Stat 201: Statistics I
Chapter 7

Metropolitan State University

Chapter 7
Estimating Parameters and Determining Samples Sizes

Confidence intervals

Recall, sample statistics can be used as estimators for population parameters. However, estimators are rarely exactly equal to the parameter.

It is more informative to calculate a **confidence interval**, a range of values that is likely to contain the value of the parameter.

Components of confidence intervals

A confidence interval, defined for a specific confidence level, is constructed with a point estimate and a margin of error.

A **confidence level** is an indication of the 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 probability that a confidence interval constructed from a random sample actually contains the true population parameter.

- Expressed as a percent, in terms of α as $(1-\alpha)\%$
- It would be 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."

Margin of error

The margin of error can be thought of as the amount 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.

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

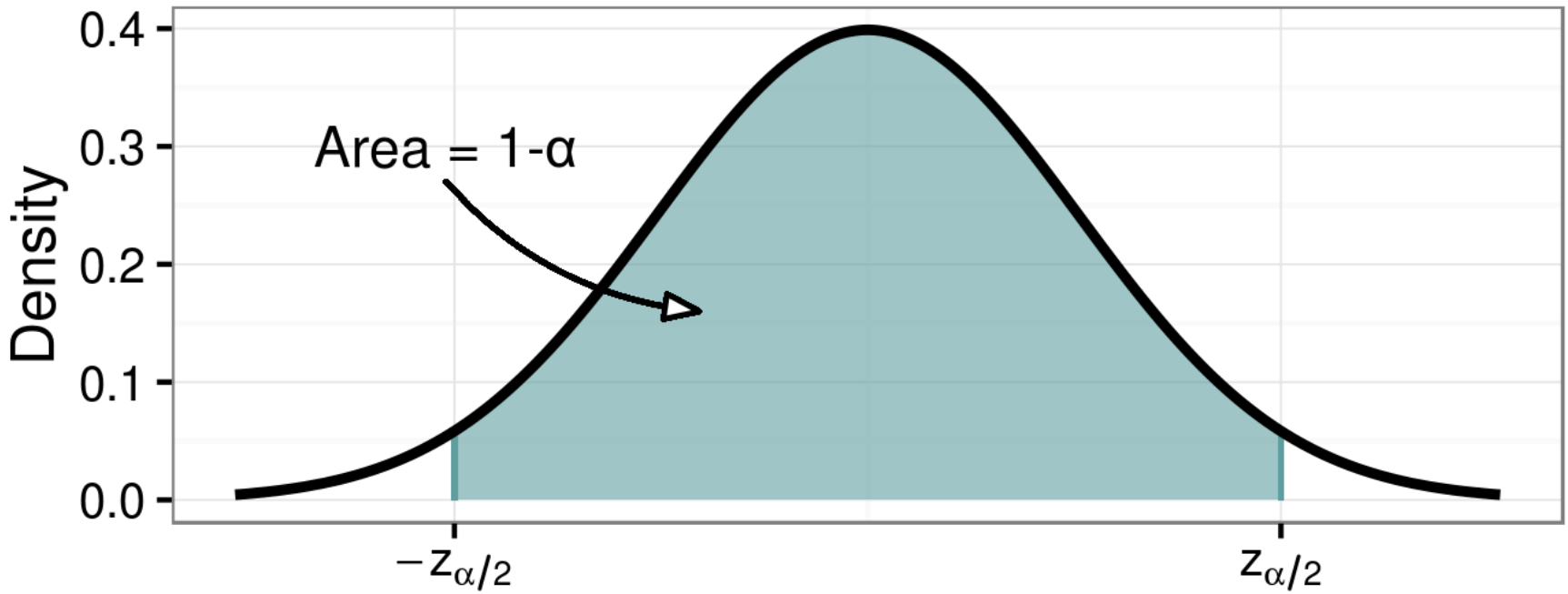
Where...

- $z_{\alpha/2}$ 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

Sig. Level (α)	Conf. 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 at confidence level $(1-\alpha)\%$, given a sample of size n with point estimate x and standard deviation s, is

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

or

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

Inference using confidence intervals

If a confidence interval does not contain a value of interest, then it can be said there is evidence that the sample was drawn from a population whose parameter is different than the value of interest.

Example

Recall, in the United States, adult men have a mean height of 69.2 in with a standard deviation of 5.79 in.

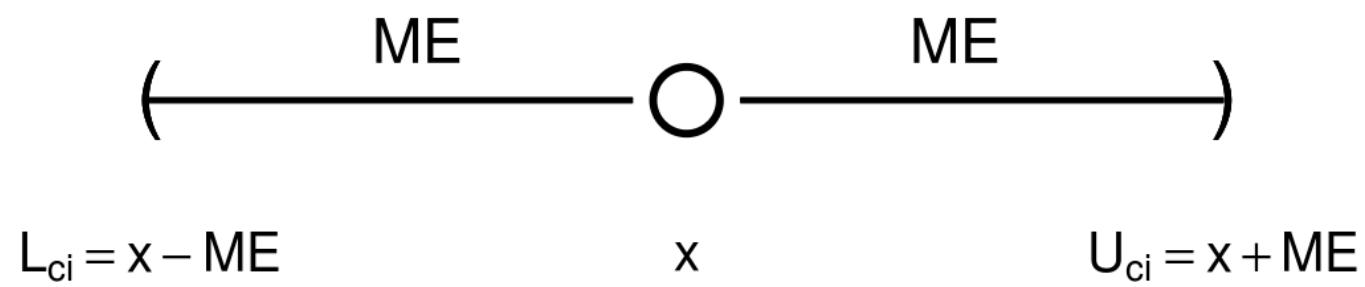
If a confidence interval for mean height calculated from a sample of male Metro State students is $(62.3, 67.9)$, then there is evidence that male Metro State students are shorter then the general US population.

Find point estimate and margin of error

Given a confidence interval (L_{ci}, U_{ci}), the point estimate and margin of error can be calculated.

$$\text{Point estimate: } x = \frac{L_{ci} + U_{ci}}{2}$$

$$\text{Margin of error: } ME = \frac{U_{ci} - L_{ci}}{2}$$



Find point estimate and margin of error, example

Example

From the previous example, we had a confidence interval for the heights of male Metro State students of $(62.3, 67.9)$ or $62.3 < \mu < 67.9$.

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

- Point estimate:

$$\bar{x} = \frac{L_{ci} + U_{ci}}{2} = \frac{62.3 + 67.9}{2} = 65.1 \text{ inches}$$

- Margin of error:

$$ME = \frac{U_{ci} - L_{ci}}{2} = \frac{67.9 - 62.3}{2} = 2.8 \text{ inches}$$

- So we can state this confidence interval as

$$\bar{x} \pm ME \Rightarrow 65.1 \pm 2.8 \text{ inches}$$

Sample size

When designing experiments, sample sizes are determined in order to achieve a desired accuracy. In other words, an acceptable margin of error is used to calculate the needed sample size.

Using the previous definition of margin of error,

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

after some algebra,

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

Section 7.1
Estimating a Population Proportion

Population proportions

Recall, 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.

- 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.

Inference of population proportion

A confidence interval can be used to make inferences about a population that a sample is drawn from by testing whether the CI contains a known value.

For example,

- To test whether a sub-population is similar to the larger population
- To test whether an intervention changed attitudes, actions or outcomes

If the known parameter value (the proportion of the larger population, or the proportion before the intervention) **is not** contained in the confidence interval, it can be said there is evidence of a change.

If the known value **is** within the interval, then there is not evidence of a difference.

Inference of population proportion, example

Example

It is thought that 30% of people will have eaten at least one taco in a given week. The Tortilla And Cheese Organization (TACO) would like to increase that. After an intensive taco promotion campaign, they survey a random sample of 55 Metro State students. 38% of them report eating a taco in the previous week. Was the campaign successful? (Test at a 95% confidence level.)

- Number of successes: $(0.38) \times 55 = 20.9 \approx 21$
- Confidence interval (from StatCrunch): $(0.253, 0.510)$
- The interval contains 30% (0.3). Thus, there is no evidence the taco promotion campaign was successful.

Point estimate and margin of error, example

Example

The survey to test the effectiveness of the taco promotion campaign found a confidence interval of $(0.253, 0.510)$.

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

- Point estimate: $\hat{p} = \dfrac{L_{ci} + U_{ci}}{2} = \dfrac{0.253 + 0.510}{2} = 0.3815$
- Margin of error: $ME = \dfrac{U_{ci} - L_{ci}}{2} = \dfrac{0.510 - 0.253}{2} = 0.1285$

Find needed sample size

The minimum sample size needed to obtain a specific margin of error is calculated by,

$$n = \left(\frac{z_{\alpha/2} \times s}{ME} \right)^2 \qquad \text{where} \qquad s = \sqrt{\hat{p}\hat{q}}$$

An estimated proportion $\hat{p}$ is needed.

- Often a $\hat{p}$ can be obtained from previous studies or expert knowledge.
- If no reasonable proportion estimate can be determined, use $\hat{p} = \hat{q} = 0.5$

Then,

$$n = \left(\frac{z_{\alpha/2} \times \sqrt{\hat{p}\hat{q}}}{ME} \right)^2$$

Find needed sample size, example

Example

TACO, disappointed by the large confidence interval of there first survey, decide to do another. This time they wish to get an estimate with a 4% margin of error with 95% confidence level. That is, they want a confidence interval of $\hat{p} \pm 0.04$.

What sample size is needed?

- The pre-intervention population proportion of 30% can be used as $\hat{p}$ for this calculation. Then, $\hat{p} = 0.3$ and $\hat{q} = 0.7$.
- $n = \left(\frac{z_{\alpha/2} \times \sqrt{\hat{p}\hat{q}}}{ME} \right)^2 = \left(\frac{1.96 \times \sqrt{(0.3)(0.7)}}{0.04} \right)^2$

  $n = 504.21 \rightarrow 505$

Section 7.2
Estimating a Population Mean

Population means

Recall, a sampling distribution of sample means is normal if the population is normally distributed or, by the Central Limit Theorem, approximately normal if the sample size is 30 or greater.

Under those conditions, the point estimate for the population mean is the sample mean and a confidence interval can be constructed.

However, the is one more factor to consider...

Population standard deviation

- If the population standard deviation σ is known, confidence intervals are calculated with critical values from the standard normal distribution and the population standard deviation. That is,

$$CI = \bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

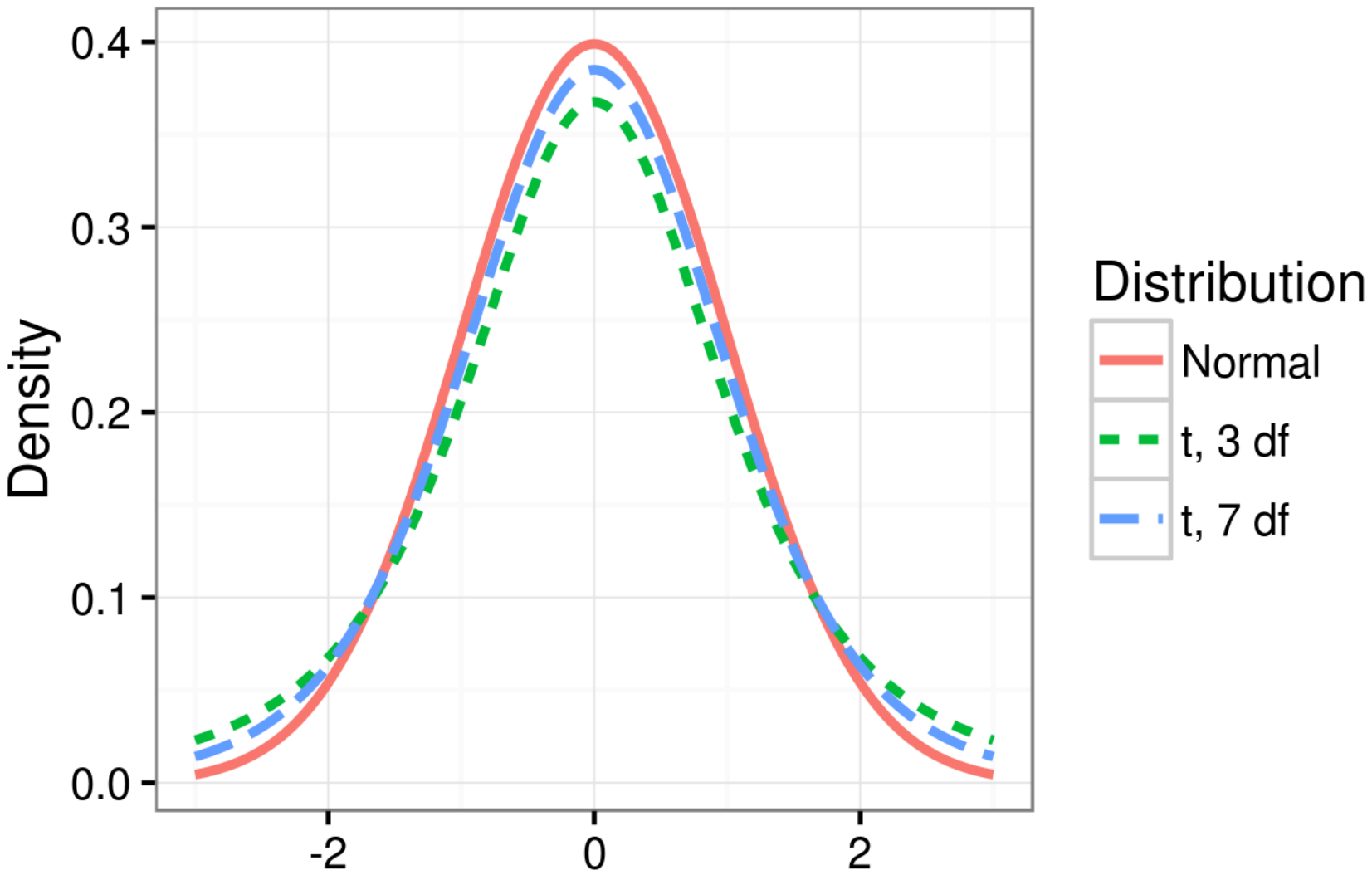
- If the population standard deviation σ is not known, the sample standard deviation is used and critical values are pulled from *Student's t distribution*.

Student's t distribution

Student's t distribution is similar to a normal distribution, except with an adjusted shape to account for an estimated standard deviation.

- The t distribution has an added parameter known as the degrees of freedom (df).
- The degrees of freedom for a sampling distribution is defined as sample size minus one ($df = n - 1$).
- As degrees of freedom increases, the t distribution approaches a normal distribution.

Student's t distribution



t distribution critical values

Critical values from t distributions can be found in tables (Table A-3) or in the StatCrunch "T" calculator.

Then, confidence interval can be calculated by,

$$CI = \bar{x} \pm t_{\alpha/2,df} \left(\frac{s}{\sqrt{n}} \right)$$

where $t_{\alpha/2,df}$ is the critical t value at $\alpha/2$ and $df = n - 1$ and s is the sample standard deviation.

Confidence intervals for means, summary

If population is normally distributed, or sample size is 30 or greater,

- If population standard deviation σ is known, confidence intervals are constructed with z distribution critical values and σ for standard deviation.
- If population standard deviation is **not** known, confidence intervals are constructed with t distribution critical values and sample standard deviation s for standard deviation.

Otherwise, if population is not normally distributed and sample size is less than 30, valid confidence intervals can not be constructed using these methods.

Confidence intervals for means, example

Example

TACO would like to know, among people who eat tacos, how many tacos per week they eat. They survey 36 taco eaters and get a sample mean of 5.4 tacos per week with a standard deviation of 2.7. What is a 90% confidence interval for the population mean number of tacos eaten?

- A 90% level of confidence means that $\alpha = 0.10$.
- It is unknown if number of tacos eaten is normally distributed (probably not), but our sample size is above 30, so we can treat the sampling distribution as normal.
- Since population standard deviation is unknown, we will use t distribution critical values with degrees of freedom of $n - 1 = 35$, $t_{\alpha/2,df} = t_{0.05,35} = 1.69$.
- $CI = \bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) = 5.4 \pm (1.69)\left(\frac{2.7}{\sqrt{36}}\right) = 5.4 \pm 0.76$

  $\mathbf{(4.64, 6.16)}$

Confidence intervals for means, example

Example

Recall, in the United States adult men have a mean height of 69.2 inches with a standard deviation of 5.79 inches.

The heights from a sample of 40 male Metro State students are measured. The mean height from the sample is 66.3 inches. What is a 95% confidence interval for the mean height of Metro State students?

- A 95% level of confidence means that $\alpha = 0.05$.
- We can assume that the heights of Metro State students have the same standard deviation as the general US population, $\sigma = 5.79$.
- Since we know the population standard deviation, we will use z distribution critical values, $z_{\alpha/2} = 1.96$.
- $CI = \bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = 66.3 \pm (1.96)\left(\frac{5.79}{\sqrt{40}}\right) = 66.3 \pm 1.794$

  $\mathbf{(64.506, 68.094)}$

Confidence intervals for means, example

Example

With a confidence interval of $(64.51, 68.01)$, can we conclude that the heights of Metro State students are different than the general US population?

- Since the US mean height of 69.2 inches is not in our interval, we can conclude (with 95% certainty) that the heights of Metro State students differ from the general population.

Find needed sample size

To calculated the sample size needed for a desired margin of error,

$$n = \left(\frac{z_{\alpha/2} \times \sigma}{ME} \right)^2$$

- Because sample size calculations are done before data is gathered, so value for standard deviation must be known or estimated.
- Because t distribution values depend of sample size, they are difficult to work with while calculating sample size. Thus, z critical values are generally used.

Find needed sample size, example

Example

Recall, in the United States, adult women have a mean height of 63.7 in with a standard deviation of 5.96 in. If we wanted to find the mean height of female students at Metro State within plus or minus 1.5 inches at a 99% confidence level, how many students would need to be included in the sample?

- A 99% level of confidence means that $\alpha = 0.01$.
- $n = \left(\frac{z_{\alpha/2} \times \sigma}{ME} \right)^2 = \left(\frac{2.576 \times 5.96}{1.5} \right)^2 = 104.76 \rightarrow 105$