Stat 401G Lab 2

Overview

In this lab you will look at analyzing data from two independent random samples using JMP. The data for this demonstration is the Body Mass Index data for men and women. The JMP data file can be found on the course web site

www.public.iastate.edu/~wrstephe/stat401.html

Computer Exercises

1. Open the “Body Mass Index” data table. This data table has a column for BMI and a column for Gender. Use the Column Info option to see the difference between these two columns.

2. To look at BMI separately for each Gender, use Analyze + Distribution and cast BMI in the role of the response variable. Then cast Gender in the role of a By variable. You should get two separate analyses, one for each Gender. For each Distribution choose Uniform Scaling and Stack.

3. Another way to compare the two Genders is to use the Analyze + Fit Y by X platform. Cast BMI in the role of Y, Response and Gender in the role of X, Factor. Click on OK. Because BMI is a continuous modeling type and Gender is a nominal modeling type, JMP will produce output with side-by-side dot plots. You can change the display to side-by-side box plots and test for a difference in population means for the two types by selecting Means/ANOVA/Pooled t. This option pools the two sample variances into a pooled estimate of the population variance. There is also a t Test option that will keep the two sample variances separate in the calculation of the t-test statistic. If sample sizes are equal the only difference between the Pooled t and t Test is in the degrees of freedom (df).

4. We will now look at residuals for the two sample problem. For the two-sample problem the model and definition of a residual are:

\[ Model : y_{1i} = \mu_1 + \varepsilon_{1i}, \quad Residual : y_{1i} - \bar{y}_{1i} \]
\[ y_{2i} = \mu_2 + \varepsilon_{2i}, \quad y_{2i} - \bar{y}_{2i} \]

JMP can automatically calculate residuals for you from the Fit Y by X platform. Go to Analyze + Fit Y by X and enter BMI as the Y, Response and Gender as the X, Factor. Click on OK. Go to the red triangle pull down in the output window and select Save + Residual. This will create a new column in your data table labeled BMI centered by Gender. These are the residuals for the two-sample problem.
You can now use **Analyze + Distribution** to analyze the residuals, BMI centered by Gender. You should have a histogram, box plot, and Normal Quantile Plot. You can use Fit Distribution + Normal, to superimpose a normal curve on the histogram. Note that even though you have two samples of BMI values (Male and Female) you analyze a combined (single) set of residuals.

5. For the two independent sample model there is also the condition of equal standard deviations, variances, for the random errors. You can check this condition both graphically and numerically.

- **Graphically:** Use Fit Y by X and cast the residuals (BMI centered by Gender) as the Y, Response and Gender as the X, Factor. You should see side-by-side dot plots of the residuals. JMP should automatically add a horizontal line at zero. Look at the spread for each Gender.

- **Numerically:** Go to the red pull down in the Fit Y by X output and select Means and Std Dev. This will provide the standard deviations for the residuals for Males and Females separately. You can compare these two sample standard deviations. The condition of equal population standard deviations is satisfied provided one sample standard deviation is less than 3 times as large as the other sample standard deviation.

Alternatively, use **Analyze + Distribution** and cast BMI centered by Gender as the Y, Columns and Gender as the By variable. This will calculate the sample standard deviations for each Gender. You can compare the two sample standard deviations.