Homework 4, due in Lab Thursday October 6 ---- **Solutions**

(assignment for these solutions can be found on the last page)

Note: There are (more than) three confidence intervals for a population proportion. In problems 4, 5, and 6 I have solved them by hand using the ptilde version. I give the intervals from Minitab for each of the three methods. If you used a different method by hand, no problem; it will match one of the three I have done in Minitab, and I will know which one by the work you do.

1. Problem 6.51 p. 215 in SW. Get the summary statistics in Minitab, construct the CI by hand, then confirm you get the same interval in Minitab.
   a. Summary statistics (note that SE Mean = StDev/sqrt(N)):

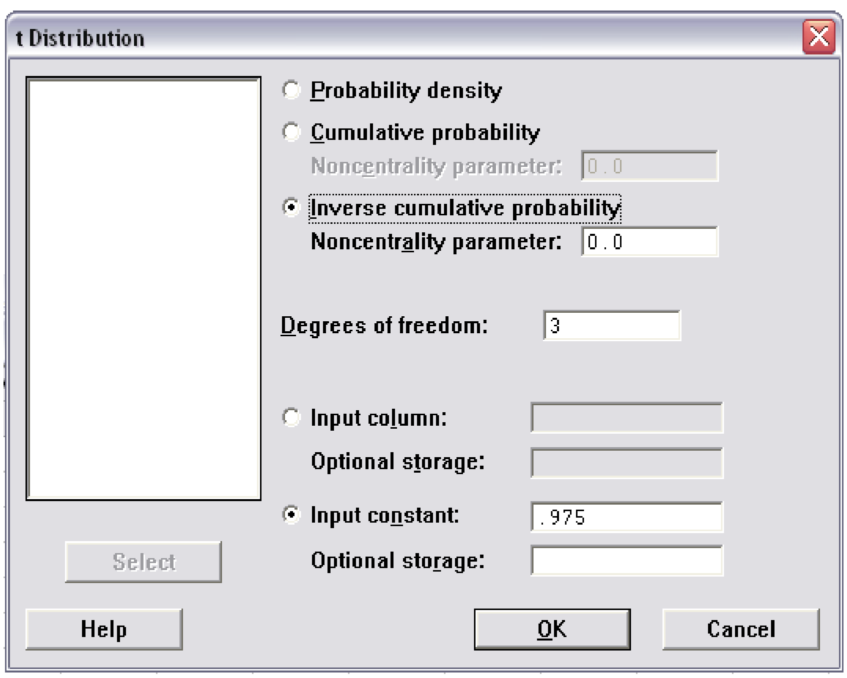
```
Variable  N   Mean  SE Mean  StDev
C1        4  51.00     1.60   3.19
```

   b. The confidence interval has this form:

95% CI: 51.00 +- t(.025,df=4-1)*1.60

To find the value of t(.025,df=4-1), we calculate the inverse cumulative probability with 3 degrees of freedom and a right-tail probability of .975, as shown below. (If you use a right tail probability of .025, you will just get –t, which is fine, just remember to drop the negative.)

Below that is found to be t = 3.18245. Returning to our interval,
95% CI: 51.00 +- 3.18*1.60 or **(45.9164, 56.0836).**

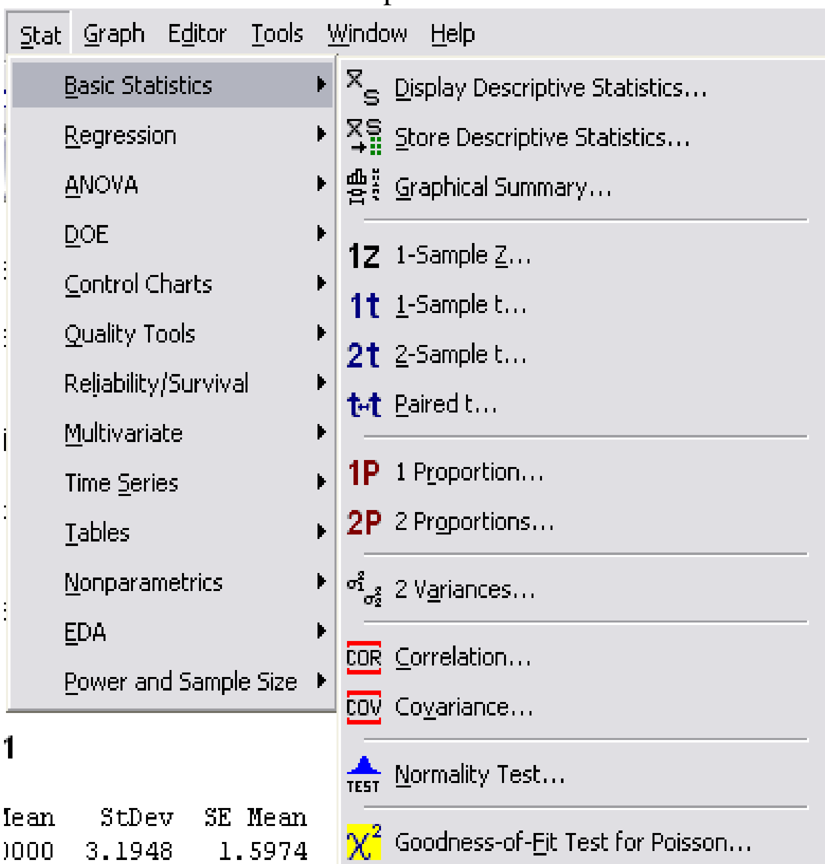


Inverse Cumulative Distribution Function

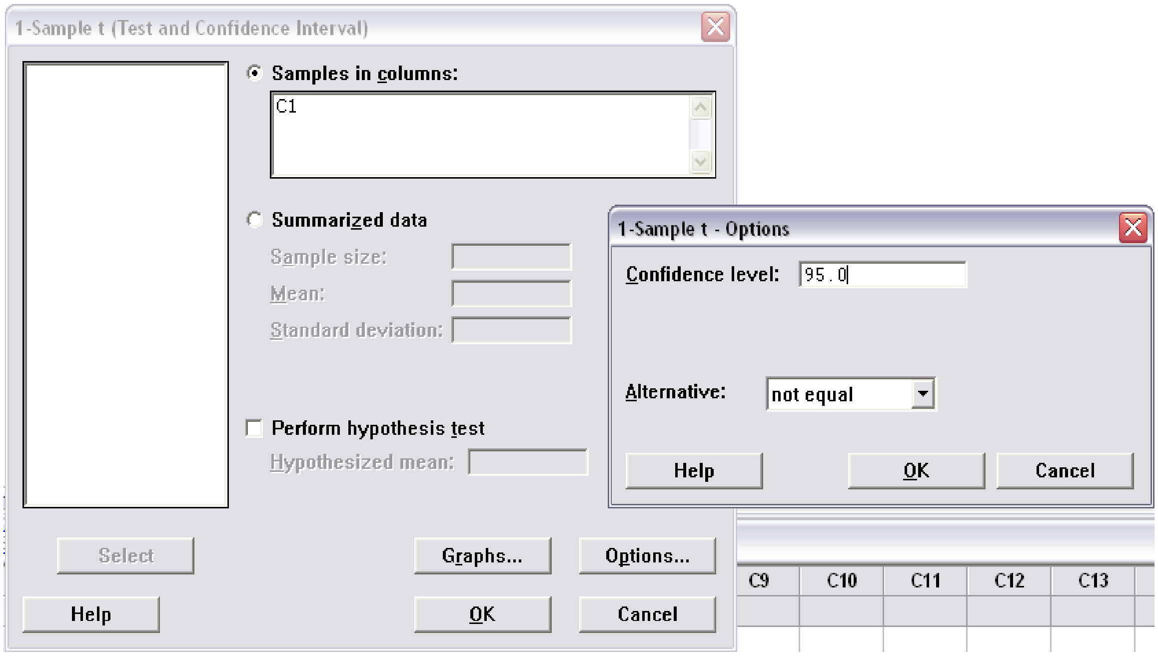
```
Student's t distribution with 3 DF
```

```
P( X <= x )          x
      0.975   3.18245
```

Minitab can calculate the Confidence interval for the mean by choosing test from the Stat/Basic Statistics/1t 1-Sample t test from the menu:



Next, either input the column name with the data (C1 below), or input the Summarized data (sample size n, mean, and standard deviation). Under the options button, input the confidence level you want (in percentage, eg. 95.0 for a 95% CI) with Alternative not equal. Below is the result, note that it is the same as above.



One-Sample T: C1

```
Variable  N     Mean   StDev  SE Mean        95% CI
C1        4  51.0000  3.1948   1.5974  (45.9164, 56.0836)
```

c. A 99% interval would be wider since the t-value would be larger to give a larger probability in the interval. The interval is here, and note it is wider.

One-Sample T: C1

```
Variable  N     Mean   StDev  SE Mean        99% CI
C1        4  51.0000  3.1948   1.5974  (41.6698, 60.3302)
```

2. 6.55 in SW p. 215. Do by hand.
   a. A 95% CI is given by

95% CI: 4.30 +- t(.025,df=20-1)*2.03/sqrt(20)
The value of t is 2.09302, giving
(3.34993, 5.25007)

   b. Mu represents the average number of puffs observed after the experiment. This confidence interval is an interval for the population (or true value of the) mean number of puffs for fruitfly larva under the same conditions.

3. 6.58 in SW p. 216. Do by hand.
   a. A 95% CI is given by

95% CI: 5.1679 +- t(.025,df=28-1)*0.6544/sqrt(28)
The value of t is 2.05183, giving
(4.91415, 5.42165)

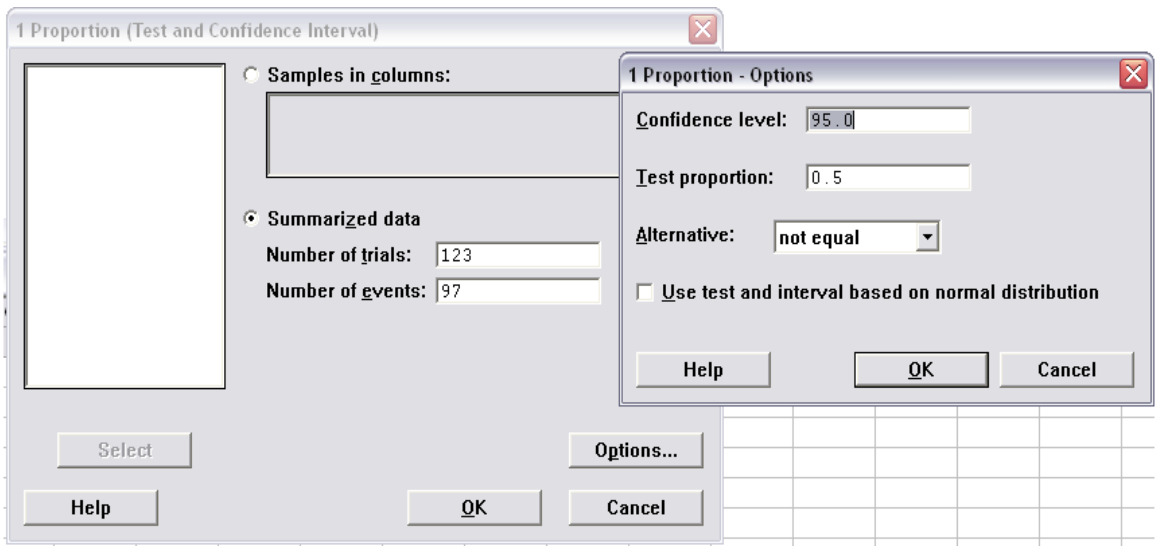
   b. A 99% CI is given by

99% CI: 5.1679 +- t(.005,df=28-1)*0.6544/sqrt(28)
The value of t is 2.77068, giving

(4.82525, 5.51055)

c. We are 95% (or 99%) confident that intervals constructed in this way will contain the true population mean.

4. 6.65 in SW p. 217. Do by hand, then confirm you get the same interval in Minitab.
   a. Using the confidence interval for p on page 208, we get ptilde = (97+2)/(123+4) = 0.7795. The standard error as defined on page 207 is then 0.0368, and 0.7795 +- 1.96*0.0368 gives **(0.7074, 0.8516)**, which is slightly wider than the interval from Minitab below.



Test and CI for One Proportion

```
Test of p = 0.5 vs p not = 0.5

                                                   Exact
Sample    X    N  Sample p        95% CI         P-Value
1        97  123  0.788618  (0.705818, 0.857030)   0.000
```

NOTE: If you select the Options button and select “Use test and interval based on normal distribution”, your interval will match the method given in the footnote of page 208, and in this case is **(0.716463, 0.860772)**.

Furthermore, if you add 4 to the number of trial, and 2 to the number of events, and choose the interval based on the normal distribution, the interval is the same as the one we calculated by hand.

```
Sample    X    N  Sample p        95% CI         Z-Value  P-Value
1        99  127  0.779528  (0.707427, 0.851628)    6.30    0.000
```

5. 6.66 in SW p. 217.
   These deer are unlikely to have been a random sample from the population. Instead, the convenience sampling was used, and perhaps it is easier to capture a pregnant deer, which would bias the proportion.

6. 6.67 in SW p. 217. Do by hand, confirm your answer with Minitab.
    a. Using the confidence interval for p on page 208, we get ptilde = (23+2)/(180+4) = 0.1359. The standard error as defined on page 207 is then 0.0253, and 0.1359 +- 1.645*0.0253 gives **(0.0868, 0.1687)**. Minitab reports the exact interval **(0.082754, 0.185538)**, or the interval based on the normal distribution **(0.086849, 0.168707)**.
    b. From page 214, the data must be a random sample from a large population, with the observation independent.
    c. 90% of confidence intervals constructed in this way will contain the true population proportion of defects in the gene coding.

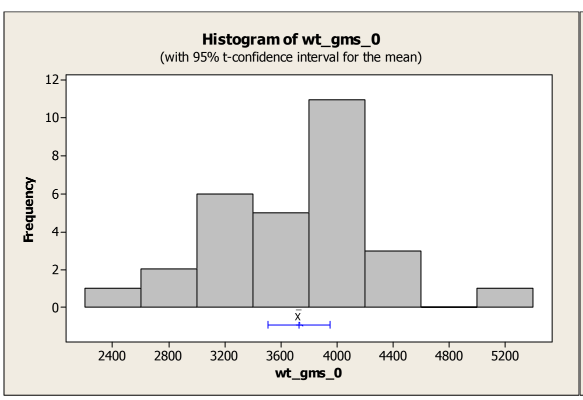
7. Use the data from problem 1 in Homework 3. Use Minitab to get 95% CI's for mean birth weights of males and separately for mean birth weights of females. Given your answer to problem 1 in HW 3, do you see difficulties with assumptions? Interpret your intervals - what do they mean?

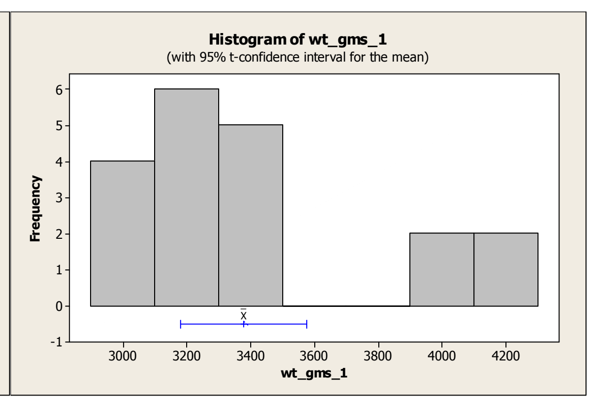
Below are the confidence intervals for the population mean of boys and girls, and histograms of the sample data. Recall from HW3 that the boys distribution was roughly normal, but the girl's distribution was clearly not normal. There is not a problem with the assumptions with the boy's confidence interval because of normality, but because the girl's data are not normal and the sample size is small (19) the assumptions are not met to be able to use this confidence interval.

The interval interpretation is that 95% of confidence intervals constructed in this way will contain the true population mean of boy's birth weight from women having their first child. Similar for the girl's confidence interval.

One-Sample T: wt_gms_0, wt_gms_1

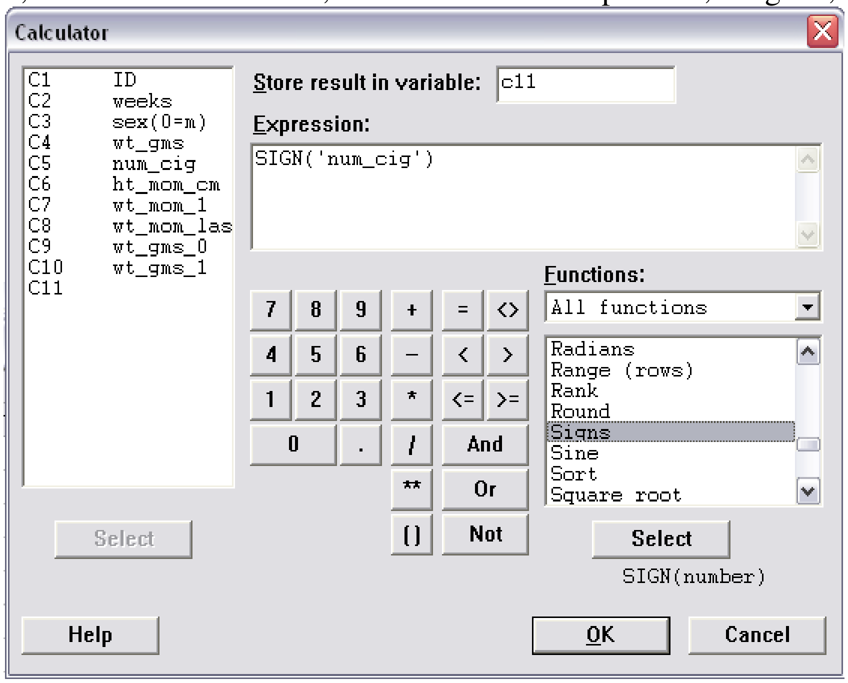
```
Variable   N     Mean   StDev  SE Mean         95% CI
wt_gms_0  29  3728.10  573.54   106.50  (3509.94, 3946.27)
wt_gms_1  19  3379.21  411.22    94.34  (3181.01, 3577.41)
```



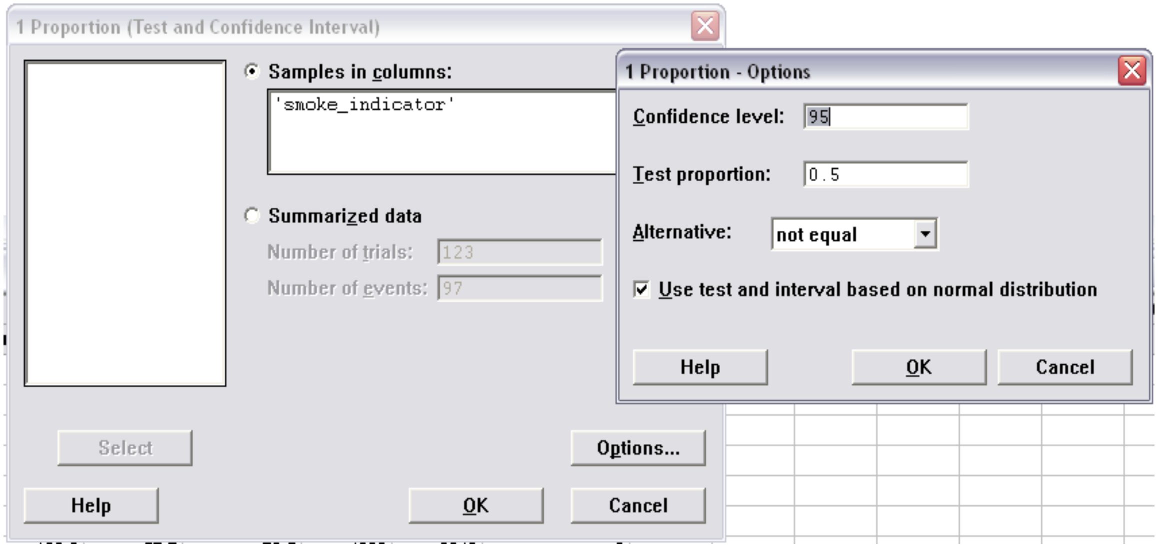


8. Use the data from problem 1 in HW 3 and use Minitab to produce a 95% CI for for the proportion of mothers who smoke during pregnancy. Use data manipulation methods to create an appropriate variable that shows if the mother smoked during pregnancy. For the CI do not use the summarized data option in Minitab. You should get this interval from "raw" data. (I want you to be able to deal with 10,000 cases where you cannot easily count by hand). Interpret the interval.

The best way to do this is to use the calculator and the SIGN function. For data that are 0, the new column will 0, but when the data is positive, we get 1, and negative we get -1.



With the new column, calculate from the Stat/Basic Statistics/1 proportion menu a 95% confidence interval from the data in the new column (which I have labeled 'smoke_indicator'). The resulting confidence interval is given below. The interpretation is that 95% of intervals constructed in this way will contain the true population proportion of mothers who smoke during their first pregancy.



Test and CI for One Proportion: smoke_indicator

Test of p = 0.5 vs p not = 0.5
Event = 1

Variable	X	N	Sample p	95% CI	Z-Value	P-Value
smoke_indicator	9	48	0.187500	**(0.077082, 0.297918)**	-4.33	0.000

```
Stat 538 - Biostatistics I

Homework 4, due in Lab Thursday October 6


1. Problem 6.51 p. 215 in SW.  Get the summary statistics in Minitab,
   construct the CI by hand, then confirm you get the same interval in
   Minitab.

2. 6.55 in SW p. 215.  Do by hand.

3. 6.58 in SW p. 216.  Do by hand.

4. 6.65 in SW p. 217.  Do by hand, then confirm you get the same
   interval in Minitab.

5. 6.66 in SW p. 217.

6. 6.67 in SW p. 217.  Do by hand, confirm your answer with Minitab.

7. Use the data from problem 1 in Homework 3.  Use Minitab to get 95%
   CI's for mean birth weights of males and separately for mean birth
   weights of females.  Given your answer to problem 1 in HW 3, do you
   see difficulties with assumptions?

   Interpret your intervals - what do they mean?

8. Use the data from problem 1 in HW 3 and use Minitab to produce a
   95% CI for for the proportion of mothers who smoke during pregnancy.
   Use data manipulation methods to create an appropriate variable that
   shows if the mother smoked during pregnancy.  For the CI do not use
   the summarized data option in Minitab.  You should get this interval
   from "raw" data. (I want you to be able to deal with 10,000 cases
   where you cannot easily count by hand).

   Interpret the interval.
```