Chapter 8.2

Estimating μ When σ is Unknown

Learning Objectives

At the end of this lecture, the student should be able to:

- State one way in which the z distribution is like the t distribution, and one way in which they are different.
- Demonstrate how to find t_c in the t table
- State how to calculate degrees of freedom
- Explain how to calculate a confidence interval when σ is not known

Introduction

- How and why the t distribution was invented
- How to use the t table when calculating confidence intervals when σ is unknown

Photograph by Wolfgang Sauber

Student’s Dilemma

What happens if you can’t use the Z table?

Problem

Known σ

Assumptions about x to consider before making confidence intervals

1. Simple random sample of size n has been drawn from a population of x values
2. The value of σ is known
3. *If x itself has a normal distribution*, then we know that x-bar will no matter how big our sample is.
4. *If we aren't sure about x's distribution*, we should get a sample of at least n=30.
5. *If we know that x's distribution is very skewed or definitely not normal*, shoot for n=50 or n=100.

Problem

Known σ

Assumptions about x to consider before making confidence intervals

1. Simple random sample of size n has been drawn from a population of x values
2. The value of σ is known
3. *If x itself has a normal distribution*, then we know that x-bar will no matter how big our sample is.
4. *If we aren't sure about x's distribution*, we should get a sample of at least n=30.
5. *If we know that x's distribution is very skewed or definitely not normal*, shoot for n=50 or n=100.

Problem

Known σ

Assumptions about x to consider before making confidence intervals

1. Simple random sample of size n has been drawn from a population of x values
2. The value of σ is known
3. *If x itself has a normal distribution*, then we know that x-bar will no matter how big our sample is.
4. *If we aren't sure about x's distribution*, we should get a sample of at least n=30.
5. *If we know that x's distribution is very skewed or definitely not normal*, shoot for n=50 or n=100.

What if σ is Unknown?

- Sometimes you don't have a σ (new topic, particular subgroup)
- And you have a small sample
- Without a σ, we have to use an estimate for σ
- Solution: use s
- But this means we are not able to use the z table unless we have a large sample because we violate the CLT

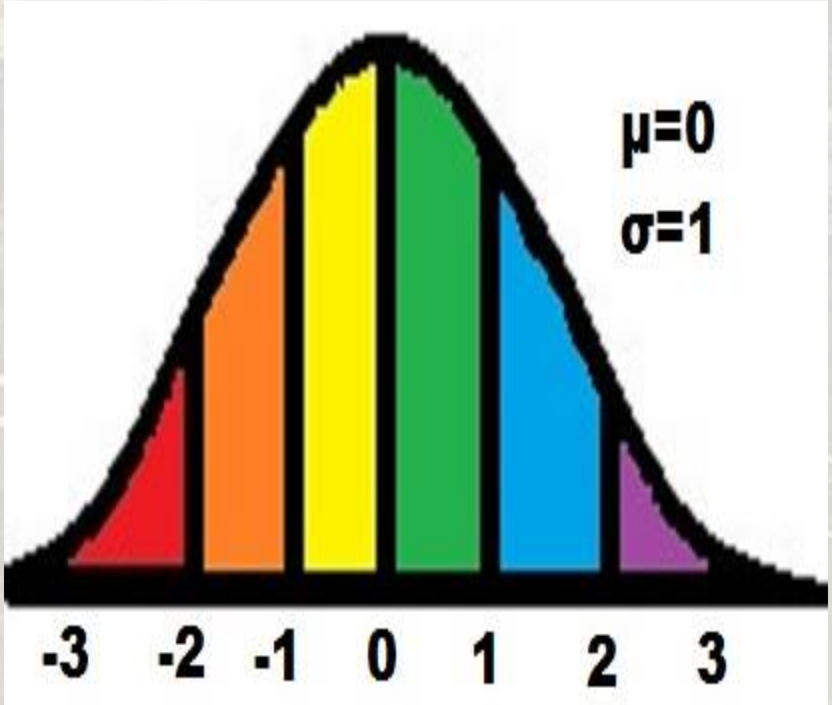
Problem

- In 1908, William Seely Gosset had this problem
- Worked at Guinness, needed to test samples of barley where he did not know the σ
- Could not test large samples
- Therefore, came up with his own distribution called the t distribution
- Due to politics, he could not publish under his name, so he published under his nickname, “Student”
- Student’s t distribution

Picture of beer by OJW. Portrait in public domain.

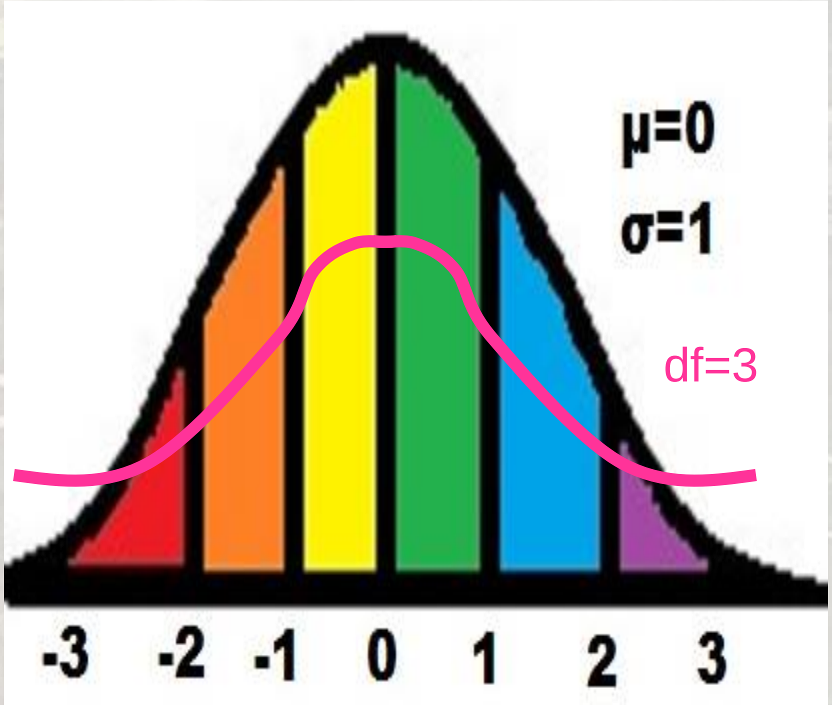
Features of the t-distribution

- Like Z, the t has its own table
- Unlike Z, t is different depending on the size of n
- Degrees of freedom (df) is calculated n-1
- You need the df to look up the t in the t-table



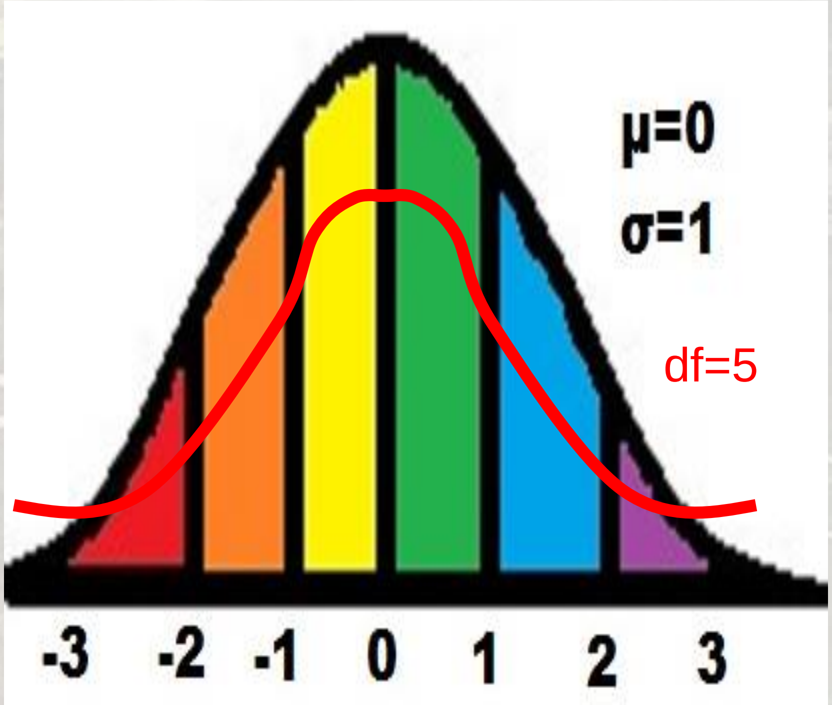
Features of the t-distribution

- Like Z, the t has its own table
- Unlike Z, t is different depending on the size of n
- Degrees of freedom (df) is calculated n-1
- You need the df to look up the t in the t-table



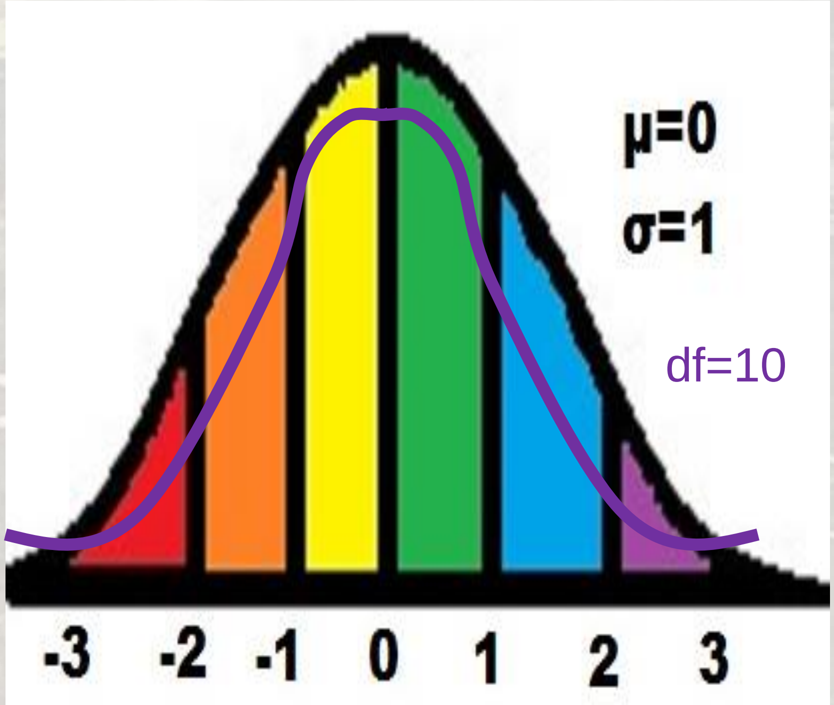
Features of the t-distribution

- Like Z, the t has its own table
- Unlike Z, t is different depending on the size of n
- Degrees of freedom (df) is calculated n-1
- You need the df to look up the t in the t-table



Features of the t-distribution

- Like Z, the t has its own table
- Unlike Z, t is different depending on the size of n
- Degrees of freedom (df) is calculated n-1
- You need the df to look up the t in the t-table



Properties of the Student's t Distribution

1. *Like Z*, the t distribution is symmetrical around 0.
2. *Unlike Z*, the distribution depends on the df (which depends on the sample size, since df = n-1)
3. *Like Z*, the distribution is bell-shaped. *Unlike Z*, the tails are thicker.
4. *As the df (sample size) increases, the t distribution approaches the Z distribution.*
5. *Like Z*, total the area under the t curve is always 1.

How to Find the Confidence Interval when σ is Unknown

New Formula

How to Find the Confidence Interval when σ is Unknown

1. Make sure you have calculated your x-bar, s, and you know your n.
2. Select your c. Calculate df, and look of t_c in the t-table.
3. Calculate E using this formula:

$$E = t_c \frac{s}{\sqrt{n}}$$

4. Subtract the E from x-bar to get the lower limit for the confidence interval
5. Add the E to the x-bar to get the upper limit for the confidence interval

How to Find the Confidence Interval when σ is Unknown

1. Make sure you have calculated your x-bar, s, and you know your n.
2. Select your c. Calculate df, and look of t_c in the t-table.
3. Calculate E using this formula:

$$E = t_c \frac{s}{\sqrt{n}}$$

4. Subtract the E from x-bar to get the lower limit for the confidence interval
5. Add the E to the x-bar to get the upper limit for the confidence interval

Class from different college! σ is unknown!

What is their μ?

How to Find the Confidence Interval when σ is Unknown

1. Make sure you have calculated your x-bar, s, and you know your n.
2. Select your c. Calculate df, and look of t_c in the t-table.
3. Calculate E using this formula:

$$E = t_c \frac{s}{\sqrt{n}}$$

4. Subtract the E from x-bar to get the lower limit for the confidence interval
5. Add the E to the x-bar to get the upper limit for the confidence interval

Class from different college! σ is unknown!

What is their μ?

1. Obtain x-bar, s and n.

x-bar = 68
s = 10
n = 30

How to Find the Confidence Interval when σ is Unknown

1. Make sure you have calculated your x-bar, s, and you know your n.
2. Select your c. Calculate df, and look of t_c in the t-table.
3. Calculate E using this formula:

$$E = t_c \frac{s}{\sqrt{n}}$$

4. Subtract the E from x-bar to get the lower limit for the confidence interval
5. Add the E to the x-bar to get the upper limit for the confidence interval

Class from different college! σ is unknown!

What is their μ?

2. Select 90% for c. df = 30 – 1 = 29. Look up t_c.

x-bar = 68
s = 10
n = 30
c = 90%
df = 29

How to Find the Confidence Interval when σ is Unknown

1. Make sure you have calculated your x-bar, s, and you know your n.
2. Select your c. Calculate df, and look of t_c in the t-table.
3. Calculate E using this formula:

$$E = t_c \frac{s}{\sqrt{n}}$$

4. Subtract the E from x-bar to get the lower limit for the confidence interval
5. Add the E to the x-bar to get the upper limit for the confidence interval

Class from different college! σ is unknown!

What is their μ?

2. Select 90% for c. df = 30 – 1 = 29. Look up t_c.

x-bar = 68
s = 10
n = 30
c = 90%
df = 29
t_c = 1.699

How to Find the Confidence Interval when σ is Unknown

1. Make sure you have calculated your x-bar, s, and you know your n.
2. Select your c. Calculate df, and look of t_c in the t-table.
3. Calculate E using this formula:

$$E = t_c \frac{s}{\sqrt{n}}$$

4. Subtract the E from x-bar to get the lower limit for the confidence interval
5. Add the E to the x-bar to get the upper limit for the confidence interval

Class from different college! σ is unknown!

What is their μ?

3. Calculating E.

x-bar = 68
s = 10
n = 30
c = 90%
df = 29
t_c = 1.699

$1.699 * (10/\sqrt{30}) = 3.1 = E$

How to Find the Confidence Interval when σ is Unknown

1. Make sure you have calculated your x-bar, s, and you know your n.
2. Select your c. Calculate df, and look of t_c in the t-table.
3. Calculate E using this formula:

$$E = t_c \frac{s}{\sqrt{n}}$$

4. Subtract the E from x-bar to get the lower limit for the confidence interval
5. Add the E to the x-bar to get the upper limit for the confidence interval

Class from different college! σ is unknown!

What is their μ?

4. Subtract E from x-bar
5. Add E to x-bar

x-bar = 68
s = 10
n = 30
c = 90%
df = 29
t_c = 1.699
E = 3.1

$68 – 3.1 = 64.9$

$68 + 3.1 = 71.1$

How to Find the Confidence Interval when σ is Unknown

1. Make sure you have calculated your x-bar, s, and you know your n.
2. Select your c. Calculate df, and look of t_c in the t-table.
3. Calculate E using this formula:

$$E = t_c \frac{s}{\sqrt{n}}$$

4. Subtract the E from x-bar to get the lower limit for the confidence interval
5. Add the E to the x-bar to get the upper limit for the confidence interval

Class from different college! σ is unknown!

What is their μ?

"I am 90% confident that the μ of the class from a different college is between 64.9 and 71.1."

x-bar = 68
s = 10
n = 30
c = 90%
df = 29
t_c = 1.699
E = 3.1

$68 - 3.1 = 64.9$

$68 + 3.1 = 71.1$

Recap

- If you know the σ, you have a very robust measurement of variation. This allows you to use the z table, as long as your x is normal, or you have a lot of sample.
  - Z scores are smaller than t scores, which means your E is smaller, so your confidence interval is smaller, so you have a more precise estimate of μ
- If you don't know the σ, then you don't have a very good measurement of variation in your data. As long as your x is normally distributed, you can use the s instead.
  - If you have a sample size smaller than 100, use the t table
  - Otherwise, you can use z table because they become the same at high numbers

Conclusion

- Student's t table useful for smaller samples and when you do not know σ
- Demonstration of how to calculate a confidence interval using s and the t table

Photo courtesy of Mulad in the Public Domain