Stat 201: Statistics I
Week 7

Metropolitan State University

Week 7
Estimating Population Parameters

Section 7.1
Sampling Distributions and Estimators

Using samples to understand populations

Recall, one of the primary functions of statistics is to use samples to learn about populations. One way this is done is by estimating population parameters from a sample.

Using samples to understand populations

Recall, one of the primary functions of statistics is to use samples to learn about populations. One way this is done is by estimating population parameters from a sample.

- A **parameter** is a value that describes a population.

Using samples to understand populations

Recall, one of the primary functions of statistics is to use samples to learn about populations. One way this is done is by estimating population parameters from a sample.

- A **parameter** is a value that describes a population.
- A **sample** is a subset of a population.

Using samples to understand populations

Recall, one of the primary functions of statistics is to use samples to learn about populations. One way this is done is by estimating population parameters from a sample.

- A **parameter** is a value that describes a population.
- A **sample** is a subset of a population.
- A **statistic** is a value calculated from the data of a sample.

Using samples to understand populations

Recall, one of the primary functions of statistics is to use samples to learn about populations. One way this is done is by estimating population parameters from a sample.

- A **parameter** is a value that describes a population.
- A **sample** is a subset of a population.
- A **statistic** is a value calculated from the data of a sample.
- An **estimator** is statistic from a sample used to estimate a population parameter.

Using samples to understand populations

Recall, one of the primary functions of statistics is to use samples to learn about populations. One way this is done is by estimating population parameters from a sample.

- A **parameter** is a value that describes a population.
- A **sample** is a subset of a population.
- A **statistic** is a value calculated from the data of a sample.
- An **estimator** is statistic from a sample used to estimate a population parameter.
  - Any statistic could be used as an estimator. The population mean could be estimated by a constant value (such as 4) or the smallest value in the sample times 2, but these are likely poor estimates.

Using samples to understand populations

Recall, one of the primary functions of statistics is to use samples to learn about populations. One way this is done is by estimating population parameters from a sample.

- A **parameter** is a value that describes a population.
- A **sample** is a subset of a population.
- A **statistic** is a value calculated from the data of a sample.
- An **estimator** is statistic from a sample used to estimate a population parameter.
    - Any statistic could be used as an estimator. The population mean could be estimated by a constant value (such as 4) or the smallest value in the sample times 2, but these are likely poor estimates.
    - A better estimate for the population mean could be the sample mean, $\bar{x}$.

Commonly used estimators

The most commonly used estimators for population parameters are often the equivalent sample statistic.

- The sample mean ($\bar{x}$) is used to estimate the population mean (μ)
- The sample standard deviation (s) is used to estimate the population standard deviation (σ)
- For binomial distributions, the sample proportion ($\hat{p}$) is used to estimate the population proportion (p)
- Since the variance and standard deviation of binomial distribution are calculated from the proportion, estimates of variance are calculated from the sample proportion

Understanding estimators

Even when using reasonable statistics as estimators, the estimates will rarely exactly match the population parameters.

- Different samples will produce different estimates.
- Therefore, it is important to understand the nature of the estimator in order to judge the quality and the meaning of the estimate.

Understanding estimators, example

Example

Suppose the data set "metro_hgts_pop.csv" on D2L contains the heights in inches for the population of male Metro State students (it doesn't). From this data, the population mean of male Metro State students is 67.42 inches.

Understanding estimators, example

Example

Suppose the data set "metro_hgts_pop.csv" on D2L contains the heights in inches for the population of male Metro State students (it doesn't). From this data, the population mean of male Metro State students is 67.42 inches.

The data set "metro_hgts_sample_stats.csv" contains the statistics from samples of 30 random heights drawn from the population. The means ($\bar{x}$) of the first 5 samples are...

Understanding estimators, example

Example

Suppose the data set “metro_hgts_pop.csv” on D2L contains the heights in inches for the population of male Metro State students (it doesn’t). From this data, the population mean of male Metro State students is 67.42 inches.

The data set “metro_hgts_sample_stats.csv” contains the statistics from samples of 30 random heights drawn from the population. The means ($\bar{x}$) of the first 5 samples are...

- 66.43333

Understanding estimators, example

Example

Suppose the data set "metro_hgts_pop.csv" on D2L contains the heights in inches for the population of male Metro State students (it doesn't). From this data, the population mean of male Metro State students is 67.42 inches.

The data set "metro_hgts_sample_stats.csv" contains the statistics from samples of 30 random heights drawn from the population. The means ($\bar{x}$) of the first 5 samples are...

- 66.43333
- 67.16667

Understanding estimators, example

Example

Suppose the data set "metro_hgts_pop.csv" on D2L contains the heights in inches for the population of male Metro State students (it doesn't). From this data, the population mean of male Metro State students is 67.42 inches.

The data set "metro_hgts_sample_stats.csv" contains the statistics from samples of 30 random heights drawn from the population. The means ($\bar{x}$) of the first 5 samples are...

- 66.43333
- 67.16667
- 68.16667

Understanding estimators, example

Example

Suppose the data set "metro_hgts_pop.csv" on D2L contains the heights in inches for the population of male Metro State students (it doesn't). From this data, the population mean of male Metro State students is 67.42 inches.

The data set "metro_hgts_sample_stats.csv" contains the statistics from samples of 30 random heights drawn from the population. The means ($\bar{x}$) of the first 5 samples are...

- 66.43333
- 67.16667
- 68.16667
- 66.03333

Understanding estimators, example

Example

Suppose the data set "metro_hgts_pop.csv" on D2L contains the heights in inches for the population of male Metro State students (it doesn't). From this data, the population mean of male Metro State students is 67.42 inches.

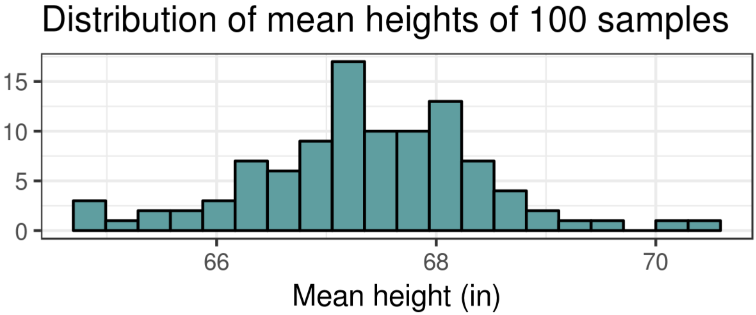
The data set "metro_hgts_sample_stats.csv" contains the statistics from samples of 30 random heights drawn from the population. The means ($\bar{x}$) of the first 5 samples are...

- 66.43333
- 67.16667
- 68.16667
- 66.03333
- 67.23333

Understanding estimators, example

Example

As more samples are gathered, the distribution of the sample means can be examined.



Sampling distributions

While it is not practical to collect many samples, much less all possible samples, a probability distribution of sample statistics can be mathematically constructed.

Sampling distributions

While it is not practical to collect many samples, much less all possible samples, a probability distribution of sample statistics can be mathematically constructed.

A **sampling distribution** is a probability distribution of a statistic from all possible samples of a certain size from a population.

Sampling distributions

While it is not practical to collect many samples, much less all possible samples, a probability distribution of sample statistics can be mathematically constructed.

A **sampling distribution** is a probability distribution of a statistic from all possible samples of a certain size from a population.

- Though sampling distributions of any estimator can be considered, the most important and commonly used is the distribution of sample means.

Sampling distributions

While it is not practical to collect many samples, much less all possible samples, a probability distribution of sample statistics can be mathematically constructed.

A **sampling distribution** is a probability distribution of a statistic from all possible samples of a certain size from a population.

- Though sampling distributions of any estimator can be considered, the most important and commonly used is the distribution of sample means.
- The distribution of sample means can be understood using the Central Limit Theorem.

Central Limit Theorem

The **Central Limit Theorem** (CLT) says, given...

Central Limit Theorem

The **Central Limit Theorem** (CLT) says, given...

- X is a random variable for a population with a mean μ and standard deviation σ

Central Limit Theorem

The **Central Limit Theorem** (CLT) says, given...

- X is a random variable for a population with a mean μ and standard deviation σ
- Sampling distribution $S_{\bar{x}}$ of sample means $\bar{x}$ for samples of size n

Central Limit Theorem

The **Central Limit Theorem** (CLT) says, given...

- X is a random variable for a population with a mean μ and standard deviation σ
- Sampling distribution $S_{\bar{x}}$ of sample means $\bar{x}$ for samples of size n

Then,

- As n increases, $S_{\bar{x}}$ approaches a normal distribution

Central Limit Theorem

The **Central Limit Theorem** (CLT) says, given...

- X is a random variable for a population with a mean μ and standard deviation σ
- Sampling distribution $S_{\bar{x}}$ of sample means $\bar{x}$ for samples of size n

Then,

- As n increases, $S_{\bar{x}}$ approaches a normal distribution
- The mean of $S_{\bar{x}}$, denoted $\mu_{\bar{x}}$, is μ

Central Limit Theorem

The **Central Limit Theorem** (CLT) says, given...

- X is a random variable for a population with a mean μ and standard deviation σ
- Sampling distribution $S_{\bar{x}}$ of sample means $\bar{x}$ for samples of size n

Then,

- As n increases, $S_{\bar{x}}$ approaches a normal distribution
- The mean of $S_{\bar{x}}$, denoted $\mu_{\bar{x}}$, is μ
- The standard deviation of $S_{\bar{x}}$, denoted $\sigma_{\bar{x}}$, is $\dfrac{\sigma}{\sqrt{n}}$

Central Limit Theorem

The **Central Limit Theorem** (CLT) says, given...

- X is a random variable for a population with a mean μ and standard deviation σ
- Sampling distribution $S_{\bar{x}}$ of sample means $\bar{x}$ for samples of size n

Then,

- As n increases, $S_{\bar{x}}$ approaches a normal distribution
- The mean of $S_{\bar{x}}$, denoted $\mu_{\bar{x}}$, is μ
- The standard deviation of $S_{\bar{x}}$, denoted $\sigma_{\bar{x}}$, is $\dfrac{\sigma}{\sqrt{n}}$
- $\sigma_{\bar{x}}$ is also known as the **standard error**

Central Limit Theorem demonstration

For a demonstration of the Central Limit Theorem in action:

- `https://seighin.shinyapps.io/clt_demo/`

Central Limit Theorem, example

Example

The fake population of heights of male Metro State students has a mean of 67.42 and a standard deviation of 5.28.

What is the probability that a sample of 30 students has a mean height of at least 69.2 inches (the mean height of adult males in the U.S.)?

Central Limit Theorem, example

Example

The fake population of heights of male Metro State students has a mean of 67.42 and a standard deviation of 5.28.

What is the probability that a sample of 30 students has a mean height of at least 69.2 inches (the mean height of adult males in the U.S.)?

- The mean of the sampling distribution ($\mu_{\bar{x}}$) is the population mean (μ) 67.42

Central Limit Theorem, example

Example

The fake population of heights of male Metro State students has a mean of 67.42 and a standard deviation of 5.28.

What is the probability that a sample of 30 students has a mean height of at least 69.2 inches (the mean height of adult males in the U.S.)?

- The mean of the sampling distribution ($\mu_{\bar{x}}$) is the population mean (μ) 67.42
- The standard deviation of the sampling distribution ($\sigma_{\bar{x}}$), or standard error, is $\dfrac{5.28}{\sqrt{30}} = 0.964$

Central Limit Theorem, example

Example

The fake population of heights of male Metro State students has a mean of 67.42 and a standard deviation of 5.28.

What is the probability that a sample of 30 students has a mean height of at least 69.2 inches (the mean height of adult males in the U.S.)?

- The mean of the sampling distribution ($\mu_{\bar{x}}$) is the population mean (μ) 67.42
- The standard deviation of the sampling distribution ($\sigma_{\bar{x}}$), or standard error, is $\dfrac{5.28}{\sqrt{30}} = 0.964$
- $P(S_{\bar{x}} > 69.2) = 0.0324$

Thoughts on CLT

- If the population X has a normal distribution, the sampling distribution $S_{\bar{x}}$ will have a normal distribution regardless of sample size n.

Thoughts on CLT

- If the population X has a normal distribution, the sampling distribution $S_{\bar{x}}$ will have a normal distribution regardless of sample size n.
- If X is not normally distributed, how normal $S_{\bar{x}}$ is, or how quickly it becomes normal as n increases, depends on how not normal X is.

Thoughts on CLT

- If the population X has a normal distribution, the sampling distribution $S_{\bar{x}}$ will have a normal distribution regardless of sample size n.
- If X is not normally distributed, how normal $S_{\bar{x}}$ is, or how quickly it becomes normal as n increases, depends on how not normal X is.
- The rule of thumb generally used is, if sample size is $n = 30$ or greater, $S_{\bar{x}}$ can be considered normal.

Thoughts on CLT

- If the population X has a normal distribution, the sampling distribution $S_{\bar{x}}$ will have a normal distribution regardless of sample size n.
- If X is not normally distributed, how normal $S_{\bar{x}}$ is, or how quickly it becomes normal as n increases, depends on how not normal X is.
- The rule of thumb generally used is, if sample size is $n = 30$ or greater, $S_{\bar{x}}$ can be considered normal.

Remember. . .

The Central Limit Theorem applies to the distribution of estimators from samples, not the distribution of individual samples.

Group work

- Do the first part of the group work.

Section 7.2
Confidence Intervals

Confidence intervals

A **confidence interval** is a range of values that is likely to contain the value of the parameter.

- A confidence interval, defined for a specific confidence level, is constructed with a point estimate and a margin of error.
- The **confidence level** is the chosen level of certainty that the interval will contain the true parameter.
- The **point estimate** is the value the estimator, the sample statistic used to estimate the population parameter. For example, $\bar{x}$.
- The **margin of error** is the amount the lower and upper bounds of the interval differ from the point estimate.

Confidence levels

The confidence level is the degree of certainty that a confidence interval constructed from a random sample actually contains the true population parameter.

Confidence levels

The confidence level is the degree of certainty that a confidence interval constructed from a random sample actually contains the true population parameter.

- The confidence level is chosen before the interval is calculated

Confidence levels

The confidence level is the degree of certainty that a confidence interval constructed from a random sample actually contains the true population parameter.

- The confidence level is chosen before the interval is calculated
- Expressed as a percent, in terms of α as $(1 - \alpha)\%$

Confidence levels

The confidence level is the degree of certainty that a confidence interval constructed from a random sample actually contains the true population parameter.

- The confidence level is chosen before the interval is calculated
- Expressed as a percent, in terms of α as $(1 - \alpha)\%$
- Higher confidence levels (i.e. more certainty) will result in larger intervals (a wider range of values)

Margin of error

The margin of error can be thought of as the size of uncertainty in an estimate. It is calculated in the context of the sampling distribution, the confidence level and the sample standard deviation and size.

Margin of error

The margin of error can be thought of as the size of uncertainty in an estimate. It is calculated in the context of the sampling distribution, the confidence level and the sample standard deviation and size.

$$ME = z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

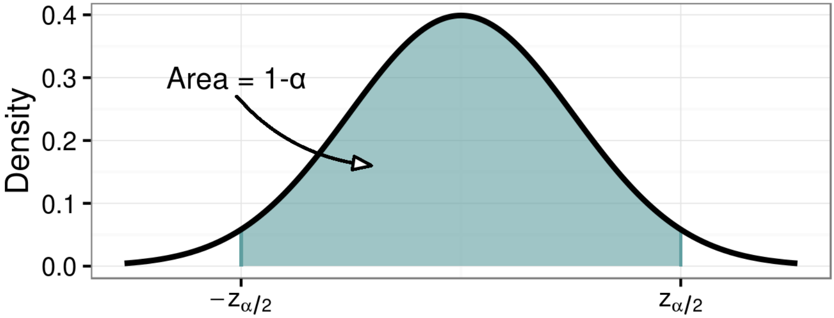
Where...

- $z_{\alpha/2}$ is the two sided critical z value at α level of significance
- s is the sample standard deviation
- n is the sample size

Critical values

Recall, for a significance level α, the critical values $z_{\alpha/2}$ and $-z_{\alpha/2}$ separate the bulk of the distribution from the lowest and highest values comprising a total proportion of α of the distribution.

Thus, between the critical values is an area or probability of $(1 - \alpha)$.



Standard normal critical values

Significance Level (α)	Confidence Level ($1-\alpha$)	Critical Value ($z_{\alpha/2}$)
0.10	90%	1.645
0.05	95%	1.96
0.01	99%	2.576

Confidence interval definition

A confidence interval describes a range of numeric values. There are two common ways to display a confidence interval:

- (L, U), where L is the lower bound and U is the upper bound
- $x \pm ME$, where x is the point estimate and ME is the margin of error

Confidence interval definition

A confidence interval describes a range of numeric values. There are two common ways to display a confidence interval:

- (L, U), where L is the lower bound and U is the upper bound
- $x \pm ME$, where x is the point estimate and ME is the margin of error

A confidence interval at confidence level $(1-\alpha)$%, given a sample of size n with point estimate x and standard deviation s, is

$$CI(1-\alpha)\% = x \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Confidence interval definition

A confidence interval describes a range of numeric values. There are two common ways to display a confidence interval:

- (L, U), where L is the lower bound and U is the upper bound
- $x \pm ME$, where x is the point estimate and ME is the margin of error

A confidence interval at confidence level $(1-\alpha)\%$, given a sample of size n with point estimate x and standard deviation s, is

$$CI(1-\alpha)\% = x \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

or

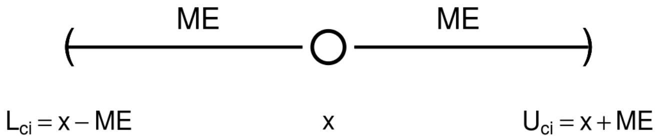
$$\left(x - z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right), x + z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)\right)$$

Find point estimate and margin of error

Given a confidence interval (L, U), the point estimate and margin of error can be calculated.

$$\text{Point estimate: } x = \frac{L+U}{2}$$

$$\text{Margin of error: } ME = \frac{U-L}{2}$$



Find point estimate and margin of error, example

Example

Suppose a confidence interval for the heights of male Metro State students of $(64.52, 68.35)$.

What is the point estimate and margin of error of this confidence interval?

Find point estimate and margin of error, example

Example

Suppose a confidence interval for the heights of male Metro State students of $(64.52, 68.35)$.

What is the point estimate and margin of error of this confidence interval?

- Point estimate:

$$\bar{x} = \frac{L+U}{2} = \frac{64.52 + 68.35}{2} = 66.44 \text{ inches}$$

Find point estimate and margin of error, example

Example

Suppose a confidence interval for the heights of male Metro State students of $(64.52, 68.35)$.

What is the point estimate and margin of error of this confidence interval?

- Point estimate:

$$\bar{x} = \frac{L+U}{2} = \frac{64.52+68.35}{2} = 66.44 \text{ inches}$$

- Margin of error:

$$ME = \frac{U-L}{2} = \frac{68.35-64.52}{2} = 1.9 \text{ inches}$$

Find point estimate and margin of error, example

Example

Suppose a confidence interval for the heights of male Metro State students of $(64.52, 68.35)$.

What is the point estimate and margin of error of this confidence interval?

- Point estimate:

$$\bar{x} = \frac{L+U}{2} = \frac{64.52 + 68.35}{2} = 66.44 \text{ inches}$$

- Margin of error:

$$ME = \frac{U-L}{2} = \frac{68.35 - 64.52}{2} = 1.9 \text{ inches}$$

- So we can state this confidence interval as

$$\bar{x} \pm ME \Rightarrow 66.44 \pm 1.9 \text{ inches}$$

Interpreting confidence intervals

- It is incorrect to say: "There is a 95% chance the true parameter is in the interval."

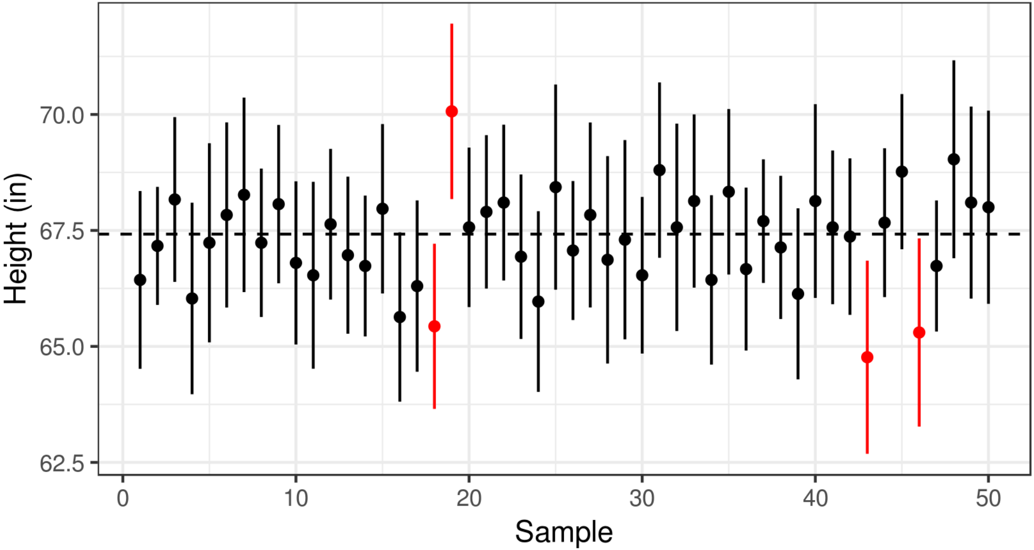
Interpreting confidence intervals

- It is incorrect to say: "There is a 95% chance the true parameter is in the interval."
- Rather: "We are 95% confident that the interval contains the true parameter."

Interpreting confidence intervals

- It is incorrect to say: "There is a 95% chance the true parameter is in the interval."
- Rather: "We are 95% confident that the interval contains the true parameter."
- Or: "When constructing intervals from random samples with this method, 95% of the time the interval will contain the true parameter."

Interpreting confidence intervals, demo



Parameters of population proportions

A population proportion p can be estimated by the point estimate of the sample proportion $\hat{p}$ which follows a normal distribution.

Parameters of population proportions

A population proportion p can be estimated by the point estimate of the sample proportion $\hat{p}$ which follows a normal distribution.

- More precisely, a sample proportion follows a binomial distribution, which approximates a normal distribution as n increases.

Parameters of population proportions

A population proportion p can be estimated by the point estimate of the sample proportion $\hat{p}$ which follows a normal distribution.

- More precisely, a sample proportion follows a binomial distribution, which approximates a normal distribution as n increases.
- The variance of $\hat{p}$ is $s^2 = \hat{p}(1 - \hat{p}) = \hat{p}\hat{q}$

Parameters of population proportions

A population proportion p can be estimated by the point estimate of the sample proportion $\hat{p}$ which follows a normal distribution.

- More precisely, a sample proportion follows a binomial distribution, which approximates a normal distribution as n increases.
- The variance of $\hat{p}$ is $s^2 = \hat{p}(1 - \hat{p}) = \hat{p}\hat{q}$
- The standard deviation of $\hat{p}$ is $s = \sqrt{\hat{p}\hat{q}}$

Confidence intervals of proportions

A confidence interval of a population proportion with confidence level $(1-\alpha)\%$ from a sample of size n and sample proportion $\hat{p}$ is

$$CI = \hat{p} \pm ME = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Confidence intervals of proportions, example

Example

Suppose 100 Metro State students were asked if they had eaten a taco in the past week. 36 students responded they had in fact eaten a taco. What is a 95% confidence interval for the proportion of all Metro State students who have eaten a taco in the past week.

Confidence intervals of proportions, example

Example

Suppose 100 Metro State students were asked if they had eaten a taco in the past week. 36 students responded they had in fact eaten a taco. What is a 95% confidence interval for the proportion of all Metro State students who have eaten a taco in the past week.

- 95% confidence level means $\alpha = 0.05$

Confidence intervals of proportions, example

Example

Suppose 100 Metro State students were asked if they had eaten a taco in the past week. 36 students responded they had in fact eaten a taco. What is a 95% confidence interval for the proportion of all Metro State students who have eaten a taco in the past week.

- 95% confidence level means $\alpha = 0.05$
- $\hat{p} = \dfrac{36}{100} = 0.36$

Confidence intervals of proportions, example

Example

Suppose 100 Metro State students were asked if they had eaten a taco in the past week. 36 students responded they had in fact eaten a taco. What is a 95% confidence interval for the proportion of all Metro State students who have eaten a taco in the past week.

- 95% confidence level means $\alpha = 0.05$
- $\hat{p} = \dfrac{36}{100} = 0.36$
- $ME = z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}} = (1.96)\sqrt{\dfrac{(0.36)(0.64)}{100}} = 0.094$

Confidence intervals of proportions, example

Example

Suppose 100 Metro State students were asked if they had eaten a taco in the past week. 36 students responded they had in fact eaten a taco. What is a 95% confidence interval for the proportion of all Metro State students who have eaten a taco in the past week.

- 95% confidence level means $\alpha = 0.05$
- $\hat{p} = \dfrac{36}{100} = 0.36$
- $ME = z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}} = (1.96)\sqrt{\dfrac{(0.36)(0.64)}{100}} = 0.094$
- $CI = \hat{p} \pm ME = 0.36 \pm 0.094 = (0.266, 0.454)$

Confidence intervals of proportions, example

Example

Suppose 100 Metro State students were asked if they had eaten a taco in the past week. 36 students responded they had in fact eaten a taco. What is a 95% confidence interval for the proportion of all Metro State students who have eaten a taco in the past week.

- 95% confidence level means $\alpha = 0.05$
- $\hat{p} = \dfrac{36}{100} = 0.36$
- $ME = z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}} = (1.96)\sqrt{\dfrac{(0.36)(0.64)}{100}} = 0.094$
- $CI = \hat{p} \pm ME = 0.36 \pm 0.094 = (0.266, 0.454)$

We are 95% confident that the true proportion of Metro State students who have eaten a taco in the last week in between 0.266 and 0.454.

Group work

- Do the second part of the group work.