# Confidence Intervals for Proportions

Suppose we wish to estimate the proportion $p$ of the population that has a certain characteristic.

# Confidence Intervals for Proportions

Suppose we wish to estimate the proportion $p$ of the population that has a certain characteristic.

More often than not it will be impossible to survey the entire population:

- Often no complete list of the members of the population exists
- There may be no way of locating all of the members of the population
- The members of the population may not want to be identified
- The members of the population may not answer truthfully to protect themselves

# Confidence Intervals for Proportions

Sampling is a practical alternative in many cases.

# Confidence Intervals for Proportions

Sampling is a practical alternative in many cases.

Recall that in this situation,

- $n$ is the size of the sample
- Each individual selected is assigned a value of zero or one
- The mean $\overline{x}$ is the sum of the $x's$ divided by $n$

# Confidence Intervals for Proportions

Sampling is a practical alternative in many cases.

# Confidence Intervals for Proportions

Sampling is a practical alternative in many cases.

Recall that in this situation,

- $n$ is the size of the sample
- Each individual selected is assigned a value of zero or one
- The mean $\overline{x}$ is the sum of the $x's$ divided by $n$

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

# Confidence Intervals for Proportions

Now the estimate of the population proportion $p$ is:

$$\hat{p} = \overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

# Confidence Intervals for Proportions

Now the estimate of the population proportion $p$ is:

$$\hat{p} = \overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

The standard deviation of $\hat{p}$ is given by:

$$SD_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p}}{n}}$$

# The Normal Approximation

We treat the sample mean as a single observation drawn from a bell curve population with:

$$\text{mean } = \hat{p}$$

# The Normal Approximation

We treat the sample mean as a single observation drawn from a bell curve population with:

$$\text{mean } = \hat{p}$$

$$\text{standard deviation} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

# Confidence Intervals

A level of confidence must be specified to determine the confidence interval.

# Confidence Intervals

A level of confidence must be specified to determine the confidence interval.

Usually, this is specified as the probability that the interval *does not* contain the true population proportion $p$.

# Confidence Intervals

A level of confidence must be specified to determine the confidence interval.

Usually, this is specified as the probability that the interval *does not* contain the true population proportion $p$.

The most common values used are $0.05$ and $0.01$.

# Confidence Intervals

A level of confidence must be specified to determine the confidence interval.

Usually, this is specified as the probability that the interval *does not* contain the true population proportion $p$.

The most common values used are $0.05$ and $0.01$.

This quantity is known as the **alpha level**

# Confidence Intervals

A level of confidence must be specified to determine the confidence interval.

Usually, this is specified as the probability that the interval *does not* contain the true population proportion $p$.

The most common values used are $0.05$ and $0.01$.

This quantity is known as the **alpha level**

The alpha level is usually denoted by the Greek letter $\alpha$.

# Confidence Intervals

A level of confidence must be specified to determine the confidence interval.

Usually, this is specified as the probability that the interval *does not* contain the true population proportion $p$.

The most common values used are $0.05$ and $0.01$.

This quantity is known as the **alpha level**

The alpha level is usually denoted by the Greek letter $\alpha$.

Once the value of $\alpha$ is determined, the **confidence level** is stated as

$$100(1 - \alpha)$$

# Confidence Intervals

So if $\alpha = 0.05$, the level of confidence is

$$100(1 - \alpha) = 100(.95) = 95$$

# Confidence Intervals

So if $\alpha = 0.05$, the level of confidence is

$$100(1 - \alpha) = 100(.95) = 95$$

In this case, the confidence interval would be described as a $95\%$ confidence interval for the population proportion.

# Confidence Intervals

So if $\alpha = 0.05$, the level of confidence is

$$100(1 - \alpha) = 100(.95) = 95$$

In this case, the confidence interval would be described as a $95\%$ confidence interval for the population proportion.

The correct interpretation of the meaning of the confidence interval is the following:

If the sample was drawn over and over again and each time a confidence interval with $\alpha = 0.05$ is constructed, this interval will contain the true population paramter $95\%$ of the time.

# Confidence Intervals

The Confidence Interval is centered at $\hat{p}$

# Confidence Intervals

The Confidence Interval is centered at $\hat{p}$

The width of the confidence interval in standard deviations is the value $z_\alpha$ from the bell curve for which the proportion of the area to the right of $z_\alpha$ is $\alpha$.

# Confidence Intervals

The Confidence Interval is centered at $\hat{p}$

The width of the confidence interval in standard deviations is the value $z_\alpha$ from the bell curve for which the proportion of the area to the right of $z_\alpha$ is $\alpha$.

This can be obtained from a spreadsheet with the formula

=NORMSINV($\alpha$)

# Confidence Intervals

The Confidence Interval is centered at $\hat{p}$

The width of the confidence interval in standard deviations is the value $z_\alpha$ from the bell curve for which the proportion of the area to the right of $z_\alpha$ is $\alpha$.

This can be obtained from a spreadsheet with the formula

=NORMSINV($\alpha$)

This quantity is known as the **alpha level**

# Confidence Intervals

The lower limit of the confidence interval is:

$$\text{Lower Limit} = \hat{p} - z_\alpha \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

# Confidence Intervals

The lower limit of the confidence interval is:

$$\text{Lower Limit} = \hat{p} - z_\alpha \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Rather than try to memorize this formula, you should code it into a spreadsheet for future use.

# Confidence Intervals

The lower limit of the confidence interval is:

$$\text{Lower Limit} = \hat{p} - z_{\alpha} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Rather than try to memorize this formula, you should code it into a spreadsheet for future use.

The upper limit of the confidence interval is:

$$\text{Lower Limit} = \hat{p} + z_{\alpha} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

# Confidence Intervals

The lower limit of the confidence interval is:

$$\text{Lower Limit} = \hat{p} - z_{\alpha} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

# Confidence Intervals

The lower limit of the confidence interval is:

$$\text{Lower Limit} = \hat{p} - z_{\alpha} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Rather than try to memorize this formula, you should code it into a spreadsheet for future use.

# Confidence Intervals

The lower limit of the confidence interval is:

$$\text{Lower Limit} = \hat{p} - z_\alpha \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Rather than try to memorize this formula, you should code it into a spreadsheet for future use.

The upper limit of the confidence interval is:

$$\text{Lower Limit} = \hat{p} + z_\alpha \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

# More General Confidence Intervals

The concept of a confidence interval extends to other kinds of estimation problems.

# More General Confidence Intervals

The concept of a confidence interval extends to other kinds of estimation problems.

The mean or expected value of a measure that has a bell curve distribution is very often of interest.

# More General Confidence Intervals

The concept of a confidence interval extends to other kinds of estimation problems.

The mean or expected value of a measure that has a bell curve distribution is very often of interest.

The same problems that usually prevent us from doing a complete survey of a population also apply to the problem of estimating a mean.

# More General Confidence Intervals

The concept of a confidence interval extends to other kinds of estimation problems.

The mean or expected value of a measure that has a bell curve distribution is very often of interest.

The same problems that usually prevent us from doing a complete survey of a population also apply to the problem of estimating a mean.

As with proportions, sampling is usually a practical way to get around these problems.

# More General Confidence Intervals

Very little changes in terms of the formulas when we consider a general mean.

# More General Confidence Intervals

Very little changes in terms of the formulas when we consider a general mean.

The biggest difference is that the mean tells us nothing about the standard deviations of either the sample mean or the population.

# More General Confidence Intervals

Very little changes in terms of the formulas when we consider a general mean.

The biggest difference is that the mean tells us nothing about the standard deviations of either the sample mean or the population.

In the case of a general bell curve mean, we have to somehow come up with a value to use for the standard deviation.

# More General Confidence Intervals

Very little changes in terms of the formulas when we consider a general mean.

The biggest difference is that the mean tells us nothing about the standard deviations of either the sample mean or the population.

In the case of a general bell curve mean, we have to somehow come up with a value to use for the standard deviation.

If we can assume we know the true standard deviation, the confidence interval for the mean has the familiar form involving adding and subtracting

$$z_\alpha \cdot SD_{\overline{x}}$$

# More General Confidence Intervals

One reason we might know the standard deviation is that we are working with a population that has been studied extensively, and the standard deviation is considered to be known (IQ tests).

# More General Confidence Intervals

One reason we might know the standard deviation is that we are working with a population that has been studied extensively, and the standard deviation is considered to be known (IQ tests).

Another reason might be that we have a very large sample and can estimate it precisely.

# More General Confidence Intervals

One reason we might know the standard deviation is that we are working with a population that has been studied extensively, and the standard deviation is considered to be known (IQ tests).

Another reason might be that we have a very large sample and can estimate it precisely.

In other cases we have to estimate the standard deviation from the sample.

# More General Confidence Intervals

One reason we might know the standard deviation is that we are working with a population that has been studied extensively, and the standard deviation is considered to be known (IQ tests).

Another reason might be that we have a very large sample and can estimate it precisely.

In other cases we have to estimate the standard deviation from the sample.

If we estimate the standard deviation from the sample, the confidence interval uses the *Student's t* distribution,

$$t_\alpha \cdot SD_{\overline{x}}$$