Using Technology to Develop Statistical Thinking—Understanding Distributions

Roxy Peck
Cal Poly, San Luis Obispo
rpeck@calpoly.edu

Distributional Thinking

- Thinking about data in terms of distributions rather than isolated data points
- Reasoning about distributions
  - Important idea in developing an understanding of statistical inference
- Technology, depending on how it is used, can help or hinder…

The plan for this workshop

- Activity to help students make the shift from looking at data points to looking at graphical displays of distributions

- Helping students distinguish between three important types of distributions that are encountered in introductory statistics: Population distributions, sample distributions, and sampling distributions.

  (Well OK, there is also a fourth type of distribution—probability distributions—that we won't be talking about in this session…)

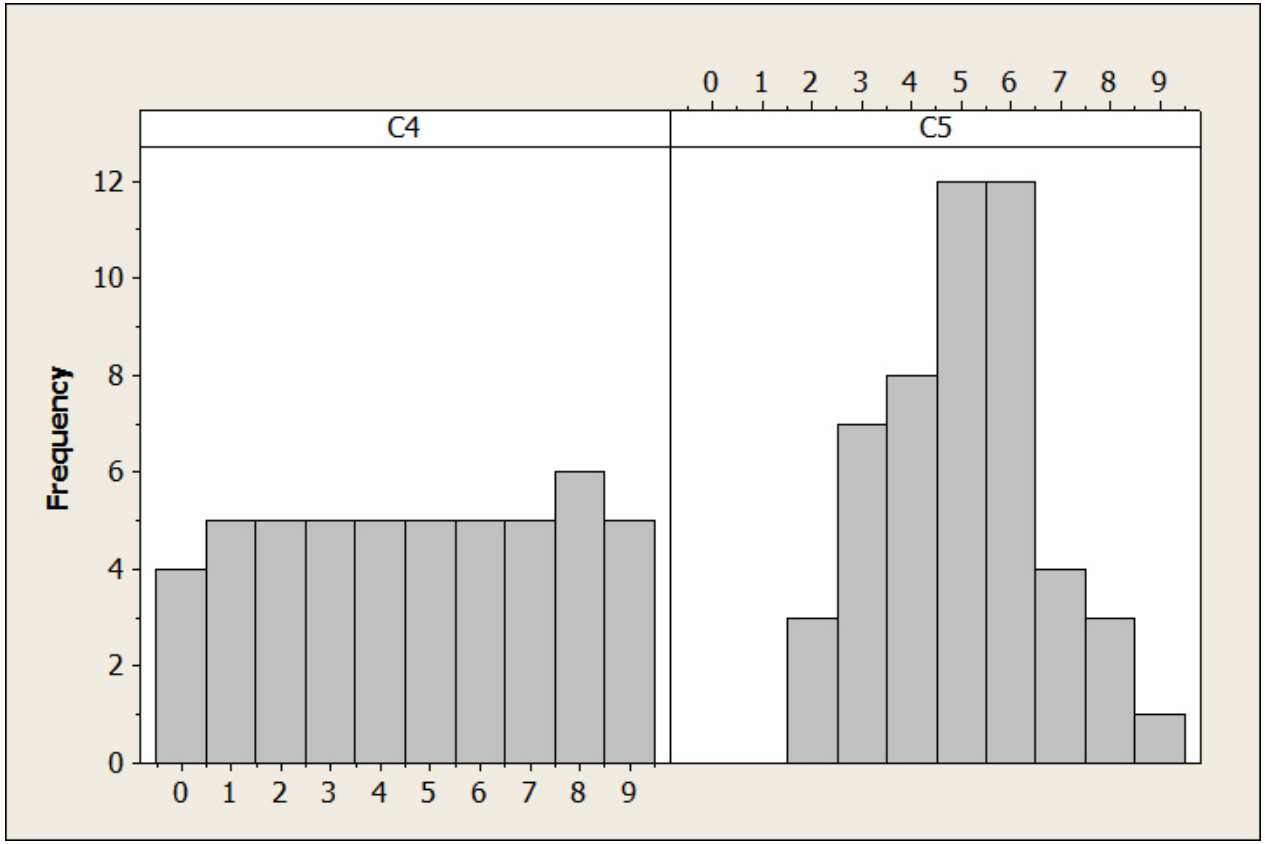
A BIG Idea

Distributions describe variability.

The First Step…

- AVOID TECHNOLOGY HERE!
- Why? Let's consider a few common student misconceptions.

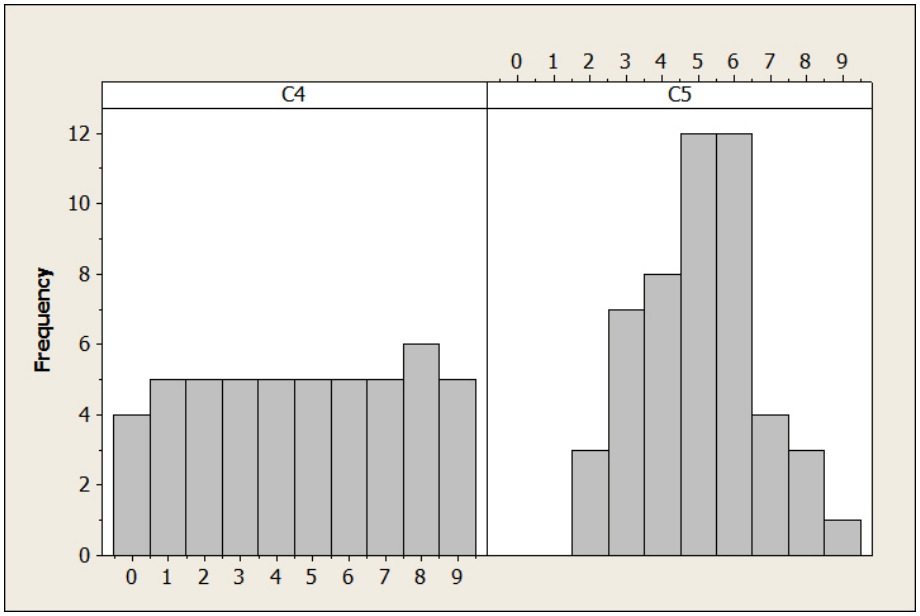
Which is more variable?

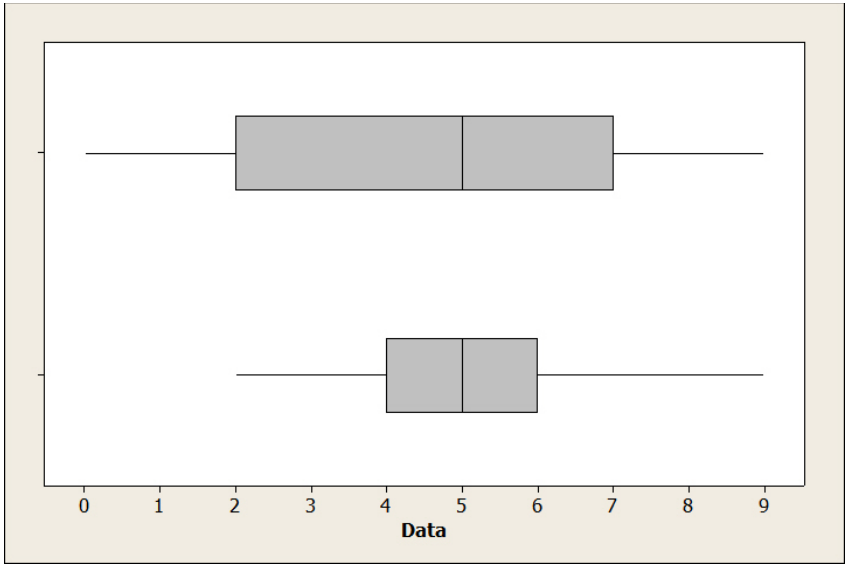


Find the mean (or median)…

Time (in minutes)	Frequency
1	4
2	8
3	14
4	6

Interpreting Graphical Displays





Why??

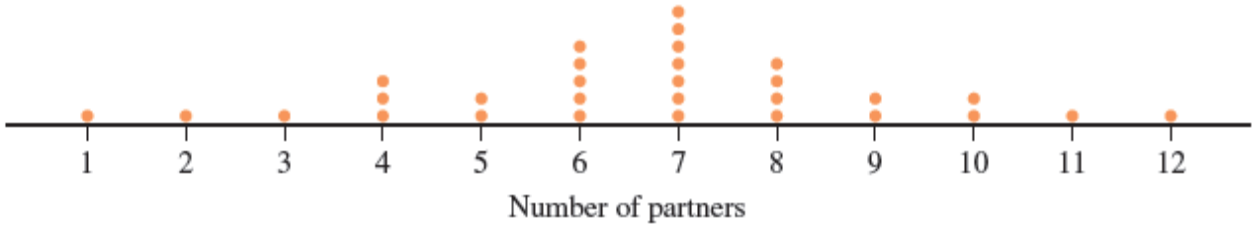
- Students aren't developing distributional thinking

- We assume that because students can construct a histogram and make statements about center, shape and spread, that they understand what the histogram represents in terms of data distribution. But if we listen carefully, we find that interpretations of center and spread are often about the GRAPH, not about the DATA. Focus is on the wrong thing.

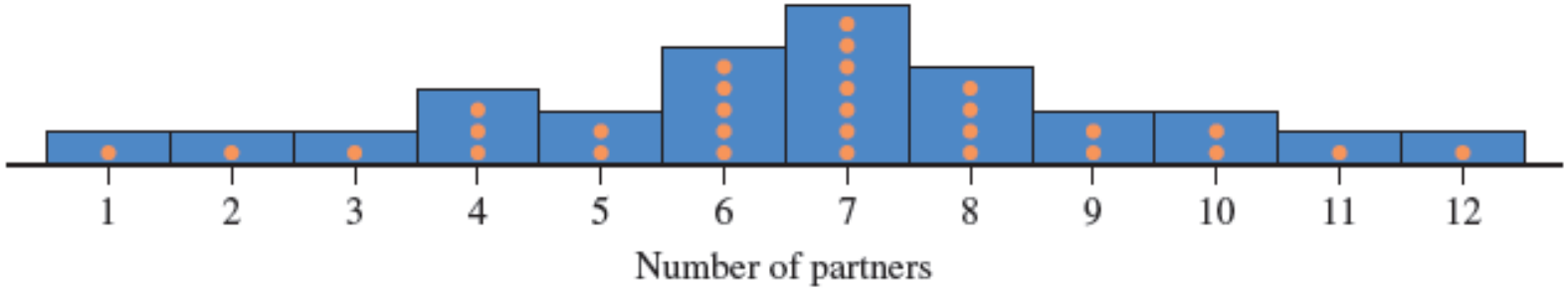
Activity 1: Developing Distributional Thinking

- A technology free activity (But don't worry—the technology will come later!)
- Starts at the very beginning with graphical displays.
- Histograms and Boxplots

FIGURE 2.13
Dotplot of queen bee number of partners data

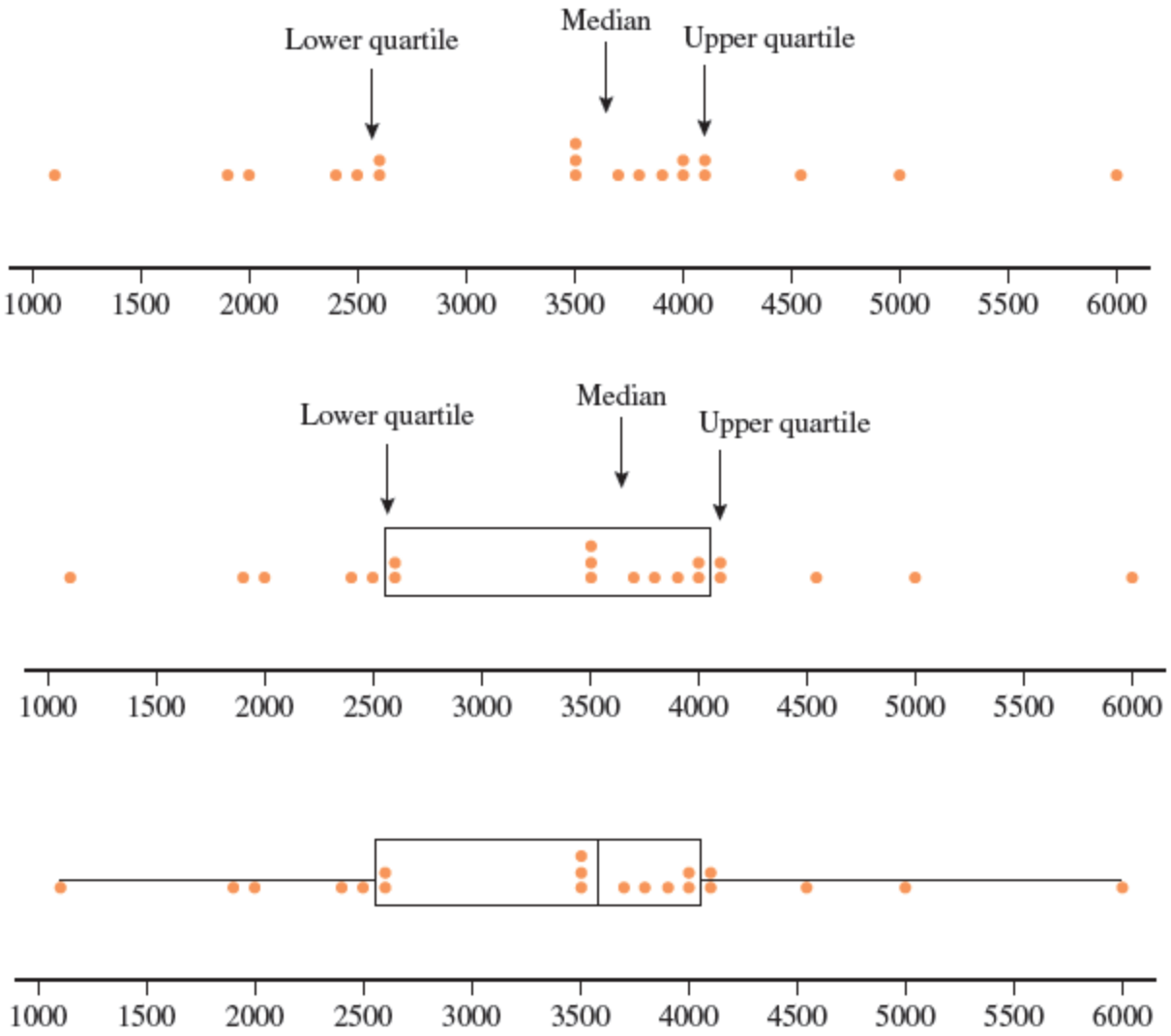


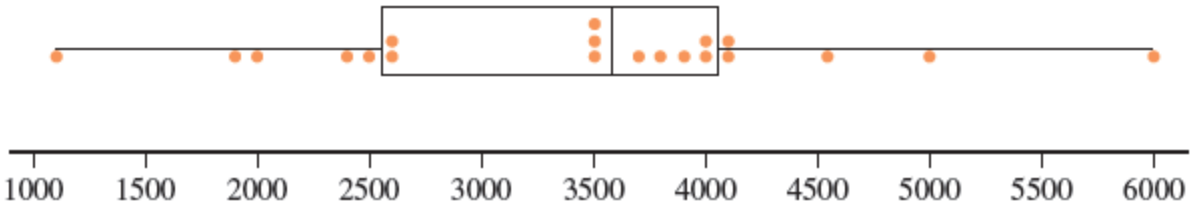
What does a dot represent?
What distinguished two dots from one another?



What does a dot represent?
What does a stack of dots represent?
What does a bar in the histogram represent?
What does a bar with a large area represent?
What does a bar with a small area represent?

When we talk about center and spread in a histogram, what does it mean in terms of the dots (observations)?


Lower quartile
Median
Upper quartile
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
Lower quartile
Median
Upper quartile
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000



What does a dot represent?
What does the box represent?
What does a narrow box represent?
What does a wide box represent?

Does a “big” area mean the same thing in a boxplot as it does in a histogram?

What does a short whisker represent?
What does a long whisker represent?

Second step

- Now it is SAFE to use technology!
- Distinguishing between population distributions and sample distributions.
- Key concept here is sampling variability and that when random sampling is used, sampling variability is predictable.

Activity 2

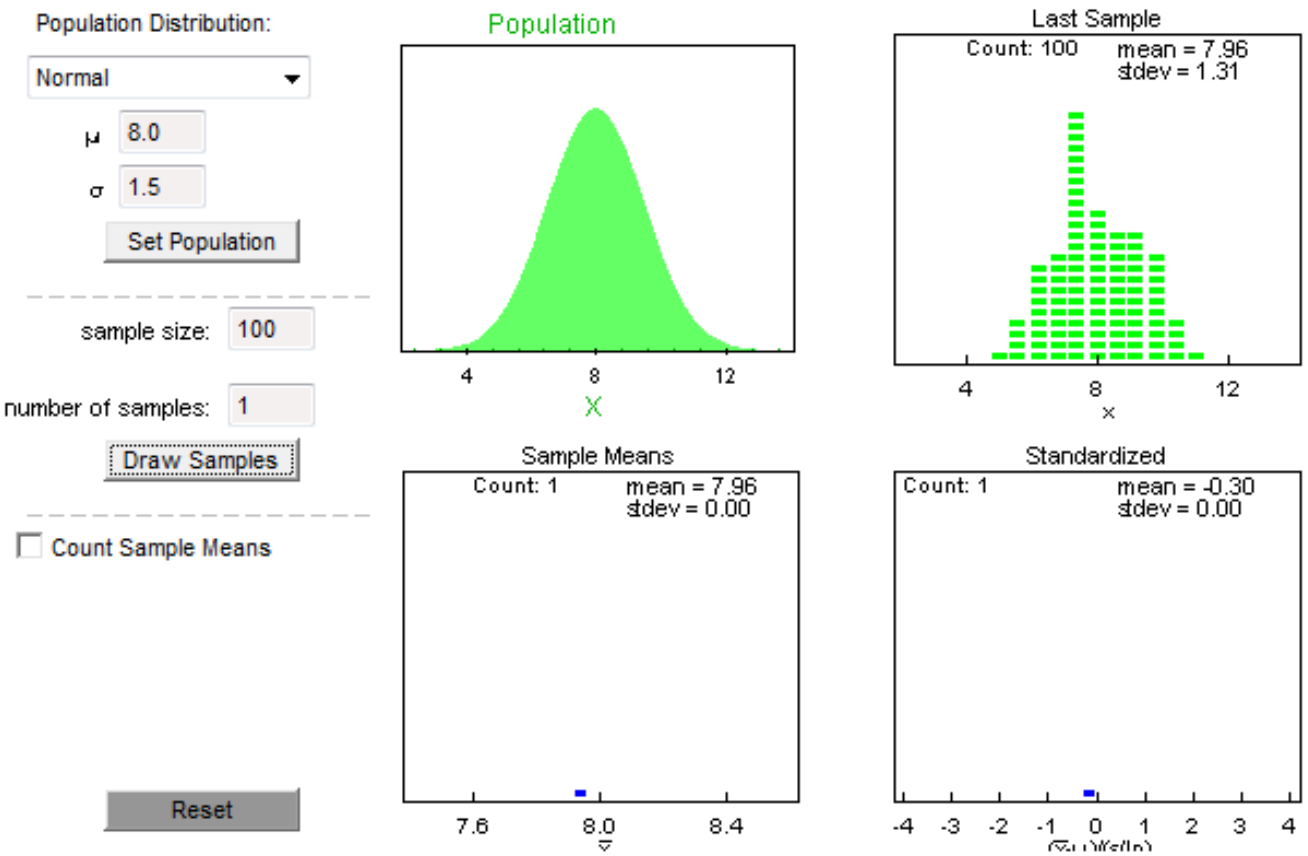
- Population distributions and sample distributions

- Technology reference:

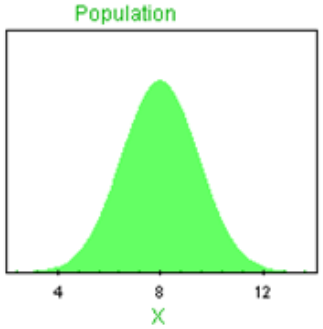
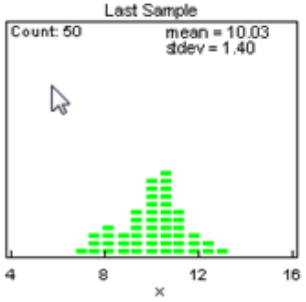
http://www.rossmanchance.com/applets/SampleMeans/SampleMeans.html

- Investigate the following question: Do random samples tend to look like the population?

Distributions of Sample Means



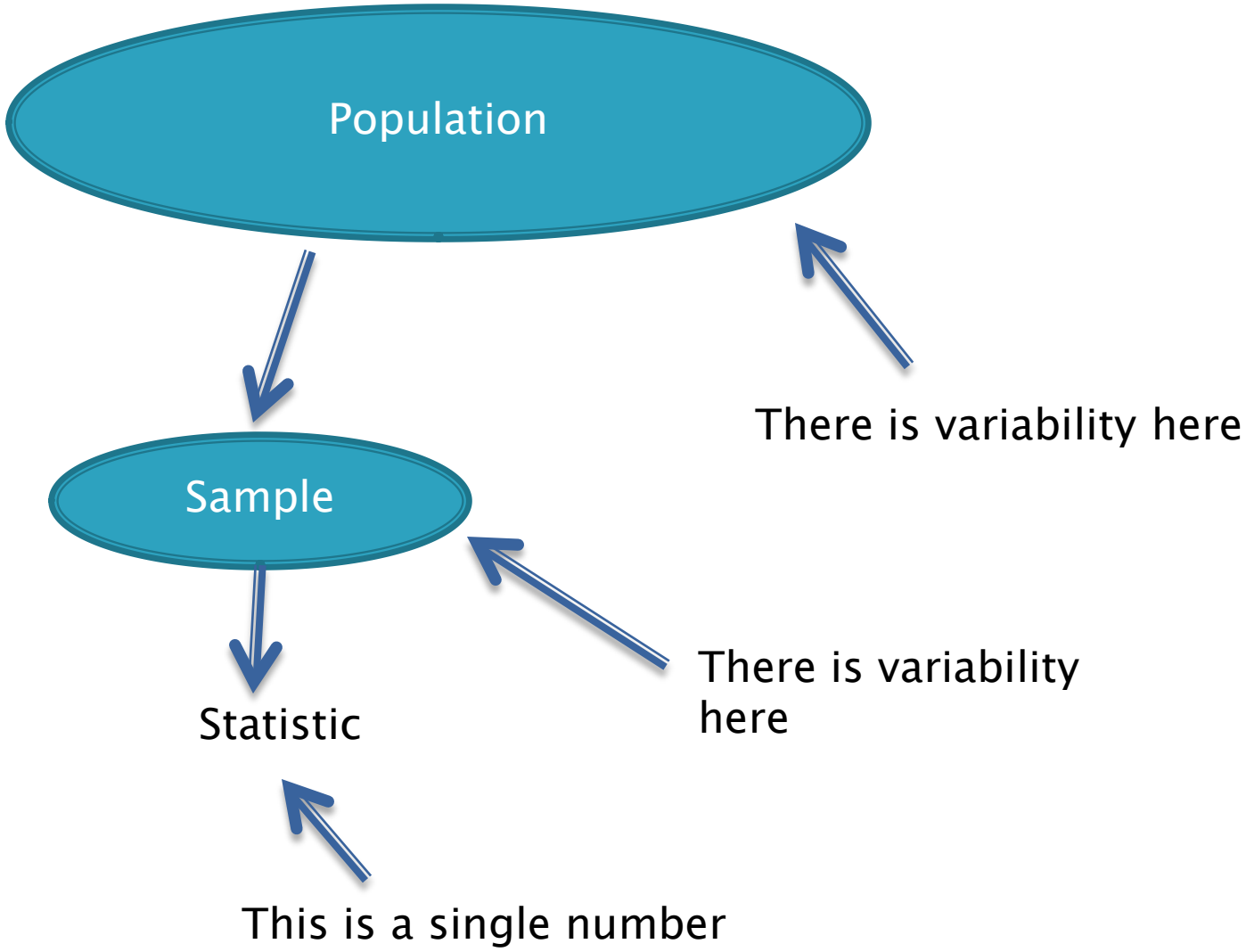
Would you be surprised if…

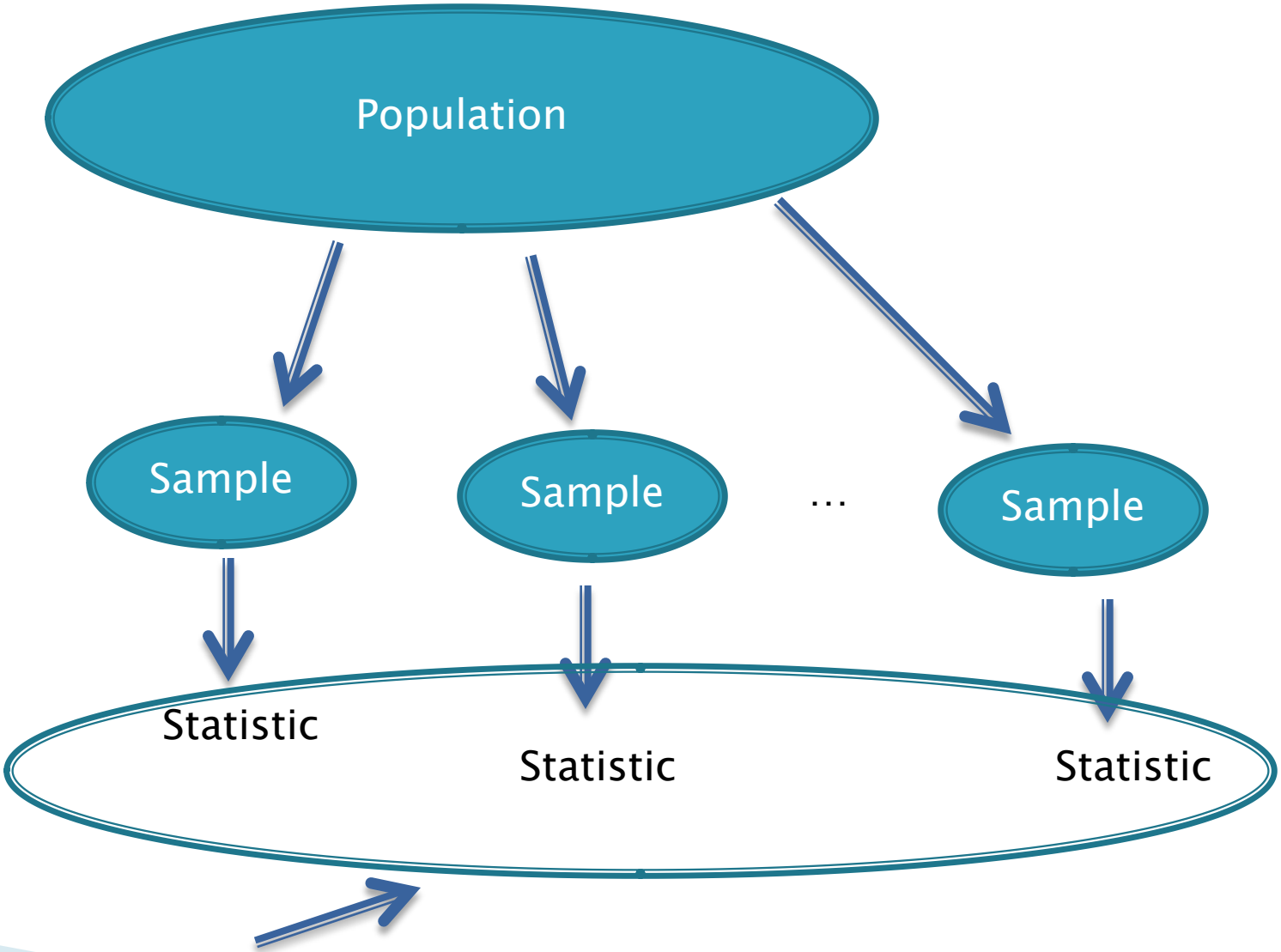
A sample looked like this	but the population looked like this?
	

That third type of distribution…

- Sampling distributions (Oh NO!!!)

- If population distributions describe variability in a population
- And Sample distributions describe variability in a sample
- What does a sampling distribution describe???


Population
There is variability here
Sample
There is variability here
Statistic
This is a single number


Population
Sample
Sample
…
Sample
Statistic
Statistic
Statistic
There is variability in these statistic values

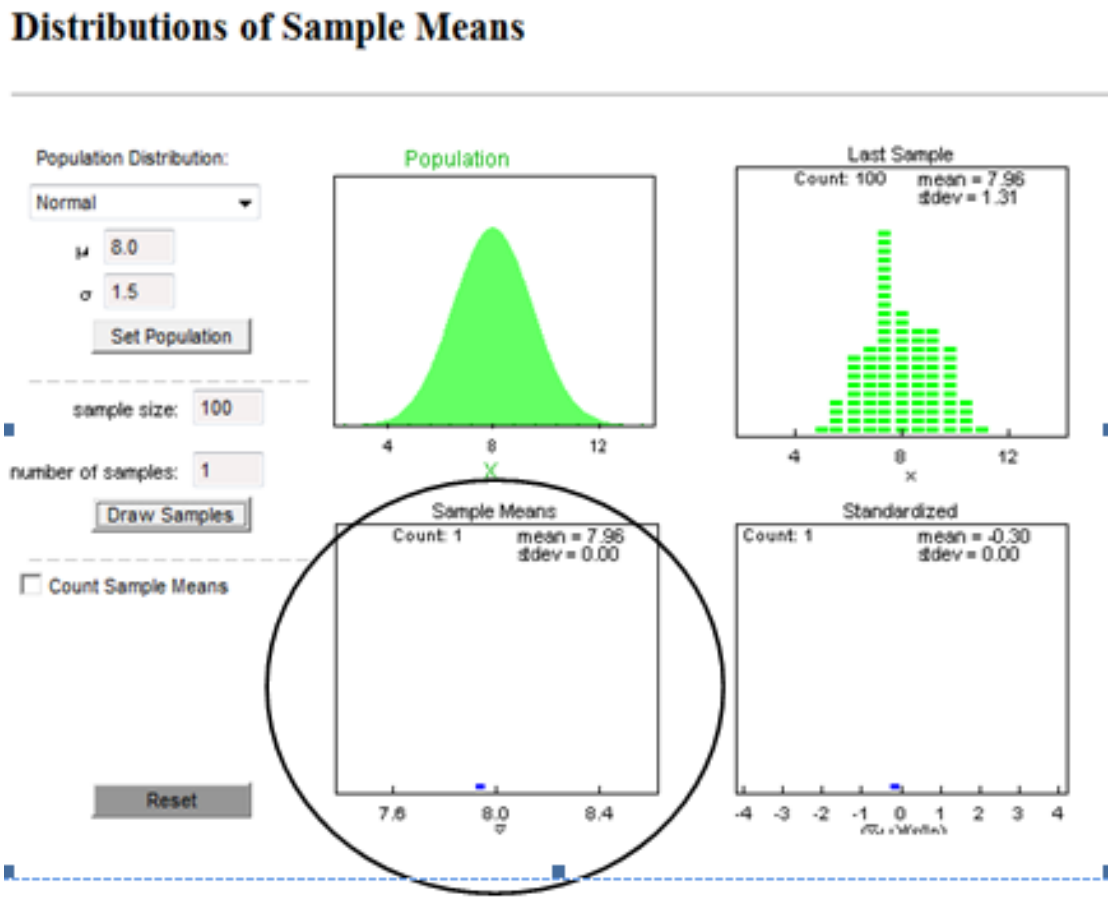
Activity 3

- Technology to the rescue!
- Technology reference:

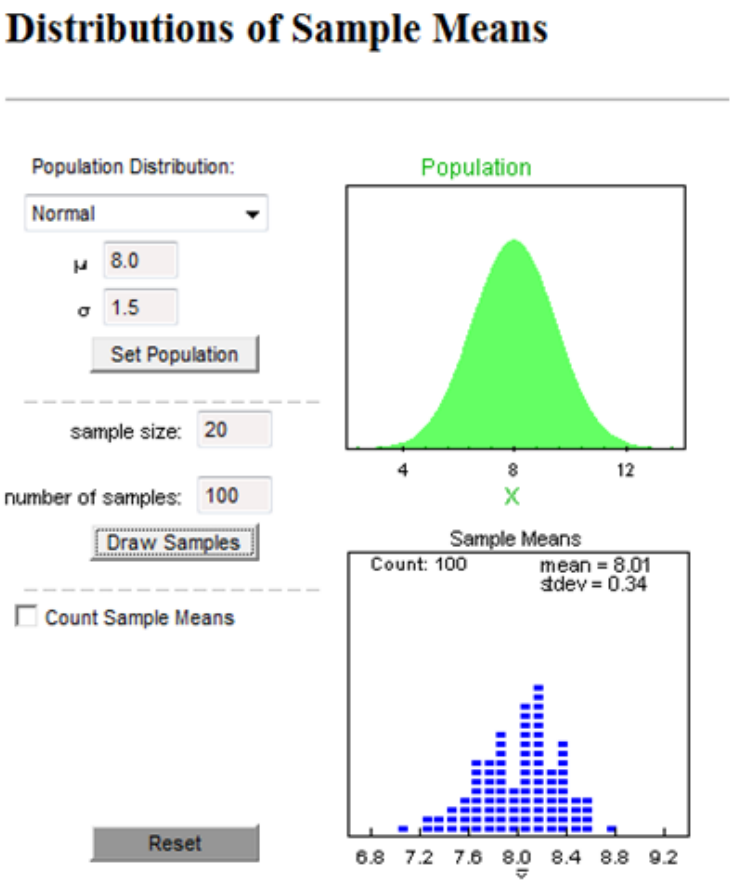
http://www.rossmanchance.com/applets/SampleMeans/SampleMeans.html

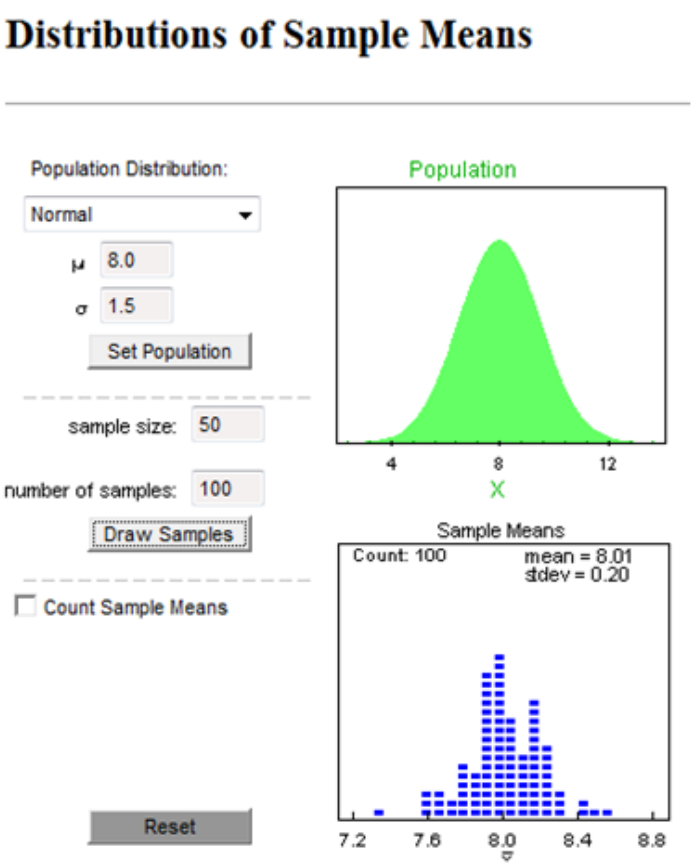
- Same applet used in Activity 2, but focus has changed.

What is happening in the lower left graph?



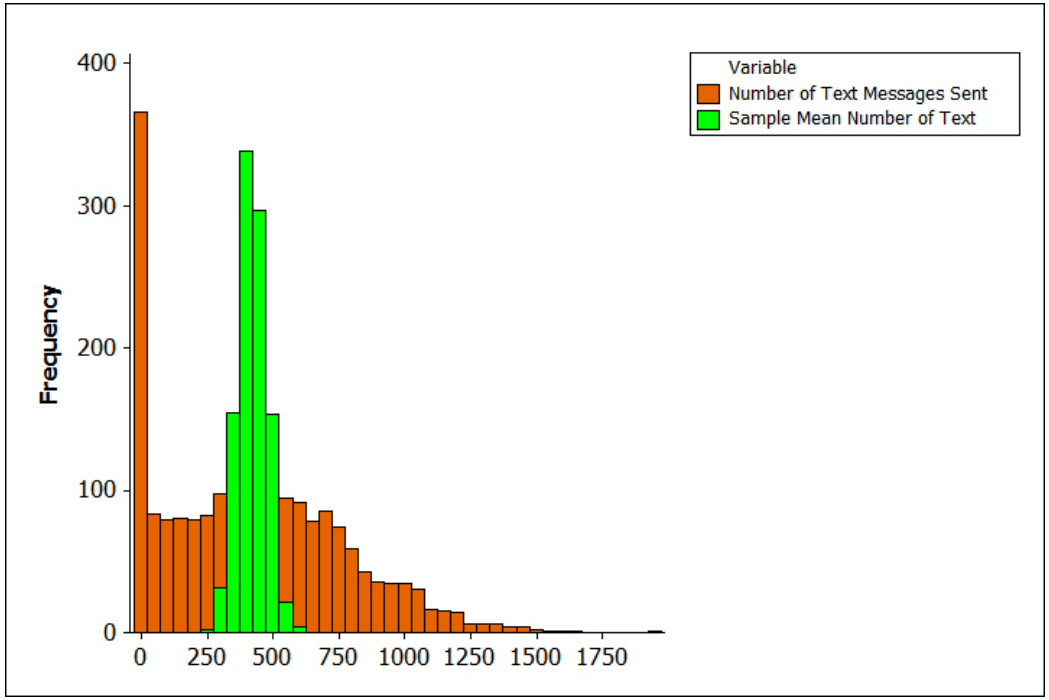
Would you be surprised if…





An important picture

For 1000 random samples



So…

- Population distributions describe
  - Variability in a population
  - Graphical displays of a population represent variability in the individual values that make up the population.
- Sample distributions describe
  - Variability in a population
  - Graphical displays of a sample represent variability in the individual values that make up the sample.
- Sampling distributions describe
  - Sample-to-sample variability in the values of a sample statistic

Thanks for coming!

- Any questions?
- Copy of slides and handout available on NCTM conference web site.
- Comments or suggestions? rpeck@calpoly.edu