

CHAPTER 7

Confidence Interval Estimation

In this Chapter we are using inferential statistics to estimate the parameters of a population given a random sample. This is the first Chapter that uses inferential statistics procedures and it is a nice introduction to this topic. We are going to compute confidence for proportions and means. The confidence interval estimator contains a range of values that attempt to capture the true value of a parameter. It is important to notice that the confidence level is not a probability associated with the parameter, instead it refers to the percentage of success in the estimation procedure. For example a 95% confidence interval means that: Under repeated sampling the confidence interval estimator will capture the true value of the parameter 95 times out of 100 (95% confidence level). The confidence level refers to the effectiveness of the procedure, and it should not be interpreted as a probability associated with the parameter (the parameter is a constant, not a random variable, thus, we can't associate a probability with a parameter).

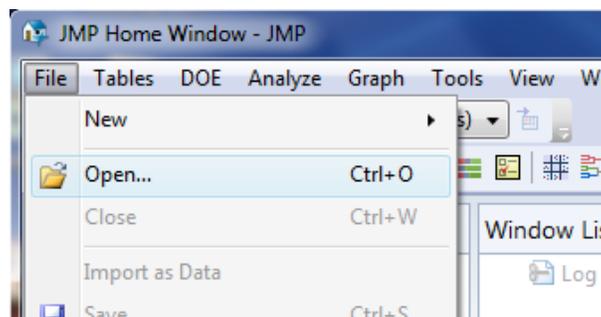
Class Exercises: Compute confidence interval estimators and verify assumptions

Class example 1:

Compute a confidence interval estimator for the proportion of males from the file "Bears"

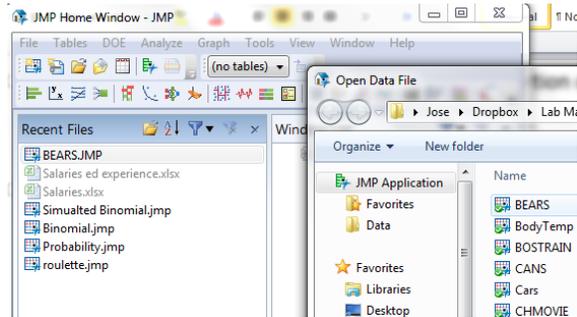
Open the file "Bears" (you can download the file at my website at sci.tamucc.edu/~jguardiola).

Figure 7.1



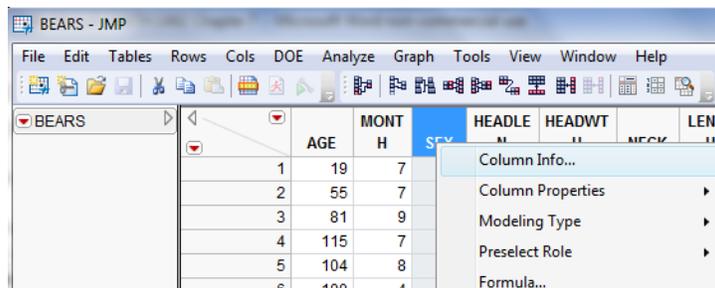
select the file "Bears" as follows:

Figure 7.2



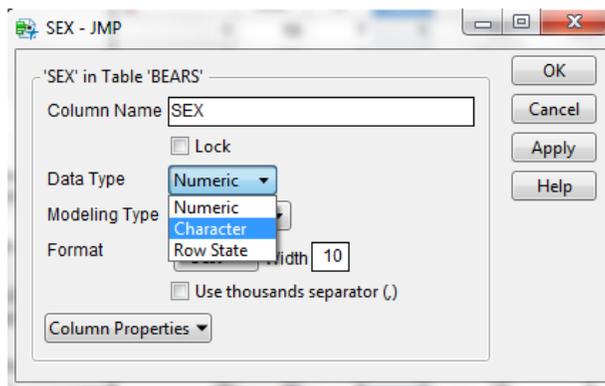
first, right click over “Sex” and select ‘Column Info...’,

Figure 7.3



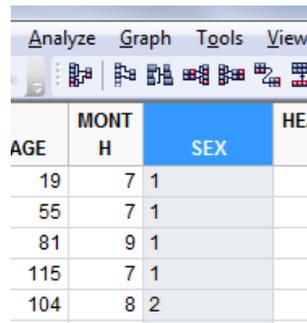
we have to change the type of variable for “Sex” as it appears as numeric variable and it should be a categorical one (notice that the variable “Sex” is a label, not a number 1=Male, 2 = Female), change the “Type of Data” from “Numeric” to “Character” as shown below (the Modeling type changes automatically to “Nominal”), and click over “OK”

Figure 7.4



you can see that the column alignment changes to the left side for column “Sex”, as it is by default for categorical (Character) variables

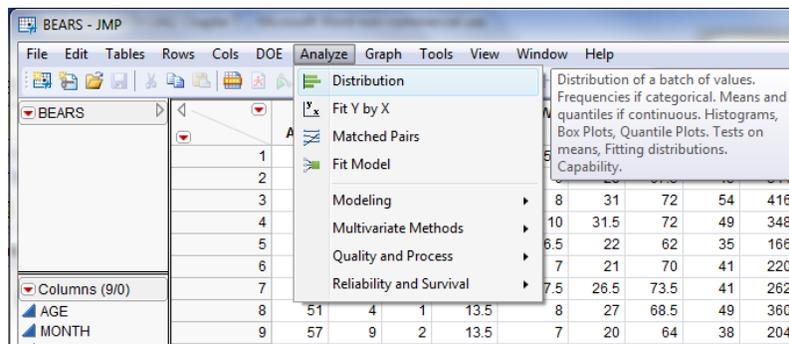
Figure 7.5



AGE	MONT H	SEX	HEADLEN
19	7	1	
55	7	1	
81	9	1	
115	7	1	
104	8	2	

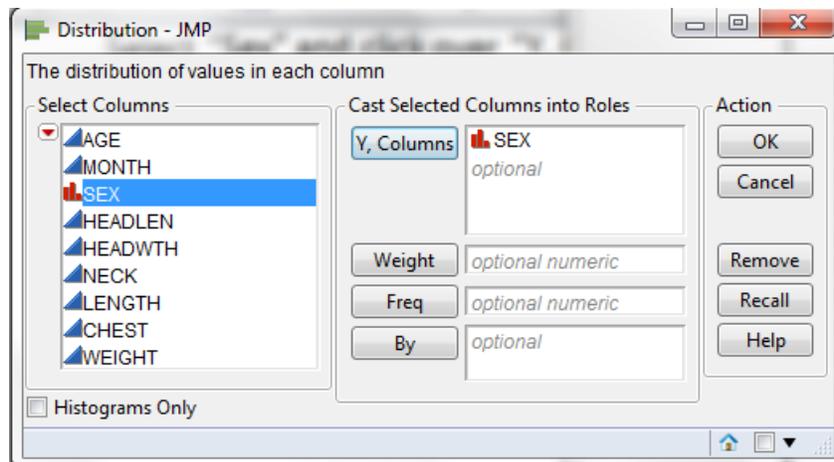
now, select “Distribution” from the “Analyze” menu

Figure 7.6



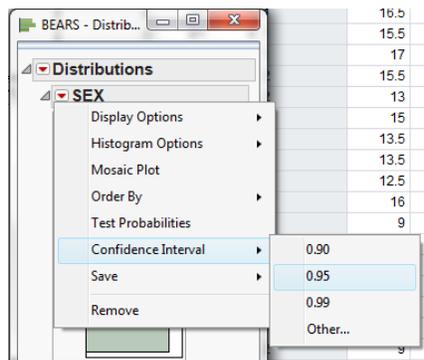
select “Sex”, click over “Y, Columns” and click over “OK”

Figure 7.7



then, at the new window, click over the second red triangle, select “Confidence Interval” and click over 0.95,

Figure 7.8



you get the following results

Figure 7.9

Confidence Intervals					
Level	Count	Prob	Lower CI	Upper CI	1-Alpha
1	35	0.64815	0.514848	0.76177	0.950
2	19	0.35185	0.23823	0.485152	0.950
Total	54				

Note: Computed using score confidence intervals.

Please notice that these confidence intervals do not match computations done by hand, using the procedure studied during lecture. The procedure used in JMP is different, but these estimators are valid and can be used for doing lab homework, (not recommended for problems assigned at MyLabsPlus). This procedure is outside the scope of this course.

The proportion for male bears is

$$0.51 < p < 0.76, 95\% \text{ CI}$$

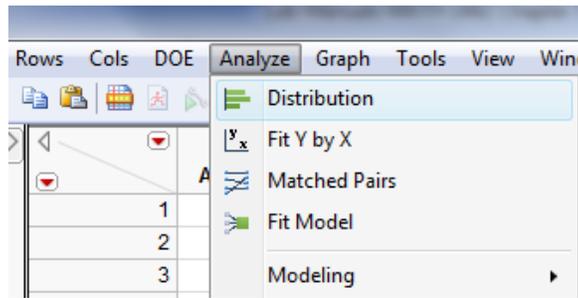
Question: From the practical (biological) point of view, do these results make sense to you?

Class Example 2:

Compute a confidence interval estimator for the mean of “weights” from the file “Bears, and separate the confidence interval for males and female bears (Coding: males =1, females = 2)

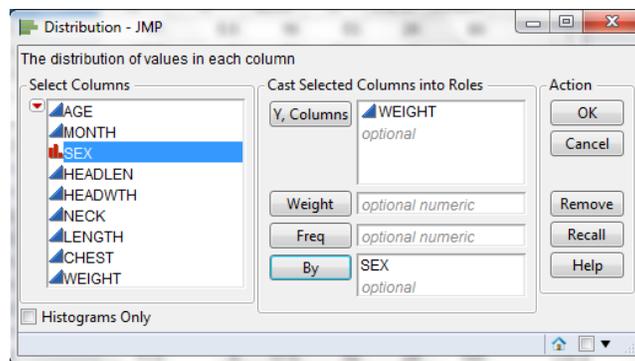
Open the file “Bears” as in Figures 7.1 and 7.2 (or ignore these steps if the file is already open), then select “Distribution” from the “Analyze” menu

Figure 7.10



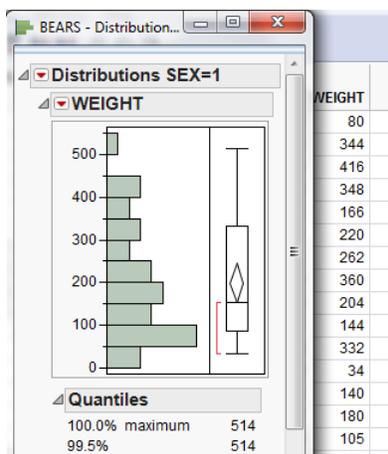
then, select “Weight” and click over “Y, Columns”, select “SEX” and click over “By”

Figure 7.11



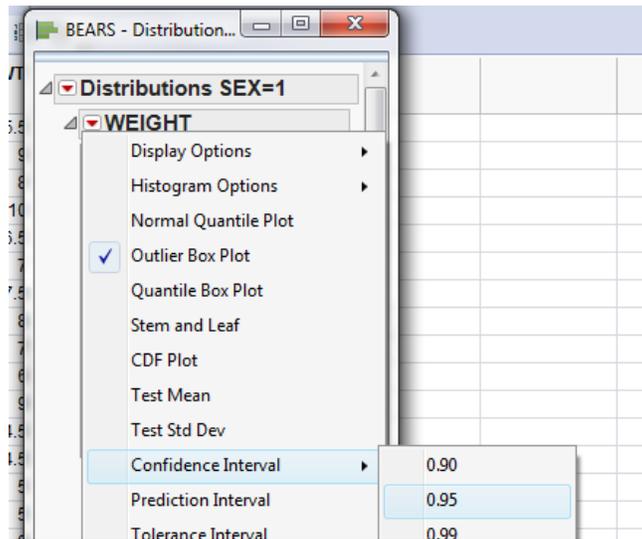
select “OK” and you can see two separate distributions for males and females, however the shapes of the distributions do not look as normally distributed, but we will discuss this issue later on

Figure 7.12



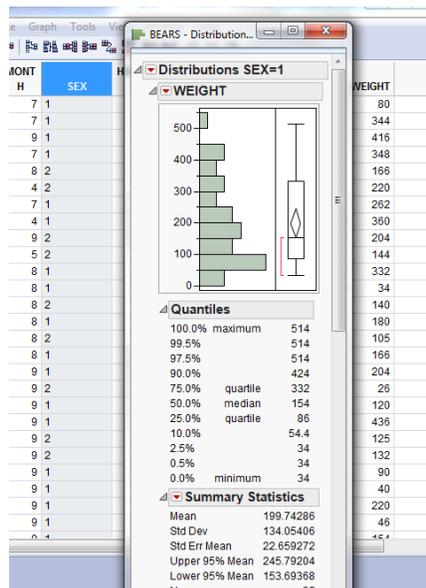
click over the second red triangle, select the option “Confidence Interval” and select “0.95”

Figure 7.13



then, you can see the confidence interval for the mean weight at the very bottom of the results window

Figure 7.13



Repeat the same procedure for female weights and the results are shown below

The confidence intervals results separated by "SEX" are:

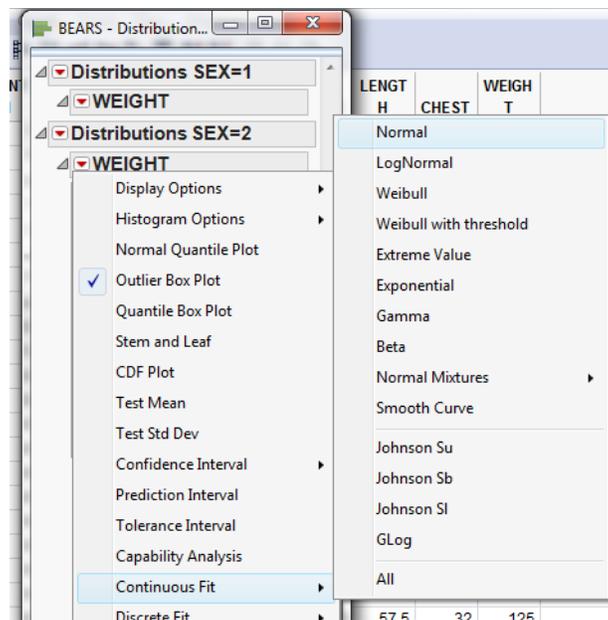
For males $153.69 < \mu < 245.79$ 95% CI

For females $108.28 < \mu < 195.40$ 95% CI

there is an overlapping between the confidence intervals for the means separated by sex, therefore, they do not look too different. Also, let's check the assumptions for constructing a confidence interval, the sample size for male bears is 35 and the sample size for female bears is 19, should we be concerned about the validity of these computations? Please discuss in class and reach a conclusion

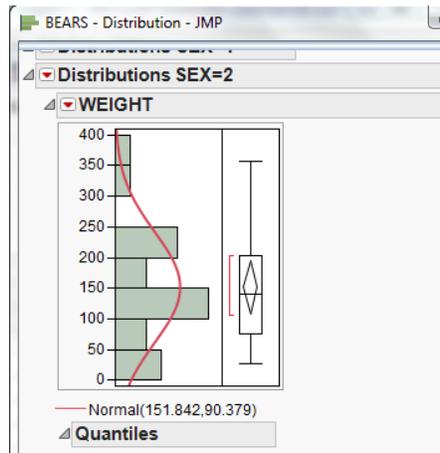
Let's check normality for female bears SEX=2, (do we need to check normality for male bears?), to do it let's select "Continuous fit" and then "Normal"

Figure 7.14



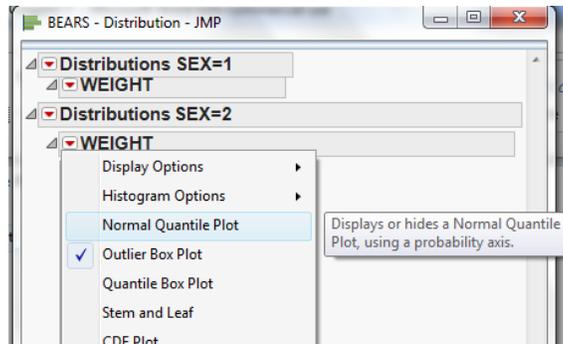
then, comparing the shape of the distribution with the overlapping normal curve it does not look very "normal"

Figure 7.15



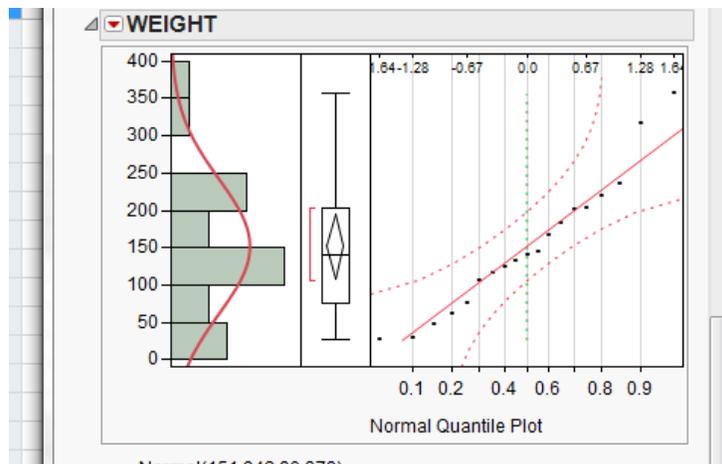
in order to be able to assess normality, we need a better tool, let's construct a Q-Q plot. First, select "Normal Quantile Plot" from the menu located at the lower right triangle as follows

Figure 7.16



the following Q-Q plot is shown

Figure 7.17



the Q-Q plot shows three outliers, but being optimistic it does not look really bad (there are some statistical tests for assessing normality but these tests are beyond the scope of this course). Looking at the Q-Q plot we can't reject the normality assumption (we are not saying that it is normally distributed, but we can't reject the assumption of normality). Conclusion: the assumption of normality is acceptable

Simulation exercise:

Go to <http://onlinestatbook.com/rvls.html> or search for "Rice Virtual Labs" and select "Simulations/Demonstrations", and click over "Confidence Intervals", and click over "begin", with the help your instructor, perform and interpret some simulations (read the instructions).

Class Exercises:

- 1- Confidence interval for proportion: Open the file "Movies.JMP" and obtain confidence intervals for the proportions for each level of the categorical variable "MPAA Rating",
- 2- Confidence intervals for the means: Open the file "Movies.JMP" and obtain confidence intervals for the means of the variables "Viewer Rating" and "Length", separated by each level of the categorical variable "MPAA Rating", discuss if you see any difference across the different MPAA Ratings. Check the assumptions for constructing these confidence intervals and discuss if they are valid.

Team Assignment: Computing confidence intervals

Use the sample that you obtained from the file "Small Town.xls" (do not use the whole dataset, if for some reason you do not have the sample with you, obtain a new sample, and save it for the next activities as described on Chapter 1) and do the following:

- 1- Obtain confidence intervals for the proportion for each level of the categorical variables SEX, RACE, ED LEVEL, and MARITAL STATUS, SMOKE separated by SEX
- 2- Obtain confidence intervals for the mean of the following continuous variables:
 - a. HEIGHT, WEIGHT and SMOKE
 - b. BMI separated by RACE
 - c. SALARY separated by SEX
 - d. SALARY separated by EDUCATION LEVEL
 - e. Verify assumptions for all previous confidence intervals, comment if the confidence intervals are valid or not.
- 3- Summarize all your confidence intervals in a table, showing each variable and its corresponding confidence interval in a column. Add a new column indicating if the confidence interval is valid or not after checking the assumptions.
- 4- Finally, comment if you find any interesting difference across groups (for example, compare the confidence intervals for separated salaries for men and women). Hypothesize if there is any difference across the different levels of the categorical variables for every continuous variable computed above in parts 2 a. to 2 d. Write your conclusions and submit a report to your lab instructor.