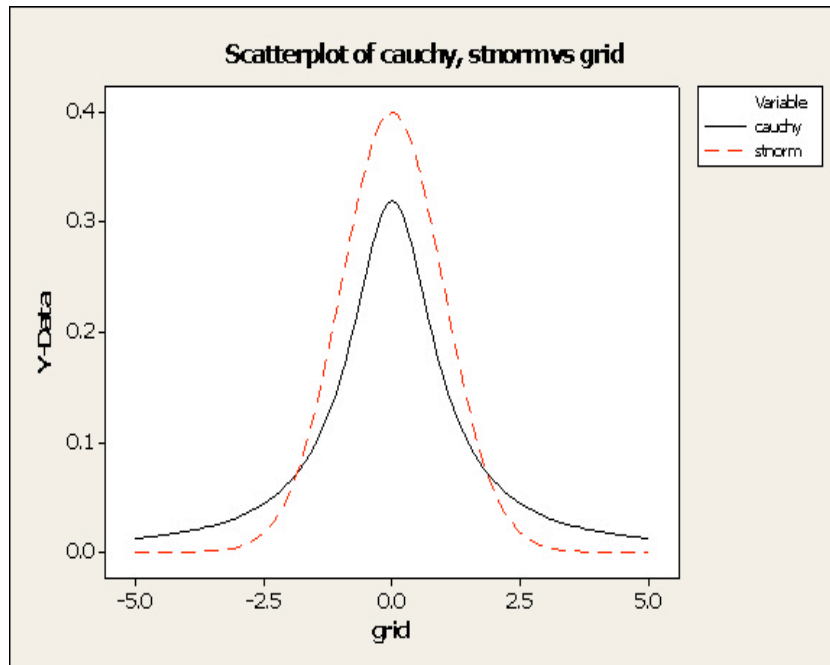


AN EXAMPLE WHERE THE CENTRAL LIMIT THEOREM FAILS

Footnote 9 on p. 440 of the text says that the Central Limit Theorem requires that data come from a distribution with finite variance. The *Cauchy distribution* (which is a special case of a t-distribution, which you will encounter in Chapter 23) is an example of a distribution that does not have a finite variance – in fact, the Cauchy distribution does not even have a finite mean. Here is a picture of the Cauchy distribution (black), along with a standard normal (dashed red) for comparison. Note that the Cauchy distribution has a shorter, narrower peak than the normal distribution, but has “fatter” tails.

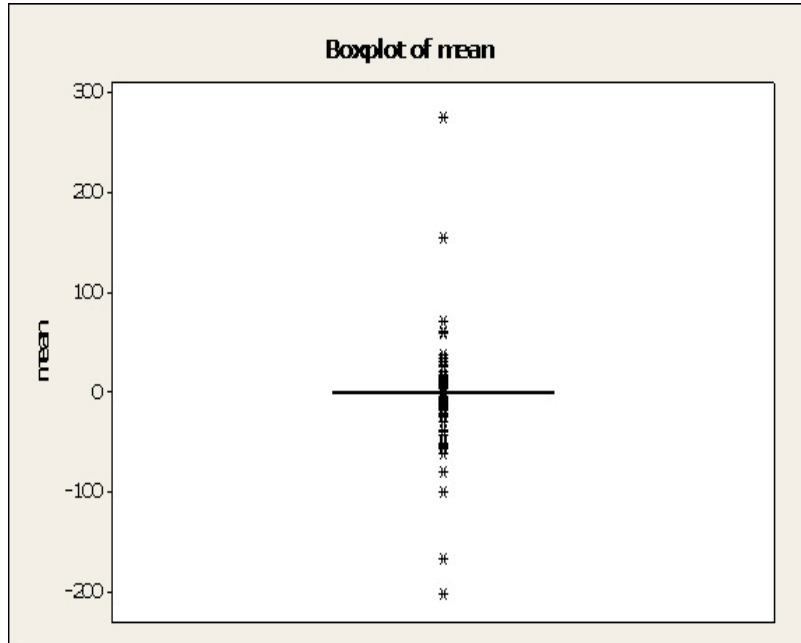


To illustrate the problem with the CLT for the Cauchy distribution, 1000 samples of size 100 were drawn from a Cauchy distribution. The mean for each sample was calculated. Looking at these 1000 sample means should give us some idea of the sampling distribution of means for samples of size 100 from the Cauchy distribution. Here are the descriptive statistics for these 1000 sample means:

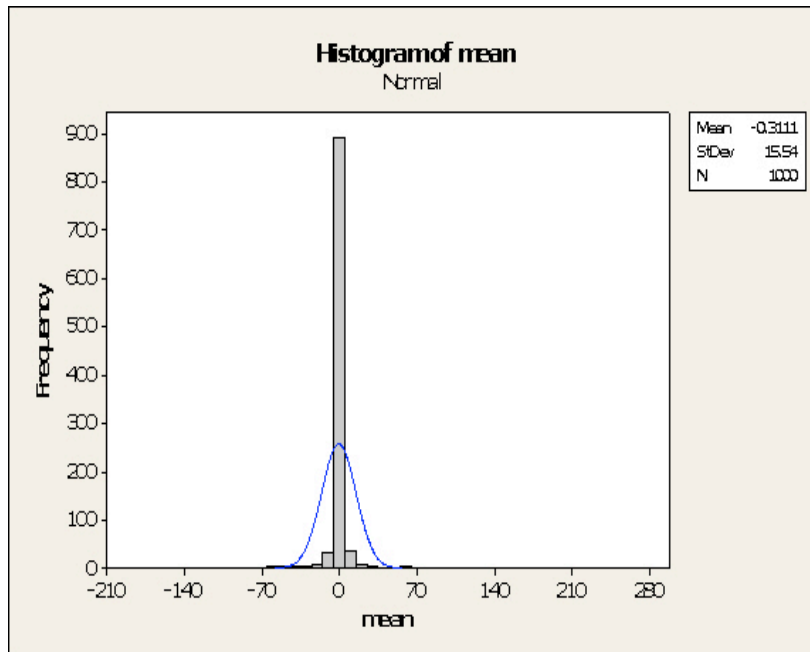
| Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|--------|--------|----------|--------|--------|-------|---------|
| -0.311 | 15.538 | -200.543 | -1.050 | -0.089 | 0.918 | 274.820 |

Notice that the standard deviation is much larger than the interquartile range, and the minimum and maximum values seem to be very extreme outliers. Before looking at the next page, think about what a box plot of these values would look like.

Here is a boxplot of the means of these 1000 samples. Quite unusual, with a thin box and many extreme outliers.



Here is a histogram of the 1000 sample means – i.e., illustrating the sampling distribution of means for samples of size 100 from the Cauchy distribution. For comparison, a normal distribution with mean and standard deviation the same as those of the 1000 sample means is overlaid. This sampling distribution is far from normal!



Note: Distributions for which the CLT fails are sometimes considered “pathological.” In many fields of study, such distributions never arise. But in other fields, they do arise in important ways. For example, in biology, the distribution of the distance that seeds end up from the parent plant may have a “fat-tailed” distribution if most fall by the plant but many are dispersed by wind, water, or animals. Standard statistical techniques are inappropriate for such distributions. This is just one example of how statistics is a very broad field, and still evolving to handle new situations such as discovering the effects of different genes in diseases. It also illustrates why it is important to check model assumptions.

Things you might want to investigate:

- Finding the formula for the Cauchy distribution and showing that both its mean and variance are infinite.
- Doing a simulation such as the one above with a smaller or larger number of samples.
- Other situations where “fat-tailed” distributions arise.