

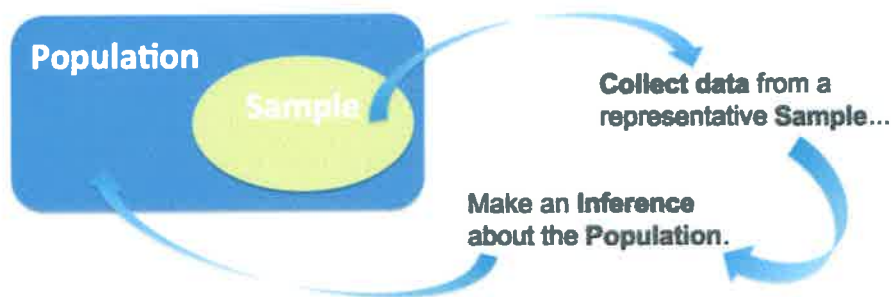
Chapter 8: Confidence Intervals

Introduction

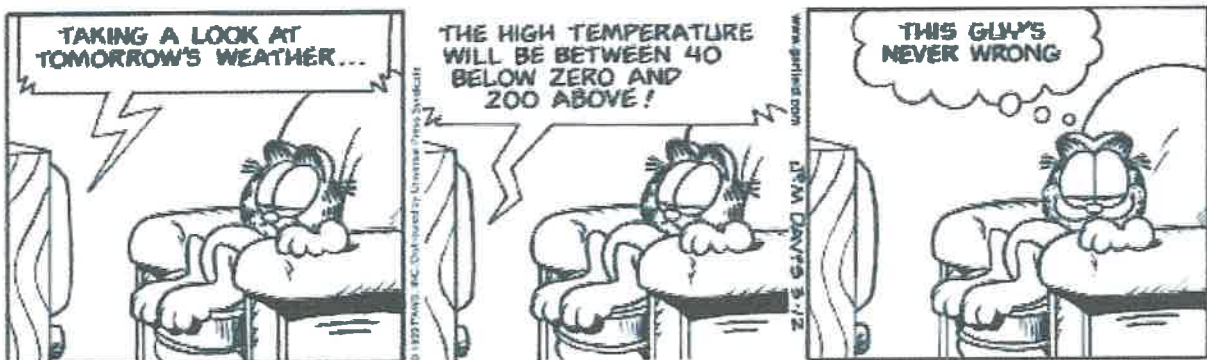
There are two major techniques for classical statistical inference: confidence intervals and hypothesis testing. We will be studying both techniques for our last unit.

Statistical Inference – The process of drawing conclusions about a population based on sample data.

- The purpose of collecting data on a sample is not simply to have data on that sample.
- Researchers take the sample to infer from that data some conclusion about the wider population represented by the sample.



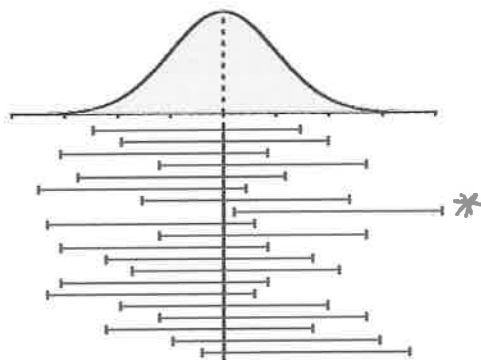
- For Confidence Intervals, we use data from the sample statistics to construct an interval that we think contains the **population parameter**. Confidence Intervals **estimate the value of a population parameter**; they do not estimate the value of a sample statistic or of an individual observation.



A **confidence interval** is an interval of values computed from sample data that is likely to include the unknown value of a population parameter. There is no guarantee that a given confidence interval does capture the parameter, but there is a predictable probability of success known as the **confidence level**. (90%, 95%, 98%, 99% - common CL's)

The confidence level is a measure of how **confident** we are that the confidence interval contains the **true** population parameter. $CL = 1 - \alpha$ OR $C = 1 - \alpha$, $\alpha =$ significance level

The **confidence level (example 95%)** tells us that, **after repeated sampling**, 95% of the confidence intervals constructed would contain the parameter of interest (and 5% would not).



- We are 95% confident that the true population parameter is between (LB, UB).

Theory: if we select many different samples of size n from the same population and construct the corresponding confidence intervals, 95% of them would contain the parameter.

Constructing Confidence Intervals (For population proportion p or population mean μ)

To construct a confidence interval, we need a **point estimate** and the **margin of error**.

A **point estimate** is a single value used to estimate an **unknown population parameter**.

- The sample proportion \hat{p} is the **best point estimate** for the population proportion, p .
- The sample mean \bar{x} is the **best point estimate** for the population mean, μ .

We realize that the point estimate is most likely not the exact value of the population parameter, but close to it. We can take those point estimates (either \hat{p} or \bar{x}) and construct interval estimates called confidence intervals for the population parameter p or μ .

The half-width of a confidence interval is often called the **margin of error, E** .

$$\text{For proportions: } E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{For means: } E = t^* \left(\frac{s}{\sqrt{n}} \right)$$

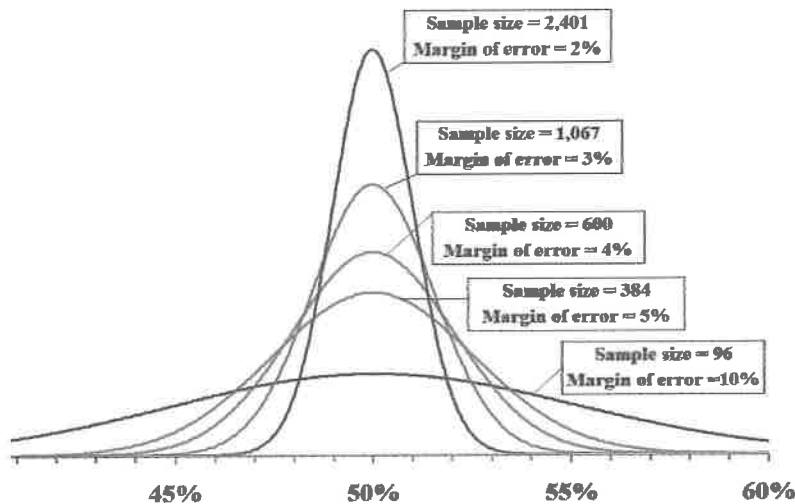
z^* and t^* are known as critical values

What affects the margin of error?

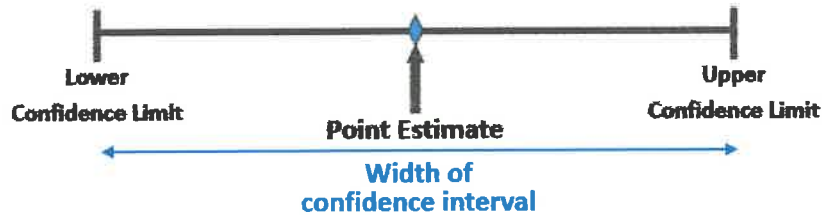
1. **The level of confidence which determines the value of z^* or t^* .** As the confidence level increases, the margin-of-error and width of the confidence level will also increase.



2. **The sample size.** A larger sample size produces a narrower confidence interval whenever other factors remain the same. In this case, we generally say our estimate of the population parameter is more precise.



A confidence interval: point estimate \pm margin of error



A confidence interval can be written as an interval or an inequality.

(Lower Bound, Upper Bound) OR Lower Bound < parameter < Upper Bound

Example 1: Find the confidence interval given the point estimate and the margin of error. Write the answer in interval notation and inequality notation.

A. Write the confidence interval for a **population proportion** p if the point estimate is 0.72 and the margin of error is 0.14.

$$\begin{aligned} & \text{p.e.} \pm E \\ & 0.72 \pm 0.14 \end{aligned}$$

$$(0.58, 0.86)$$

$$0.58 < p < 0.86$$

B. Write the confidence interval for a **population mean** μ if the point estimate is 48.7 and the margin of error is 5.24.

$$\begin{aligned} & \text{p.e.} \pm E \\ & 48.7 \pm 5.24 \end{aligned}$$

$$(43.46, 53.94)$$

$$43.46 < \mu < 53.94$$

8.3: A Population Proportion

To construct a confidence interval for a single unknown population proportion, p , we need a point estimate for p and the margin of error, E .

- The sample proportion, \hat{p} , is the **best point estimate** of the population proportion, p .
- The **margin of error**, E , is the critical value times the standard error for the sample proportion.
- A **critical value** is a table value based on sampling distribution of the point estimate and the desired confidence level. z^* represents the **critical value** for a population proportion, p .
- The **standard error** is the standard deviation of the sample statistic. $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Confidence Interval for a Population Proportion, p

Point Estimate \pm Margin of Error

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

Confidence Interval for p : (LB, UB) or $LB < p < UB$

Interpretation: "We are [confidence level] confident the population proportion is between [lower bound] and [upper bound]."

true

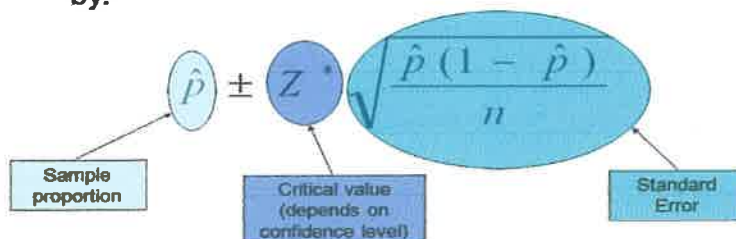
Conditions for Normality:

1. Representative sample
2. X , the number of successes, follows a binomial distribution
3. Both $n\hat{p}$ and $n\hat{q}$ are at least 5 (at least 5 successes and 5 failures)

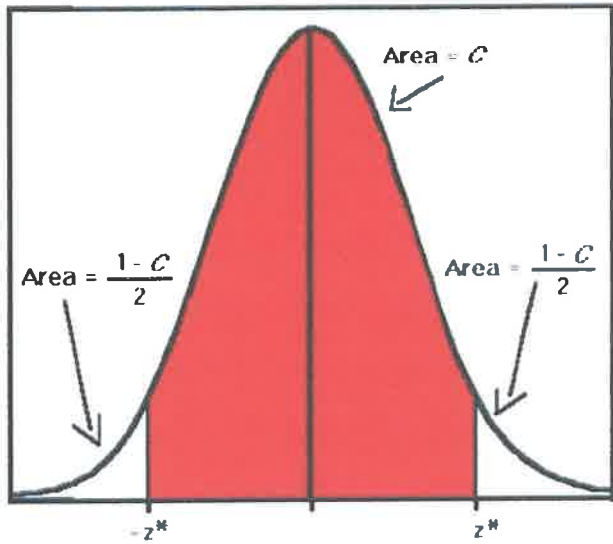


Confidence Interval Formula

- A confidence interval for the population p is given by:



Notice the \hat{p} is used in the standard error. That is because we do not know what p is (remember our purpose is to estimate p).

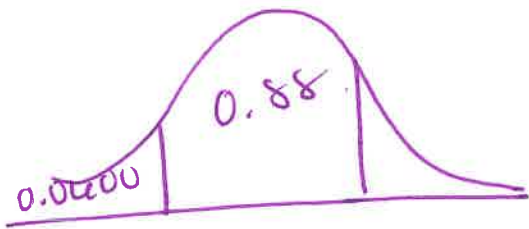


One strategy for determining z^* , or $z_{\alpha/2}$, for a general confidence level is to use the fact that $-z^*$ is the z-score that has area $(1-C)/2$ below it. This can easily be read from the z-table provided. Search within the body of the table and identify the corresponding value of $-z^*$.

pr. $\pm E$
 \uparrow

The table at the bottom left of our z-chart gives the critical values z^* for the common confidence levels of (90%, 95%, 98% and 99%)

Example 2. Find the critical value for a confidence interval in which we are 88% confidence.



$$1 - 0.88 = 0.12$$

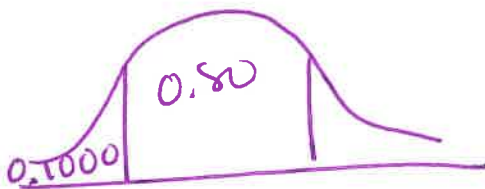
$$\frac{0.12}{2} = 0.0600$$

look up 0.0600 in MIDDLE of z-table

$$\frac{-1.55 + -1.50}{2} = -1.555$$

$$z^* = 1.555$$

Example 3. Find the critical value for a confidence interval in which we are 80% confident.

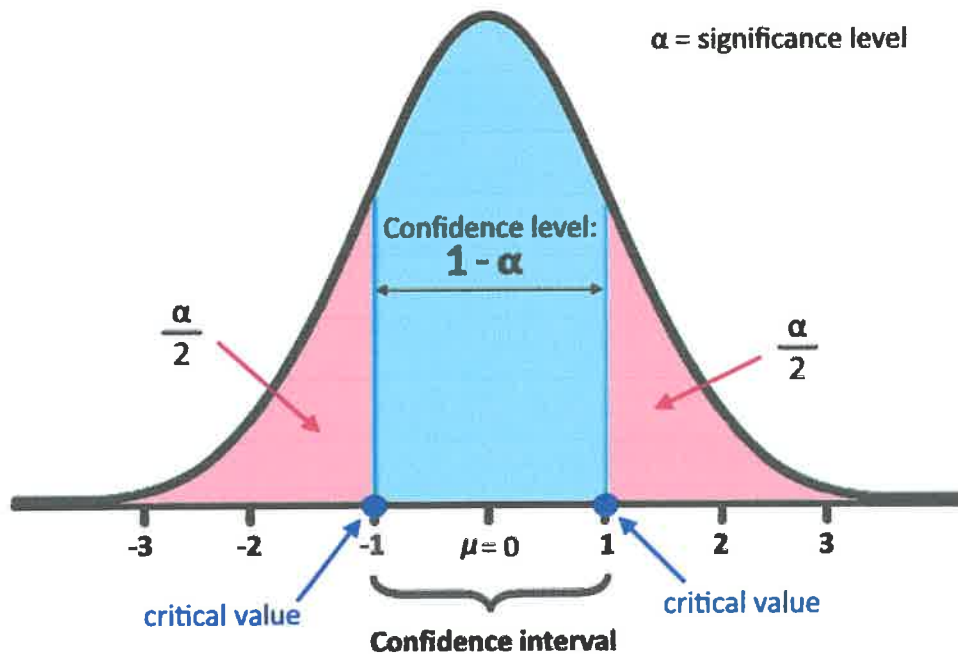


$$1 - 0.80 = 0.20$$

$$\frac{0.20}{2} = 0.1000$$

look up 0.1000 in MIDDLE of z-table

$$z^* = 1.28$$



α = significance level

Example 4. The Marist Institute for Public Opinion surveyed 883 randomly selected American adults about allergies. According to a report posted at the Institute's website, 36% of the sample answered "yes" to the question "Are you allergic to anything?"

- (a) Construct a 95% confidence interval to estimate the population parameter p = proportion of all American adults who are allergic to something.

point estimate $\hat{p} = 0.36$

sample size $n = 883$

critical value is $z^* = 1.96$ (to achieve 95% confidence)

$$\text{margin of error } E = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \sqrt{\frac{0.36(1-0.36)}{883}} = 0.03166$$

$$n = 883$$

$$\hat{p} = 0.36$$

The 95% confidence interval is point estimate \pm margin of error

$$\hat{p} \pm E$$

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

$$0.36 \pm 0.0316639$$

ans $(0.3283, 0.3917)$ or $32.83\% < p < 39.17\%$

- (a) Interpret the interval.

ans We are **95% confident** the **population proportion** of American adults who are allergic to something is between **32.83%** and **39.17%**.

- (b) Suppose a researcher claimed that less than 40% of the population would answer "yes." Address this claim based on our interval.

ans Because the value 0.40 does NOT lie in the interval, and our interval is LESS THAN 0.40, we can conclude the claim is true.

- (c) Address all assumptions and the validity of your results.

ans sampling methodology is sound

X follows a binomial distribution with

$$n\hat{p} = 883 * 0.36 = 317.88 \geq 5 \text{ and}$$

$$n\hat{q} = 883 * (1 - 0.36) = 565.12 \geq 5.$$

The needed assumption that $\hat{p} \sim$ normal is confirmed.

Example 5. A survey of 500 airline passengers found that 338 , or 68% , were satisfied with the service they received from the flight attendants. Calculate and interpret a 95% confidence interval for the proportion of passengers who are satisfied with the service from flight attendants.

(a) Find the estimate for p :

$$\hat{p} = 0.68$$

$$n = 500$$

$$\hat{p} = 0.68$$

(b) Calculate the 95% CI for p :

$$p.e. \pm E$$

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

$$0.68 \pm 1.96 \sqrt{\frac{0.68(1-0.68)}{500}}$$

$$95\%$$

$$z^* = 1.96$$

$$(0.6391, 0.7209)$$

(c) Interpret the confidence interval:

We are 95% confident the proportion of ALL passengers who are satisfied with the service from flight attendants is between 63.91% and 72.09%.

(d) Check the assumptions:

- sound sampling method
- $X \sim \text{binomial}$
- $n\hat{p} = 500(0.68) = 340 \geq 5$
- $n\hat{q} = 500(0.32) = 160 \geq 5$

The assumptions for $\hat{p} \sim \text{normal}$ is confirmed

(e) Suppose a researcher claimed that less than 68% of airline passengers were satisfied with the service from flight attendants. Address this claim according to our interval.

Because 0.68 is in our interval, we do not agree with this claim.

Working Backwards to Find the Margin of Error or Sample Mean

When we calculate a confidence interval, we find the sample mean, calculate the margin of error, and use them to calculate the confidence interval. (point estimate \pm margin of error)

However, sometimes when we read statistical studies, the study may state the confidence interval only. If we know the confidence interval, we can **work backwards** to find both the margin of error and the sample mean.

Finding the Margin of Error and Point Estimate When Given the Confidence Interval (LB, UB)

- Margin of Error: $E = \frac{UB-LB}{2}$
- Point Estimate: \bar{x} or $\hat{p} = \frac{UB+LB}{2}$

Example 6. Suppose you have the following confidence interval: (10.21, 10.59).

- (a) Find the margin of error.

$$E = 0.19$$

$$\frac{10.59 - 10.21}{2} = \frac{0.38}{2} = 0.19$$

- (b) Find the point estimate.

$$p.e. = 10.40$$

$$\frac{10.59 + 10.21}{2} = \frac{20.8}{2} = 10.40$$

$$10.21 + 0.19 = 10.40$$

Example 7. Suppose you have the following confidence interval: (8, 20).
Find the following without using a calculator.

- (a) The margin of error.

$$E = 6$$

$$\frac{20-8}{2} = \frac{12}{2} = 6$$

- (b) The point estimate.

$$p.e. = 14$$

$$\frac{20+8}{2} = \frac{28}{2} = 14$$

8.2: A Single Population Mean Using the Student t Distribution

To construct a confidence interval for a single unknown population mean, μ , where the population standard deviation is **UNKNOWN**, we need a point estimate for μ and the margin of error, E .

If we DO NOT KNOW the population standard deviation, σ , we use the t -distribution for means (instead of the z -distribution) to find the critical value needed.

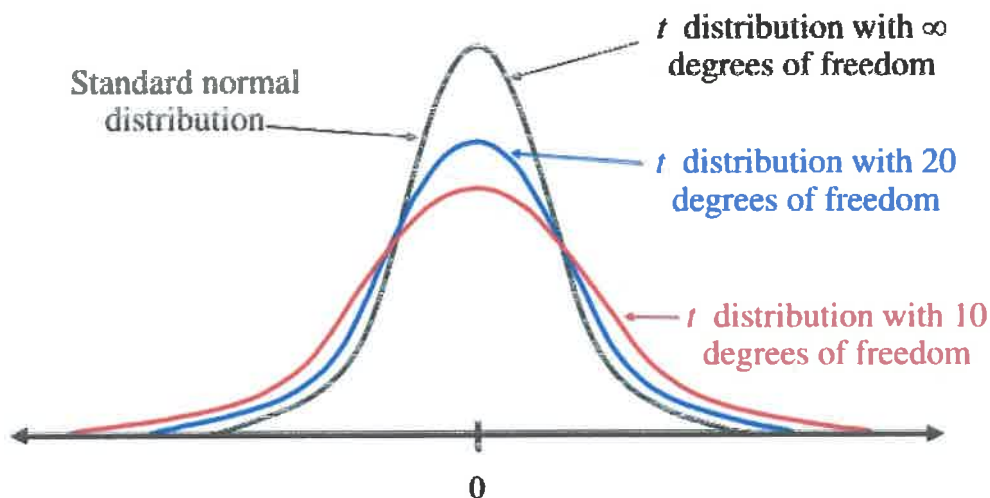
- The sample mean, \bar{x} , is the best point estimate of the population mean, μ .
- The margin of error, E , is the critical value (t^*) times the standard error for the sample mean, $\left(\frac{s}{\sqrt{n}}\right)$
- t^* represents the critical value from the t -distribution for the confidence level desired, with $n - 1$ degrees of freedom ($df = n - 1$).

The degrees of freedom is the number of sample observations that can vary after certain restrictions have been imposed. For the t distribution, the degrees of freedom is equal to $n - 1$.

NOTE: Your textbook uses $t_{\alpha/2}$ as the critical value, where we use t^* .

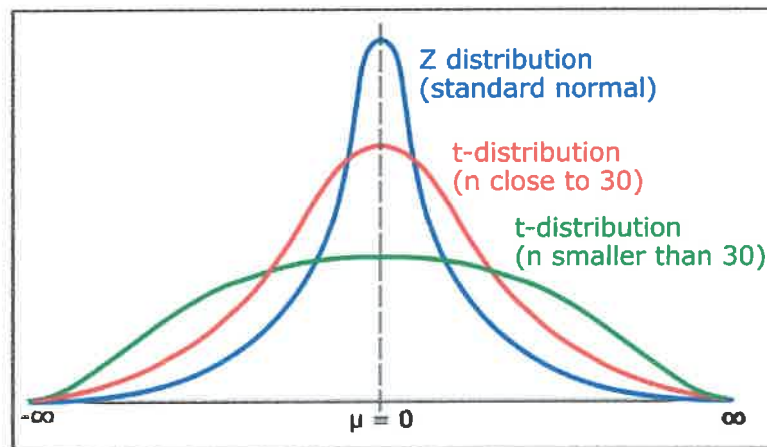
t Distribution

The t -distribution is used when n is small and σ is unknown.



Properties of the t -Distribution

- There are an infinite number of t distributions, one for each possible value for degrees of freedom. ($df = n - 1$)
- The t distribution has the same symmetric bell shape as the z distribution, but it reflects the greater variability that is expected with estimating σ .
- The t distribution has a mean of 0.
- The standard deviation of the t distribution varies with the sample size but is always greater than 1 (unlike the z).
- As the sample size n gets larger, the t distribution gets closer to the z distribution.



Example 8. Find the t^* critical value for the following:

- a. 95% confidence, $n = 15$.

$$df = n - 1 = 14$$

$$t^* = 2.145$$

- b. 99% confidence, $n = 43$

$$df = n - 1 = 42$$

$$t^* = 2.704$$

Excel: $\rightarrow 1 - CL$
 $= t.inv.2t(\alpha, df)$
 $= t.inv.2t(0.01, 42)$

- c. 80% confidence, $n = 30$

$$df = n - 1 = 29$$

$$t^* = 1.311$$

Confidence Interval for a Population Mean, μ

Point Estimate \pm Margin of Error

$$\bar{x} \pm t^* \left(\frac{s}{\sqrt{n}} \right)$$

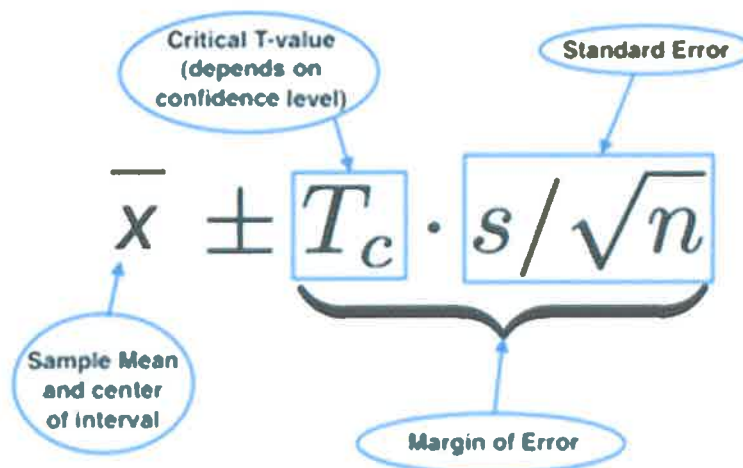
Confidence Interval for μ : (LB, UB) or $LB < \mu < UB$

Interpretation: "We are [confidence level] confident the population(true) mean is between [lower bound] and [upper bound]."

Conditions for Normality:

1. Representative Sample
2. One of the following:
 - (a) The population must be normally distributed, $X \sim \text{normal}$, OR
 - (b) The sample size needs to be *large* enough, $n \geq 30$.

- distribution of population
- bias?



Example 9. A physician wants to test a medicine to lower blood glucose concentrations in diabetic patients. The drug is considered "successful" if it can lower blood glucose concentrations below 110mg/dL. He uses the drug on 30 patients, and at the end of the trial the sample mean is 102 mg/dL with a sample standard deviation of 3.2. Assume values are normally distributed.

Construct a 90% confidence interval for the true value of the population mean blood glucose concentration.

means

$$n = 30$$

$$\bar{x} = 102$$

$$s = 3.2$$

(a) Check the conditions for the validity of the test.

- representative sample
- $n \geq 30$

BUT... we're told values are normally dist ✓

(b) Determine the critical value.

$$df = n - 1 = 29$$

$$90\%$$

$$t^* = 1.699$$

(c) Determine the point estimate and margin of error.

$$p.e. = \bar{x} = 102$$

$$E = t^* \cdot \frac{s}{\sqrt{n}} = 1.699 \cdot \frac{3.2}{\sqrt{30}} = 0.9926$$

(d) Construct the confidence interval.

$$p.e. \pm E$$

$$102 \pm 0.9926$$

$$(101.01, 102.99)$$

(e) Interpret the interval found in part (d).

We are 90% confident the true mean blood glucose concentration is between 101.01 mg/dL and 102.99 mg/dL.

< 110

2. The physician claims the drug is successful. Is this claim reasonable?

Yes!

Example 10. The Dallas Stars active roster has 24 players on it, and the average player weight (in pounds) is 200.2, while the standard deviation (in pounds) is 18.2.

- (a) Assuming that the data is somewhat symmetric and treating the Stars as a representative sample of NHL players, construct a 95% confidence interval on the average weight of hockey players.

$$\begin{aligned}
 & \text{P.e. } \pm E \\
 & \bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}} \\
 & 200.2 \pm 2.069 \cdot \frac{18.2}{\sqrt{24}} \\
 & 200.2 \pm 7.682 \\
 & (192.51, 207.89)
 \end{aligned}$$

means

$$\begin{aligned}
 n &= 24 \\
 \bar{x} &= 200.2 \\
 s &= 18.2 \\
 df &= n-1 = 23 \\
 t^* &= 2.069
 \end{aligned}$$

- (b) Interpret the interval.

We are 95% confident the (true) population mean weight of hockey players is between 192.51 pounds and 207.89 pounds.

- (c) Address the claim that the mean weight of all hockey players is over 195 pounds.

The entire interval is NOT over 195 pounds, we cannot agree with this claim.

* We can conclude the mean weight is over 190 lbs. *

- (d) Address all assumptions and the validity of your results.

- Representative sample ✓
- $n < 30$ and we don't know if the population is normal *

* We do know the data is "somewhat symmetric" ✓

Mixed Examples

Example 11. The Berkman Center for Internet & Society at Harvard conducted a study analyzing the privacy management habits of teen internet users. In a group of 50 teens, 13 reported having more than 500 Facebook friends on Facebook. Find a 90% confidence interval for the true proportion of teens who would report having more than 500 Facebook friends.

- (a) Obtain a point estimate for the proportion of teens who have more than 500 Facebook friends.

$$\hat{p} = \frac{x}{n} = \frac{13}{50} = 0.26$$

$$n = 50 \\ \hat{p} = 0.26$$

- (b) Verify that the requirements for constructing a confidence interval for p are satisfied.

- sound sampling method ✓

- $X \sim \text{binomial}$ ✓

$$- n\hat{p} = 50(0.26) = 13 \geq 5 \quad \checkmark$$

$$- n\hat{q} = 50(0.74) = 37 \geq 5 \quad \checkmark$$

- (c) Construct a 90% confidence interval for the true proportion of teens who have more than 500 Facebook friends.

$$z^* = 1.645$$

P.C. $\pm E$

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

$$0.26 \pm 1.645 \sqrt{\frac{0.26(0.74)}{50}}$$

$$(0.1580, 0.3620)$$

- (d) Interpret the confidence interval.

We are 90% confident the true proportion of teens who report having more than 500 Facebook friends is between 15.80% and 36.20%.

- (e) What is the effect of increasing the level of confidence on the width of the interval?

smaller interval
wider

80% \rightarrow 95%
() ())

Example 12: A study was done on acupuncture to determine how effective it is in relieving pain. Sensory rates for 15 subjects were measured which resulted in a mean of 8.23 and standard deviation of 1.67.

a) Construct a 95% confidence interval for the mean sensory rate for the population.

P.e. $\pm E$
 $\bar{x} \pm t^* \frac{S}{\sqrt{n}}$
 $8.23 \pm 2.145 \cdot \frac{1.67}{\sqrt{15}}$
 $(7.305, 9.155)$

means
 $n = 15$
 $\bar{x} = 8.23$
 $S = 1.67$
 $df = 14$
 $t^* = 2.145$

b) Interpret the confidence interval.

We are 95% confident the ^{true} mean sensory rate ~~for~~ is between 7.31 and 9.15.

c) Checks the assumptions for normality.

- sound sampling method ✓
- $n \geq 30$? X
- population normal? X

The assumptions for normality are not confirmed; we should be cautious with these results

Example 13. Suppose you have the following confidence interval: (0.247, 0.304).

(a) Find the margin of error.

$$E = 0.0285$$

$$\frac{0.304 - 0.247}{2}$$

(b) Find the point estimate.

$$p.e. = 0.2755$$

~~pva.~~

$$0.247 + 0.0285$$

$$\begin{aligned} & \frac{0.304 + 0.247}{2} \\ &= \frac{0.551}{2} = 0.2755 \end{aligned}$$

Example 14. What are the two (2) ways to reduce the width of a confidence interval?

- a. Larger sample size and higher level of confidence
- b. Smaller sample size and lower level of confidence
- c. Larger sample size and lower level of confidence
- d. Unable to determine without seeing the data

↓ CL ↓ int

Example 15. Suppose you compute an interval with a sample size of 77. What will happen to the confidence interval if the sample size increases to 90?

- a. The confidence interval will widen.
- b. The confidence interval will narrow.
- c. The width of the confidence interval will stay the same.

Example 16. A student was asked to find a 99% confidence interval for the proportion of students who take notes using data from a random sample of $n = 77$. Which of the following is a correct interpretation of the interval $0.12 < p < 0.3$?

- a. There is a 99% probability that the proportion of the population is between 0.12 and 0.3.
- b. The proportion of all students who take notes is between 0.12 and 0.3, 99% of the time.
- c. With 99% confidence, a randomly selected student takes notes in a proportion of their classes that is between 0.12 and 0.3.
- d. With 99% confidence, the proportion of all students who take notes is between 0.12 and 0.3.

Example 17. Answer the following.

- a) Given a confidence interval of (3.4, 5.6), can we conclude that the parameter is no more than 3.8? **no**
- b) Given a confidence interval of (3.4, 6.2), can we conclude that the parameter is different from 5.8? **no**
- c) Given a confidence interval of (3.4, 5.6), can we conclude that the parameter is at least 3.2? **yes**