STA 2023

Module 11A
Confidence Intervals for One Population Mean

Learning Objectives

Upon completing this module, you should be able to:

1. Obtain a point estimate for a population mean.
2. Construct and interpret a confidence interval for a population mean when the population standard deviation is known.
3. Compute and interpret the margin of error for the estimate of the population mean.
4. Describe the relationship between sample size, standard deviation, confidence level, and margin of error for a confidence interval for the population mean.
5. Determine the sample size required for a specified confidence level and margin of error for the estimate of the population mean.

Learning Objectives

6. Describe the difference between the standardized and studentized versions of the sample mean.
7. State the basic properties of *t*-curves.
8. Explain how to find the *t*-value for df = n-1 and selected values of alpha.
9. Construct and interpret a confidence interval for a population mean when the population standard deviation is unknown.
10. Decide whether it is appropriate to use the *z*-interval procedure, *t*-interval procedure, or neither.

Common Problem in Statistics

A common problem in statistics is to obtain information about the mean of a population. If the population is large, taking a census is generally impractical, extremely expensive, or impossible.

For example: the mean cost of our vehicle insurance, the mean salary of a manager, or the mean debt of a household. Take a census?

One way to obtain information about a population mean without taking a census is to estimate it by a sample mean.

Estimating a Population Mean

In this first part of Module 11, we are going to examine methods for estimating the population mean by using the sample mean.

Since the sample mean will be utilized to estimate the population mean, we cannot expect the population mean will equal to the sample mean exactly. Why? This is because the existing of sampling error.

What is a Point Estimate?

Remember, a sample mean is a statistic, whereas a population mean is a parameter.

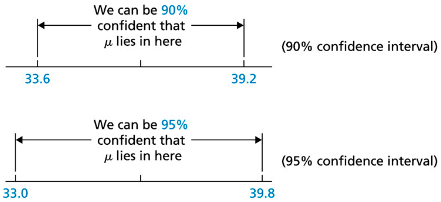
When we try to make our best guess or estimate the population mean with a sample mean, a statistic is being used to estimate the parameter.

In short, the value obtained from the sample data (the sample mean) is the point estimate of the population mean.

Point Estimate

A **point estimate** of a parameter is the value of a statistic used to estimate the parameter.

Let's Look at Some Examples of Confidence Intervals



What is a Confidence-Interval Estimate?

Confidence-Interval Estimate

Confidence interval (CI): An interval of numbers obtained from a point estimate of a parameter.

Confidence level: The confidence we have that the parameter lies in the confidence interval (i.e., that the confidence interval contains the parameter).

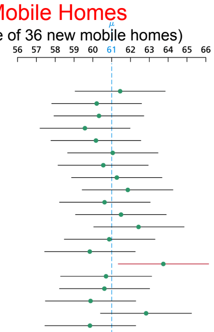
Confidence-interval estimate: The confidence level and confidence interval.

In short, a confidence-interval estimate for a parameter provides a range of numbers along with a percentage confidence that the parameter lies in the range.

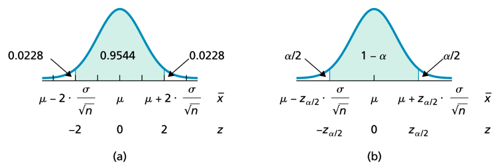
Twenty Confidence Intervals for the Mean Price of All New Mobile Homes

(each based on a sample of 36 new mobile homes)

Sample	$\bar{x}$	95.44% CI	μ in CI?
1	61.45	59.06 to 63.85	yes
2	60.21	57.81 to 62.61	yes
3	60.33	57.93 to 62.73	yes
4	59.59	57.19 to 61.99	yes
5	60.17	57.77 to 62.57	yes
6	61.07	58.67 to 63.47	yes
7	60.56	58.16 to 62.96	yes
8	61.28	58.88 to 63.68	yes
9	61.87	59.48 to 64.27	yes
10	60.61	58.22 to 63.01	yes
11	61.51	59.11 to 63.91	yes
12	62.45	60.05 to 64.85	yes
13	60.88	58.48 to 63.28	yes
14	59.85	57.45 to 62.25	yes
15	63.73	61.33 to 66.13	no
16	60.70	58.30 to 63.10	yes
17	60.60	58.20 to 63.00	yes
18	59.88	57.48 to 62.28	yes
19	62.82	60.42 to 65.22	yes
20	59.84	57.45 to 62.24	yes



Confidence Interval for a Population Mean



What does it mean?

(a) 95.44% of all samples have means within 2 standard deviations of the population mean.

(a) 100(1-alpha)% of all samples have means within a certain standard deviations of the population mean.

How to find a One-Mean z-Interval?

One-Mean z-Interval Procedure

Purpose To find a confidence interval for a population mean, μ

Assumptions

1. Simple random sample
2. Normal population or large sample
3. σ known

STEP 1 **For a confidence level of $1 - \alpha$, use Table II to find $z_{\alpha/2}$.**

STEP 2 **The confidence interval for μ is from**

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{to} \quad \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

where $z_{\alpha/2}$ is found in Step 1, n is the sample size, and $\bar{x}$ is computed from the sample data.

STEP 3 **Interpret the confidence interval.**

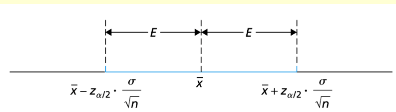
The confidence interval is exact for normal populations and is approximately correct for large samples from nonnormal populations.

Check Out the Margin of Error

Margin of Error for the Estimate of μ

The **margin of error** for the estimate of μ is

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$



In short, the margin of error is half of the length of the confidence interval.

What is the Required Sample Size?

Sample Size for Estimating μ

The sample size required for a $(1-\alpha)$-level confidence interval for μ with a specified margin of error, E, is given by the formula

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

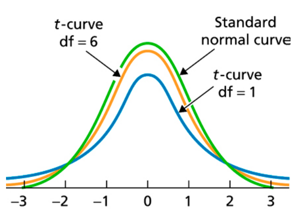
rounded up to the nearest whole number.

Note that increasing the sample size improves the precision, and vice versa.

What is a t-curve or Student's *t*-model?

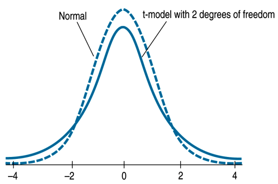
- William S. Gosset, an employee of the Guinness Brewery in Dublin, Ireland, worked long and hard to find out what the sampling model was.
- The sampling model that Gosset found has been known as Student's *t*.
- The Student's *t*-models form a whole *family* of related distributions that depend on a parameter known as degrees of freedom.
  - We often denote degrees of freedom as *df*, and the model as t_{df}.

Standard Normal Curve and Two *t*-curves



Note that increasing the sample size, increasing the df, and the *t*-curve is getting closer to a normal curve.

Degrees of Freedom and *t*-curve



- As the degrees of freedom increase, the *t*-curves look more and more like the Normal.
- In fact, the *t*-curve or *t*-model with infinite degrees of freedom is exactly Normal.

Student's *t*-models

- Student's *t*-models are unimodal, symmetric, and bell shaped, just like the Normal.
- But *t*-models with only a few degrees of freedom have much fatter tails than the Normal.

Basic Properties of *t*-Curves

Basic Properties of *t*-Curves

Property 1: The total area under a *t*-curve equals 1.

Property 2: A *t*-curve extends indefinitely in both directions, approaching, but never touching, the horizontal axis as it does so.

Property 3: A *t*-curve is symmetric about 0.

Property 4: As the number of degrees of freedom becomes larger, *t*-curves look increasingly like the standard normal curve.

How to Use Table to find *t*-value?

df	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	df
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Step 1: Obtain the df, df = n-1, where n is the sample size.

Step 2: Go down the outside column to locate df = 13, go across the row to the column label with alpha=0.05 and reach 1.771. This number is the *t*-value having area 0.05 to its right and df =13.

How to find One-Mean *t*-Interval?

One-Mean *t*-Interval Procedure

Purpose To find a confidence interval for a population mean, μ

Assumptions

1. Simple random sample
2. Normal population or large sample
3. σ unknown

STEP 1 **For a confidence level of $1-\alpha$, use Table IV to find $t_{\alpha/2}$ with df $= n-1$, where n is the sample size.**

STEP 2 **The confidence interval for μ is from**

$$\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{to} \quad \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ is found in Step 1 and $\bar{x}$ and s are computed from the sample data.

STEP 3 **Interpret the confidence interval.**

The confidence interval is exact for normal populations and is approximately correct for large samples from nonnormal populations.

What are the Assumptions and Conditions?

- Gosset found the *t*-model by simulation.
- Years later, when Sir Ronald A. Fisher showed mathematically that Gosset was right, he needed to make some assumptions to make the proof work.
- We will use these assumptions when working with Student's *t*.

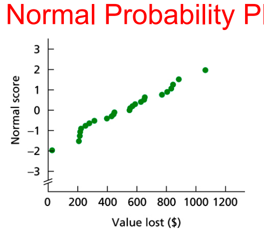
Assumptions and Conditions

- Independence Assumption:
  - Independence Assumption: The data values should be independent.
  - Randomization Condition: The data arise from a random sample or suitably randomized experiment. Randomly sampled data (particularly from an SRS) are ideal.
  - 10% Condition: When a sample is drawn without replacement, the sample should be no more than 10% of the population.

Assumptions and Conditions (cont.)

- Normal Population Assumption:
  - We can never be certain that the data are from a population that follows a Normal model, but we can check the
  - Nearly Normal Condition: The data come from a distribution that is unimodal and symmetric.
    - Check this condition by making a histogram or Normal probability plot.

A Normal Probability Plot



We use this normal probability plot to check the normality and potential outlier(s) of the data. As we can see, this plot reveals no potential outlier and falls roughly in a straight line.

Cautions About Interpreting Confidence Intervals

Remember that interpretation of your confidence interval is key.
What **NOT** to say:

- "*90% of all the vehicles* on Triphammer Road drive at a speed between 29.5 and 32.5 mph."
  - The confidence interval is about the *mean* not the *individual* values.
- "We are 90% confident that *a randomly selected vehicle* will have a speed between 29.5 and 32.5 mph."
  - Again, the confidence interval is about the *mean* not the *individual* values.

Cautions About Interpreting Confidence Intervals (cont.)

- What **NOT** to say:
  - "The mean speed of the vehicles is 31.0 mph *90% of the time.*"
    - The true mean (population mean) does not vary—it's the confidence interval that would be different had we gotten a different sample.
  - "*90% of all samples* will have mean speeds between 29.5 and 32.5 mph."
    - The interval we calculate does not set a standard for every other interval—it is no more (or less) likely to be correct than any other interval.

Make a Picture

- Pictures tell us far more about our data set than a list of the data ever could.
- The only reasonable way to check the Nearly Normal Condition is with graphs of the data.
  - Make a histogram of the data and verify that its distribution is unimodal and symmetric with no outliers.
  - You may also want to make a Normal probability plot to see that it's reasonably straight.

What Can Go Wrong?

Ways to Not Be Normal:

- Beware multimodality.
  - The Nearly Normal Condition clearly fails if a histogram of the data has two or more modes.
- Beware skewed data.
  - If the data are very skewed, try re-expressing the variable.
- Set outliers aside—but remember to report on these outliers individually.

What Can Go Wrong? (cont.)

- Watch out for bias—we can never overcome the problems of a biased sample.
- Make sure data are independent.
  - Check for random sampling and the 10% Condition.
- Make sure that data are from an appropriately randomized sample.
- Interpret your confidence interval correctly.
  - Many statements that sound tempting are, in fact, misinterpretations of a confidence interval for a mean.
  - A confidence interval is about the mean of the population, not about the means of samples, individuals in samples, or individuals in the population.

What have we learned?

We have learned to:

1. Obtain a point estimate for a population mean.
2. Construct and interpret a confidence interval for a population mean when the population standard deviation is known.
3. Compute and interpret the margin of error for the estimate of the population mean.
4. Describe the relationship between sample size, standard deviation, confidence level, and margin of error for a confidence interval for the population mean.
5. Determine the sample size required for a specified confidence level and margin of error for the estimate of the population mean.

What have we learned? (Cont.)

6. Describe the difference between the standardized and studentized versions of the sample mean.
7. State the basic properties of *t*-curves.
8. Explain how to find the *t*-value for df = n-1 and selected values of alpha.
9. Construct and interpret a confidence interval for a population mean when the population standard deviation is unknown.
10. Decide whether it is appropriate to use the *z*-interval procedure, *t*-interval procedure, or neither.

Credit

Some of these slides have been adapted/modified in part/whole from the slides of the following textbooks.

- Weiss, Neil A., Introductory Statistics, 8th Edition
- Bock, David E., Stats: Data and Models, 3rd Edition