CHAPTER 8: CONFIDENCE INTERVAL ESTIMATES for Means and Proportions

Introduction: We want to know the value of a parameter for a population. We don't know the value of this parameter for the entire population because we don't have data for the entire population. (If we did already know the value of the parameter, we wouldn't need to do any statistical investigation or calculations.) We will use sample statistics to estimate population parameters.

Recall from chapter 2: A parameter is ____

____

If we don't know the value of a population parameter, we can estimate it using a sample statistic.

Recall from chapter 2: A statistic is____

____

Using data from a sample to draw a conclusion about a population is called ______________ statistics.

Two Types of Estimates for population parameters:

POINT ESTIMATE: A population parameter can be estimated by one number: the sample statistic. This is called a **point estimate.**

(Statistical theory has identified desirable properties of point estimates, which are studied in more depth in upper level statistics classes. One desirable property is that a point estimate be "unbiased", meaning that the average of the point estimates from all possible samples would equal the true value of the population parameter.)

- The "best" point estimate of a population mean μ is______________________________
- The "best" point estimate of a population proportion p is____________________________
- The "best" point estimate of a population standard deviation σ is__________________________

CONFIDENCE INTERVAL ESTIMATE:

- The population parameter is estimated by an **interval of numbers** that we believe contains the true (unknown) value of the population parameter.
- We can state how confident we are that the interval estimate contains the true value of the parameter.
- This **confidence interval estimate** is built using two items: a point estimate, and margins of error
- The margins of error are also called error bounds.
- We will use confidence interval estimates based on sample data to estimate a population average (mean) and to estimate a population proportion.
- Confidence intervals for means and proportions are symmetric; the point estimate is at the center of the interval. The endpoints of the interval are found as.

  (point estimate – error bound , point estimate + error bound)
- For some other parameters, such as standard deviation, a confidence interval may not be symmetric about the point estimate, moving different distances above and below the point estimate to the ends of the interval estimate.

We'll start in class by examining a jar with beads to determine the proportion of beads in the jar that are blue; after we explore the concepts, then we'll move on to the mathematical calculations.
In Examples 1, 2, 3 we learn to calculate the point estimates and the error bounds and what they mean.
The last 3 pages of these notes has a concise summary of formulas, procedures, and interpretations.

CHAPTER 8 EXAMPLE 1:	CONFIDENCE INTERVAL ESTIMATE for an unknown POPULATION PROPORTION p

a. Statistics and data in this example are based on information from :
http://sf.streetsblog.org/2014/08/15/car-free-households-are-booming-in-san-francisco/
http://en.wikipedia.org/wiki/List_of_U.S._cities_with_most_households_without_a_car

A trend in urban development is to reduce the need for residents to have a car; city neighborhoods are often ranked for "walkability'.

- The US city with the lowest car ownership rate is New York City; a majority (56%) of households are "car-free" with only 44% of households owning any vehicles.
- San Jose has the highest car ownership rate of large US cities; only about 6% of households "car-free".
- San Francisco's percent of "car-free" households has changed rapidly in recent years.

Suppose a recent study of 1200 households in San Francisco showed that 372 households were "car-free". Construct and interpret a 95% confidence interval for the true proportion of households in San Francisco that are "car-free". Use a 95% confidence level.

population parameter: p = ______________________________

____

random variable p′ = ______________________________

We are using sample data to estimate an unknown proportion for the whole population

HOW TO CALCULATE THE CONFIDENCE INTERVAL		
Point Estimate = **p ′**	Confidence Level **CL** is area in the middle	Standard Error
Error Bound = (Critical Value)(Standard Error) **EBP** = $Z_{\alpha/2}\sqrt{\frac{p'q'}{n}}$	Critical Value is $Z_{\alpha/2}$ is the Z value that creates area of CL in the middle; Z ~N(0,1)	$\sqrt{\frac{p'q'}{n}}$
Confidence Interval = Point Estimate + Error Bound Confidence Interval = p ′ $\pm$ EBP	Use POSITIVE value of Z invnorm(*area to left*, 0, 1)	

<u>Calculations and interpretation in context of the problem:</u>

CHAPTER 8 EXAMPLE 1: Using Confidence Intervals

1b. In Chapter 9 we'll learn a more appropriate technique for asking and answering questions about a particular possible value of a parameter.
At this point, however, we'll use our confidence interval estimate of the parameter to see if it can provide any insight into the parameter's value.

In Example 1 the point estimate is ____ = _______ ; the confidence interval is ___________________

(i) Can we conclude with 95% confidence that more than 25% of SF households are "car-free"? Explain.

(ii) Can we conclude with 95% confidence that more than 30% of SF households are "car-free"? Explain.

1c. **What does it mean when we say the confidence level is 95% or that "we are 95% confident"?**

CHAPTER 8 EXAMPLE 2: CONFIDENCE INTERVAL ESTIMATE for unknown POPULATION MEAN μ when the POPULATION STANDARD DEVIATION σ is KNOWN

A soda bottling plant fills cans labeled to contain 12 ounces of soda. The filling machine varies and does not fill each can with exactly 12 ounces. To determine if the filling machine needs adjustment, each day the quality control manager measures the amount of soda per can for a random sample of 50 cans. Experience shows that its filling machines have a known population standard deviation of 0.35 ounces.

In today's sample of 50 cans of soda, the sample average amount of soda per can is 12.1 ounces.

Construct and interpret a 90% confidence interval estimate for the true population average amount of soda contained in all cans filled today at this bottling plant. Use a 90% confidence level.

X = ____

population parameter: μ = ____

random variable $\overline{X}$ = ____

We are using sample data to estimate an unknown mean (average) for the whole population

<table>
<tr><th colspan="3">HOW TO CALCULATE THE CONFIDENCE INTERVAL for μ</th></tr>
<tr><td colspan="3">When σ IS known, use the Standard normal distribution Z ~ N(0,1)</td></tr>
<tr><td>Point Estimate = $\overline{X}$
Error Bound = (Critical Value)(Standard Error)
EBM = $Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$
Confidence Interval = Point Estimate + Error Bound
Confidence Interval = $\overline{X} \pm$ EBM</td><td>Confidence Level CL is area in the middle
Critical Value is $Z_{\alpha/2}$ is the Z value that creates area of CL in the middle; Z ~N(0,1)
Use POSITIVE value of Z
invnorm(area to left, 0, 1)</td><td>Standard Error
$\frac{\sigma}{\sqrt{n}}$</td></tr>
</table>

Calculations and interpretation in context of the problem:

CHAPTER 8 EXAMPLE 3: CONFIDENCE INTERVAL ESTIMATE for unknown POPULATION MEAN μ when the POPULATION STANDARD DEVIATION σ is NOT KNOWN

a. The traffic commissioner wants to know the average speed of all vehicles driving on River Rd. Police use radar to observe the speeds for a sample of 20 vehicles on River Rd.

For the vehicles in the sample, the average speed is 31.3 miles per hour with standard deviation 7.0 mph.

Construct and interpret a 98% confidence interval estimate of the true population average speed of all vehicles on River Rd. Use a 98% confidence level.

X = ____________________

population parameter: μ = ____________________

random variable $\overline{X}$ = ____________________

We are using sample data to estimate an unknown mean (average) for the whole population

HOW TO CALCULATE THE CONFIDENCE INTERVAL for μ		
When σ is NOT known, use the t distribution with degrees of freedom = sample size – 1 : t with df = n – 1)		
Point Estimate = $\overline{X}$	Confidence Level **CL** is area in the middle	
Error Bound = (Critical Value)(Standard Error) **EBM** = $t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$	Critical Value $t_{\alpha/2}$ is the t value that creates an area of CL in the middle; Use t distribution with df = n – 1	Standard Error $\frac{s}{\sqrt{n}}$
Confidence Interval = Point Estimate + Error Bound Confidence Interval = $\overline{X} \pm$ EBM	Use POSITIVE value of t TI-84: $t_{\alpha/2}$ = invT(*area to left, df*)	

Calculations and interpretation in context of the problem:

b. In Example 3, suppose that you were not given the sample mean and sample standard deviation and instead you were given a list of data for the speeds (in miles per hour) of the 20 vehicles.

19 19 22 24 25 27 28 37 35 30 37 36 39 40 43 30 31 36 33 35

How would you use the data to do this problem?

NOTE: Use of t-distribution requires the underlying population of individual values to be approximately normally distributed. It is OK if this assumption is violated a little, but if the underlying population of individual values has a distribution that differs too much from the normal distribution, then this confidence method is not appropriate, and statisticians would use other techniques that we do not study in Math 10.

CHAPTER 8: Confidence Interval for a Proportion: Calculating the Sample Size needed in a Study

Given a desired confidence level and a desired margin of error, how large a sample is needed?

$$\text{EBP} = Z_{\alpha/2}\sqrt{\frac{p'q'}{n}}$$

We know the error bound EBP that we want.

We know the confidence level CL we want, so we can find $Z_{\alpha/2}$ corresponding to the desired CL.

We don't know p' or q' until we do the study, so we will assume for now that p' = q' = ½ = 0.5

Then we can substitute all these values into EBP = $Z_{\alpha/2}\sqrt{\frac{p'q'}{n}}$ and solve for n.

Solving EBP = $Z_{\alpha/2}\sqrt{\frac{p'q'}{n}}$ for n gives . n = $\left(\frac{Z_{\alpha/2}}{EBP}\right)^2 p'q'$.

Sample Size Formula to determine sample size n needed when estimating a population proportion p	$n = \left(\frac{Z_{\alpha/2}}{EBP}\right)^2 (.25)$	*The 0.25 appears in the formula because we are assuming that* p' = q' = ½ = 0.5 **ALWAYS ROUND UP to the next higher integer**

EXAMPLE 4: **Finding the Sample Size:**

a. Public opinion and political polls often do surveys with a 95% confidence level and 3% margin of error. Find the sample size needed.

$$n = \left(\frac{Z_{\alpha/2}}{EBP}\right)^2 (.25) =$$

b. Suppose a margin of error of 2% was wanted with a 95% confidence level. Find the sample size needed.

$$n = \left(\frac{Z_{\alpha/2}}{EBP}\right)^2 (.25) =$$

c. Suppose a margin of error of 3% was wanted with a 90% confidence level. Find the sample size needed.

$$n = \left(\frac{Z_{\alpha/2}}{EBP}\right)^2 (.25) =$$

d. Suppose a poll uses a sample size of n=100, and a confidence level of 95%.
Estimate the expected error bound using p' = q' = ½ = 0.5

$$\text{EBP} = Z_{\alpha/2}\sqrt{\frac{p'q'}{n}}$$

Note the actual error bound will differ after the study is done because we will know p' and q' and will no longer be estimating that p'=q'=0.5

e. Is the error bound in part d large or small compared to the examples in parts a, b, c?
Explain why this happened.

CHAPTER 8 EXTRA PRACTICE EXAMPLES : CALCULATING CONFIDENCE INTERVALS

Sources for Practice Examples: #6 based on Chapter 8 Practice in Introductory Statistics from OpenStax available for download for free at http://cnx.org/content 11562/latest/. ; #6 based on car rental rate information from www.tripadvisor.com on 1/8/2019; #7 based on information from http://www.pewinternet.org/2016/02/11/15-percent-of-american-adults-have-used-online-dating-sites-or-mobile-dating-apps/ ; #8 based on information from http://www.pewsocialtrends.org/2016/05/24/for-first-time-in-modern-era-living-with-parents-edges-out-other-living-arrangements-for-18-to-34-year-olds/;

PRACTICE EXAMPLE 5: A supermarket chain is deciding what produce providers to purchase from. A sample of 20 heads of lettuce is selected to estimate the average weight of the lettuce from this provider. The population standard deviation for the weight is known to be 0.2 pounds. The sample of 20 heads of lettuce had a mean weight of 2.2 pounds with a sample standard deviation of 0.1 pounds.

Calculate and interpret a confidence interval estimate for the true average weight of all heads of lettuce from this provider. Use a 90% confidence level.

PRACTICE EXAMPLE 6: If you rent a car from Enterprise Car Rental, the daily rental rate for the same car in the same city can vary depending on which Enterprise location ~~location~~ in the city you rent the car from. On January 8, 2019, daily rental rates at a sample of 8 San Francisco locations for a Hyundai Elantra were quoted as $30, $31, $31, $32, $35, $36, $36, $43.

Calculate and interpret a 96% confidence interval estimate for the true average daily rental rate for a Hyundai Elantra at all Enterprise locations in San Francisco on that date. Use a 96% confidence level.

PRACTICE EXAMPLE 7: Suppose a survey of 600 young adults age 18 to 24 showed that 162 had used online dating. Calculate and interpret a confidence interval estimate for the true proportion of all young adults age 18 to 24 who ever used online dating. Use a 95% confidence level.

PRACTICE EXAMPLE 8: Suppose a survey of 500 people age 18 to 34 indicated that 32.2% of them live with one or both of their parents. Calculate and interpret a confidence interval estimate for the true proportion of all people age 18 to 34 who live with one or both parents. Use a 94% confidence level.

CHAPTER 8: FLOW CHART VIEW OF FORMULAS FOR CONFIDENCE INTERVAL ESTIMATES

For more details, see the presentation of the formulas in boxes for each case on the next page.

- **Mean μ (average)**
  - Mean μ (average) if population standard deviation **σ IS KNOWN**
    - Point Estimate: $\overline{X}$
      Error Bound:
      $$EBM = Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$
    - Distribution: Normal
  - Mean μ (average) if population standard deviation **σ IS NOT KNOWN**
    - Point Estimate $\overline{X}$
      Error Bound:
      $$EBM = t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$
    - Distribution: Students **t**
      t_{n-1}
      degrees of freedom $= n-1$
- **Proportion p**
  - Point Estimate p′
    Error Bound:
    $$EBP = Z_{\alpha/2}\sqrt{\frac{p'q'}{n}}$$
    $$= Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$
  - Distribution: Normal

Confidence Interval for a Proportion p

To estimate a population proportion p (binomial probability of success).

Point Estimate $\pm$ Margin of Error (Margin of Error is also called Error Bound)

Point Estimate: Sample Proportion: $p' = \frac{x}{n} = \frac{\text{number of successes in sample}}{\text{total number in sample}}$

Error Bound: EBP = (critical value)(standard error) $= Z_{\alpha/2}\sqrt{\frac{p'q'}{n}}$

The critical value $Z_{\alpha/2}$ depends on the confidence level. $Z_{\alpha/2}$ is the Z value that puts an area equal to the confidence level (CL) in the middle of standard normal distribution N(0,1)

$Z_{\alpha/2}$ tells us how many "appropriate standard deviations" to enclose about the point estimate, where the "standard error" $\sqrt{\frac{p'q'}{n}}$ is the appropriate standard deviation for a proportion

Confidence Interval: $p' \pm EBP$ which is $p' \pm Z_{\alpha/2}\sqrt{\frac{p'q'}{n}}$

Confidence Interval for a Mean μ when σ is known

To estimate the population average μ if we already know the population standard deviation σ.

Point Estimate $\pm$ Margin of Error (Margin of Error is also called Error Bound)

Point Estimate: Sample Average (Sample Mean) $\overline{X}$

Error Bound: EBM = (critical value)(standard error) $= Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

The critical value $Z_{\alpha/2}$ depends on the confidence level. $Z_{\alpha/2}$ is the Z value that puts an area equal to the confidence level (CL) in the middle of standard normal distribution N(0,1)

$Z_{\alpha/2}$ tells us how many "appropriate standard deviations" to enclose about the point estimate, where the "standard error" $\frac{\sigma}{\sqrt{n}}$ is the appropriate standard deviation for the sample mean

Confidence Interval: $\overline{X} \pm EBM$ which is $\overline{X} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

Confidence Interval for a Mean μ when σ is NOT known

To estimate the population average μ and we do not know the population standard deviation σ.
Use the sample standard deviation s to estimate the population standard deviation σ

Point Estimate $\pm$ Margin of Error (Margin of Error is also called Error Bound)

Point Estimate: Sample Average (Sample Mean) $\overline{X}$

Error Bound: EBM = (critical value)(standard error) $= t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$

The critical value $t_{\alpha/2}$ depends on the confidence level. $t_{\alpha/2}$ is the t value that puts an area equal to the confidence level (CL) in the middle of the student t-distribution with $n - 1$ degrees of freedom

$t_{\alpha/2}$ tells us how many "appropriate standard deviations" we need to move away from the point estimate, where $\frac{s}{\sqrt{n}}$ is an *estimat*e of the standard error ("appropriate standard deviation") for the sample mean

Confidence Interval: $\overline{X} \pm EBM$ which is $\overline{X} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$

YOU ARE <u>NOT</u> PERMITTED TO BRING A <u>PRINTOUT</u> OF THIS PAGE AS YOUR NOTES FOR AN EXAM OR QUIZ.
You CAN write whatever information you want from this page into your handwritten notes for exams or quizzes.

Interpreting the Confidence Interval for a PROPORTION (*2 ways to word it*)

We estimate with _____% confidence that the true proportion of the population that *describe the population parameter in the situation of this problem* is between ______ and ________

We estimate with _____% confidence that between _______% and _______% of the population *describe the population parameter in the situation of this problem*

Interpreting the Confidence Interval for a MEAN (average)

We estimate with _____% confidence that the true population average (or mean) *describe the population parameter in the situation of this problem* is between _______ and _________

What is the meaning of the Confidence Level? What does it mean to be CL% confident?

The confidence level represents an "expected accuracy rate" for the confidence interval process. It tells us what percent of confidence intervals calculated in this manner would be "good estimates".

If we took repeated samples and calculated many confidence interval estimates (one for each sample), we expect that CL% of the confidence interval estimates would be "good estimates" that successfully enclose (capture) the true value of the population parameter that we want to estimate.

If we took repeated samples), we expect that 100% − CL% of the confidence interval estimates would be "bad estimates" that would NOT enclose (capture) the true value of the population parameter.

Note that the confidence interval does not estimate individual data values. It estimates proportions or averages. It does NOT imply that CL% of the data lies within the confidence interval.

Finding the Point Estimate and Error Bound (Margin of Error) if we know the Confidence Interval:

The interval is (lower bound, upper bound)

Point Estimate = (lower bound + upper bound)/2 Error Bound = (upper bound − lower bound)/2

To find Z that puts the area equal to the confidence level "in the middle"

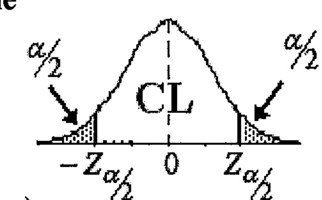
CL specifies the area in the middle

$\alpha = 1 - CL$ is area "outside", split equally between both tails

$\alpha/2$ is area in one tail.

To find $Z_{\alpha/2}$: invnorm($1-\frac{\alpha}{2}$, 0, 1)

OR use invnorm($\frac{\alpha}{2}$, 0, 1) and take absolute value (drop the "−" sign)



Without calculator: Use a standard normal probability table to find Z.

To find t that puts the area equal to the confidence level "in the middle"

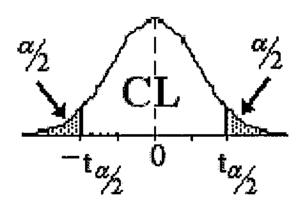
CL specifies the area in the middle

$\alpha = 1 - CL$ is area "outside", split equally between both tails

$\alpha/2$ is area in one tail. df = degrees of freedom = n−1

To find $t_{\alpha/2}$: TI-84+: invT($1-\frac{\alpha}{2}$, df)

OR use invT ($\frac{\alpha}{2}$, df) and take absolute value (drop the "−" sign)



TI-83,83+: Use INVT program; ask instructor to download it to your calculator: PRGM INVT
enter area to the left and df after prompts: area to left is $1-\frac{\alpha}{2}$; (if using $\frac{\alpha}{2}$ as area to left , drop the "−" sign)

Without calculator or if calculator does not have inverse t: Use a student's-t distribution probability table. Value of t is found at the intersection of the column for the confidence level and row for degrees of freedom

YOU ARE NOT PERMITTED TO BRING A PRINTOUT OF THIS PAGE AS YOUR NOTES FOR AN EXAM OR QUIZ. You CAN write whatever information you want from this page into your handwritten notes for exams or quizzes.