SAMPLING DISTRIBUTION
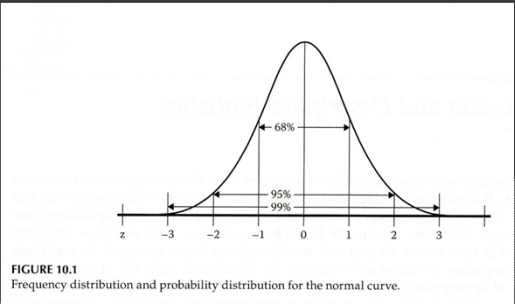
Sampling Distribution   1

Simulation

⊙ Areas under the normal curve: can be divided in terms of standard deviations (68-95-99%). These in turn can be related to probabilities

Normal distribution

Law of large numbers simulation

FIGURE 10.1
Frequency distribution and probability distribution for the normal curve.

Sampling Distribution   2

The sampling distribution

- The distribution that is approached as the number of samples approaches infinity.
  - With 5,000 to 10,000 you get a pretty good approximation
- Any statistic you can compute in a sample has a sampling distribution.
  - Mean, Median
  - Standard deviation, Variance, etc.

Sampling Distribution   3

At-home assignment

- This assignment is based on the simulations in: http://onlinestatbook.com/stat_sim/sampling_dist/index.html
- Conventions:
- The top distribution is *the parent population*.
- The second distribution (third graph from the top) is *sampling distribution 1*.
- The third distribution (last graph from the top) is *sampling distribution 2*

Sampling Distribution   4

Part 1

Set the two sampling distributions to estimate the mean. The sampling distribution 1 will be set to a sample size of 2 (i.e., N=2). Set the sampling distribution 2 to a sample size of 10 (i.e., N=10). Run 10,000 replications

- What is the value of the mean of the sampling distribution 1? Compare that value to the mean value of the parent population.
- What is the value of the standard deviation of the sampling distribution 1? Compare that value to the standard deviation of the parent population.
- What is the value of the mean of the sampling distribution 2? Compare that value to the mean value of the parent population. And to the value of the sampling distribution 1.
- What is the value of the standard deviation of the sampling distribution 2. Compare that value to the standard deviation of the parent population? Compare that value to the standard deviation of the sampling distribution 1.

Sampling Distribution   5

Part 2

Set the two sampling distributions to estimate the mean. The sampling distribution 1 will be set to a sample size of 10 (i.e., N=10). Set the sampling distribution 2 to a sample size of 25 (i.e., N=25). Run 10,000 replications

- What is the value of the mean of the sampling distribution 2? Compare that value to the mean value of the parent population? And to the value of the sampling distribution 1?
- What is the value of the standard deviation of the sampling distribution 2? Compare that value to the standard deviation of the parent population? Compare that value to the standard deviation of the sampling distribution 1

Sampling Distribution   6

Part 3

Set the two sampling distributions to estimate the mean. The sampling distribution 1 will be set to a sample size of 10 (i.e., N=10). Set the sampling distribution 2 to a sample size of 25 (i.e., N=25). Run 100,000 replications

◉ What is the value of the mean of the sampling distribution 2? Compare that value to the mean value of the parent population. And to the value of the sampling distribution 1.

◉ What is the value of the standard deviation of the sampling distribution 2? Compare that value to the standard deviation of the parent population. Compare that value to the standard deviation of the sampling distribution 1.

◉ How are these values different from those obtained when you ran the same example using 10,000 replications (problem 2)?

Sampling Distribution   7

Part 4

Set the parent population to skewed, and two sampling distributions to estimate the mean. The sampling distribution 1 will be set to a sample size of 2 (i.e., N=2). Set the sampling distribution 2 to a sample size of 25 (i.e., N=25). Run 100,000 replications

◉ What is the value of the mean of the sampling distribution 2? Compare that value to the mean value of the parent population? And to the value of the sampling distribution 1?

◉ What is the value of the standard deviation of the sampling distribution 2? Compare that value to the standard deviation of the parent population? Compare that value to the standard deviation of the sampling distribution 1?

◉ What are the values for skewness for: a) the parent population, b) the sampling distribution 1? The sampling distribution 2?

Sampling Distribution   8

Standard Error

 ◉ It is the standard deviation of the sampling distribution.
 ◉ Can you determine if it is affected by the sample size? how it is affected?

Bias

 ◉ A statistic is unbiased if the mean of the sampling distribution of the statistic (e.g., M) is the parameter (e.g., μ).
 ◉ Does the shape of the original distribution affects bias?

Sampling Distribution   9

The central limit theorem

 ◉ Define in your own words the central limit theorem, by answering this question:

 ◉ **How does the shape of the distribution affects the shape of the sampling distribution?**

Sampling Distribution   10