Homework

6, 8, 10, 12, 18, 20, 32, 34, 36, 42, 44, 56, 58, 60, 64, 66, 72, 74

Confidence Intervals: The Basics

Confidence Intervals: The Basics

If you had to give one number to estimate an unknown population parameter, what would it be? If you were estimating a population mean μ, you would probably use $\bar{x}$. If you were estimating a population proportion p, you might use p. In both cases, you would be providing a **point estimate** of the parameter of interest.

Definition:

A **point estimator** is a statistic that provides an estimate of a population parameter. The value of that statistic from a sample is called a **point estimate**. Ideally, a point estimate is our "best guess" at the value of an unknown parameter.

We learned in Chapter 7 that an ideal point estimator will have no bias and low variability. Since variability is almost always present when calculating statistics from different samples, we must extend our thinking about estimating parameters to include an acknowledgement that repeated sampling could yield different results.

The Idea of a Confidence Interval

The big idea: The sampling distribution of $\bar{x}$ tells us how close to μ the sample mean $\bar{x}$ is likely to be. All confidence intervals we construct will have a form similar to this:

estimate $\pm$ margin of error

Definition:

A **confidence interval** for a parameter has two parts:

- An interval calculated from the data, which has the form:

  estimate $\pm$ margin of error

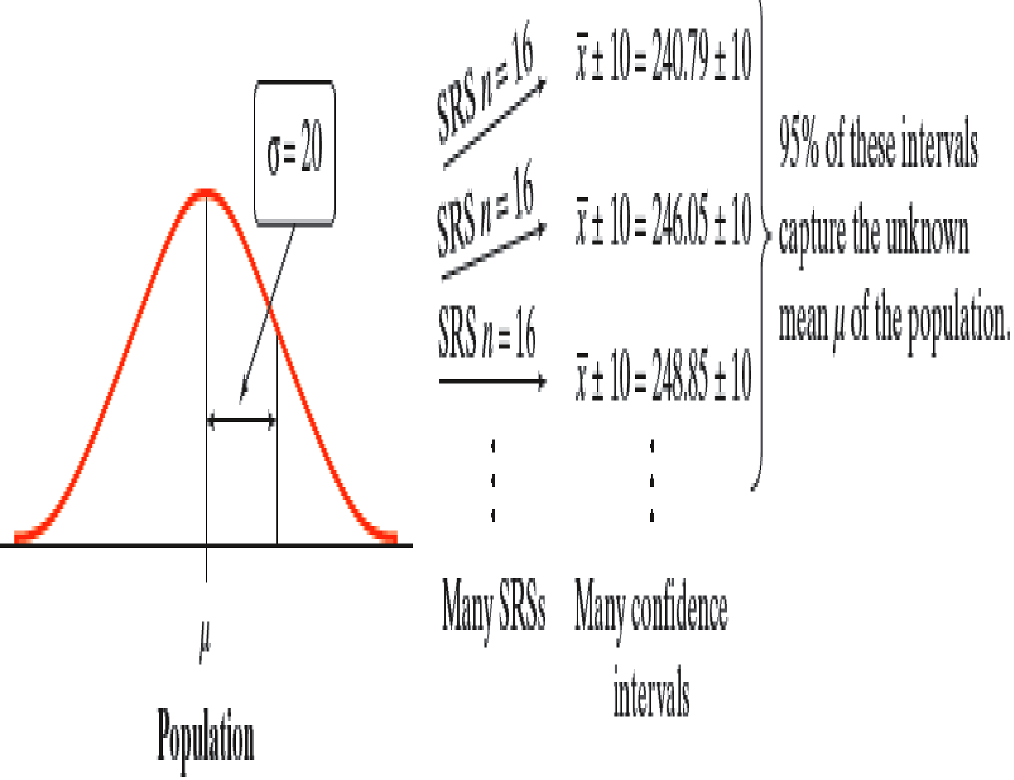
- The **margin of error** tells how close the estimate tends to be to the unknown parameter in repeated random sampling.
- A **confidence level *C***, the overall success rate of the method for calculating the confidence interval. That is, in *C*% of all possible samples, the method would yield an interval that captures the true parameter value.

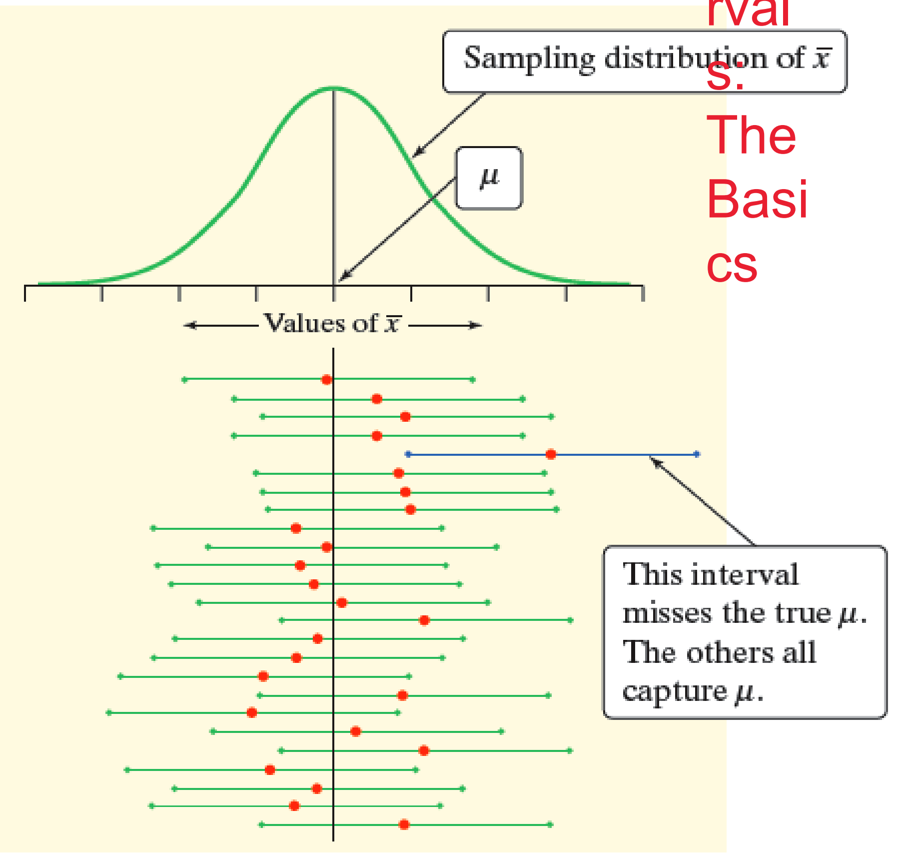
We usually choose a confidence level of 90% or higher because we want to be quite sure of our conclusions. The most common confidence level is 95%.

Confidence Intervals: The Basics

nterpreting onfidence evels and onfidence ntervals

confidence level is e overall capture ate if the method is sed many times. tarting with the opulation, imagine king many SRSs of 6 observations. The ample mean will vary om sample to ample, but when we se the method *stimate* ± *margin of rror* to get an interval ased on each ample, 95% of these tervals capture the nknown population ean μ.





- ***Confidence level:*** To say that we are 95% *confident* is shorthand for "***95% of*** **all possible samples of a [*given size*] from this population will result in an interval that captures the [unknown parameter]**."

- ***Confidence interval:*** To interpret a *C*% confidence interval for an unknown parameter, say, "**We are *C%* confident that the interval from ______ to ______ captures the actual value of the [population parameter in context].**"

The 95% confidence interval for the true proportion of West Johnston students who hate math was 0.65 $\pm$ .05. Interpret the confidence interval and confidence level.

95% of all possible samples of a given size from WJHS will result in an interval that captures the true proportion of students who hate math.

We are 95 % confident that the interval from .60 to .70 captures the actual proportion of WJHS students who hate math.

The 85% confidence interval for the true proportion of North Carolinians who are terrified of snow is 0.76 $\pm$ .02. Interpret the confidence interval and confidence level.

85% of all possible samples of a given size from North Carolinians will result in an interval that captures the true proportion of people who are terrified of snow.

We are 85 % confident that the interval from .74 to .78 captures the actual proportion of North Carolinians who are terrified of snow.

Interpreting Confidence Levels and Confidence Intervals

- The confidence level tells us how likely it is that the method we are using will produce an interval that captures the population parameter *if we use it many times.*

The confidence level does not tell us the chance that a particular confidence interval captures the population parameter.

Instead, the confidence interval gives us a set of plausible values for the parameter.

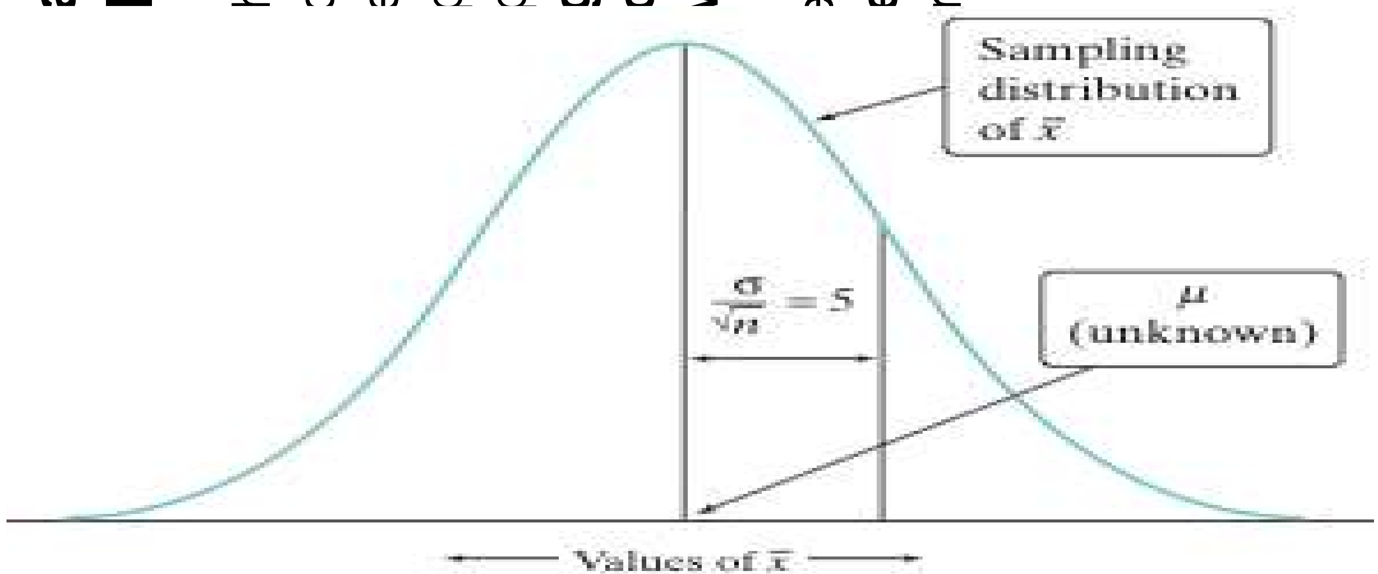
We interpret confidence levels and confidence intervals in much the same way whether we are estimating a population mean, proportion, or some other parameter.

Confidence Intervals: The Basics

Constructing a Confidence Interval

hy settle for 95% confidence when estimating a parameter? The price we pay for **greater confidence is a wider interval.**

e start with **estimate ± margin of error**



This leads to a more general formula for confidence intervals:

statistic ± (critical value) • (standard deviation of statistic)

1) **Random**: The data should come from a well-designed random sample or randomized experiment.

2) **Normal:** The sampling distribution of the statistic is approximately Normal.

For means: The sampling distribution is exactly Normal if the population distribution is Normal. When the population distribution is not Normal, then the central limit theorem tells us the sampling distribution will be approximately Normal if n is sufficiently large ($n \geq 30$).

For proportions: We can use the Normal approximation to the sampling distribution as long as $np \geq 10$ and $n(1 - p) \geq 10$.

3) **Independent**: Individual observations are independent. When sampling without replacement, the sample size n should be no more than 10% of the population size N (the *10% condition*) to use our formula for the standard deviation of the statistic.

Calculating a Confidence Interval

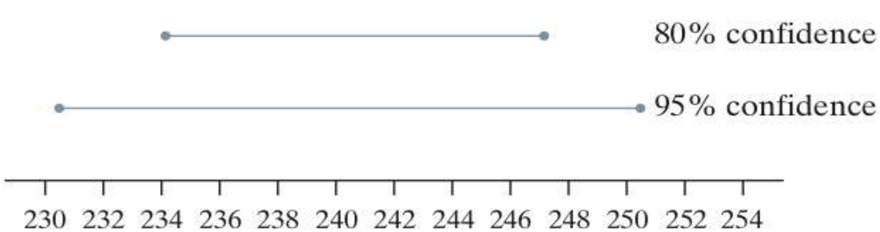
The confidence interval for estimating a population parameter has the form

statistic ± (critical value) • (standard deviation of statistic)

where the statistic we use is the point estimator for the parameter.

Properties of Confidence Intervals:

- The "margin of error" is the (critical value) • (standard deviation of statistic)
- **The user chooses the confidence level**, and the margin of error follows from this choice.
- The critical value depends on the confidence level and the sampling distribution of the statistic.
- Greater confidence requires a larger critical value
- The standard deviation of the statistic depends on the sample size *n*



The margin of error gets smaller when:

✓ **The confidence level decreases**

✓ **The sample size *n* increases**

+

Section 8.1
Confidence Intervals: The Basics

Summary

In this section, we learned that…

- ✓ To estimate an unknown population parameter, start with a statistic that provides a reasonable guess. The chosen statistic is a **point estimator** for the parameter. The specific value of the point estimator that we use gives a **point estimate** for the parameter.
- ✓ A **confidence interval** uses sample data to estimate an unknown population parameter with an indication of how precise the estimate is and of how confident we are that the result is correct.
- ✓ Any confidence interval has two parts: an interval computed from the data and a confidence level C. The interval has the form

$$\text{estimate} \pm \text{margin of error}$$

- ✓ When calculating a confidence interval, it is common to use the form

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

Section 8.1
Confidence Intervals: The Basics

Summary

In this section, we learned that…

- ✓ The **confidence level C** is the success rate of the method that produces the interval. If you use 95% confidence intervals often, in the long run 95% of your intervals will contain the true parameter value. You don't know whether a 95% confidence interval calculated from a particular set of data actually captures the true parameter value.
- ✓ Other things being equal, the **margin of error** of a confidence interval gets smaller as the confidence level C decreases and/or the sample size *n* increases.
- ✓ Before you calculate a confidence interval for a population mean or proportion, be sure to check conditions: **Random** sampling or random assignment, **Normal** sampling distribution, and **Independent** observations.
- ✓ The margin of error for a confidence interval includes only chance variation, not other sources of error like nonresponse and undercoverage.

+

Section 8.2
Estimating a Population Proportion

Learning Objectives

After this section, you should be able to…

- ✓ CONSTRUCT and INTERPRET a confidence interval for a population proportion
- ✓ DETERMINE the sample size required to obtain a level *C* confidence interval for a population proportion 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*

Finding a Critical Value

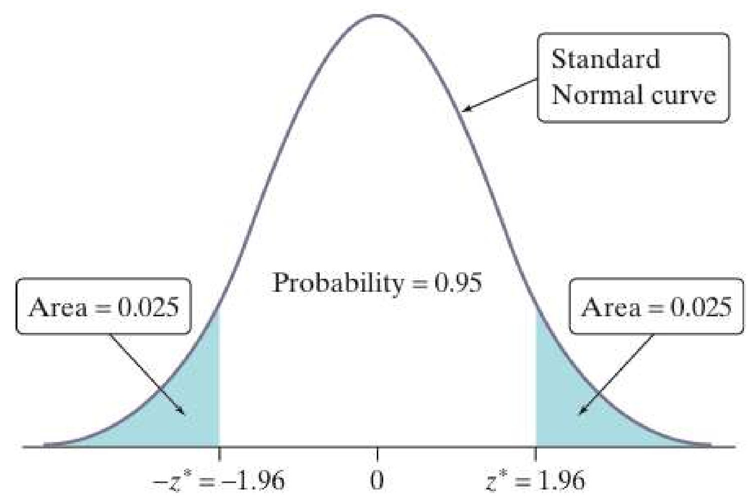
How do we find the critical value for our confidence interval?

Estimating a Population Proportion

statistic ± (critical value) · (standard deviation of statistic)

If the Normal condition is met, we can use a Normal curve. To find a level *C* confidence interval, we need to catch the central area *C* under the standard Normal curve.

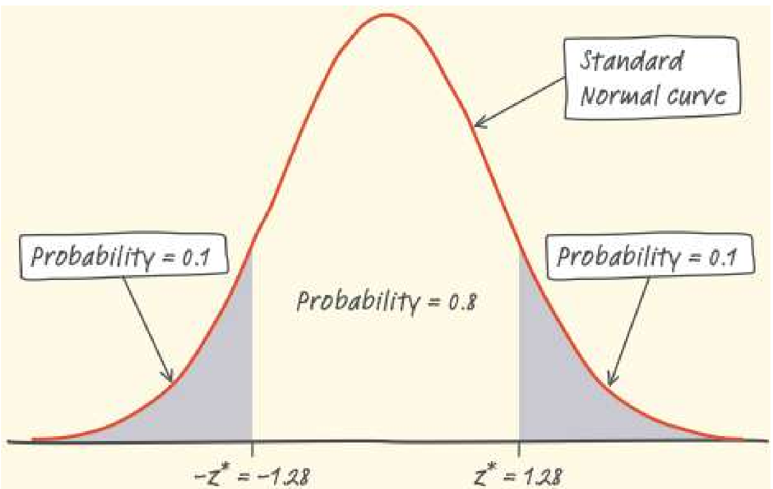
For example, to find a 95% confidence interval, we use a critical value of 2 based on the 68-95-99.7 rule. Using Table A or a calculator, we can get a more accurate critical value.
Note, the **critical value z^*** is actually 1.96 for a 95% confidence level.



Finding a Critical Value

Use Table A to find the critical value z^* for an 80% confidence interval. Assume that the Normal condition is met.

Estimating a Population Proportion



Since we want to capture the central 80% of the standard Normal distribution, we leave out 20%, or 10% in each tail.

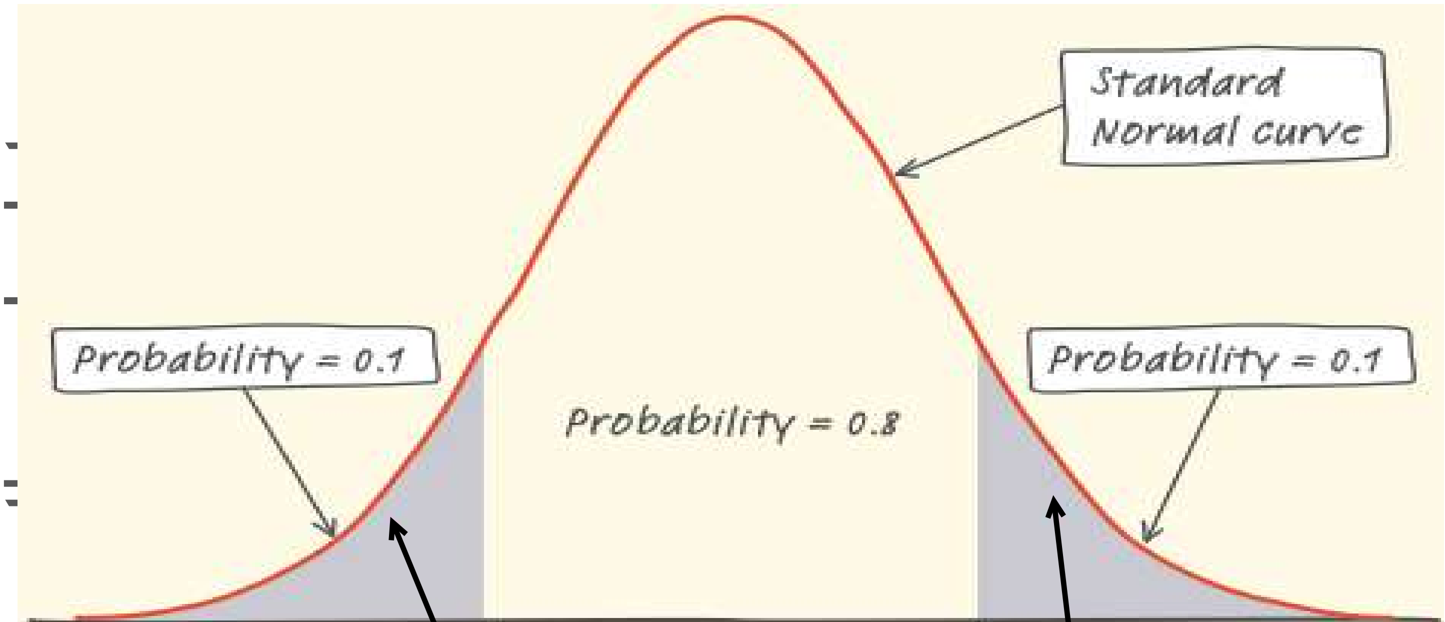
Search Table A to find the point z^* with area 0.1 to its left.

z	.07	.08	.09
– 1.3	.0853	.0838	.0823
– 1.2	.1020	.1003	.0985
– 1.1	.1210	.1190	.1170

The closest entry is $z = -1.28$.

So, the **critical value z^*** for an 80% confidence interval is $z^* = 1.28$.

Find the z-score



Remember both of the s

invNorm (.90,0,1) =

Constructing a Confidence Interval for *p*

We can use the general formula from Section 8.1 to construct a confidence interval for an unknown population proportion p:

statistic $\pm$ **(critical value)** $\cdot$ **(standard deviation of statistic)**

The sample proportion p is the statistic we use to estimate p.
When the Independent condition is met, the standard deviation of the sampling distibution of p is

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

Since we don't know p, we replace it with the sample proportion p.
This gives us the **standard error (SE)** of the sample proportion:

$$\sqrt{\frac{p(1-p)}{n}}$$

Definition:

When the standard deviation of a statistic is estimated from data, the results is called the **standard error** of the statistic.

One-Sample z Interval for a Population Proportion

Once we find the critical value z^*, our confidence interval for the population proportion p is

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

$$= \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

One-Sample z Interval for a Population Proportion

Choose an SRS of size *n* from a large population that contains an unknown proportion *p* of successes. An approximate level *C* **confidence interval for *p* is**

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where z^* is the critical value for the standard Normal curve with area *C* between $-z^*$ and z^*.

Use this interval only when the numbers of successes and failures in the sample are both at least 10 and the population is at least 10 times as large as the sample.

1. Define parameter
2. State the confidence level we are using. “We want to estimate p at [C] % confidence level.
3. State type of test. We are using a 1-sample z interval for proportions
4. Check conditions of the sample.
   (Random, Normal (use $\hat{p}$), Independent)
5. State p-hat value
6. find critical value (z^*)
“invNorm(decimal of shaded area, 0, 1)
7. Use formula: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
8. Conclude. “We are [C] % confidence that the interval from [here] to [there] captures the true proportion of [whatever the context of this problem is].

Estimating a Population Proportion

Conditions for Estimating *p*

Suppose there are 3000 beads in a jar. Suppose one SRS of beads resulted in 107 red beads and 144 beads of another color. The point estimate for the unknown proportion *p* of red beads in the population would be

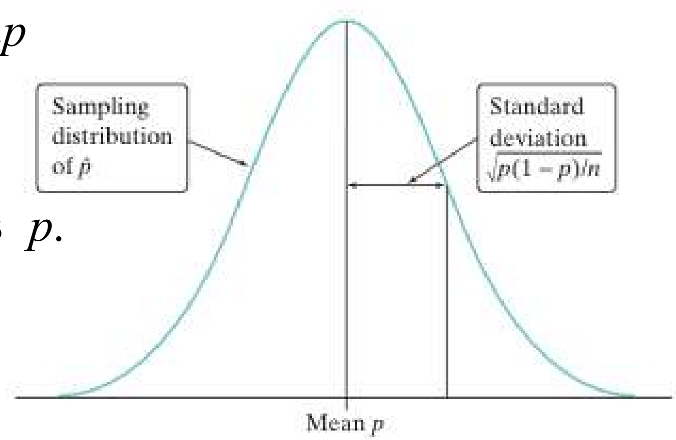
$$\hat{p} = \frac{107}{251} = 0.426$$

How can we use this information to find a confidence interval for *p*?

- If the sample size is large enough that both np and $n(1-p)$ are at least 10, the sampling distribution of p is approximately Normal.

- The mean of the sampling distribution of p is p.

- The standard deviation of the sampling distribution of p is $\sigma_{p} = \sqrt{\frac{p(1-p)}{n}}$.



In practice, we do not know the value of p. If we did, we would not need to construct a confidence interval for it! In large samples, p will be close to p, so we will replace p with p in checking the Normal condition.

Conditions for Estimating *p*

Check the conditions for estimating *p* from our sample.

$$\hat{p} = \frac{107}{251} = 0.426$$

Random: The class took an SRS of 251 beads from the container.

Normal: Both *np* and *n*(1 – *p*) must be greater than 10. Since we don't know *p*, we check that

$$n\hat{p} = 251\left(\frac{107}{251}\right) = 107 \quad \text{and} \quad n(1-\hat{p}) = 251\left(1 - \frac{107}{251}\right) = 144$$

The counts of successes (red beads) and failures (non-red) are both ≥ 10.

Independent: Since the class sampled without replacement, they need to check the 10% condition. At least 10(251) = 2510 beads need to be in the population. The teacher reveals there are 3000 beads in the container, so the condition is satisfied.

Since all three conditions are met, it is safe to construct a confidence interval.

One-Sample z Interval for a Population Proportion

Calculate and interpret a 90% confidence interval for the proportion of red beads in the container. Your teacher claims 50% of the beads are red.

Estimating a Population Proportion

z	.03	.04	.05
– 1.7	.0418	.0409	.0401
– 1.6	.0516	.0505	.0495
– 1.5	.0630	.0618	.0606

- ✓ sample proportion = 107/251 = 0.426
- ✓ We checked the conditions earlier.
- ✓ For a 90% confidence level, z^* = 1.645

statistic ± (critical value) • (standard deviation of the statistic)

$$p \pm z^* \sqrt{\frac{p(1-p)}{n}}$$

$$= 0.426 \pm 1.645\sqrt{\frac{(0.426)(1-0.426)}{251}}$$

$$= 0.426 \pm 0.051$$

$$= (0.375,\ 0.477)$$

We are 90% confident that the interval from 0.375 to 0.477 captures the actual proportion of red beads in the container.

Since this interval gives a range of plausible values for p and since 0.5 is not contained in the interval, we have reason to doubt the claim.

- **Problem:** Ms. Smith has a jar with over 2000 pennies in it. Her class took an SRS of 102 pennies and discovered that 57 of the pennies were more than 10 years old.
- (a) Calculate and interpret a 99% confidence interval for p = the true proportion of pennies from the collection that are more than 10 years old.

Solution:
Let p = true proportion of pennies in jar older than 10 years

We want to estimate p at a 99% confidence interval. We are using a 1-sample z interval for proportions.
Problem states a SRS was taken. 102(57/102)=57 and 102(1 – 57/102) = 45, both are greater than or equal to 10,so distribution is approximately Normal.
102(10) < 2000, so 10% condition is met for independence.
All conditions are met.

p-hat is 57/102 = .559.
The critical value for 99% confidence interval is
invNorm(shaded area, mean, standard deviation)
invNorm(.995,0,1)=2.576.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad 0.559 \pm 2.576 \sqrt{\frac{(.559)(1-.559)}{102}} = \quad (0.432, 0.686)$$

We are 99% confident that the interval from 0.432 to 0.686 captures the actual proportion of pennies in Mrs. Smith’s collection that are more than 10 years old.

b. Yes it is plausible, since 0.6 is included in the confidence interval, it is plausible that 60% of all the pennies in the collection are more than 10 years old.

According to an article in the *San Gabriel Valley Tribune* (2-13-03), "Most people are kissing the 'right way'." That is, according to the study, the majority of couples tilt their heads to the right when kissing. In the study, a researcher observed a random sample 124 couples kissing in various public places and found that 83/124 (66.9%) of the couples tilted to the right. Construct and interpret a 95% confidence interval for the proportion of all couples who tilt their heads to the right when kissing.

State: We want to estimate p = the true proportion of couples that tilt their heads to the right when kissing at the 95% confidence level.

Plan: We will use a one-sample z interval for p if the following conditions are satisfied.

Random: The researcher observed a random sample of couples.

Normal: $(.669)(124) = 83 \geq 10$ and $(.331)(124) = 41 \geq 10$

Independent: The number of couples in the population is more than 10(124) = 1240.

Do: $\hat{p} + z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.669 \pm 0.083 = (0.586, 0.752)$

Conclude: We are 95% confident that the interval from 0.586 to 0.752 captures the true proportion of couples that tilt their heads to the right when kissing.

What to do when p-hat value is not given
Fill in table.

p	.1	.2	.3	.4	.5	.6	.7	.8	.9
$\sqrt{p(1-p}$	.30	.40	.46	.49	.5	.49	.46	.40	.30

If we have to guess the value for p-hat we use 0.5 because it give us the largest margin of error (safest choice).

Choosing the Sample Size

In planning a study, we may want to choose a sample size that allows us to estimate a population proportion within a given margin of error.

Estimating a Population Proportion

The margin of error (*ME*) in the confidence interval for *p* is

$$ME = z^* \sqrt{\frac{p(1-p)}{n}}$$

✓ *z** is the standard Normal critical value for the level of confidence we want.

Because the margin of error involves the sample proportion p, we have to guess the latter value when choosing n. There are two ways to do this:

Use a guess for p based on past experience or a pilot study

Use $p = 0.5$ as the guess. ME is largest when $p = 0.5$

Sample Size for Desired Margin of Error

To determine the sample size *n* that will yield a level *C* confidence interval for a population proportion *p* with a maximum margin of error *ME*, solve the following inequality for *n*:

$$z^* \sqrt{\frac{p(1-p)}{n}} \le ME$$

where p is a guessed value for the sample proportion. The margin of error will always be less than or equal to ME if you take the guess p to be 0.5.

Example: Customer Satisfaction

Determine the sample size needed to estimate *p* within 0.03 with 95% confidence.

Estimating a Population Proportion

✓ The critical value for 95% confidence is $z^* = 1.96$.

✓ Since the company president wants a margin of error of no more than 0.03, we need to solve the equation

$$1.96\sqrt{\frac{p(1-p)}{n}} \leq 0.03$$

Multiply both sides by square root *n* and divide both sides by 0.03.

$$\frac{1.96}{0.03}\sqrt{p(1-p)} \leq \sqrt{n}$$

Square both sides.

$$\left(\frac{1.96}{0.03}\right)^2 p(1-p) \leq n$$

Substitute 0.5 for the sample proportion to find the largest *ME* possible.

$$\left(\frac{1.96}{0.03}\right)^2 (0.5)(1-0.5) \leq n$$

$$1067.111 \leq n$$

We round up to 1068 respondents to ensure the margin of error is no more than 0.03 at 95% confidence.

An Environmental Protection Agency (EPA) investigator wants to know the proportion of fish that are inedible because of chemical pollution downstream of an offending factory. If the answer must be within $\pm$ 3% at the 96% confidence level, how many fish should be in the sample tested?

$$ME \geq z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.03 \geq 2.054 \sqrt{\frac{.5(1-.5)}{n}}$$

$$(2.054/0.03)^2 (.25) \leq n$$

$$1167.36 \leq n$$

So, therefore we should choose a sample of 1168 fish to give us the edible proportion to within $\pm$ 0.03 at the 96% confidence level.

+

Section 8.2
Estimating a Population Proportion

Summary

In this section, we learned that…

- ✓ Confidence intervals for a population proportion p are based on the sampling distribution of the sample proportion p. When n is large enough that both np and $n(1-p)$ are at least 10, the sampling distribution of p is approximately Normal.

- ✓ In practice, we use the sample proportion p to estimate the unknown parameter p. We therefore replace the standard deviation of p with its standard error when constructing a confidence interval.

  The level C confidence interval for p is: $p \pm z^* \sqrt{\dfrac{p(1-p)}{n}}$

+

Section 8.2
Estimating a Population Proportion

Summary

In this section, we learned that…

- ✓ When constructing a confidence interval, follow the familiar four-step process:
  - ✓ STATE: What parameter do you want to estimate, and at what confidence level?
  - ✓ PLAN: Identify the appropriate inference method. Check conditions.
  - ✓ DO: If the conditions are met, perform calculations.
  - ✓ CONCLUDE: Interpret your interval in the context of the problem.
- ✓ The sample size needed to obtain a confidence interval with approximate margin of error *ME* for a population proportion involves solving

$$z^* \sqrt{\frac{p(1-p)}{n}} \le ME$$

for n, where p is a guessed value for the sample proportion, and z^* is the critical value for the level of confidence you want. If you use $p = 0.5$ in this formula, the margin of error of the interval will be less than or equal to ME.

Section 8.3
Estimating a Population Mean

Learning Objectives

After this section, you should be able to…

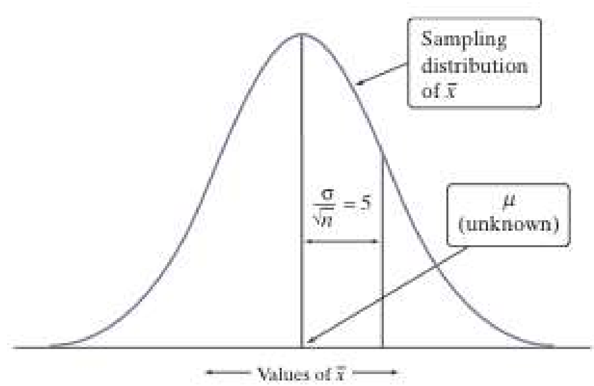
- ✓ CONSTRUCT and INTERPRET a confidence interval for a population mean
- ✓ DETERMINE the sample size required to obtain a level *C* confidence interval for a population mean with a specified margin of error
- ✓ DESCRIBE how the margin of error of a confidence interval changes with the sample size and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

The One-Sample z Interval for a Population Mean

Estimating a Population Mean

To calculate a 95% confidence interval for μ , we use the familiar formula:

estimate ± (critical value) • (standard deviation of statistic)



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.

Choosing the Sample Size

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

Estimating a Population Mean

We determine a sample size for a desired margin of error when estimating a mean in much the same way we did when estimating a proportion.

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*:

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

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. How many monkeys should be used in this experiment?

- ✓ The critical value for 95% confidence is $z^* = 1.96$.
- ✓ We will use $\sigma = 5$ as our best guess for the standard deviation.

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

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

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

Square both sides.

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

$$96.04 \leq n$$

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

- As part of their final project in AP Statistics, Christina and Rachel randomly selected 18 rolls of a generic brand of toilet paper to measure how well this brand could absorb water. You can assume the population is approximately Normal. To do this, they poured 1/4 cup of water onto a hard surface and counted how many squares it took to completely absorb the water. Here are the results from their 18 rolls:
- 292025292124272524
- 292427282125262223
- **State:** We want to estimate μ = the mean number of squares of generic toilet paper needed to absorb 1/4 cup of water with 99% confidence.

WHY IS THIS WRONG?

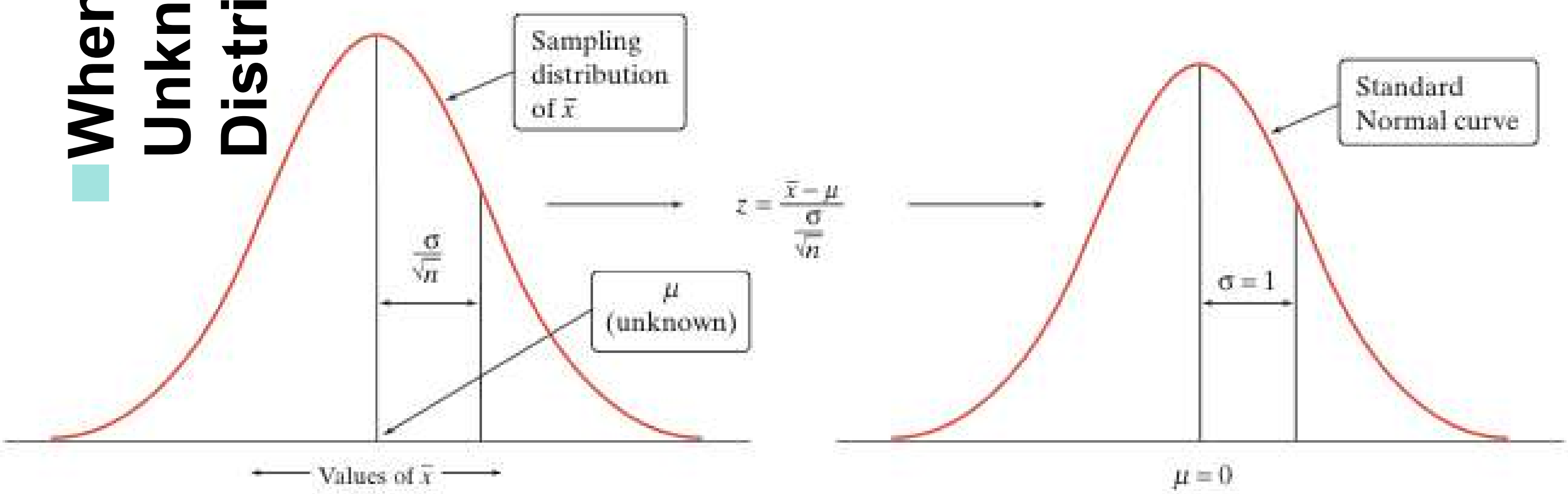
$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

When σ is Unknown: The *t* Distributions

Estimating a Population Mean

When the sampling distribution of $\bar{x}$ is close to Normal, we can find probabilities involving $\bar{x}$ by standardizing:

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



When we don't know *σ*, we can estimate it using the sample standard deviation s_x. What happens when we standardize?

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

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

Estimating a Population Mean

When σ is Unknown: The *t* Distributions

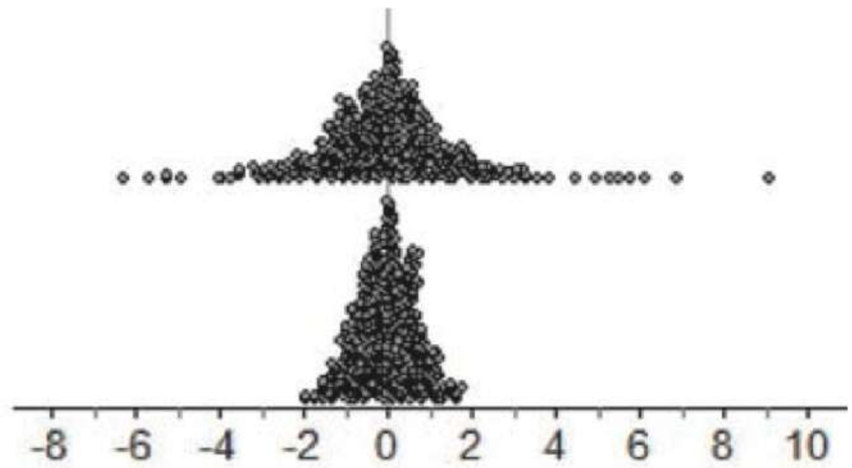
en we standardize ased on the sample tandard deviation s_x, ur statistic has a new istribution called a ***t* istribution**.

as a *different shape* an the standard ormal curve:

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

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

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



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)**.

e t
stributions;
grees of
eedom

we perform
rence about a
ulation mean μ
ng a t distribution,
appropriate
rees of freedom
found by
tracting 1 from the
nple size n, making
n - 1. We will write
t distribution with n
degrees of freedom
n-1.

The *t* Distributions; Degrees of Freedom

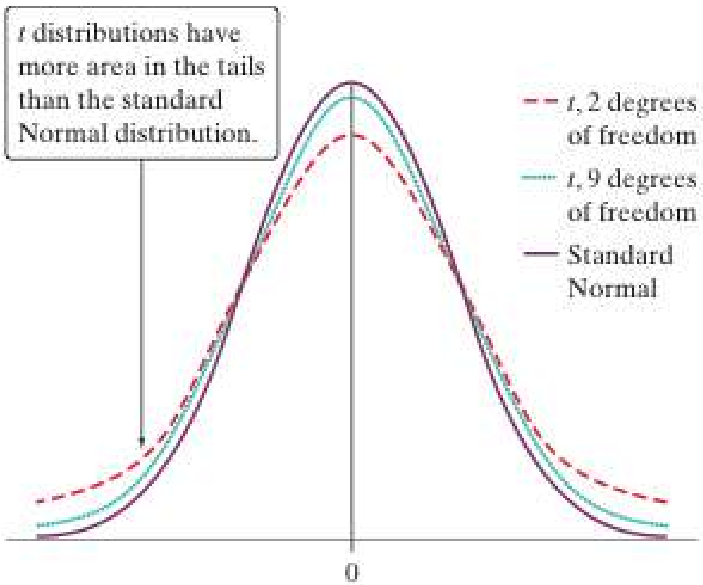
Draw an SRS of size *n* from a large population that has a Normal distribution with mean *μ* and standard deviation *σ*. The statistic

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

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.

butions; ees of lom

mparing the curves of the rd Normal ition 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.

Estimating a Population Mean

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.

- As part of their final project in AP Statistics, Christina and Rachel randomly selected 18 rolls of a generic brand of toilet paper to measure how well this brand could absorb water. To do this, they poured 1/4 cup of water onto a hard surface and counted how many squares it took to completely absorb the water. Here are the results from their 18 rolls:
- 292025292124272524
- 292427282125262223
- **State:** We want to estimate = the mean number of squares of generic toilet paper needed to absorb 1/4 cup of water with 99% confidence.

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,

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

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

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.

Estimating a Population Mean

A manufacturer of high-resolution video terminals must control the tension on the mess of fine wires that lies behind the surface of the video screen. Here is the simple random sample readings for 22 screens from a single day's production:

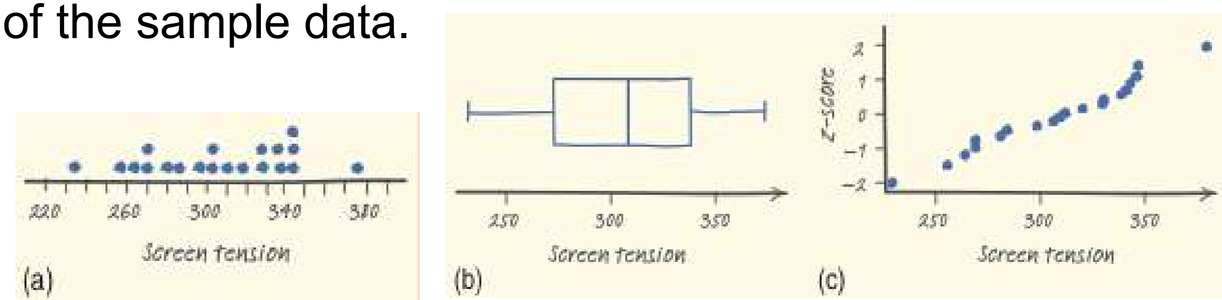
269.5	297.0	269.6
283.3	304.8	280.4
233.5	257.4	317.4
317.5	327.4	
264.7	307.7	310.0
343.3	328.1	342.6
342.6	338.8	340.1
374.6	336.1	

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 22 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(22) = 220 video terminals were produced this day.

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

$$\bar{x} = 308.47 \text{ mV} \qquad \text{and} \qquad s_x = 35.35 \text{ mV}$$

Upper-tail probability *p*

df	**.05**	**.025**	**.02**
18	1.130	1.734	2.101
19	1.328	1.729	2.093
20	1.325	1.725	2.086
21	1.721	2.080	2.189
	90%	95%	96%

Confidence level *C*

Since n = 22, we use the t distribution with df = 21 to find the critical value.

From Table B, we find t^* = 1.721.

Therefore, the 90% confidence interval for μ is:

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}} = 308.47 \pm 1.721 \frac{35.35}{\sqrt{22}}$$
$$= 308.47 \pm 12.97$$
$$= (295.5, 321.44)$$

CONCLUDE: We are 90% confident that the interval from 295.5 to 321.44 mV captures the true mean tension in the entire batch of video terminals produced that day.

- The principal at a large high school claims that students spend at least 10 hours per week doing homework on average. To investigate this claim, an AP Statistics class selected a random sample of 250 students from their school and asked them how many long they spent doing homework during the last week. The sample mean was 10.2 hours and the sample standard deviation was 4.2 hours.

- **Problem:**

- (a) Construct and interpret a 95% confidence interval for the mean time spent doing homework in the last week for students at this school.

- (b) Based on your interval in part (a), what can you conclude about the principal's claim?

Using *t* Procedures Wisely

- The stated confidence level of a one-sample *t* interval for μ is exactly correct when the population distribution is exactly Normal. 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.

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 *t* Procedures Wisely

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 $n \geq 30$.

Section 8.3
Estimating a Population Mean

Summary

In this section, we learned that…

- ✓ **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)**.

Section 8.3
Estimating a Population Mean

Summary

- ✓ 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.