Chapter 7: Estimating Parameters and Determining Sample Sizes

Section 7.1: Estimating a Population Proportion Stat 50

INTRO

DESCRIPTIVE STATISTICS	+	PROBABILITY *Find a level of RELIABILITY*	=	INFERENTIAL STATISTICS

Look at the tray of beans that was brought to class today. We are going to try and estimate the percentage of black beans in the tray.

What percentage of the tray's contents do you feel are black beans? Write your guess below.

a. How could we find out if we are right?

b. Based on our sample, the instructor's best guess is that there is between _____% and _____% of black beans in the tray. That's a proportion of between _____ and _____.

What level of agreement do you have with this guess? (*Circle one*)

Totally agree Somewhat agree Somewhat disagree No Way!

c. Your instructor has revised their guess! Now they believe that there is between _____% and _____% of black beans in the tray. That's a proportion of between _____ and _____.

What level of agreement do you now have with the instructor's new guess? (*Circle one*)

Totally agree Somewhat agree Somewhat disagree No Way!

d. One last revision! Now they believe that there is between _____% and _____% of black beans in the tray. That's a proportion of between _____ and _____.

What level of agreement do you now have with the instructor's new guess? (*Circle one*)

Totally agree Somewhat agree Somewhat disagree No Way!

e. Discuss:

- What would happen if we took another sample? Would we get the same sample proportion?

- Would you like to know the true proportion of black beans in the entire bag?

Def ***Point Estimate*** A single value used to approximate a population parameter.

Note: *The sample proportion $\hat{p}$ is the point estimate of the population proportion p.*

SAMPLE PROPORTION: $\hat{\boldsymbol{p}} = \frac{\textbf{\textit{successes}}}{\textbf{\textit{total}}} =$ _____

Ex: I took a sample of 40 PCC students and asked them "Do you think Sir Pugsly Farnsworth Esquire III is cute or not?" The results were as follows: **24** said **YES** and **16** said **NO**.

a. What is the variable? Is it qualitative or quantitative?

b. This is a _____ problem since there are only two outcomes.

c. What would $\hat{p}$ be?

d. Which statement do you think we can make from this point estimate?

1. ***Exactly*** _____ of ALL PCC students think Sir Pugsly Farnsworth Esquire III is cute.
2. ***About*** _____ of ALL PCC students think Sir Pugsly Farnsworth Esquire III is cute.

From this example, we can imagine that there may be some _____ associated with taking a sample from a population. We will create an interval around $\hat{p}$ with the hope that the *true proportion* p lies within that interval and we state a certain level of confidence we have.

$$(\hat{p} \pm ERROR)$$

Def A ***confidence interval*** (abbreviated CI) is a range (or interval) of values used to estimate the true value of a population parameter.

Def ***Margin of Error*** The maximum likely difference between the observed sample proportion $\hat{p}$ and the true value of the population proportion p.

MARGIN OF ERROR FORMULA: $$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

Ex: If the margin of error is 8%, then what interval would you expect the **true proportion** of PCC students to be who think Sir Pugsly Farnsworth Esquire III is cute?

EFFECTS ON THE WIDTH OF YOUR CONFIDENCE INTERVAL

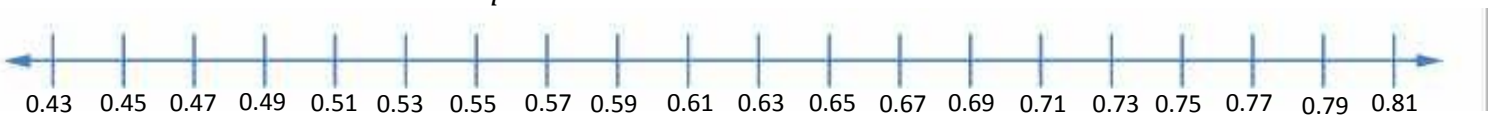
1)

- If n is small,
- If n is large,

2)

- The lower the confidence level,
- The higher the confidence level,

Let's use our first $\hat{p}$ = _____ and then take two more samples where the $\hat{p}$'s were .51 and then .71. Using the same margin of error, if the true proportion of PCC who thinks Sir Pugsly is cute is 0.65, then which $\hat{p}$'s had confidence intervals that contained p?



It looks like our probability of success is _____ right now, but are there only three samples of 40 PCC students that can be taken?

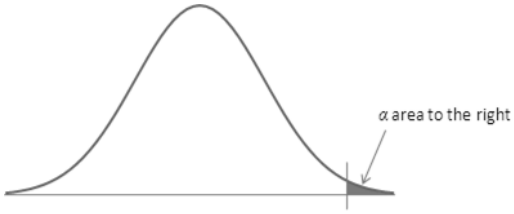
Def ***Confidence Level*** The probability equal to the proportion of times that the confidence interval actually contains the true population parameter.

Note: If we were to repeat the estimation process a large number of times, the confidence level is the percentage of confidence intervals that would actually contain the true population proportion.

Notation: The confidence level is denoted by $1 - \alpha$

Note: The ***critical value*** *(denoted* $z_{\alpha/2}$*) is the positive* z *score that separates an area of* $\alpha/2$ *in the right tail of the standard normal distribution. It is determined by the confidence level.*

Ex: Find $z_{0.02}$



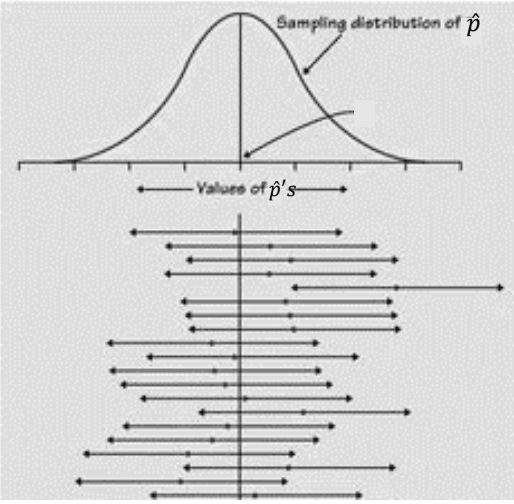
Ex: Find the margin of error if $\hat{p} = 0.48$, $n = 200$, and the confidence level is as follows:

a. 96% confidence level

b. 80% confidence level

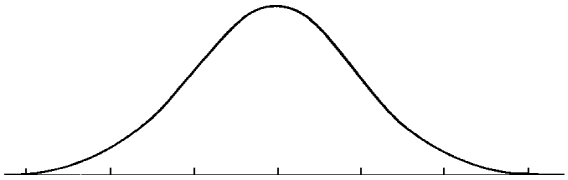
Ex: In the example to the right we took twenty different samples of size n, and found the corresponding confidence intervals with a 95% level of confidence.

A 95% confidence interval indicates that _____ out of **20** samples from the same population will produce confidence intervals that contain p.



Ex: In a 2018 random survey of 160 American Democrats, 136 said that they support Medicare-for-all, also known as single-payer healthcare. Find the 95% confidence interval estimate for the true proportion of Democrats in America who support Medicare-for-all.

Identify point estimate (sample proportion)



Determine critical value $z_{\alpha/2}$

Find margin of error

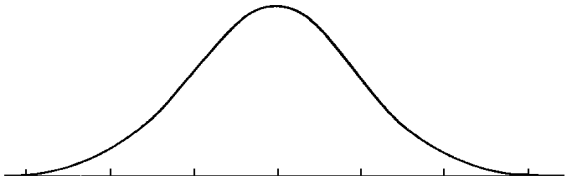
Construct confidence interval

Interpretation of CI

Does the proportion of Democrats who support Medicare-for-all appear to be substantially different than the 70% rate for the general population? If so, why do you think that is?

Ex: In a 2018 survey on climate change, 626 out of 1278 Americans (18 years or older) were "extremely" or "very sure" it is happening. Construct a 90% confidence interval for the true percentage of Americans who are "extremely" or "very sure" climate change is happening.

Identify point estimate (sample proportion)



Determine critical value $z_{\alpha/2}$

Find margin of error

Construct confidence interval

Interpretation of CI

GRAPHING CALCULATOR (TI-83 OR 84)

Instructions: (a) STAT $\Rightarrow$ TESTS $\Rightarrow$ 1-PropZInt

(b) Enter $\begin{cases} x = \text{number of successes} \\ n = \text{number of trials} \\ C-Level = \text{confidence level} \end{cases}$

Determining Sample Size

Point Estimate $\hat{p}$ Known	*Point Estimate $\hat{p}$ Unknown*
$n = \dfrac{[\, z_{\alpha/2} \,]^2 \, \hat{p}\,\hat{q}}{E^2}$	$n = \dfrac{[\, z_{\alpha/2} \,]^2}{4E^2}$

Ex: An economist wants to know if the proportion of the children in the United States that speak a language other than English at home has changed since 2016, when 22% of children did. How many children would need to be surveyed if the economist wants to be within 2 percentage points of the true proportion with 92% confidence?

Ex: Before US presidential elections potential candidates have their exploratory teams work to determine the percentage of people who will vote for their candidate. Kamala Harris has such a team looking into what percentage of people would vote for her. If they want to construct a 98% confidence interval with a 3% margin of error, how many people do they need to poll to achieve this?

CONFIDENCE INTERVAL FOR THE POPULATION MEAN

Alternative Forms: $\bar{x} - E < \mu < \bar{x} + E$ or $\bar{x} \pm E$ or $(\bar{x} - E,\ \bar{x} + E)$

Requirements

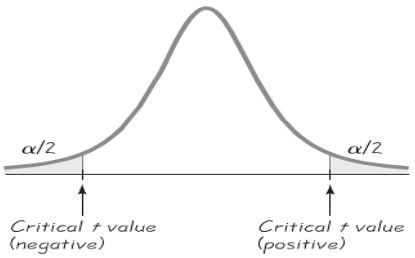
1. The sample is a simple random sample.
2. The value of the population standard deviation σ is not known.
3. Either or both of the given conditions are satisfied: $\begin{cases} \text{The population is normally distributed} \\ \text{or} \\ n > 30 \end{cases}$

To create these confidence intervals, we won't have σ but we can get a sample standard deviation ____. The larger the sample, the closer the sample s will get you to σ. We will need to use a different distribution.

STUDENT T DISTRIBUTION

STUDENT'S t- DISTRIBUTION RULES	• _____ - shaped but the tails will be _____ than a standard normal curve. • Centered around $\mu_t =$ ____ and $\sigma_t = \frac{}{\sqrt{n}}$ • Characterized by the *degree of freedom* FORMULA: $df =$ • We change $\boldsymbol{z_{\frac{\alpha}{2}}}$ to *_____

Ex: Find the critical t −value that corresponds to 98% confidence with 15 degrees of freedom.



Def ***Point Estimate*** A single value used to approximate a population parameter.

Note: The sample mean $\bar{x}$ is the point estimate of the population mean μ.

SAMPLE MEAN: $\bar{x} = \dfrac{\sum x}{n}$

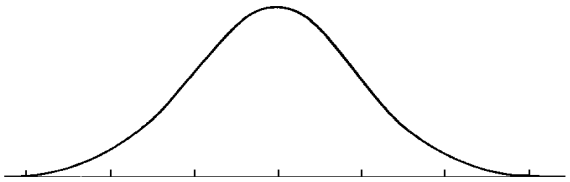
Def ***Margin of Error*** The maximum likely difference between the observed sample mean $\bar{x}$ and the true value of the population mean μ.

MARGIN OF ERROR FORMULA: $E = t_{\alpha/2} \cdot \dfrac{s}{\sqrt{n}}$

Ex: A 2017 sample of 34 cities in the US found that the average cost of a wedding was $36,000 with a standard deviation of $14,114. Determine a 99% confidence interval for the corresponding population mean cost of a US wedding.

Identify point estimate (sample mean)

Determine critical value $t_{\alpha/2}$



Find margin of error

Construct confidence interval

Interpretation of CI

GRAPHING CALCULATOR (TI-83 OR 84)

Instructions: (a) STAT $\Rightarrow$ TESTS $\Rightarrow$ TInterval

(b) Enter $\begin{cases} \text{Data} \rightarrow \text{ if individual values are entered in list} \\ \text{Stats} \rightarrow \text{ if using summary statistics} \end{cases}$

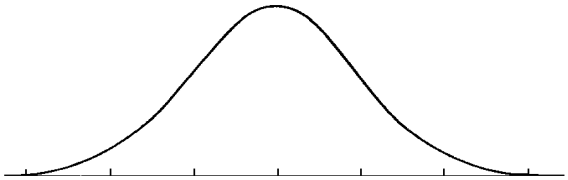
Ex: A Statistics instructor wanted to find out the mean amount of time PCC students spent on social media per day. She randomly sampled 12 students and found out how many minutes per day they used social media. Those results are below:

80	112	95	72	150	120	85	67	92	102	50	115

Assuming that such times for all PCC students are normally distributed, make a 95% confidence interval for the corresponding population mean amount of time spent on social media for all students at PCC.

Find point estimate (sample mean)

Determine critical value $t_{\alpha/2}$



Find margin of error

Construct confidence interval

Interpretation of CI

Ex: The last example produced an error within _____, and with the fact that we were 95% confident, we believed that the true _____ number of minutes spent on social media by PCC students is between _____ minutes and _____ minutes.

Let's say we want to be 95% confident that the true _____ will be within 5 minutes. What sample size is needed?

Remember: $E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$. Need to solve for _____.

Determining Sample Size

Population Standard Deviation σ Known
$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$

Ex: The president of PCC Erika Endrijonas is concerned about the amount of time that her students spend on jobs. She would like to estimate the mean number of hours worked per week by these students. She knows that the standard deviation of the times spent per week on such jobs by all students is 2.5 hours. What sample size should she choose so that she can be 90% confident that the estimate is within .75 hours of the population mean?