Unit 5: Estimating with Confidence

Section 8.3
Estimating a Population Mean

The Practice of Statistics, 4th edition – For AP*

STARNES, YATES, MOORE

Unit 5
Estimating with Confidence

- 8.1 Confidence Intervals: The Basics
- 8.2 Estimating a Population Proportion
- **8.3 Estimating a Population Mean**

+

Section 8.3
Estimating a Population Mean

Learning Objectives

After this section, you should be able to…

- ✓ CONSTRUCT and INTERPRET a confidence interval for a population mean
- ✓ DETERMINE the sample size required to obtain a level *C* confidence interval for a population mean with a specified margin of error
- ✓ DESCRIBE how the margin of error of a confidence interval changes with the sample size and the level of confidence *C*
- ✓ DETERMINE sample statistics from a confidence interval

When σ is Unknown: The *t* Distributions

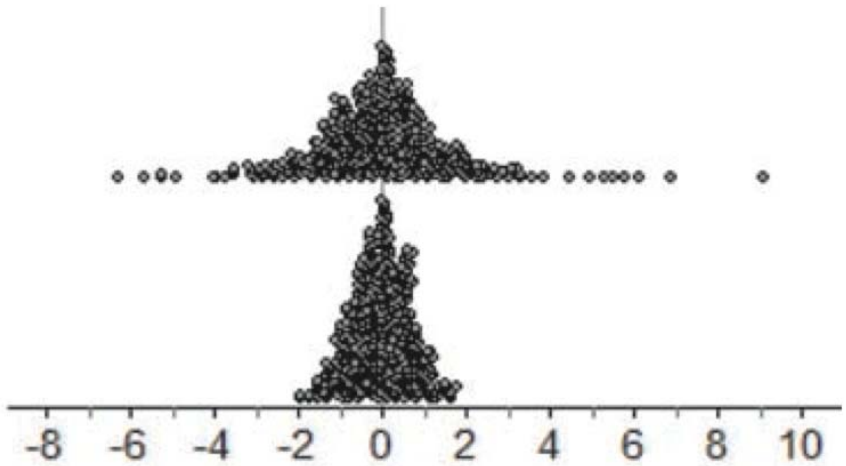
When we standardize based on the sample standard deviation s_x, our statistic has a new distribution called a ***t* distribution**.

It 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.

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

$$z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$$



Like any standardized statistic, t tells us how far $\overline{x}$ 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

When we perform inference about a population mean μ using a t distribution, the appropriate degrees of freedom are found by subtracting 1 from the sample size n, making df = n - 1. We will write the t distribution with n - 1 degrees of freedom as t_{n-1}.

The t Distributions; Degrees of Freedom

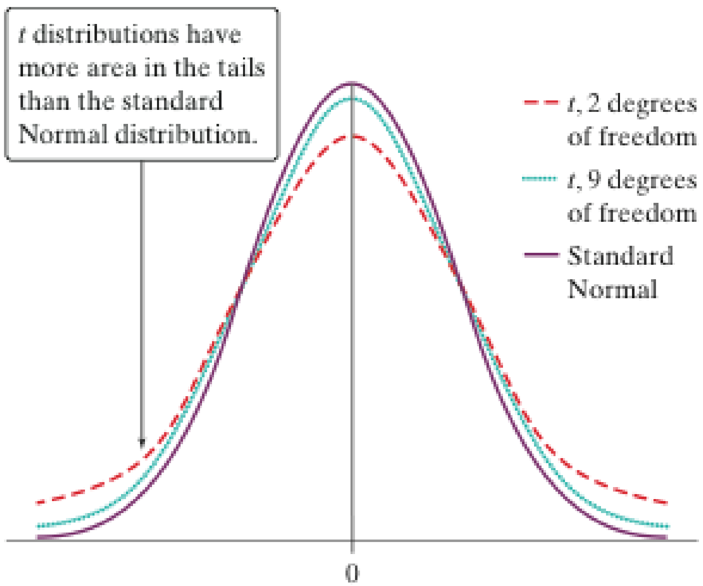
Draw an SRS of size n from a large population that has a Normal distribution with mean μ and standard deviation σ. The statistic

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

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.

The *t* Distributions; 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 in the back of the book to determine critical values t^* for *t* distributions with different degrees of freedom.

Using 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.

Constructing a Confidence Interval for μ

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:

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,

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

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

One-Sample *t* Interval for a Population Mean

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 three 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 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.

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

Construct and interpret a 90% confidence interval for the mean tension μ of all the screens produced on this day.

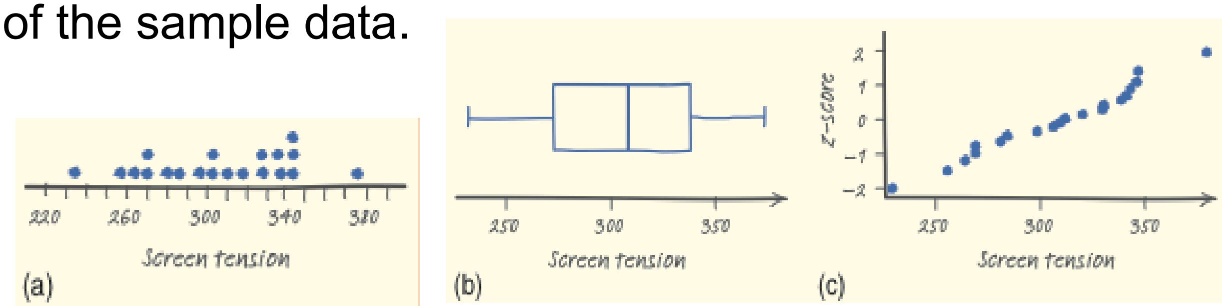
Example: Video Screen Tension

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.



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

DO: Using our calculator, we find that the mean and standard deviation of the 20 screens in the sample are:

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

Upper-tail probability *p*

df	.10	.05	.025
18	1.130	1.734	2.101
19	1.328	1.729	2.093
20	1.325	1.725	2.086
	80%	90%	95%

Confidence level *C*

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

From Table B, we find $t^* = 1.729$.

Therefore, the 90% confidence interval for μ is:

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

$$= 306.32 \pm 14$$

$$= (292.32,\ 320.32)$$

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.

Using *t* Procedures Wisely

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.

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.

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 $n \geq 30$.

Choosing the Sample Size

The margin of error *ME* of the confidence interval for the population mean μ is

$$t^* \cdot \frac{S_x}{\sqrt{n}}$$

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 *t** 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*:

$$t^* \frac{S_x}{\sqrt{n}} \leq ME$$

+

Homework

Textbook – Chapter 8 #'s, 56, 65, 67, 70