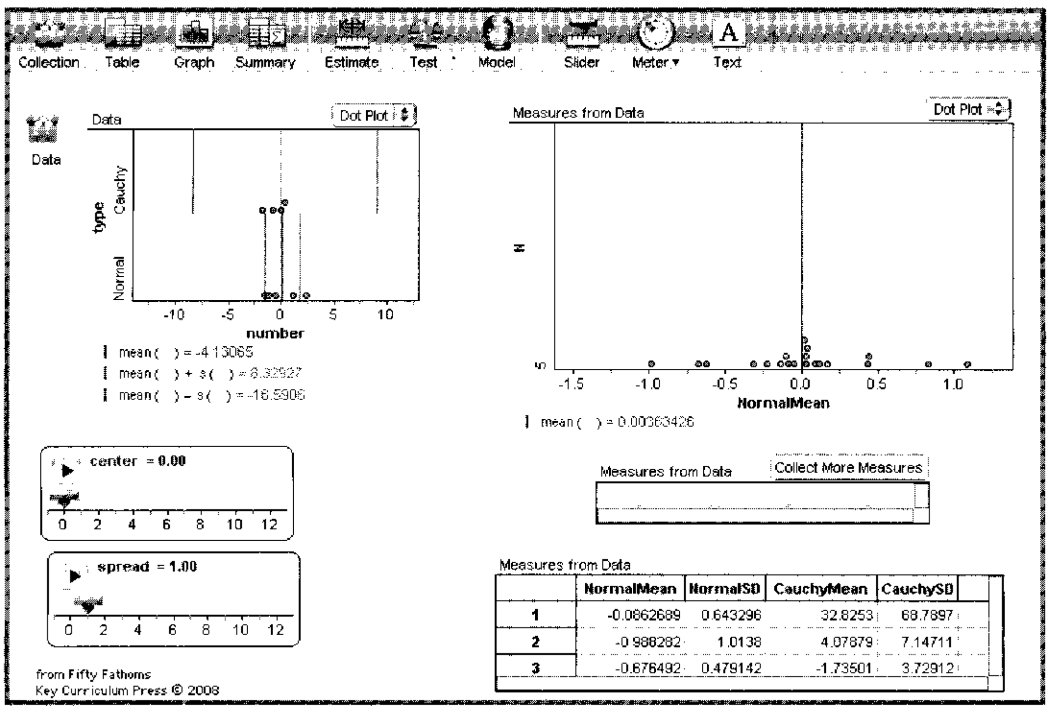
Demo 50: The Bizarre Cauchy Distribution

The Cauchy distribution • The meaning of mean and standard deviation; how it's possible for a distribution to have neither

Suppose you stand close to a very long wall and spin around. To make matters worse, you have a pea shooter, and at random times—while you're spinning—you shoot peas. Half the time, of course, the peas go away from the wall. Of the peas that hit, many will hit near the point on the wall closest to you. But some—at angles close to ±90°—will hit the wall very far away.

The positions of the peas that hit the wall form a *Cauchy* distribution. It has a hump in the middle just like some very well-behaved distributions we know, but it has an interesting catch: It *has no standard deviation.*

What can that mean? Of course it has a standard deviation! All you have to do is take the points in the sample and compute it. But I wasn't talking about the sample; I was talking about the *population*. Let's see what I mean.



What To Do

▹ Open **Cauchy.ftm**. It will look something like the illustration.

In the upper left, you see a collection and graph that shows a sample of five points drawn from a Cauchy distribution, and five more drawn from a Normal distribution for comparison. You control their centers and spreads with the sliders below (though we will not have to here).

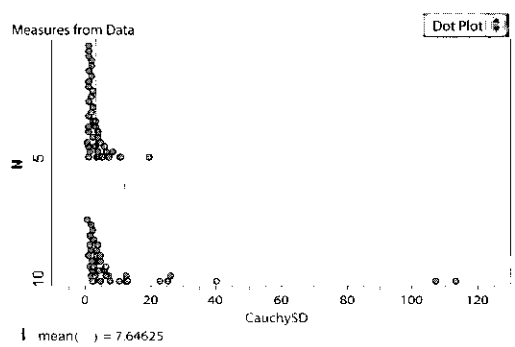
On the right, you see a measures collection (**Measures from Data**) in the middle, with a case table below it and a graph above. The graph shows the mean of the normals (the attribute **NormalMean**) from the collection on the left, from 20 separate resamplings.

▹ Click once on the upper-left (**Data**) collection to select it. Then choose **Rerandomize** from the **Collection** menu (or use its shortcut: ⌘+Y on the Mac or **Control+Y** in Windows.) The numbers in the graph will change, as well as the lines that show the mean and ±1 standard deviation of the *sample*.

- ▷ Repeat several times. Note how the Cauchy values jump around more.
- ▷ Let's add cases. With the **Data** collection selected, choose **New Cases** from the **Collection** menu; add 10 cases and press **OK** to close the box.

Note: The way we have this set up, adding 10 cases adds 5 Normals and 5 Cauchys.

- ▷ Again, **Rerandomize** to see how the distributions change.
- ▷ On the right-hand side, press the **Collect More Measures** button in the **Measures from Data** collection. Fathom rerandomizes the **Data** collection 20 more times and adds the means and standard deviations to that right-hand collection; you'll see two distributions now: **NormalMean** for **N = 5** and the same for **N = 10.**
- ▷ On the right-hand side, drag other attributes (**NormalSD, CauchyMean,** and **CauchySD**) from the case table at the bottom to the horizontal axis of the right-hand graph, replacing **NormalMean.** Observe the characteristics of each graph. The graph for **CauchySD** will look something like the illustration.



You can already see that there's something very strange about the Cauchy data. In the illustration, for example, while most of the resamplings have a small standard deviation, two have huge spreads when **N = 10.** Your statistical instincts should be saying, "We should increase the sample size; that will make these wild swings even out." Ordinarily, you would be right. But this is the Cauchy distribution.

- ▷ Alternate now between adding cases on the left and collecting measures on the right. (Remember that on the left, the number of cases in the collection is the number of Cauchys *plus* the number of Normals.) So add 20 cases (for a total of 40, or 20 of each type); then 60 (to make 100, or 50 of each type); then 100 (for 200, or 100 of each type).
- ▷ Again, drag different attributes to the horizontal axis of the right-hand graph and compare. Some of the Cauchy data will be so wild that you'll need to rescale your axes by hand to see the trends.

What You Should Take Away

You should see that, as sample size increases, the means (the red lines) of **NormalMean** nicely converge to zero and that **NormalSD** nicely converges to one. But the **CauchyMean**s swing more wildly as **N** increases, and **CauchySD** *systematically increases with sample size.*

Because of this, official mathematical statisticians admit that the Cauchy—as an abstract population distribution, a probability density—*actually has no mean nor standard deviation,* because the various sums or integrals you would use to compute them do not converge. For this reason, its parameters are named **mode** and **spread** in Fathom and in many statistics books.

Extension

Medians are supposed to tame distributions with outliers, right? Modify the document to collect medians and IQRs as well as means and SDs. You'll need to add new measures to the **Data** collection, delete the cases in the measures collection, and then re-collect the measures for different sample sizes.

Do the median and the IQR seem to converge as sample size increases?