Section 9.3
Estimating a Population Mean

Topics Covered in PRESENTATION

- ✓ find confidence interval for a **population mean** when the **population standard deviation is KNOWN**
- ✓ find confidence interval for a **population mean** when the **population standard deviation is UNKNOWN**
- ✓ **Understand the new distribution – the t-distribution**
- ✓ determine the **sample size** required to obtain a level *C* confidence interval for a population mean with a specified margin of error when the **population standard deviation is KNOWN and UNKNOWN**

Topics NOT Covered in PRESENTATION –

- **READ and STUDY the ENTIRE packet and see me with any questions.**

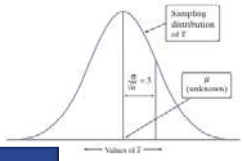
▪The One-Sample *z* Interval for a Population Mean (known σ)

Recall 9.1, the "mystery mean *μ*," *was from a population with a Normal distribution and σ=20. We took a SRS of n=16 and calculated* the sample mean= 240.79.

We want to calculate a 95% confidence interval for *μ* .

We use the familiar formula:
estimate ± (critical value) • (standard deviation of statistic)

$$\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}} = 240.79 \pm 1.96 \cdot \frac{20}{\sqrt{16}}$$
$$= 240.79 \pm 9.8$$
$$= (230.99, 250.59)$$



One-Sample *z* Interval for a Population Mean

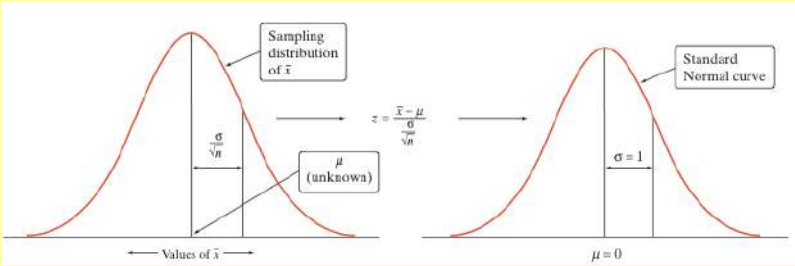
Choose an SRS of size *n* from a population having unknown mean *μ* and known standard deviation *σ*. As long as the Normal and Independent conditions are met, a level *C* confidence interval for *μ* is $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$

The critical value *z** is found from the standard Normal distribution.

Note, the margin of error *ME* of the confidence interval for the population mean *μ* is $z^* \cdot \frac{\sigma}{\sqrt{n}}$

When σ is Unknown: Use the *t* Distributions

Remenber: When the sampling distribution of $\bar{x}$ is close to Normal, we can find probabilities involving $\bar{x}$ by standardizing: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$



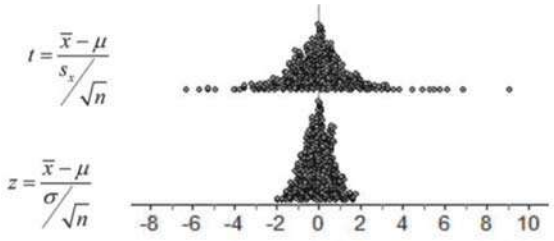
When we do NOT know σ, we can estimate it using the sample standard deviation s_x. **And we get a new statistics:** $t = \dfrac{\bar{x} - \mu}{s_x/\sqrt{n}}$

This new statistic "t" does *not* have a Normal distribution!

When σ is Unknown : The *t* Distributions

The new ***t* distribution** has a ***different shape* than the standard Normal curve**:

- ✓ It is symmetric with a single peak at 0,
- ✓ However, it has much more area in the tails.



Like any standardized statistic, *t* tells us how far $\bar{x}$ is from its mean μ in standard deviation units.

However, there is a different *t* distribution for each sample size, specified by its **degrees of freedom (df)**.

The *t* Distributions ; Degrees of Freedom

The *t* Distributions; Degrees of Freedom

Draw an SRS of size *n* from a large population that has a Normal distribution with mean μ and standard deviation σ. The statistic

$$t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$$

has the ***t* distribution** with **degrees of freedom** df = n – 1.

The statistic will have approximately a t_{n-1} distribution as long as the sampling distribution is close to Normal.

How to Find Critical *t Values**

- Suppose you want to construct a 95% confidence interval for the mean μ of a Normal population based on an SRS of size $n = 12$. What critical *t** should you use?
  - 95% confidence interval with df=11
  - **Calculator command: invT(.025,11)=-2.20 or invT(.975,11)=+2.20**
  - **t*=2.20**

One-Sample *t* Interval for a Population Mean (σ is Unknown)

- The **one-sample *t* interval for a population mean is similar in both reasoning and computational detail to the one-sample *z* interval for a population proportion**.
- As before, we have to verify 3 important conditions before we estimate a population mean.

Conditions for Inference about a Population Mean

- **Random**: The data come from a random sample of size *n* from the population of interest or a randomized experiment.
- **Normal**: The **population has a Normal distribution (either told or you must graph it)** **or** the sample size is large ($n \geq 30$).
- **Independent**: The method for calculating a confidence interval assumes that individual observations are independent. To keep the calculations reasonably accurate when we sample without replacement from a finite population, we should check the *10% condition*: verify that the sample size is no more than 1/10 of the population size.

The One-Sample *t* Interval for a Population Mean

Choose an SRS of size *n* from a population having unknown mean μ. A level *C* confidence interval for μ is

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$$

where *t is the critical value for the t_{n-1} distribution.**

Use this interval ONLY when the above conditions have been met.

One-Sample *t* Interval for a Population Mean (known σ)

Example: Video Screen Tension A manufacturer of high-resolution video terminals must control the tension on the mesh of fine wires that lies behind the surface of the viewing screen. Too much tension will tear the mesh, and too little will allow wrinkles. The tension is measured by an electrical device with output readings in millivolts (mV). Some variation is inherent in the production process.

Here are the tension readings from a random sample of 20 screens from a single day's production:

269.5	297.0	269.6	283.3	304.8	280.4	233.5	257.4	317.5	327.4
264.7	307.7	310.0	343.3	328.1	342.6	338.8	340.1	374.6	336.1

What do you need to do?

1. **Create a list with the data**
2. **Graph it**
3. **Check conditions**
4. **Find the mean and standard deviation**
5. **Find the confidence interval**
6. **Make your conclusion in the context of the question asked**

Choosing the Sample Size (known σ)

Example: How Many Monkeys? Researchers would like to estimate the mean cholesterol level μ of a particular variety of monkey that is often used in laboratory experiments. They would like their estimate to be within 1 milligram per deciliter (mg/dl) of the true value of μ at a 95% confidence level. A previous study involving this variety of monkey suggests that the standard deviation of cholesterol level is about 5 mg/dl.

Choosing Sample Size for a Desired Margin of Error When Estimating μ

To determine the sample size n that will yield a level C confidence interval for a population mean with a specified margin of error ME:

- Get a reasonable value for the population standard deviation σ from an earlier or pilot study.
- Find the critical value z^* from a standard Normal curve for confidence level C.
- Set the expression for the margin of error to be less than or equal to ME and solve for n:

$$z^* \frac{\sigma}{\sqrt{n}} \leq ME$$

Topics NOT Covered in PRESENTATION –

- **READ and STUDY the Following Points and see me with any questions.**

How to Use Table B to Find Critical *t** Values

Suppose you want to construct a 95% confidence interval for the mean *μ* of a Normal population based on an SRS of size *n* = 12. What critical *t should you use?**

	Upper-tail probability *p*			
df	**.05**	**.025**	**.02**	**.01**
10	1.812	2.228	2.359	2.764
11	1.796	2.201	2.328	2.718
12	1.782	2.179	2.303	2.681
z*	1.645	1.960	2.054	2.326
	90%	95%	96%	98%
		Confidence level *C*		

In Table B, we consult the row corresponding to df = *n* – 1 = 11.

We move across that row to the entry that is directly above 95% confidence level.

The desired critical value is *t = 2.201.**

Calculator command: invT(.025,11)=-2.20 or invT(.975,11)=+2.20

■ Review Constructing a Confidence Interval for *μ* (*σ* is Unknown)

When the conditions for inference are satisfied, the sampling distribution for $\bar{x}$ has roughly a Normal distribution. Because we don't know σ, we estimate it by the sample standard deviation s_x.

Definition Reviewed (covered in prior section):

The **standard error of the sample mean** $\bar{x}$ is $\frac{s_x}{\sqrt{n}}$, where s_x is the sample standard deviation. It describes how far $\bar{x}$ will be from μ, on average, in repeated SRSs of size n.

To construct a confidence interval for μ,

- ✓ Replace the standard deviation of $\bar{x}$ by its standard error in the formula for the one-sample z interval for a population mean.
- ✓ Use critical values from the *t* distribution with *n* - 1 degrees of freedom in place of the *z* critical values. That is,

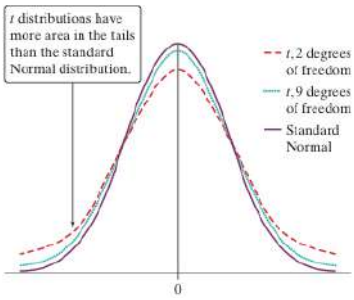
$$\text{statistic} \pm (\text{critical value}) \cdot (\text{standard deviation of statistic})$$

$$= \bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$$

Things to know about the *t* Distribution

- **Their Density Curve – shape, center, & spread; Degrees of Freedom**

When comparing the density curves of the standard Normal distribution and *t* distributions, several facts are apparent:



✓The density curves of the *t* distributions are similar in shape to the standard Normal curve.

✓The spread of the *t* distributions is a bit greater than that of the standard Normal distribution.

✓The *t* distributions have more probability in the tails and less in the center than does the standard Normal.

✓As the degrees of freedom increase, the *t* density curve approaches the standard Normal curve ever more closely.

We can use Table B on your AP green sheet to determine critical values *t** for *t* distributions with different degrees of freedom.

Things to know about the *t* Distribution

Understand Why it is Considered Robust

Definition:

An inference procedure is called **robust** if the probability calculations involved in the procedure remain fairly accurate when a condition for using the procedures is violated.

- The stated confidence level of a one-sample *t* interval for μ is exactly correct when the population distribution is exactly Normal.
- But NO population of real data is exactly Normal.
- The usefulness of the *t* procedures in practice therefore depends on how strongly they are affected by lack of Normality.
- Fortunately, the t procedures are quite robust against non-Normality of the population **except** when outliers or strong skewness are present.
- Larger samples improve the accuracy of critical values from the t distributions when the population is not Normal.

Things to know about the *t* Distribution
When to use it – check the shape and sample size first

- Except in the case of small samples, the condition that the data come from a random sample or randomized experiment is more important than the condition that the population distribution is Normal. Here are practical guidelines for the Normal condition when performing inference about a population mean.

Using One-Sample *t* Procedures: The Normal Condition

- *Sample size less than 15*: Use *t* procedures if the data appear close to Normal (roughly symmetric, single peak, no outliers). If the data are clearly skewed or if outliers are present, do not use *t*.
- *Sample size at least 15*: The *t* procedures can be used except in the presence of outliers or strong skewness.
- *Large samples*: The *t* procedures can be used even for clearly skewed distributions when the sample is large, roughly $n \geq 30$.

Estimating a Population Mean

Summary of Key Concepts

- ✓ **Confidence intervals for the mean μ of a Normal population** are based on the sample mean of an SRS.
- ✓ If we somehow know σ, we use the *z* critical value and the standard Normal distribution to help calculate confidence intervals.
- ✓ The sample size needed to obtain a confidence interval with approximate margin of error ME for a population mean involves solving $z^* \frac{\sigma}{\sqrt{n}} \leq ME$
- ✓ In practice, we usually don't know σ. Replace the standard deviation of the sampling distribution with the **standard error** and use the *t* distribution with $n - 1$ **degrees of freedom (df)**.

Estimating a Population Mean

Summary of Key Concepts

- ✓ There is a *t* distribution for every positive degrees of freedom. All are symmetric distributions similar in shape to the standard Normal distribution. The *t* distribution approaches the standard Normal distribution as the number of degrees of freedom increases.
- ✓ A level C confidence interval for the mean μ is given by the one-sample *t* interval

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$$

- ✓ This inference procedure is approximately correct when these conditions are met: Random, Normal, Independent.
- ✓ The *t* procedures are relatively robust when the population is non-Normal, especially for larger sample sizes. The *t* procedures are not robust against outliers, however.

Determine Sample Size (known σ)

Example: How Many Monkeys?

- ✓ They would like their estimate to be within 1 milligram per deciliter (mg/dl) of the true value of μ at a 95% confidence level. The critical value for 95% confidence is $z^* = 1.96$.
- ✓ A previous study involving this variety of monkey suggests that the standard deviation of cholesterol level is about 5 mg/dl. We will use $\sigma = 5$ as our best guess for the standard deviation.
- ✓ And solve for n using the formula for margin of error (ME): $z^* \frac{\sigma}{\sqrt{n}} \le ME$

Multiply both sides by square root *n* and divide both sides by 1.

$$1.96 \frac{5}{\sqrt{n}} \le 1$$

Square both sides.

$$\frac{(1.96)(5)}{1} \le \sqrt{n}$$

$$(1.96 \cdot 5)^2 \le n$$

$$96.04 \le n$$

We round up to 97 monkeys to ensure the margin of error is no more than 1 mg/dl at 95% confidence.

Appendix – Answers to Problems in the Presentation

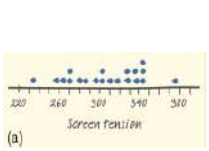
■ Example: Video Screen Tension

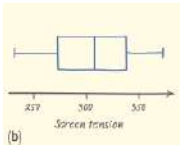
STATE: We want to estimate the true mean tension *μ* of all the video terminals produced this day at a 90% confidence level.

PLAN: If the conditions are met, we can use a one-sample *t* interval to estimate *μ*.

Random: We are told that the data come from a random sample of 20 screens from the population of all screens produced that day.

Normal: Since the sample size is small ($n < 30$), we must check whether it's reasonable to believe that the population distribution is Normal. Examine the distribution of the sample data.





Plot a Histogram
- Stat Plot
- Type 1st row 3rd graph
- ZoomStat

These graphs give no reason to doubt the Normality of the population

Independent: Because we are sampling without replacement, we must check the 10% condition: we must assume that at least 10(20) = 200 video terminals were produced this day.

■ Example: Video Screen Tension (continued)

We want to estimate the true mean tension *μ* of all the video terminals produced this day at a 90% confidence level.

MECHANICS: Using our calculator, we find that the mean and standard deviation of the 20 screens in the sample are **(STAT – CALC – 1-Var-Stats):**

$$\bar{x} = 306.32 \text{ mV} \quad \text{and} \quad s_x = 36.21 \text{ mV}$$

Note: This is the Standard Deviation

Since *n* = 20, we use the *t* distribution with df = 19 to find the critical value.

Calc: invT(.05,19)=-1.729 **We find *t**=1.729.** Or use Table B (AP green sheet)

Therefore, the 90% confidence interval for *μ* is:

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}} = 306.32 \pm 1.729 \frac{36.21}{\sqrt{20}} = 306.32 \pm 14$$

$$= (292.32, 320.32)$$

Note: This is the Margin of Error (ME)

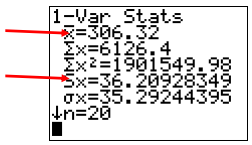
Note: This is the Standard Error of the Sample Mean = 8.097

CONCLUDE: ***We are 90% confident that the interval from 292.32 to 320.32 mV captures the true mean tension in the entire batch of video terminals produced that day.***

Example: Video Screen Tension (continued)

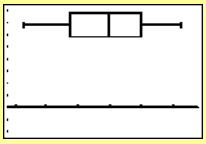
Screen shots from the calculator

Find that the mean and standard deviation



Create Histogram to see the shape of the date

Remember Box Plots do NOT show the shape of the distribution! But Box Plots do show outliers or strong skewness.



After you know how to do the mechanics to find a CI by hand, you can use the calculator. Here you can find 90% confidence interval for a mean point estimator using a t distribution