Significance and confidence interval practice problems
STAT200

Means

Alfalfa problem:

Give a 90% confidence interval for the mean cellulose content in the population.

$$\bar{x} \pm t_{14}^{*}\left(\frac{s_x}{\sqrt{n}}\right)$$

$$145 \pm 1.761\left(\frac{8}{\sqrt{15}}\right)$$

We are 90% confident that the true mean cellulose content is in the interval (141.36,148.64)

a. A previous study claimed that the mean cellulose content was $\mu = 140$ mg/g, but the agronomist believes that the mean is higher than that figure. State H_o and H_a and carry out the significance test to see if the data support this.
   μ: true mean cellulose content of the alfalfa
   H_0: $\mu=140$: the true mean alfalfa content is 140
   H_A: $\mu>140$: the true mean alfalfa content is greater than 140

n is not large enough so we must assume population is normal
we must also assume that the sample was a random to use the CLT

$\alpha = .05$

$$T = \frac{x - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{145 - 140}{\frac{8}{\sqrt{15}}}$$

T = 2.42: p-value = 0.030

Since p-value < α we reject H_0, i.e. the true mean alfalfa content is greater than 140.

c. The statistical procedures used in (a) and (b) are valid when some assumptions are made – what are these assumptions? ***Assumptions are mentioned in b*.**

<u>Income Problem</u>

a. Would a 95% confidence interval be wider or narrower? Be able to explain this. ***95% would be narrower (the Z-critical is 1.96 for 95% and 2.58 for 99% all other parts of margin of error calculation are the same.) intuitively: We are less sure so the area would be smaller.***
b. Would the null hypothesis that the 1994 mean family income was \$33,000 be rejected at the 1% significance level in favor of the alternative hypothesis that the mean income is not \$33,000?
   $397 = \text{ME} = t_{60000}^{*}\left(\frac{s_x}{\sqrt{n}}\right) = 1.96(SE_{\bar{x}}): SE_{\bar{x}} = \frac{397}{1.96} = 202$
   $T = \frac{32264-33000}{202} = 3.64$: p-value < .01 for 1000 df therefore we would reject

---

Significance and confidence interval practice problems
STAT200

NHL Problem

155	136	131	138	140
139	134	146	132	149

a. Estimate the true average length of a hockey using a confidence of 95%.

$$\bar{x} = 140$$
$$s_x = 7.77$$
$$SE_{\bar{x}} = \left(\frac{s_x}{\sqrt{n}}\right) = \frac{7.77}{\sqrt{10}} = 2.46$$
$$\bar{x} \pm t_9^*\left(\frac{s_x}{\sqrt{n}}\right) = 140 \pm 2.262(2.46) = (134.44, 145.56)$$

We are 95% confident that the true mean hockey game in 2004 is in the interval (134.44, 145.56)

b. the average length of a game in 2002 was 155 minutes. Based on your answer in a., is there reason to believe the average length of an NHL game has truly shortened from last year? (explain)

Since 155 is not in the interval this value is more than likely not the true mean.

c. Do a complete appropriate Hypothesis test to test the claim that the lengths of NHL games have shortened this year compared to 2002 as a result of the rule changes. *(SHOW ALL RELEVANT WORK!)*

μ: true mean NHL Game after rule changes in 2004
H_0: μ=155: the true mean NHL Game after rule changes in 2004 is 155
H_A: μ<155: the true mean NHL Game after rule changes in 2004 is less than 155

n is not large enough , looking at a dotplot of the data there is no reason to believe the data is not normal therefore n may be less than 30.
Data was chosen from randomly chosen games of that season, CLT Holds.
$\alpha = .05$

$$T = \frac{x - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{140 - 155}{\frac{7.77}{\sqrt{10}}}$$

T = -6.10: p-value = 0.00 (essentially)

Since p-value < α we reject H_0, i.e. the true mean NHL game length in 2004 is less than 140.

Significance and confidence interval practice problems
STAT200

Proportions

Church Synagogue problem

a. Give a 99% confidence interval for the proportion of adults who attended church or synagogue during the week preceding the poll.

$\hat{p} = \frac{750}{1785}$ =.42; .42*1785 = 750>10; (1-.42)*1785 = 1035 >10

Trusting Gallup the poll should be random from this particular week.

$$\hat{p} \pm Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .42 \pm 2.58 \sqrt{\frac{.42(1-.42)}{1785}} = (.39, .45)$$

We are 99% confident that true proportion of people going to church this particular week is in the interval (.39, .45)

b. Do the results provide good evidence that less than half of the population attended church or synagogue?

Yes (Corrected 12/12/08) .5 is not in the interval.

(I mistakenly said no previously)

c. How large a sample would be required to obtain a margin of error of .01 in a 99% confidence interval for the proportion who attend church or synagogue? (use a conservative guess of pi = .5)

$$ME = 2.58 \sqrt{\frac{.5(1-.5)}{n}} = .01$$

$$\frac{.01}{2.58} = \sqrt{\frac{.25}{n}}; \left(\frac{.01}{2.58}\right)^2 = \frac{.25}{n}; n = \frac{.25}{\left(\frac{.01}{2.58}\right)^2} = \frac{.25}{(.003876)^2} = \frac{.25}{.000015} = 16{,}641$$

d. Can we use this data for this one week and extrapolate and say that this gives us a reasonable estimate of what proportion of people go to church or synagogue every week. Explain why or why not.

No we do the nature of religious attendance it is very dependent on the week, and whether there is a holiday close or during a vacation period so no this week can not be representative a typical week.

Significance and confidence interval practice problems
STAT200

Cafeteria Problem

a. State the null and alternative hypotheses based on the above information

b. Perform a test of significance on the three samples and the combined sample of all three sets of data above at an $\alpha = .1$ level. Indicate whether or not each of the surveyors reject their null hypothesis.

x = number of people in survey that responded that they enjoy college food
n = number of people questioned on whether they enjoy college food
p-hat = proportion in study that enjoy college food.
π=the true proportion that enjoy college food
H_0: π=1/3: one third of college students enjoy college food
H_A : $\pi \neq 1/3$: the proportion of college students that enjoy college food is not 1/3.

Subscript T indicates Tonya, F indicates Frank, S indicates Sarah, and C indicates Combined

$\theta_o = 1/3$

$x_T = 18; n_T = 50; \hat{p}_T = \frac{18}{50} = .36;$ using 1PropZTest p-value = $.689; Z_T = .4$

$x_F = 98; n_F = 350; \hat{p}_F = \frac{98}{350} = .28;$ using 1PropZTest p-value = $.034; Z_F = -2.117$

$x_s = 140; n_S = 500; \hat{p}_S = \frac{140}{500} = .28;$ using 1PropZTest p-value = $.0114; Z_F = -2.53$

$x_C = 256; n_F = 900; \hat{p}_F = \frac{256}{900} = .28444;$ using 1PropZTest p-value = $.0019; Z_F = -3.111$

For Tonya I meant to put 14 (not 18) which would have changed the problem as follows:

$x_T = 14; n_T = 50; \hat{p}_T = \frac{14}{50} = .28;$ using 1PropZTest p-value = $.424; Z_T = -.8$

$x_F = 98; n_F = 350; \hat{p}_F = \frac{98}{350} = .28;$ using 1PropZTest p-value = $.034; Z_F = -2.117$

$x_s = 140; n_S = 500; \hat{p}_S = \frac{140}{500} = .28;$ using 1PropZTest p-value = $.0114; Z_F = -2.53$

$x_C = 252; n_F = 900; \hat{p}_F = \frac{252}{900} = .28;$ using 1PropZTest p-value = $.0007; Z_F = -3.394$

c. Make sure you can explain why a similar p-hat provided essentially 3 different results for the assuredness of the results.

As n increases the normal curve indicated by the CLT gets narrower, therefore the same p-hat will be more standard deviations from the mean proportion making it more significant.

Significance and confidence interval practice problems
STAT200

Teacher Problem

π=true proportion of students that believe team or group work is important.

H_0:π=.5: true proportion of students that believe team or group work is important is .5

H_A:π>.5: true proportion of students that believe team or group work is important is greater than .5

.5 * 735 =376.5 >10 and (1-.5) * 735 =376.5 >10 (notice π_0 is used for hyp test)

The sample is random from these nine schools. So must be careful applying to all students.

May Apply CLT for 9 schools, with caution for all students.

α=.05; Z* = 1.96

$$Z = \frac{\hat{p} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{\frac{423}{735} - .5}{\sqrt{\frac{.5(1-.5)}{735}}} = \frac{.576 - .5}{.01844} = 4.222$$

p-value = 0 essentially

Since p-value < α (or alternatively Z > Z*) we will reject the null hypothesis, i.e. The true proportion of people who think that group or teamwork is important is greater than .5