CHAPTER 9

Two Samples Hypothesis Testing

In this Chapter we are using inferential statistics to make a decision about a certain statement (claim) when comparing two datasets. This Chapter includes exercises for testing hypothesis for comparing the means of two independent samples.

Class Example 1:

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 the mean weight of female bears at the 0.05 significance level.

Null hypothesis: the mean weight of male bears is equal to the mean weight of female bears

$$H_0: \mu_1 = \mu_2$$

Alternative hypothesis: the mean weight of male bears is greater than the mean weight of female bears
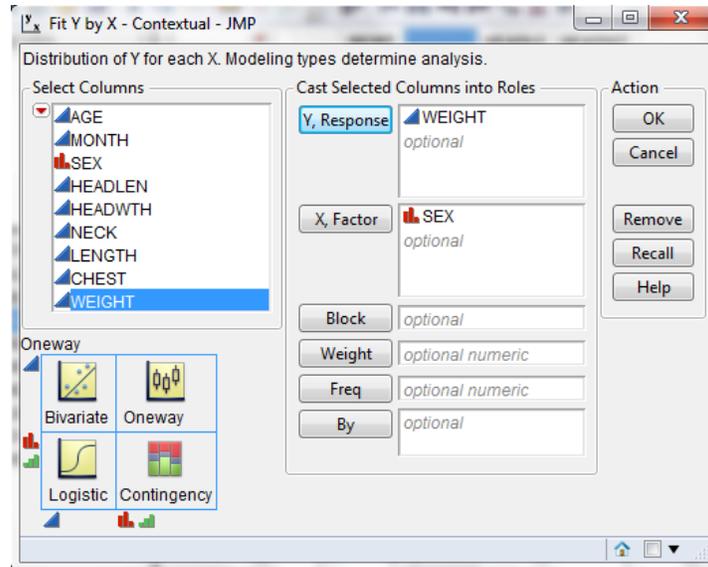$$H_1: \mu_1 > \mu_2$$

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 "Fit Y by X" from the "Analyze" menu as follows,
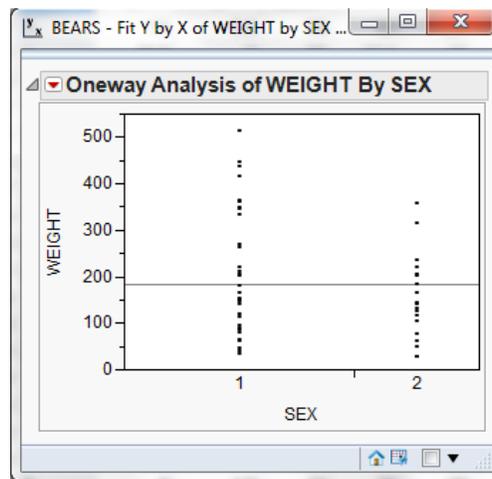
Figure 9.1
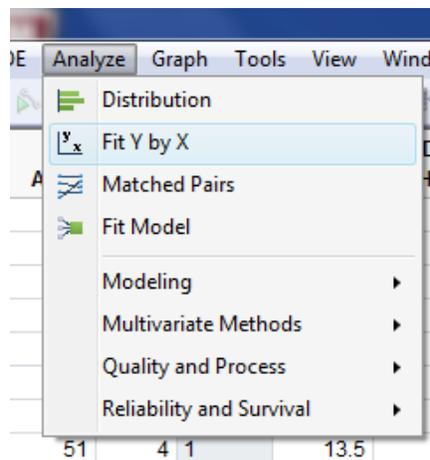
then, select the variable "SEX" and click over "X, Factor" and select the variable "WEIGHT" and click over "Y, Response" as shown below
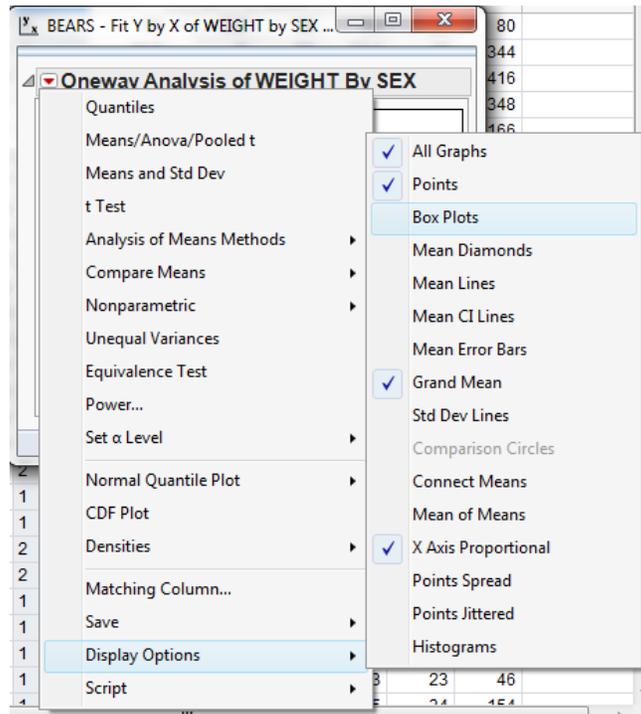
Figure 9.2

Then click over "OK", and you will see the following screen,

Figure 9.3

As can be seen, each dot corresponds to a data point, it seems that male bears are heavier than their female counterparts, but we still have to perform a statistical test, and show additional graphs. In order to do that, click over the red triangle and select "Display Options", then select "Box Plots" in the next window
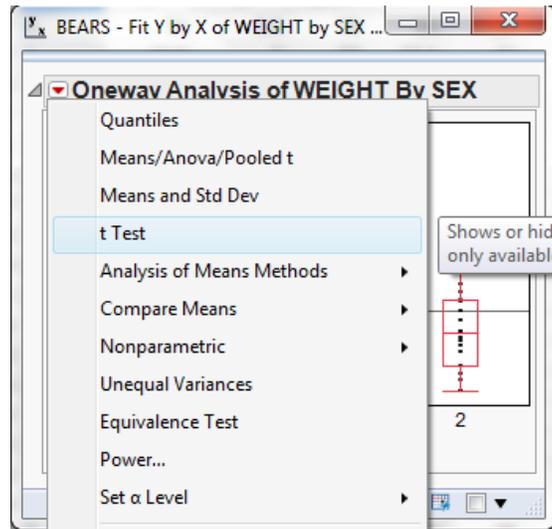
Figure 9.4

At this point, you can see parallel box plots for males and females as below,
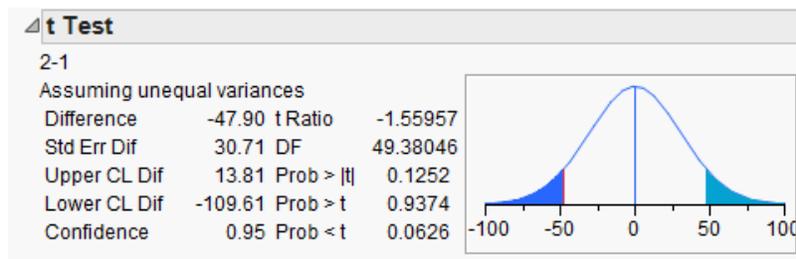
Figure 9.5
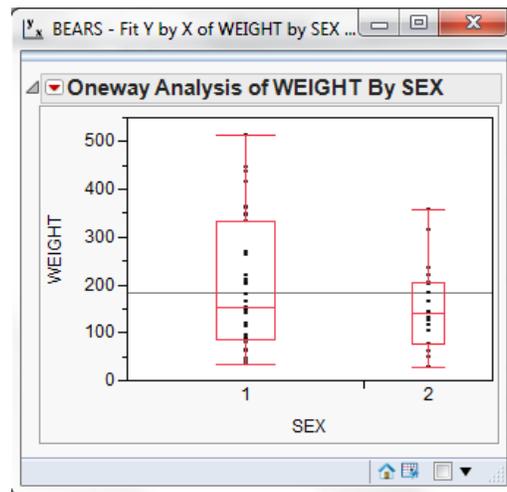
Next, click again on the red triangle and select "t Test"

Figure 9.6

this option produces a *t*-test for unequal variances, and the results are shown below,

Figure 9.7

The results shown above indicate that the hypothesis that is being tested is $\mu_2-\mu_1$, therefore we need to rewrite the null and alternative hypothesis as follows:

$$\text{Ho: } \mu_2 - \mu_1 = 0 \qquad (8.1)$$

$$\text{H1: } \mu_2 - \mu_1 < 0$$

The computer output shown in Figure 9.7 include the three possible choices for alternative hypothesis, "not equal" (Prob > |t|), "greater than" (Prob > t) and "less than" (Prob <t) , the p-value for each alternative is shown next to these expressions. For testing this hypothesis, we need to look at the corresponding *p*-value that is shown at the bottom (Prob < t) for a left tailed test, and this number is 0.0626. This *p*-value is not small enough to reject the null hypothesis, as this *p*-value is greater than α. Then the plain English conclusion is: There is not enough evidence to conclude that male bears are heavier than female bears. This test was performed under the assumption that the variances are unequal.

The rule that we studied in class is that if the ratio of the two variances is greater than 3, the variances are considered unequal. This is a rule of thumb, not a formal test. Let's check the ratio of the two variances. We need to compute the variances for each sample. To do this we need to click on the red triangle and select "Means and Std Dev" as shown below,

Figure 9.8
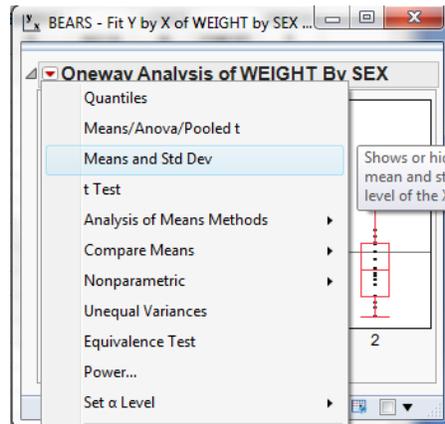
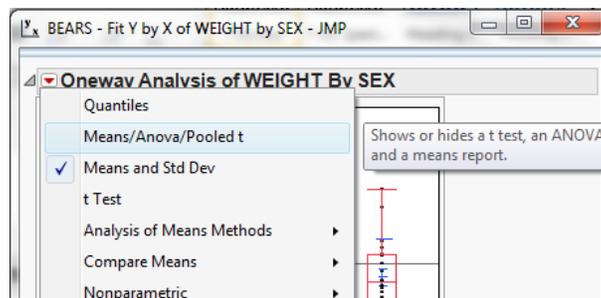The results are shown below,

Figure 9.9

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
1	35	199.743	134.054	22.659	153.69	245.79
2	19	151.842	90.379	20.734	108.28	195.40

Then the ratio of the largest variance to the smallest variance is:

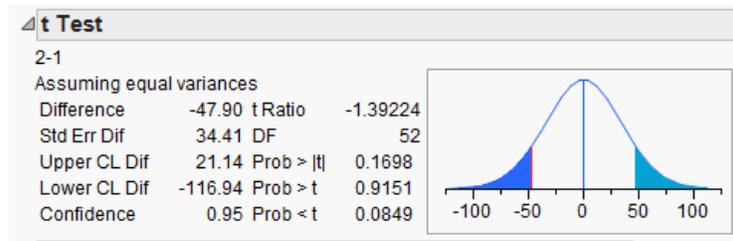$$\frac{(134.05)^2}{(108.28)^2}=2.2$$

Thus, applying the rule of thumb a t-test for equal variances is appropriate. Then, click on the red triangle and select "Means/Anova/Pooled t" as follows, (a pooled t-test is a t-test using the assumption of equal variances).

Figure 9.10

Then, the results are shown below, (just ignore the ANOVA results for now),

Figure 9.11

The table above shows a *t*-test results using the assumption of equal variances, again, given that the null hypothesis is expression 8.1, this is a left tailed test and the corresponding *p*-value is the last number on the table above, in this case, the p-value is 0.0849, that again is not small enough to reject the null hypothesis, therefore the conclusion is still the same.

Statistical Conclusion: Reject the null hypothesis

Conclusion in plain English: There is not enough evidence to support the claim that the mean weight of male bears is greater than the mean weight for female bears.

Class Exercises:

1- Hypothesis test for the means: Open the file "Bears.JMP" and test the claim that the mean length of male bears is greater than the mean length of female bears. Write the null and alternative hypothesis, obtain the statistical conclusion and write the plain English conclusion.
2- Hypothesis test for the means: Open the file "Bears.JMP" and test the claim that the mean neck measurement of male bears is different from the mean neck measurement of female bears. Write the null and alternative hypothesis, obtain the statistical conclusion and write the plain English conclusion.

Team Assignment: Two samples 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 the mean:
   a. Test the claim that the mean salary for males is greater than the mean salary for females.
   b. Test the claim the mean BMI for males is less than the BMI for females.
   c. Test the claim the mean height for males is greater than the mean height for females.
   d. Verify assumptions for all previous hypothesis tests, comment if the tests are valid or not (based on the requirements for each test), and support your reasoning.
2- Write a report that summarizes your findings and submit it to your lab instructor.