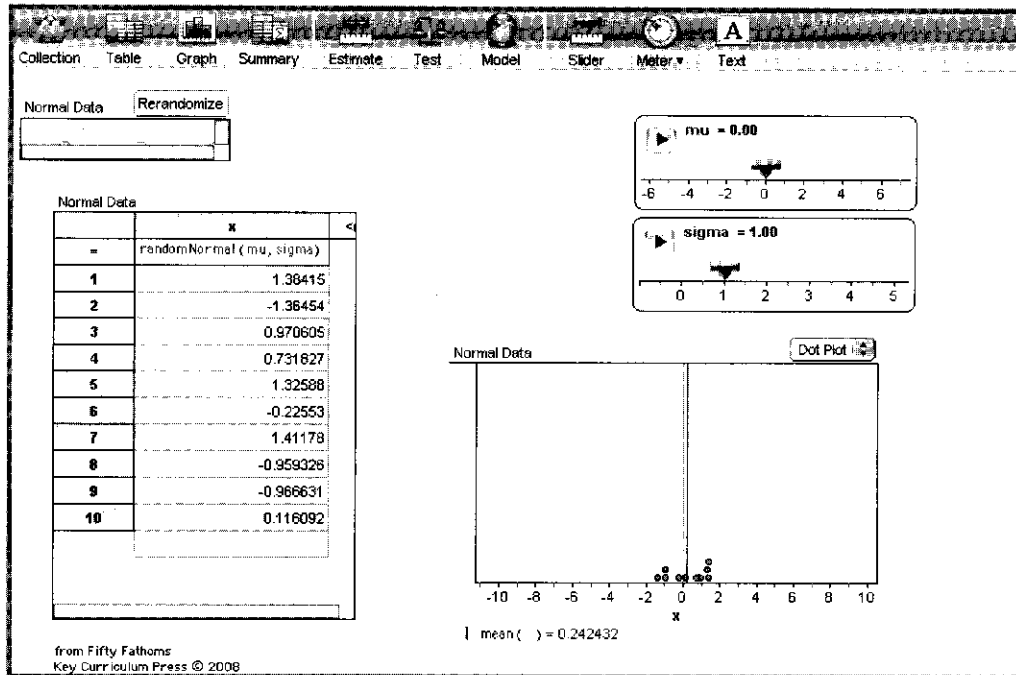


Demo 3: What Do Normal Data Look Like?

Normally distributed data • The effect of changing the mean and standard deviation

A lot of things in traditional statistics depend on data being roughly normal. But what do such data look like? We may have a mental image of the ideal bell curve, but a real sample from a normally distributed population may look very different.



What To Do

- ▶ Open **Normal Data.ftm**. It will look something like the illustration.

In the upper left is the collection (the box of balls), opened to show its **Rerandomize** button. Two sliders, **mu** and **sigma**, control the mean and standard deviation of the population, respectively. The case table and the graph show the data.

- ▶ Press the **Rerandomize** button repeatedly to see how the data change. Observe how the mean of this sample of 10 numbers jumps around.
 - ⇒ The data also rerandomize whenever you move the sliders.

- ▶ Change the mean by dragging the **mu** slider.
- ▶ Change the standard deviation by changing **sigma**.
- ▶ Change the dot plot to a histogram by choosing **Histogram** from the pop-up menu in the graph.
- ▶ Again change **mu** and **sigma** (you may want to rescale axes; see “Rescaling Graph Axes”) to see how this looks.
- ▶ Change the graph to a **Normal Quantile Plot** (use the pop-up menu). In this plot, if the data are perfectly normal, they will lie on a straight line.
- ▶ Rerandomize the data repeatedly to get an idea of the variety of “lines” you can get with a sample of 10 points drawn from a genuine normal distribution.

Questions

For these questions, display the data as a histogram or a dot plot.

- 1 With $\mu = 0$ and $\sigma = 1$ (this is called the *standard normal distribution*), what's a typical range of values for the points in your sample?
- 2 What's a typical range of *means* for your sample of 10 from a standard normal distribution?
- 3 Set $\mu = 1$. Now what's a typical range of values?
- 4 Set $\mu = 0$ and $\sigma = 2$. Again—what's a typical range of values?
- 5 About what proportion of the time do the data look fairly symmetrical and have a hump in the middle?

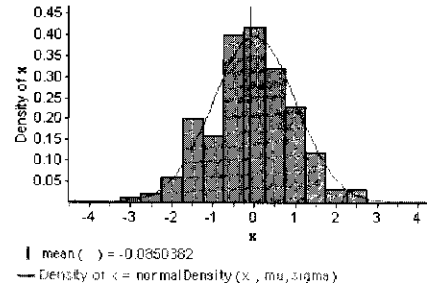
You should see that with a sample of 10, the graphs often look far from normal—they often don't have a hump in the middle, and they aren't symmetrical. The normal quantile plots can be pretty far from straight lines. So let's add some points.

Onward!

- ▶ Click on the collection (or case table or graph) once to select it. Then choose **New Cases** from the **Collection** menu. Enter 190 and press **OK**. Now you have 200 cases in your distribution.
- ▶ Play with the displays and sliders as before. Observe how different it is from the sample of 10.
- ▶ Let's put the curve on the graph. First, make the graph a histogram with the pop-up menu.
- ▶ Choose **Scale | Density** from the **Graph** menu. The vertical axis will change to a density scale.

Note: With a *density scale*, the vertical axis shows the proportion that are in the bin per unit on the *x*-axis. This way, the total area under the curve is 1.

- ▶ Choose **Plot Function** from the **Graph** menu. The formula editor will open.
- ▶ Enter **NormalDensity(x, mu, sigma)**. Press **OK** to close the editor. You should see something like the graph in the illustration.



- ▶ Again, play with the sliders to see how the graph changes.

More Questions

- 6 With 200 cases, about what proportion of the time do the data look fairly symmetrical with a hump in the middle?
- 7 In the normal quantile plots (choose **Normal Quantile Plot** from the pop-up menu in the graph), where do the points deviate the most from the straight line? **Sol**

Extension

Finally, let's see how this normal curve looks with fewer cases.

- ▶ Drag a rectangle to select most of the bars in the graph. They will turn red.
- ▶ Choose **Delete Cases** from the **Edit** menu. Those bars will vanish.
- ▶ Drag the sliders or rerandomize to see how the normal curve looks with fewer data.