8 Lecture 9 Notes: Confidence Intervals & Sample Size Determination

I'm not a 'Business-Man'! I'm a Business... man! Let me handle my business, damn! Sean Carter

Experiment: What percent of students respond to an email from a professor within the first day that the email was sent at UCSC?

I did this!

So, I sent an email on Saturday at 10:00 pm and counted the number of individuals who emailed me before 10:00 pm on Sunday.

What is my random variable? The number of students that responded.

What distribution does the random variable follow? _________

- __ is the number of students in the class
- __ is unknown (but we can can make inferences about it)
- Mutually Exclusive (Respond or not Responded)
- Each student is independent from each other

I found that 0.67 responded to my email. I will call this my sample proportion. A better name is called a **point estimate** (__).

Is this STATISTIC conclusive about UCSC students responding to emails within the first day?

____

This tells me that

$$\hat{p} \neq p$$

But I can still use $\hat{p}$ to make inferences about p.

How? You may ask. & How? You may wonder.
Great questions.

Remember that if $np \geq 5$ and $nq \geq 5$ we can assume the following

$$\hat{p} \sim N(\mu_{\hat{p}}, \sigma_{\hat{p}})$$

where $\mu_{\hat{p}} = np$ and $\sigma_{\hat{p}} = \sqrt{npq}$

$\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$ assume that I am doing the experiment more than once

Therefore, we are going to divided both parameters by n, because we only sampled once, which can be expressed as

$$\mu_{\hat{p}} = \frac{np}{n} = p$$

$$\sigma_{\hat{p}} = \frac{\sqrt{npq}}{n} = \quad = \sqrt{\frac{pq}{n}}$$

However, we do not have p, we have $\hat{p}$, therefore the mean and **standard error** (__) are

$$\mu_{\hat{p}} = \hat{p} \tag{26}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} \tag{27}$$

(Please note that we are still under the assumption that $n\hat{p} \geq 5$ and $n\hat{q} \geq 5$)

When we have the mean and standard error from a sample we can do some

POWERFUL THINGS

8.1 Confidence Interval for Sample Proportion

We can create an interval that we are *confident* p is within, in other words we can be confident that the true proportion is between a lower bound and an upper bound.

Our confidence will be based on a percentage, which we will call a confidence level.

Since a percentage is in between 0 and 1 we want to have a high level of confidence. Close to 1.

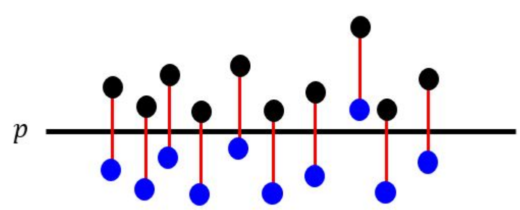
Let's say we want to be 95% confident that the true mean is in between the lower and upper bound. We call this a ______.

This means that if we conducted this experiment 100 times, 95 out of the 100 experiments will have a confidence interval in which the true proportion p will be within the lower and upper bounds. However, the other 5 of the experiments are when the interval does not contain the population proportion.

This means that 5% of the samples will not capture the true proportion within our lower and upper bound. This can happen where our confidence interval is below or above the true proportion.

Example 1

Figure 2: Black Circles refer to the Upper Bound & Blue Circles refer to the Lower Bound



1. How many samples were taken?
2. What is the proportion of intervals that contained the true proportion within its bounds?

This example shows that 90% of the intervals contained the true proportion within the bounds, whereas, 10% of the intervals do not.

The percentage that denotes the samples in which the true proportion is not within the confidence interval has a special symbol, α. Therefore, or **Confidence Level** is determined from $1 - \alpha$.

To find the confidence interval based on the data and the amount the confidence, we must do the following:

1. Define α
2. Find $\alpha/2$, because the confidence interval can be above or below the true proportion so we divide the region in which the confidence intervals DOES NOT obtain the true proportion to be above and below the estimate
3. Find the critical value ($CV = z_{\alpha/2}$) that corresponds to $\alpha/2$ which is also the z-score
4. Find Margin of Error ($ME = SE\text{x}CV$)
5. Lower Bound $LB = PE - ME$
6. Upper Bound $UB = PE + ME$
7. **Interpretation:** We are $1 - \alpha\%$ that the true proportion is within this interval. **OR.** We are $1 - \alpha\%$ that the true proportion is within the LB and UB.

(E = ME) in book

RECAP: To create a confidence interval for a proportion we must have the following:

1. CALCULATE/DETERMINE: Point Estimate (PE) $\hat{p} = \sum X_i/n$
2. CALCULATE: Standard Error: (SE) $\sqrt{(\hat{p}\hat{q})/n}$
3. STATE Confidence Level: $1 - \alpha\%$
4. DETERMINE Critical Value (CV) from $\alpha/2$
5. CALCULATE Lower Bound: $LB = PE - (CV)SE$
6. CALCULATE Upper Bound: $UB = PE + (CV)SE$
7. **Interpretation:** We are $1 - \alpha\%$ that the true proportion is within the LB and UB.

Example 2:

29. Smoking and College Education The tobacco industry closely monitors all surveys that involve smoking. One survey showed that among 785 randomly selected subjects who completed four years of college, 18.3% smoke (based on data from the American Medical Association).

a. Construct the 98% confidence interval for the true percentage of smokers among all people who completed four years of college.

b. Based on the result from part (a), does the smoking rate for those with four years of college appear to be substantially different than the 27% rate for the general population?

1. $PE =$
2. $SE =$
3. $1 - \alpha\% =$
4. $CV =$ from Z-table using $\alpha/2$
5. $LB = PE - (CV)SE =$
6. $UB = PE + (CV)SE =$
7. **Interpretation:** We are $1 - \alpha\%$ that the population proportion is within the $LB =$ and $UB =$.

8.2 Sample Size Determination (we take for granted knowing n)

We take things for granted at times, for example sample size.

If we want to find the number of students we must sample in order for our results to show a specific proportion, say 75% of students respond to emails within the first day, we can implement the following equation:

$$n = \frac{(CV^2)\hat{p}\hat{q}}{ME^2}$$

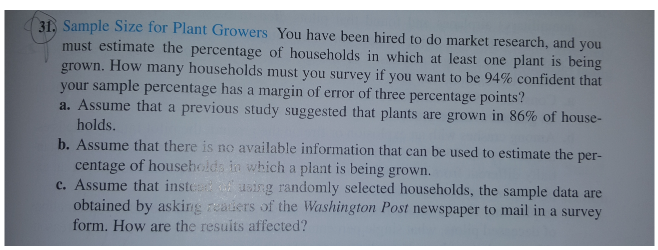
If we do not know the specific proportion, we use the following equation:

$$n = \frac{(CV^2)0.25}{ME^2}$$

RECAP: To find the sample size based on a specific proportion and margin of error:

1. DETERMINE: Point Estimate (PE)
2. DETERMINE: Margin of Error: (ME)
3. STATE Confidence Level: $1 - \alpha\%$
4. DETERMINE Critical Value (CV) from $\alpha/2$
5. CALCULATE Sample Size: $n = \frac{(CV^2)\hat{p}\hat{q}}{ME^2}$
6. NOTE: Always, I mean always round up to the nearest whole number
7. **Interpretation:** The sample size required to have a specific proportion and margin of error is n.

Example 3:



31. Sample Size for Plant Growers You have been hired to do market research, and you must estimate the percentage of households in which at least one plant is being grown. How many households must you survey if you want to be 94% confident that your sample percentage has a margin of error of three percentage points?

a. Assume that a previous study suggested that plants are grown in 86% of households.

b. Assume that there is no available information that can be used to estimate the percentage of households in which a plant is being grown.

c. Assume that instead of using randomly selected households, the sample data are obtained by asking readers of the *Washington Post* newspaper to mail in a survey form. How are the results affected?

1. $PE =$
2. $ME =$
3. $1 - \alpha\% =$
4. $CV =$ from Z-table using $\alpha/2$
5. CALCULATE Sample Size: $n = \frac{(CV^2)\hat{p}\hat{q}}{ME^2}$
6. NOTE: Always, I mean always round up to the nearest whole number
7. **Interpretation:** The sample size required to have a specific proportion and margin of error is $n =$.

8.3 Confidence Interval for Sample Mean

What if we analyze the heights of those that responded to my email within the first day.

The sample mean was $\bar{x} = 5.667$, $(0.667 = 8/12)$ and standard deviation was $s = 2.532$.

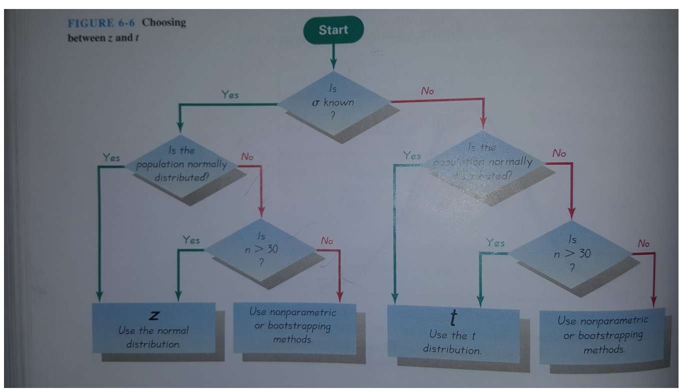
What is the random variable?

What distribution does it follow?

$$\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}})$$

where $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

Since we do not have ___ we will approximate it with ___. With this information, we can create another interval. (BIG)



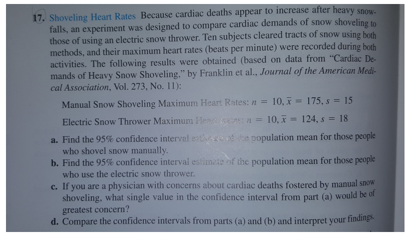
We can create an interval that we are *confident* μ is within, in other words we can be confident that the population mean is between a lower bound and an upper bound.

Skipping the ________we can conclude the following.

RECAP: To create a confidence interval for a sample mean we must have the following:

1. CALCULATE/DETERMINE: Point Estimate (PE) $\bar{x} = \sum X_i/n$
2. CALCULATE: Standard Error: (SE) $\frac{s}{\sqrt{n}}$
3. STATE Confidence Level: $1 - \alpha\%$
4. DETERMINE Critical Value (CV) from $\alpha/2$
5. CALCULATE Lower Bound: $LB = PE - (CV)SE$
6. CALCULATE Upper Bound: $UB = PE + (CV)SE$
7. **Interpretation:** We are $1 - \alpha\%$ that the population mean is within the LB and UB.

Example 4:



17. Shoveling Heart Rates Because cardiac deaths appear to increase after heavy snowfalls, an experiment was designed to compare cardiac demands of snow shoveling to those of using an electric snow thrower. Ten subjects cleared tracts of snow using both methods, and their maximum heart rates (beats per minute) were recorded during both activities. The following results were obtained (based on data from "Cardiac Demands of Heavy Snow Shoveling," by Franklin et al., *Journal of the American Medical Association*, Vol. 273, No. 11):

Manual Snow Shoveling Maximum Heart Rates: $n = 10, \bar{x} = 175, s = 15$

Electric Snow Thrower Maximum Heart Rates: $n = 10, \bar{x} = 124, s = 18$

a. Find the 95% confidence interval estimate of the population mean for those people who shovel snow manually.
b. Find the 95% confidence interval estimate of the population mean for those people who use the electric snow thrower.
c. If you are a physician with concerns about cardiac deaths fostered by manual snow shoveling, what single value in the confidence interval from part (a) would be of greatest concern?
d. Compare the confidence intervals from parts (a) and (b) and interpret your findings.

1. $PE =$
2. $SE =$
3. $1 - \alpha\% =$
4. $CV =$ from Z-table using $\alpha/2$
5. $LB = PE - (CV)SE =$
6. $UB = PE + (CV)SE =$
7. **Interpretation:** We are $1 - \alpha\%$ that the population proportion is within the $LB =$ and $UB =$.

What if we did have σ

Thursday: Test Hypothesis