My Notes

Ch 7 - Estimates & Sample Sizes

- Develop estimates of population parameters
- Use sample results to estimate the population
- See how estimates are found
- Test some claims (hypothesis) about a population
- Determine sample sizes necessary to estimate a parameter

7.2 Estimating a Population Proportion

$p = population\ proportion$

$\hat{p} = \frac{x}{n} = sample\ proportion\ of\ x\ successes\ in\ a\ sample\ size\ n$

$\hat{q} = 1 - \hat{p}$

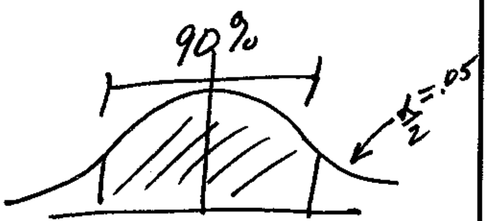
Def: Point Estimate = $\hat{p} = \frac{x}{n}$ = single value (pt) used to approx a population parameter

$\hat{p}$ is the best point estimate for the population proportion p

Confidence Interval is just an interval or range of values that gives an estimate for the population parameter we are discussing.

<u>Confidence Level</u> is the probability $(1-\alpha)$ that the proportion of times that the confidence interval actually does contain the population parameter.

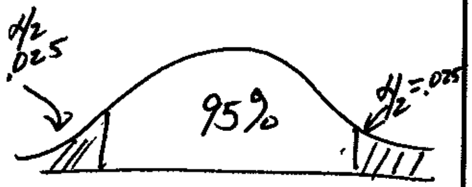
90% $\alpha = .10$



$-z_{\alpha/2}$ $z_{\alpha/2}$

-1.645 1.645

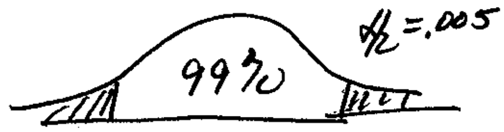
95% $\alpha = .05$



$-z_{\alpha/2}$ $z_{\alpha/2}$

-1.96 1.96

99% $\alpha = .01$



$-z_{\alpha/2}$ $z_{\alpha/2}$

-2.575 2.575

Margin of Error (E) = maximum likely difference between the observed sample proportion $\hat{p}$ & the true p

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Thus Confidence Interval (CI) =

$$\hat{p} - E < p < \hat{p} + E$$

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

$$(\hat{p} - E, \hat{p} + E)$$

Look at M&M problem handout

handout - next page

MTH 115 Name ____________________

Finding a Confidence Interval for the Proportion of a certain color of M&M.

Results from mms.com with respect to their claimed proportions of colors.

	Regular m&ms	Peanut m&ms
This is brown ⟶	13%	12%
This is yellow ⟶	14%	15%
This is red ⟶	13%	12%
This is blue ⟶	24%	23%
This is orange ⟶	20%	23%
This is green ⟶	16%	5%

Mars Company (the maker of m&m's) claims that each bag of m&m's contains the given proportion of colors. (http://us.mms.com/us/about/products/milkchocolate/) We are going to test their claim to see if it is indeed correct.

What is your color? _______ What is m&m's claim regarding the proportion of that color? _________

Count the number of m&m's in your bag that are your color.(x) _________

Count the total number of m&m's in your bag. (n) _________

What is the proportion of m&m's in your bag that are your color? ($\hat{p}$) _______

Find the 95% confidence interval on the proportion of the color you chose.

n = _______ $\hat{p}$ = _______ $Z_{\alpha/2}$ = ________

x = _______ $\hat{q}$ = _______ $E = Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} =$

95% CI for population proportion is: $(\hat{p}-E, \hat{p}+E) =$

$= \boxed{}$

Find the 99% confidence interval on the proportion of the color you chose.

n = _______ $\hat{p}$ = _______ $Z_{\alpha/2}$ = ________

x = _______ $\hat{q}$ = _______ $E = Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} =$

99% CI for population proportion is: $(\hat{p}-E, \hat{p}+E) = \boxed{}$

Sample Size for estimating p

$$n = \frac{\left[z_{\alpha/2}\right]^2 \hat{p}\hat{q}}{E^2} \quad or \quad n = \frac{\left[z_{\alpha/2}\right]^2 0.25}{E^2}$$

Ex: 500 students survey and 375 of them have twitter accounts.

a) Find the point estimate for the true population proportion of all students that have twitter accounts.

$$\hat{p} = \frac{x}{n} = \frac{375}{500} = 0.75$$

b)Find 99% CI for p

$$(\hat{p} - E, \ \hat{p} + E)$$

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 2.575\sqrt{\frac{(0.75)(.25)}{500}}$$

$$= 0.04986$$

$$\approx 0.05$$

$$(0.75 - .05, \ 0.75 + .05)$$

$$(.70, \ .80)$$

$$\boxed{(70\%, \ 80\%)}$$ ← 99% conf. that the true pop prop of all college students that have a twitter acct is between 70% + 80%

c) We want to estimate, with a margin of error of 3%, the true proportion of all college students who have twitter accounts and we want to be 95% confident in our results. How many students should we survey?

$E = 3\% = .03$

95% conf.

* Use the sample data above

$$n = \frac{(1.96)^2(.75)(.25)}{(.03)^2} = 800.33 \Rightarrow \boxed{801}$$

* Use no prior knowledge or sample data

$$n = \frac{1.96^2(.25)}{(.03)^2} = 1067.11 \Rightarrow \boxed{1068}$$

7.3

Estimating a Population Mean - ~~7.3 for Ed 4, 7.4 for Ed. 5~~

Point Estimate for mean = $\bar{x}$

Confidence Interval $\bar{x} - E < \mu < \bar{x} + E$ where $E = t_{\alpha/2} \frac{s}{\sqrt{n}}$

t distribution
student t (Gosset)
Table A-3

Requirements -
1) normal dist
OR
2) $n > 30$

Ex: A common claim is that garlic lowers cholestrerol levels. In a test of the effectiveness of garlic, 49 subjects were treated with doses of raw garlic and their cholesterol levels were measured before and after treatment. Their changes in their levels of LDL (gm/dl) cholesterol have a mean of 0.4 and a standard deviation of 21.0. Use the sample statistics to construct a 95% confidence interval of the mean net change in LDL after garlic treatment. Based on the confidence interval that you constructed, does the garlic treatment appear to lower cholesterol levels?

$n = 49$ (df = 48 / 50)
$\bar{x} = 0.4$
$s = 21.$

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}} = 2.009 \frac{21}{\sqrt{49}} = 6.027$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$0.4 - 6.03 \qquad 0.4 + 6.03$$

$$-5.63 < \mu < 6.43$$

- we are 95% conf this interval contains true pop mean of change

Notice that this interval contains 0. Thus, it is very possible there is NO change in cholesterol levels
∴ does not appear the garlic treament is lowering cholesterol levels

6.84 lbs

1.53 lbs

Ex: A random sample of birth weights of 186 babies has a mean of 3103 g and a standard deviation of 696 g (based on data from "Cognitive Outcomes of Preschool Children with Prenatal Cocaine Exposure," by Singer et al., *Journal of the American Medical Association*, Vol. 291, No 20). These babies were born to mothers who did not use cocaine during their pregnancies.

a. What is the best point estimate of the mean weight of babies born to mother who did not use cocaine during their pregnancies?

b. Construct a 95% confidence interval estimate of the mean birth weight for all such babies.

c. Compare the confidence interval from part (b) to this confidence interval obtained from birth weights of babies born to mother who used cocaine during pregancy: 2608 g < μ < 2792 g. Does cocaine use appear to affect the birth weight of a baby?

$n = 186$
$\bar{x} = 3103$
$s = 696$

a) pt. est of $\mu = \bar{x} = 3103 g$.

b) $\bar{x} - E < \mu < \bar{x} + E$

$E = t_{\alpha/2} \frac{s}{\sqrt{n}} = 1.972 \frac{694}{\sqrt{186}}$
$= 100.64$

$3103 - 100.64 < \mu < 3103 + 100.64$

$\boxed{3002.36 g < \mu < 3203.64 g}$

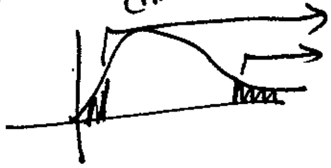
Note: Solve for n Ex

c) Since $2608 < \mu < 2792$ is completely lower than CI in b) then

Yes Cocaine does affect birth weight.

7.4 **Estimating a Population Variance or St. Deviation**

χ^2 = Chi-Squared Dist. Table A-4 (Area Right)



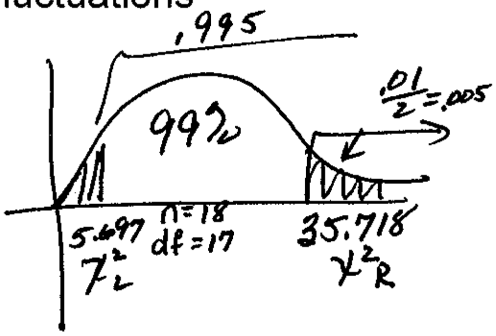
Confidence Interval for Variance: $\frac{(n-1)s^2}{X_R^2} < \sigma^2 < \frac{(n-1)s^2}{X_L^2}$

Confidence Interval for St. Deviation: $\sqrt{\frac{(n-1)s^2}{X_R^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{X_L^2}}$

Ex: A container of car antifreeze is suppose to hold 3785 ml of liquid. The Control Manager wants to be sure that the standard deviation is less than 30 ml; otherwise some of the containers would overflow or the consumer may feel cheated if the container did not contain at least 3785 ml. He selects a sample of 18 containers and finds the sample mean to be 3787 ml with a standard deviation of 55.4 ml.

a. Construct a 99% confidence interval for the true variance.

b. Does the confidence interval suggest that the fluctuations are at an acceptable level?



Sample
$n = 18$
$\bar{x} = 3787$
$s = 55.4$

a) 99% CI variance

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

$$\frac{17(55.4)^2}{35.718} < \sigma^2 < \frac{17(55.4)^2}{5.697}$$

$$\boxed{1460.8 < \sigma^2 < 9158.5}$$ ← CI for variance

b) $\sqrt{1460.8} < \sigma \text{(st.dev)} < \sqrt{9158.5}$

$38.22 < \sigma < 95.70$ ← CI for st.dev

want st dev < 30 ml —
Not Acceptable since CI is much larger.

Ex: (from Triola 7th ed.

The Husdon Valley Bakery makes doughnuts that are packaged in boxes with labels stating that there are 12 doughnuts weighing a total of 42 oz. If the variation among the doughnuts is too large, some boxes will underweight (cheating consumers) and others will be overweight (lowering profit). A consumer would not be happy with a doughnut so large that it resembles a tractor tire (or would they?) The quality control supervisor has found that he can stay out of trouble if the doughnuts have a mean of 3.5 oz. and a standard deviation of 0.06 oz or less. Twelve doughnuts are randomly selected from the production line and weighed, with the results given below (in ounces). Construct a 95% confidence interval for μ and a 95% confidence interval for σ. Then determine whether the quality control supervisor is in trouble.

3.43	3.37	3.58	3.50	3.68	3.61
3.42	3.52	3.66	3.50	3.36	3.42

Sample

$n = 12$ $df = 11$

$\bar{x} = 3.504$

$s = 0.109$

use Excel or Calculator

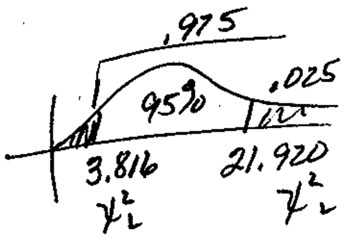
a) $\bar{x} - E < \mu < \bar{x} + E$

$E = t_{\alpha/2} \frac{s}{\sqrt{n}} = 2.201 \left(\frac{0.109}{\sqrt{12}}\right)$

$= .069$

$3.504 - .069 < \mu < 3.504 + .069$

$\boxed{3.435_{oz} < \mu < 3.573\ oz}$ mean CI



b) $\sqrt{\frac{(n-1)s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_L}}$

$\sqrt{\frac{11(0.109)^2}{21.920}} < \sigma < \sqrt{\frac{11(0.109)^2}{3.816}}$

$\sqrt{0.00596} < \sigma < \sqrt{0.034}$

$\boxed{0.077_{oz} < \sigma < 0.185_{oz}}$ CI st.dev

8

c) mean OK
too high ⇒ manager in trouble