CHAPTER 3

Summary Statistics

It is possible to compute many summary statistics for continuous variables using the appropriate menu options in JMP. We can also generate contingency tables by counting the number of observations that match different levels and combinations of categorical variables.

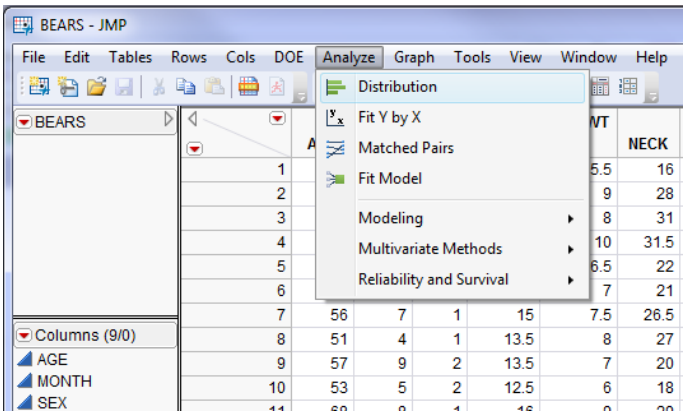
Class Exercise: Practice computing some summary statistics from the file BEARS.

First, open the JMP file BEARS located on the website for this course at sci.tamucc.edu/~jguardiola and click over your corresponding course, then select "Online Datasets" near the top of the page.

We are going to compute some summary statistics using a numeric-continuous variable, for example let's use the variable "Weight" for that purpose.

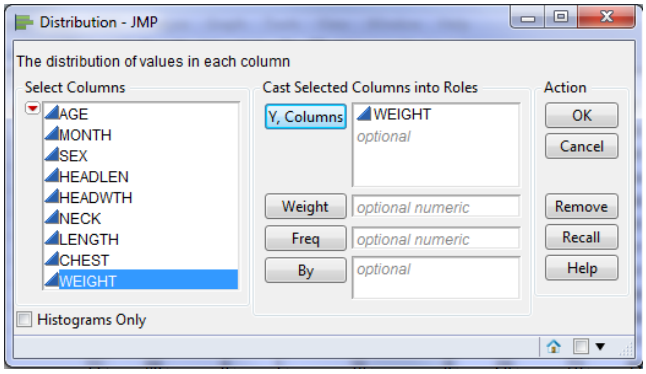
Use the "Analyze" menu and then choose "Distributions" as follows:

Figure 3.1



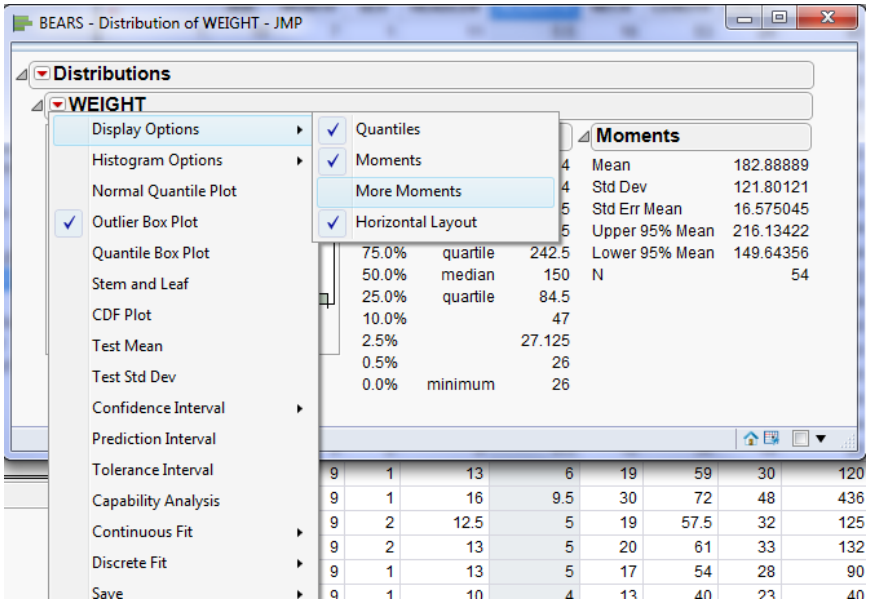
Then select the variable WEIGHT and click over "Y Columns" as described below:

Figure 3.2



click over OK, and you get a new window with a histogram and the summary statistics described below. You can change the orientation of the histogram and summary statistics as described in Chapter 2 (Clicking over the second red triangle and choosing “Display Options”, then select “Horizontal Layout”). You can also add additional information to this window by selecting again the option “Display Options” from the second red triangle and then click over “More Moments” as shown below:

Figure 3.3

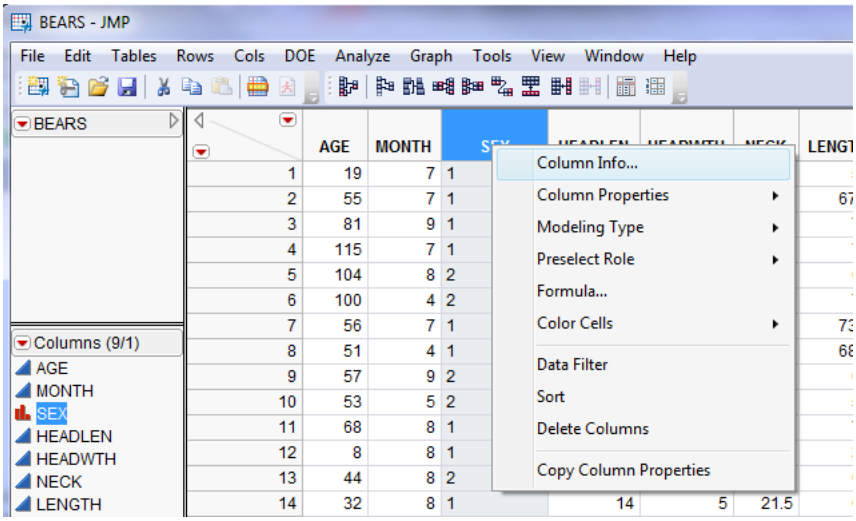


You can see in the new window the Variance, the Skewness, the Kurtosis and the coefficient of variation (CV). Please refer to your text book for an explanation of these measurements.

You can also split the dataset in two of more parts by using a classification variable, for example, you can obtain summary statistics of WEIGHT, separated by SEX. First, let’s change the type of variable in SEX to

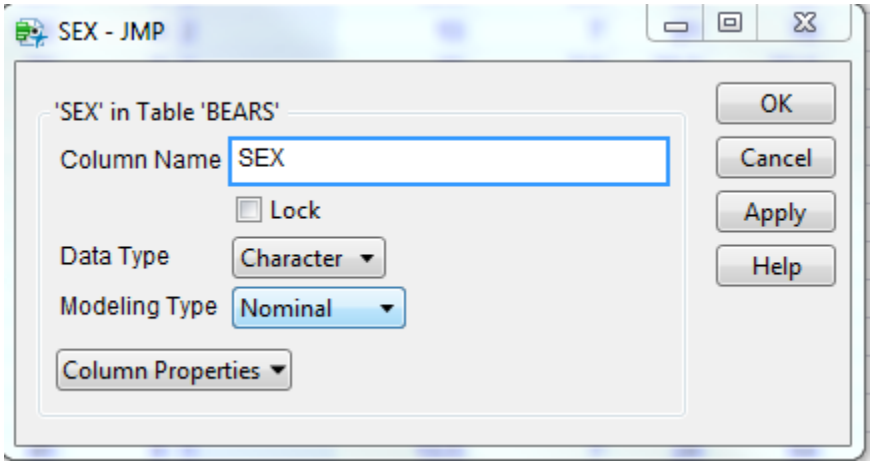
define it as a categorical variable. To do that, right click over the top of the column over “SEX” and then select the option “Column Info” as follows:

Figure 3.4



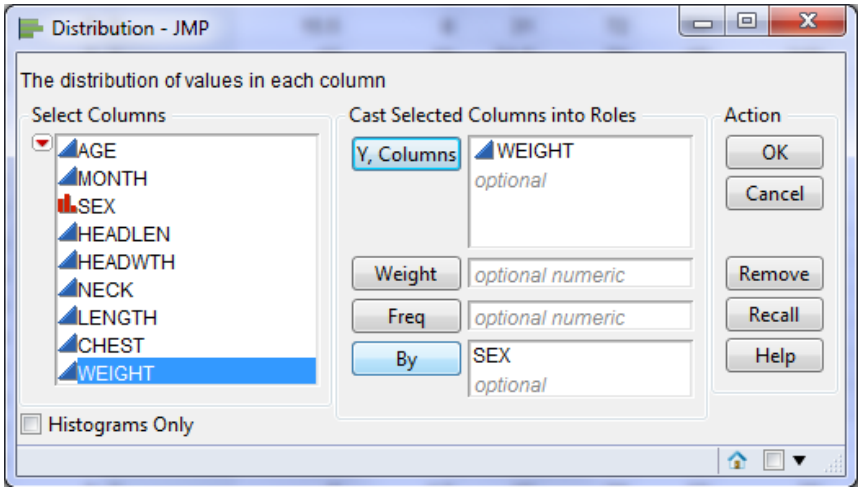
in a new window, select “Character” and “Nominal” as the type of variable. Then click over “OK” and close the window

Figure 3.5



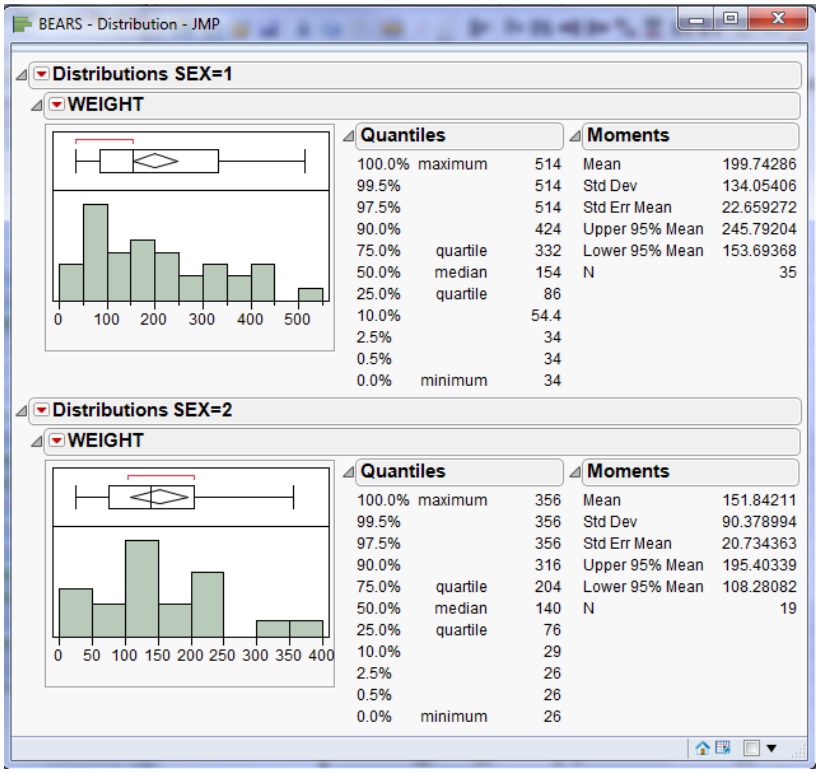
Next, let’s start again the procedure to obtain the summary statistics, first select the “Analyze” menu and then choose “Distributions”, then select WEIGHT and click over “Y Columns”, next, select the variable SEX and click over “By” as indicated in the next figure:

Figure 3.6



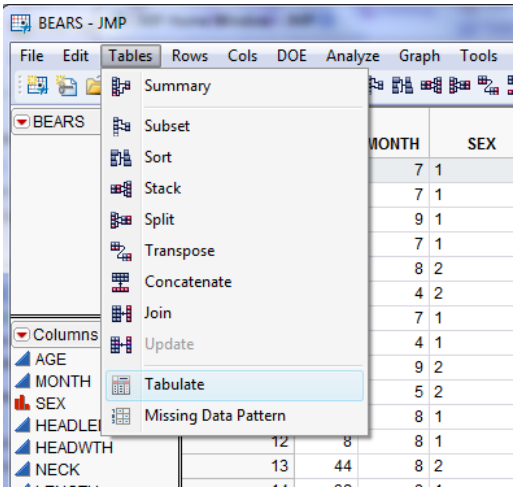
select ”OK” and you will obtain separated windows for their corresponding summary statistics according to the classification variable, (here both windows are shown in a horizontal layout):

Figure 3.7



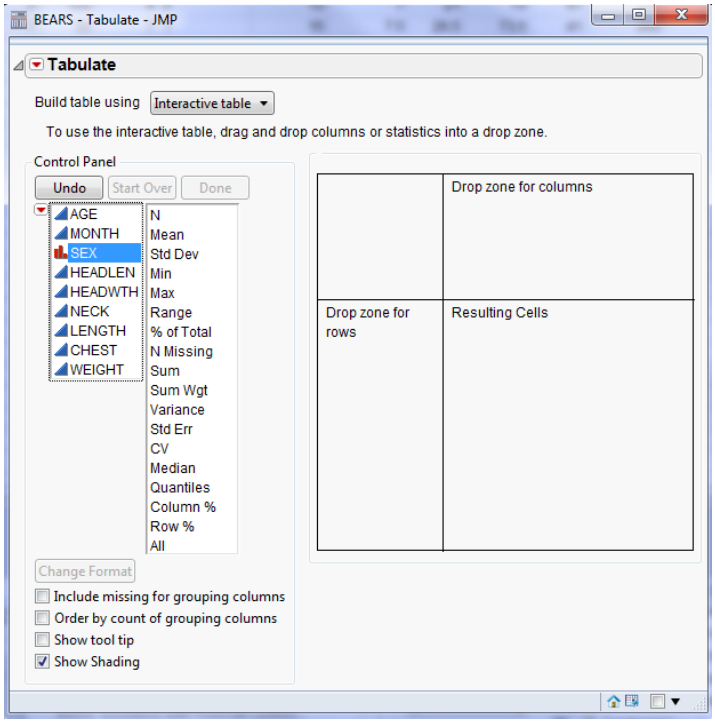
You can also obtain summary statistics using the menu “Tables”, for example let’s select the menu “Tables”, then select “Tabulate” as follows,

Figure 3.8



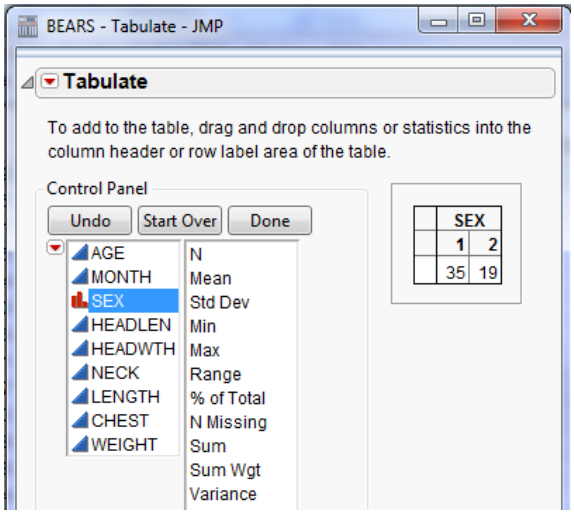
in the new window, drop the variable “SEX” over the drop zone for columns”

Figure 3.9



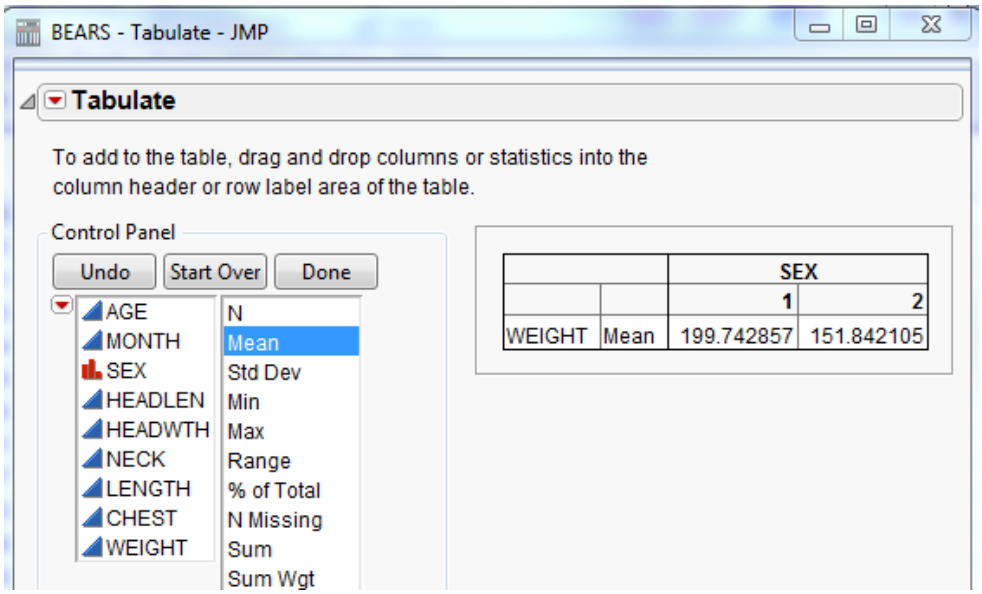
You get a new window with a categorical variable at the top and the counts for the number of observations that correspond to each categorical variable as follows:

Figure 3.10



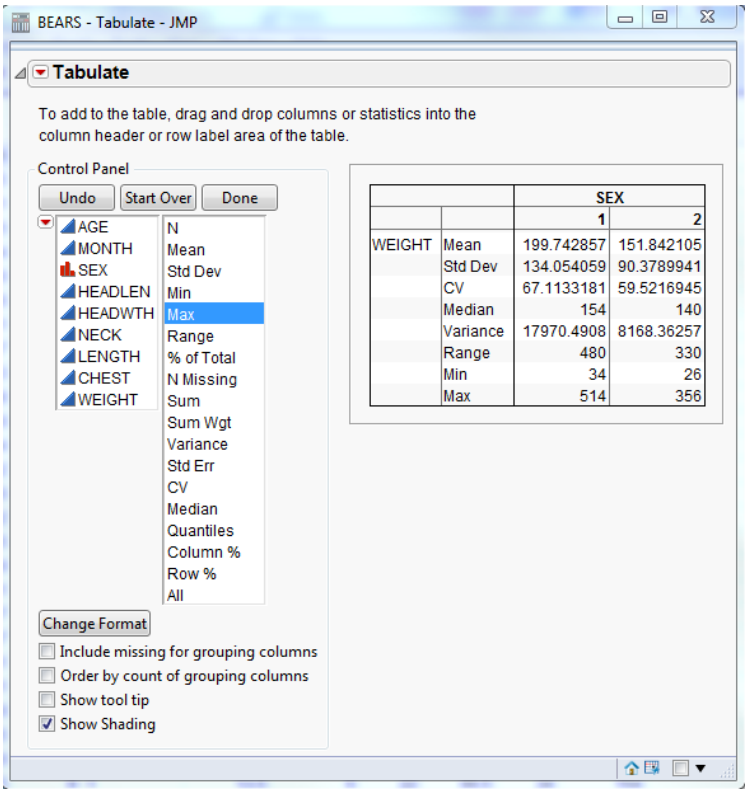
We can also obtain some summary statistics for a numeric variable, to do that, drag the variable WEIGHT over the rows area and choose the option "Add Analysis Columns", then you get the sum of the WEIGHTS separated by SEX for each cell, probably this is not that statistics that you wanted, but we can also compute the mean and the standard deviation, in order to do that, let's drag "Mean" over the rows area over the place where "Sum" is located, then you get the next window

Figure 3.11



Using the same procedure we can also add, the standard deviation, CV, Median, Variance, Range, Min, and Max. To do that, just drag the name of the statistic just below the "Mean" on the table, and you can get the following table,

Figure 3.12

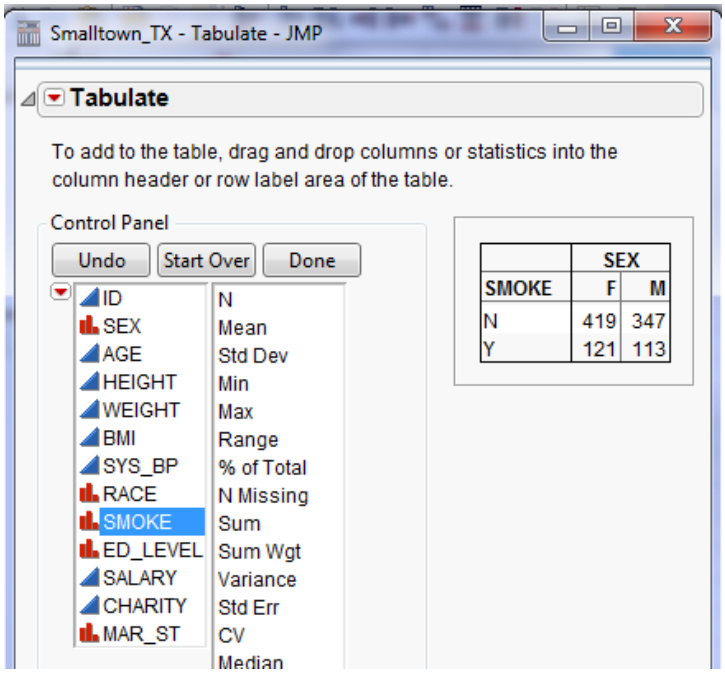


As you can see, you can design your own table containing only the summary statistics that you want or need. You can also obtain counts for a contingency table (remember the contingency table for the Titanic example that we did in class). In that case you only drag one categorical variable over the top of the table and the other categorical variable on the left side of the table. JMP counts for you the number of observations that correspond to each cell and the contingency table will be generated.

Class Example:

Open the Excel file "Smalltown_TX.xls" located on the website for this course at sci.tamucc.edu/~jguardiola and click over your corresponding course. Then click over the menu "Tables" and select "Tabulate" as in Figure 3.8, then select the variables SEX and drag it over the top of the table, next, select the variable SMOKE and drag it over the left side of the table as follows:

Figure 3.13



you can see a contingency table containing the number of observations in each cell that corresponds to each combination of the classification variables. As you can see, this is a powerful tool!

Class Exercises:

1- Repeat all the summary statistics and tables that have been done before using the variables AGE, LENGTH and CHEST, from the same file (BEARS), obtain summary statistics for each variable, separated by SEX.
2- Open the file "Movies" and obtain the summary statistics for each numeric variable, separated by the variable "MPAA Rating", also, develop tables that contain summary statistics for all of these variables, separated again by "MPAA Rating".

Team Assignment:

At this point you know how to compute summary statistics using JMP.

1- Open your sample from the Excel file "Smalltown_TX.xls" that you obtained in Chapter 1. If you have not done so, please refer to Chapter 1 to obtain your random sample from this file and save it in a safe place.
2- Obtain summary statistics for all the numeric variables, highlight and make comments about any important characteristics that you may find on some of these variables.

3- Obtain summary statistics for the variables, SALARY, CHARITY, HEIGHT, and BMI, separated by SEX. Do you notice any differences?
4- Obtain summary statistics for the variables, SALARY, CHARITY, HEIGHT and BMI, separated by ED_LEVEL. Do you notice any differences?
5- Obtain summary statistics for the variables, SALARY, CHARITY, HEIGHT, and BMI, separated by RACE. Do you notice any differences?
6- Create a contingency table to count the number of observations of SEX vs. ED_LEVEL. Hint: Use the interactive tables at the menu "Tables" and then select "Tabulate", after that just drag the name of one the variables at the top, and the second variable on the left side of the table, a contingency table will be generated.
7- Create a contingency table to count the number of observations of ED_LEVEL vs. SMOKE.
8- Create a contingency table to count the number of observations of SEX vs. RACE.
9- Write a report for your lab instructor summarizing all your findings and highlight and make comments on any interesting differences. Organize tables containing the summary statistics that were obtained before. Write your report from a point of view of a consultant reporting findings on a real research project.

Notes: You do not have to obtain all possible summary statistics, just obtain those that you consider relevant.

Chapter Summary

You already practiced obtaining summary statistics during this chapter:

- Use the "Analyze" menu to obtain summary statistics for univariate data
- Separate the summary statistics using a classification variable (categorical variable)
- Use tables to organize and separate summary statistics according to a classification variable (categorical variable).
- Create contingency tables using two categorical variables and the number of observations that correspond to each level of the categorical variables.