling distributions look normal—and when they don't.

Challenges

1 Select the population collection, then choose **Revert Collection** from the **File** menu to get our uniform distribution back. Now, instead of a sampling distribution of the mean, make a sampling distribution of the *standard deviation* (define it in the **Measures** panel in the inspector for the **population**). Try different sample sizes and population distributions. What do you find?

2 Imagine making an asymmetrical, sharply bimodal distribution (just zeros and ones, but a lot more zeros). Think about what we found out about what it takes to make the sampling distribution of the mean look normal. How does that relate to the rule of thumb ($np > 10$) for using the normal approximation when you calculate confidence intervals for proportions? (See Demo 31, "Why $np > 10$ Is a Good Rule of Thumb.") In fact, don't just *imagine* it; do it! **Sol**

3 Imagine (or make) a sharply bimodal distribution. Now consider the sampling distribution for the *median,* and try it out. What do you find?

4 We never changed the *number* of samples in the sampling distribution. What would be different if we collected more points?

