The Sample Mean

Perhaps the most important statistic we compute from a sample is the **sample mean**

The Sample Mean

Perhaps the most important statistic we compute from a sample is the **sample mean**

If you have the sample values entered in a spreadsheet, you can use the **AVERAGE** function to compute the sample mean.

The Sample Mean

Perhaps the most important statistic we compute from a sample is the **sample mean**

If you have the sample values entered in a spreadsheet, you can use the **AVERAGE** function to compute the sample mean.

An important idea in statistics is to think of the sample mean as a single observation from the population consisting of the *means of all possible samples*

The Sample Mean

Perhaps the most important statistic we compute from a sample is the **sample mean**

If you have the sample values entered in a spreadsheet, you can use the **AVERAGE** function to compute the sample mean.

An important idea in statistics is to think of the sample mean as a single observation from the population consisting of the *means of all possible samples*

So, we can think of the sample mean as having its own distribution, which will generally be different from that of the underlying population from which we collected the sample.

The Sample Mean

Fortunately, we can apply much of what we know about normal or bell curve populations because the distribution of the sample mean will also be a bell curve.

The Sample Mean

Fortunately, we can apply much of what we know about normal or bell curve populations because the distribution of the sample mean will also be a bell curve.

We can assume that the sample mean has a bell curve distribution in the following cases:

- 1. The population from which the sample was taken has a bell curve distribution
- 2. The sample size is at least 30

The Sample Mean

In the first case, the sample size does not matter. If the underlying population that we sampled is normal, the sample mean also has a bell curve distribution.

The Sample Mean

In the first case, the sample size does not matter. If the underlying population that we sampled is normal, the sample mean also has a bell curve distribution.

In the second case, the **central limit theorem** tells us that the distribution of the sample mean will be close enough to a bell curve distribution to treat it as one, without knowing what the underlying distribution is.

The Sample Mean

There is one important difference in the bell curve for the sample mean: the standard deviation is smaller than that of the underlying population.

The Sample Mean

There is one important difference in the bell curve for the sample mean: the standard deviation is smaller than that of the underlying population.

To get the correct standard deviation for the bell curve of the sample mean, we always divide the standard deviation of the underlying population by the **square root** of the sample size.

The Sample Mean

There is one important difference in the bell curve for the sample mean: the standard deviation is smaller than that of the underlying population.

To get the correct standard deviation for the bell curve of the sample mean, we always divide the standard deviation of the underlying population by the **square root** of the sample size.

Once we have done this, we treat the sample mean as a **single observation** from a population of sample means.

The bell curve for this population has:

- The same mean as the underlying population
- The standard deviation of the underlying population divided by $\sqrt{n}$

The Sample Mean

Once we have made the correction to the standard deviation (dividing by $\sqrt{n}$), we treat the sample mean as a single observation from a bell curve distribution just as we have done with individual observations.

The Sample Mean

Once we have made the correction to the standard deviation (dividing by $\sqrt{n}$), we treat the sample mean as a single observation from a bell curve distribution just as we have done with individual observations.

This means everything we said about areas under bell curves applies to sample means as well, once we adjust the standard deviation.

The Sample Mean

Once we have made the correction to the standard deviation (dividing by $\sqrt{n}$), we treat the sample mean as a single observation from a bell curve distribution just as we have done with individual observations.

This means everything we said about areas under bell curves applies to sample means as well, once we adjust the standard deviation.

This gives us a very powerful tool that we can use to make statements about sample means.

Normal Approximation to Binomial

When the number of trials in a binomial experiment is large, the probability distribution of the number of successes can be approximated by a normal distribution.

Normal Approximation to Binomial

When the number of trials in a binomial experiment is large, the probability distribution of the number of successes can be approximated by a normal distribution.

If n is the number of trials and p is the probability of success, the distribution of the number of successes is approximately normal with:

$$\text{mean } \mu = np \quad \text{and} \quad \sigma = \sqrt{n \cdot p(1-p)}$$

Normal Approximation

Suppose 68 percent of people have Rh positive blood types. In a sample of 1000 donors, find the approximate probability that 660 or fewer are Rh positive.

Normal Approximation

Suppose 68 percent of people have Rh positive blood types. In a sample of 1000 donors, find the approximate probability that 660 or fewer are Rh positive.

Approximate this as a normal distribution with

mean $= 1000 \cdot 0.68 = 680$ and standard deviation $= \sqrt{1000 \cdot 0.6$

so the probability is =NORMDIST(660,680,14.75,TRUE) which is 0.0875

Normal Approximation

Suppose 68 percent of people have Rh positive blood types. In a sample of 1000 donors, find the approximate probability that between 660 or and 700 are Rh positive.

Normal Approximation

Suppose 68 percent of people have Rh positive blood types. In a sample of 1000 donors, find the approximate probability that between 660 or and 700 are Rh positive.

Approximate this as a normal distribution with

mean $= 1000 \cdot 0.68 = 680$ and standard deviation $= \sqrt{1000 \cdot 0.6}$

so the probability is =NORMDIST(660,680,14.75,TRUE) -NORMDIST(660,680,14.75,TRUE) which is 0.825

Normal Approximation

Suppose 68 percent of people have Rh positive blood types. What is the 95th percentile of the distribution of the number of Rh positive individuals in samples of size 1000?

Normal Approximation

Suppose 68 percent of people have Rh positive blood types. What is the 95th percentile of the distribution of the number of Rh positive individuals in samples of size 1000?

Approximate this as a normal distribution with

mean $= 1000 \cdot 0.68 = 680$ and standard deviation $= \sqrt{1000 \cdot 0.}$

so the percentile is =NORMINV(0.95,680,14.75)=704