

# CHAPTER 8

## *Hypothesis Testing*

In this Chapter we are using inferential statistics to help us to make a decision about a certain statement. The decision that we will make is either to “reject” or “fail to reject” a null hypothesis. For more details, please check your textbook and your class notes. This Chapter includes exercises for testing hypothesis for one sample proportion, and for one sample mean when sigma is unknown.

Class Exercise 1:

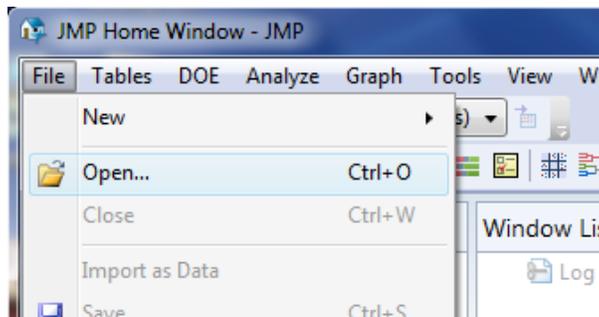
Compare the proportion of female bears versus male bears. Test the claim that the proportion of female bears is equal to 0.5. Perform this test at the 0.05 significance level.

Null hypothesis  $H_0: p=0.5$

Alternative hypothesis  $H_1: p \neq 0.5$

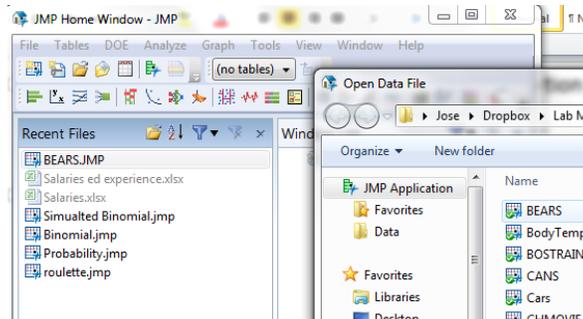
Open the file “Bears” (you can download the file at my website at [sci.tamucc.edu/~jguardiola](http://sci.tamucc.edu/~jguardiola)). The initial steps listed here for opening the file and preparing the data for analysis are the same as in Chapter 7.

Figure 8.1



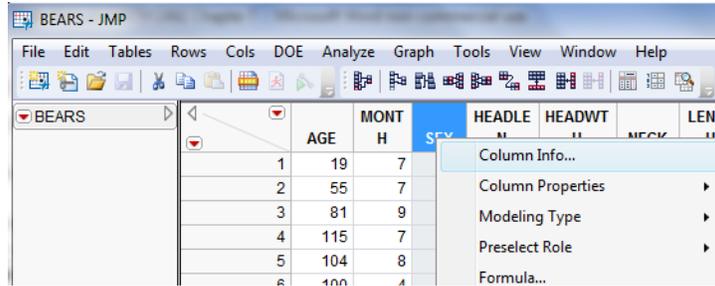
select the file “Bears” as follows:

Figure 8.2



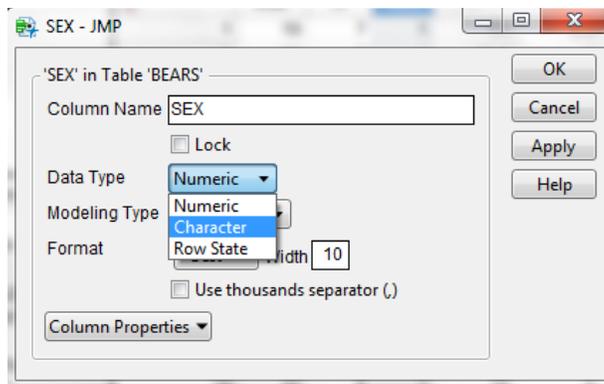
first, right click over “Sex” and select “Column Info...”,

Figure 8.3



we have to change the type of variable for “SEX”, as it appears as a numeric variable and it should be changed to a categorical one (notice that the variable “SEX” should be 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 8.4



you can see that the column alignment changes to the left for column “SEX”, as this is the alignment is used by default for categorical (Character) variables.

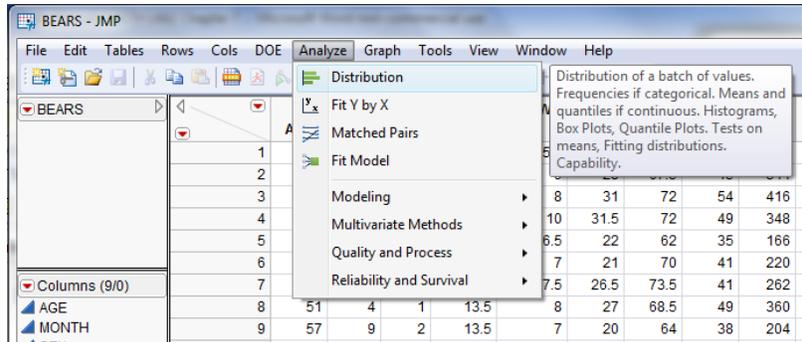
Figure 8.5

A screenshot of the JMP data table showing the 'SEX' column. The columns are AGE, MONT H, SEX, and HEADLE. The 'SEX' column is highlighted in blue and has a left alignment. The data rows are: (19, 7, 1), (55, 7, 1), (81, 9, 1), (115, 7, 1), and (104, 8, 2).

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

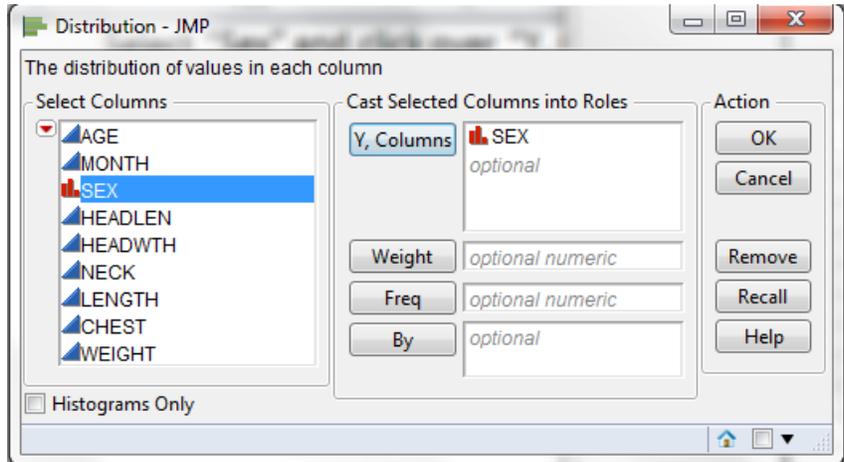
now, select “Distribution” from the “Analyze” menu

Figure 8.6



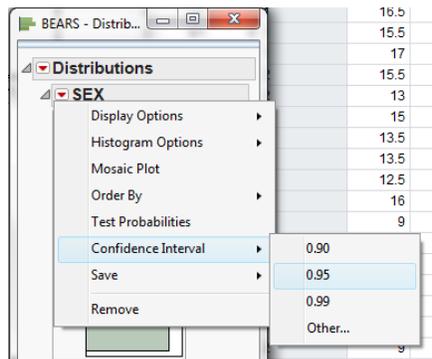
select “SEX”, click over “Y, Columns” and click over “OK”

Figure 8.7



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

Figure 8.8



you get the following results

Figure 8.9

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

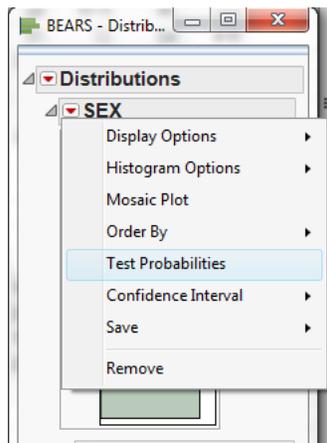
The confidence interval for the proportion of female bears is

$$0.238 < p < 0.485, 95\% \text{ CI}$$

We can use the confidence interval shown above to make a decision about the hypothesis testing for proportion. As you can see, the 95 % confidence interval for female bears does not include the hypothesized value of 0.5. This means that we can reject the null hypothesis (when the hypothesized value shown in the null hypothesis is not included inside the confidence interval, we reject the null hypothesis). Thus, we can't support the claim that the proportion of female bears is equal to 0.5.

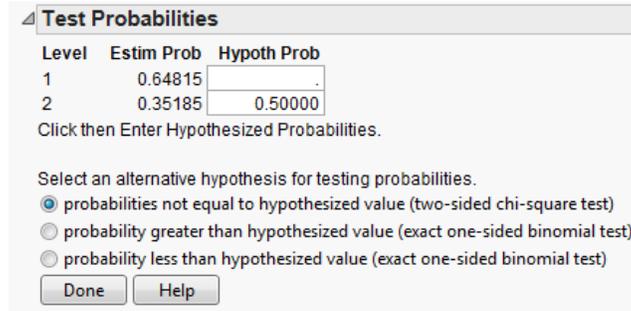
Moreover, we can also test the hypothesis for proportions using a formal hypothesis testing procedure. To do this, click over the lower red triangle and select "Test Probabilities" as follows:

Figure 8.10



A new window opens and you need to type the hypothesized value of the proportion for the null hypothesis, in this case it is 0.5 for female bears (labeled as 2), and then select the option that matches the alternative hypothesis, in this case: “probabilities not equal to hypothesized value (two sided chi-square tests)” as follows, then select “Done”:

Figure 8.11



This statistical test shown here differs from the procedures studied during lecture, as it is an alternative procedure. The interpretation of the computer output and  $p$ -values are as usual.

Figure 8.12

Level	Estim Prob	Hypoth Prob
1	0.64815	0.50000
2	0.35185	0.50000

Test	ChiSquare	DF	Prob>Chisq
Likelihood Ratio	4.8127	1	0.0283*
Pearson	4.7407	1	0.0295*

Method: Fix hypothesized values, rescale omitted

As can be seen from the computer output, testing a hypothesis at the 0.05 significance level, the null hypothesis is rejected because the  $p$ -value is less than 0.05. We can't support the claim that the proportion of female bears is equal to 0.5.

**Class Example 2:**

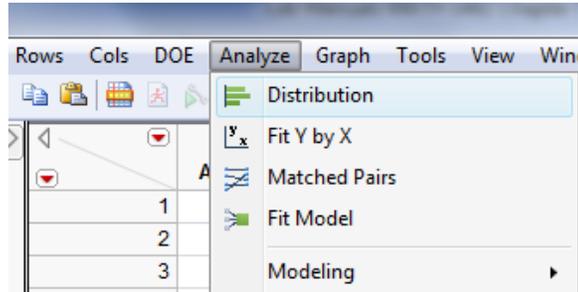
Use the variable “WEIGHT” from the file “Bears”, and separate male bears from female bears using the variable “SEX” (Coding: males =1, females = 2). Test the claim that the mean weight of male bears is greater than 150 lbs at the 0.05 significance level.

Null hypothesis: the mean weight of male bears is equal to 150 lbs  $H_0: \mu = 150$  lbs

Alternative hypothesis: the mean weight of male bears is greater than 150 lbs  $H_1: \mu > 150$  lbs

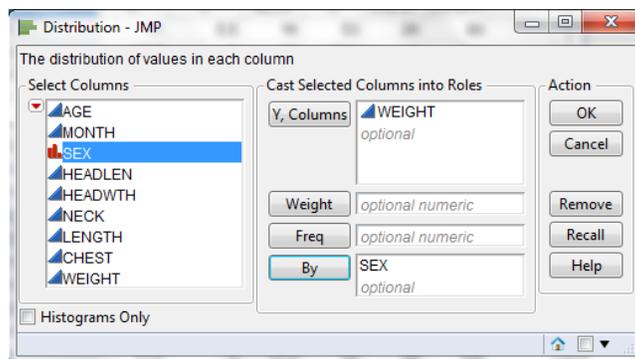
Open the file “Bears” as in Figures 8.1 to 8.2 (or ignore these steps if the file is already open), and change the variable “SEX” to character, as in Figures 8.3 and 8.4. Then select “Distribution” from the “Analyze” menu

Figure 8.13



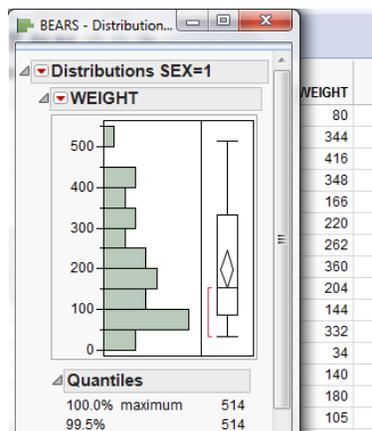
Select “WEIGHT” and click over “Y, Columns”, select “SEX” and click over “By”

Figure 8.14



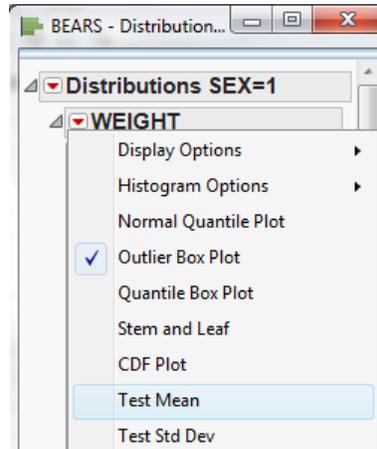
click over “OK” and you obtain two separate distributions for males and females, however the shapes of the distributions do not as look normally distributed, but this issue was already discussed on Chapter 6. Here we are assuming that weights are normality distributed.

Figure 8.15



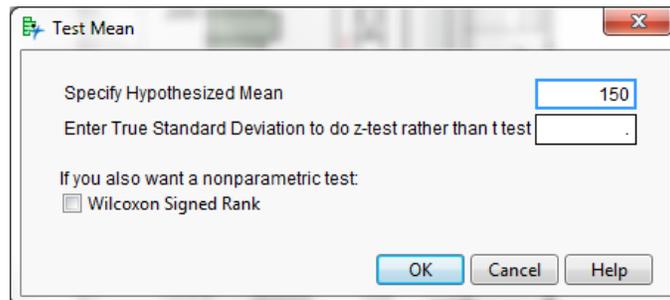
click over the lower red triangle, select the option “Test Mean”

Figure 8.16



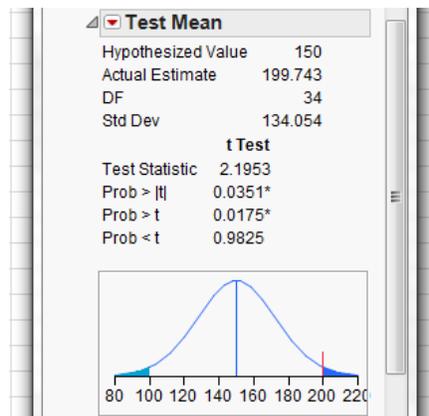
then, type the hypothesized value of the mean, and leave blank the standard deviation (we are assuming that the standard deviation is unknown, which is the most common practical situation)

Figure 8.17



then, look at the computer output,

Figure 8.18



you can see that this is a right tail test (please take a look at your class notes or textbook) and the  $p$ -value for a right tail test is indicated at the computer output as “Prob >  $t$ ”, equal to 0.0175, which is statistically significant at the 0.05 significance level (0.0175 is less than  $\alpha = 0.05$ ).

Conclusion: Reject the null hypothesis

Conclusion in plain English: There is enough evidence to support the claim that the mean weight of male bears is greater than 150 lbs.

Alternatively, a  $z$ -test can also be performed assuming that  $\sigma$  is known (not recommended in practice), by specifying the population standard deviation on the box shown in Figure 8.17. In this case a  $z$ -test will be performed (instead of a  $t$ -test), but this procedure is of little practical value as the population standard deviation is rarely known.

#### Class Exercises:

- 1- Hypothesis test for proportion: Open the file “Movies.JMP” and test the claim at the 0.05 level that the proportions of movies classified as PG-13 is greater than 50 %.
- 2- Hypothesis test for the means: Open the file “Bears.JMP” and test the claim that the mean length of male bears, is greater than 50 inches. Write the null and alternative hypothesis, obtain the statistical conclusion and write the plain English conclusion. Verify assumptions.
- 3- Hypothesis test for the means: Open the file “Bears.JMP” and test the claim that the mean neck measurement of male bears, is less than 25 inches. Write the null and alternative hypothesis, obtain the statistical conclusion and write the plain English conclusion. Verify assumptions.

#### Team Assignment: One sample hypothesis testing

Use the sample that you obtained at the beginning of the semester 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) and do the following:

- 1- Test the following hypothesis for proportions:
  - a. Test the claim that the proportion of Hispanics is greater than 50%.
  - b. Test the claim that the proportion of smokers is less than 30%.
  - c. Test the claim that the proportion of people with a bachelor’s degree (BD) is equal to 20%.
- 2- Test the following hypothesis for the mean:
  - a. Test the claim that the mean salary for females is less than \$30,000.
  - b. Test the claim the mean BMI for females is greater than 20.
  - c. Test the claim the mean weight for males is equal to 160 lbs.
  - d. Verify assumptions for all previous hypothesis tests, comment if the tests are valid or not, and support your reasoning.
- 3- Present a written report that summarizes your findings.