**Stat 401 - Fitting an ANOVA and ``After the ANOVA’’**

**Goals:** In this lab, we will see how to fit an ANOVA model and how to do various ``After the ANOVA’’ analyses.

We will use the case0501.csv data set. This is the mouse diet / longevity study discussed in lecture.

Download the case0501.csv file from the JMP or the datasets page and load it into JMP.

**Set up to fit an ANOVA model – JMP modeling type for a variable:**

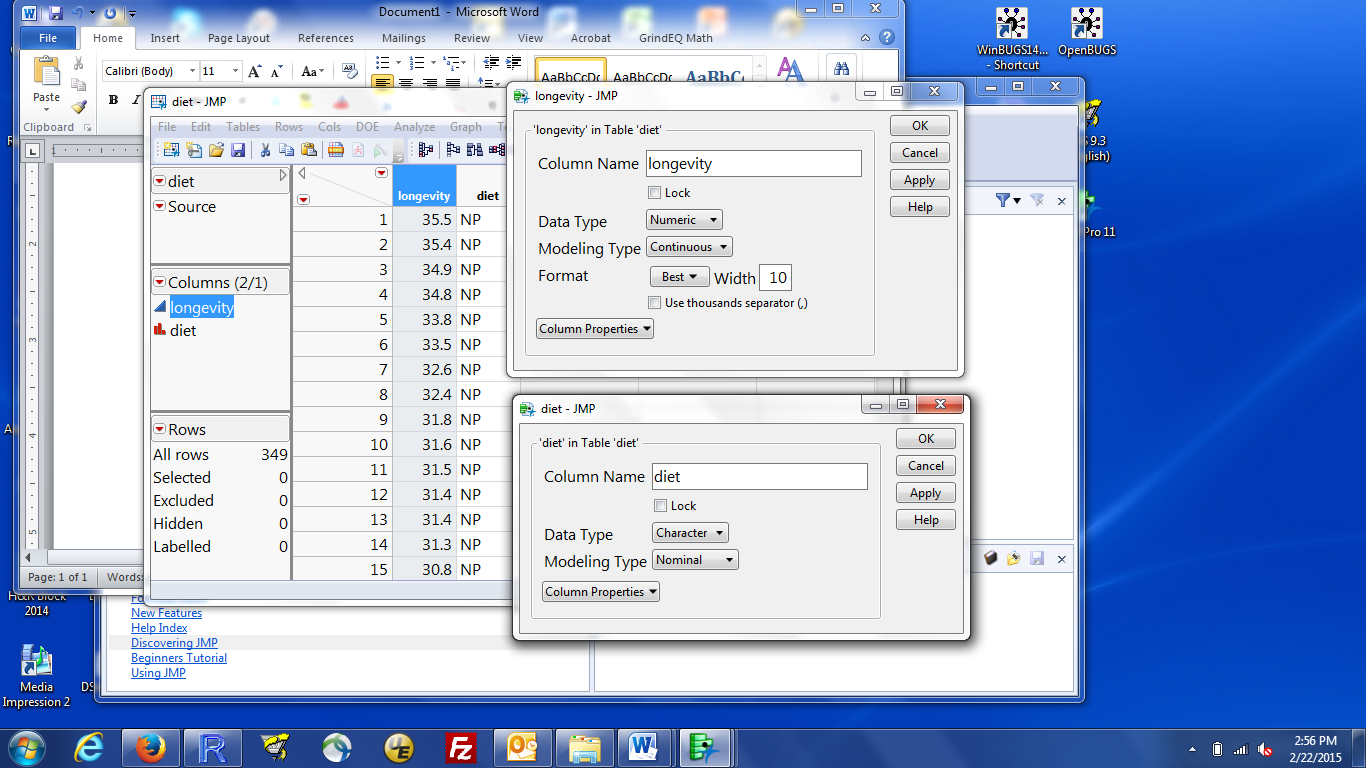
The hardest thing about fitting an ANOVA model is making sure JMP decides to do what you want. Each variable in a JMP data set has a “modeling type”. If the results don’t look appropriate for an ANOVA, check that JMP is using the correct modeling type. The Analyze / Fit Model platform fits many different types of models. It is the JMP equivalent of the SAS proc glm or the R lm() function. JMP uses the modeling type to determining whether it will fit an ANOVA model (means for groups) or a regression model (next week's topic).

To check the modeling type, look at the little icons by each variable in the Columns box (mid left) of the data window and most other places where you can choose variables. The little icons to the left of the variable name portray the modeling type:

a blue ramp for a continuous variable

red bars for a nominal variable.

To change the modeling type, either right click on the variable name in the columns box, or left click on the column to highlight it then select Cols from the main menu. Choose Column Info. The box that appears tells you information about that column of data. Below is that information for both the longevity and diet variables.



Modeling type is continuous for longevity and nominal for diet. Nominal is JMP’s name for groups. Change the values in the Modeling type box if necessary.

Note: The data type indicates what JMP sees in the column; the modeling type tells JMP what to do with the variable. A data type: Numeric variable can be modeled as either Continuous (fit a regression) or Nominal (fit group means). A data type: Character can only be modeled as Nominal (fit group means).

To fit an ANOVA model, the response variable needs to be continuous and the predictor variable needs to be nominal.

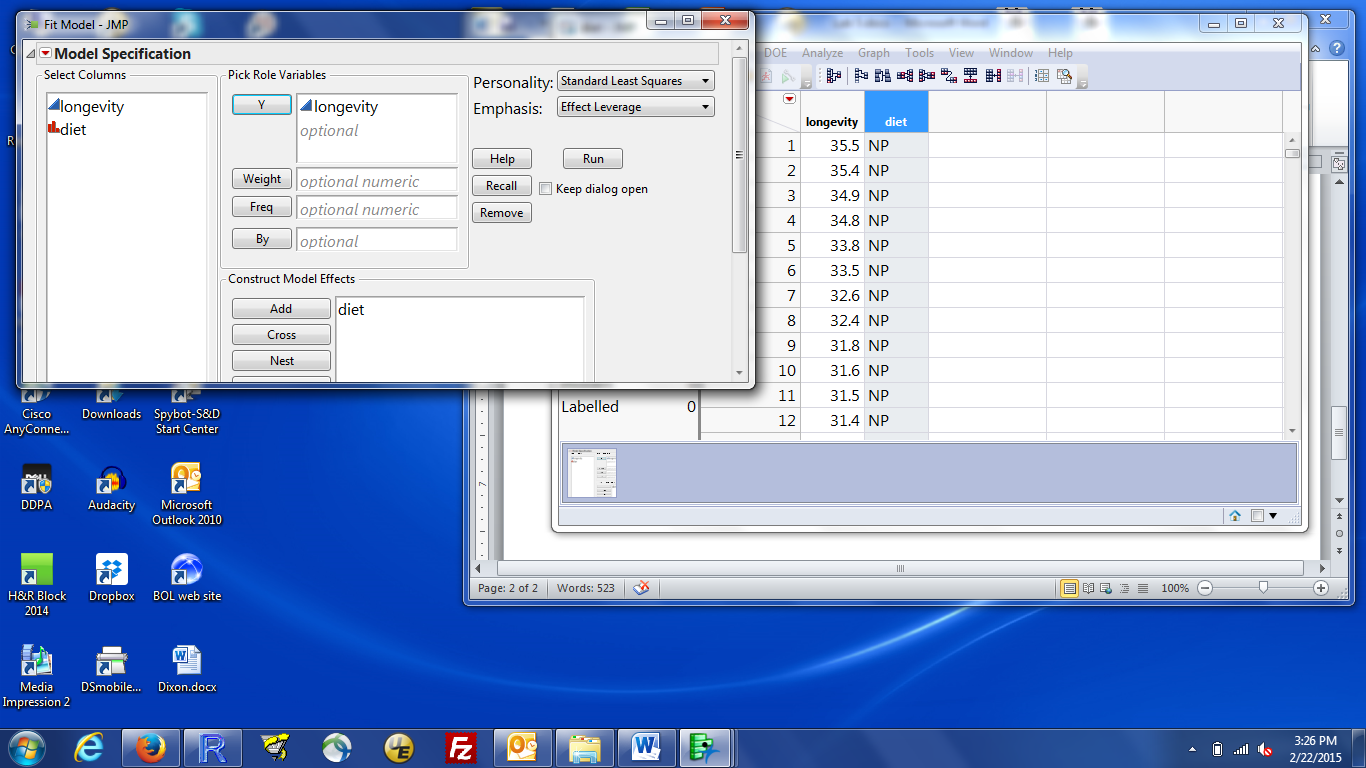
If the modeling types are not correct, you can also change them by clicking on the black triangle and selecting your desired type in the column info box. You can also left click on the blue ramp or red bars to directly change the modeling type. JMP will always let you go from continuous to nominal but it will only let you go from nominal to continuous if the variable is numeric (see detail below if you have this problem).

Detail that can be ignored unless you have this problem: If you believe a variable should be numeric but JMP read it as character, the most likely reason is that one or values are not numbers (e.g. missing value codes that were not read as missing). Either re-read the data using text import preview and specifying the appropriate missing value code, or just convert fields to numbers by changing the data type to numeric. You can then change the modeling type to numeric.

The most likely reason you need to change the modeling type when the data set has a grouping variable with numeric values. JMP will assume these should be modeled as continuous variables (and will fit a regression model).

**Fitting an ANOVA model:**

From the main menu: Analyze / Fit Model. The Analyze / Fit Y by X platform can also fit the ANOVA models we are currently discussing. Fit Y by X does not let you fit more complicated models and does not let you do some of the "after ANOVA" items. So, we'll just use Analyze / Fit Model.

Select the response variable (here, longevity) and click Y to put longevity in the Y box. Select the grouping variable (here, diet) and click Add to put diet in the Construct Model Effects box. The model specification dialog box should look like this:

Fit the ANOVA by left clicking the Run box.

Pieces of the output that are especially useful:

pooled sd: Root Mean Square Error value in the Summary of Fit box

ANOVA table: Analysis of Variance box. Ask if you don’t understand how these pieces relate to what was discussed in class.

Means for each group: Least Sq Mean in the Least Squares Means box. For a one-way ANOVA, the Mean values are the same as the LS means. For more complicated designs, the LS means are more generally useful.

se for each group mean: Std Error values in the Least Squares Means box. These are based on the pooled sd.

Residual by Predicted Plot: the basic plot for diagnosing problems with assumptions

Additional things you can get:

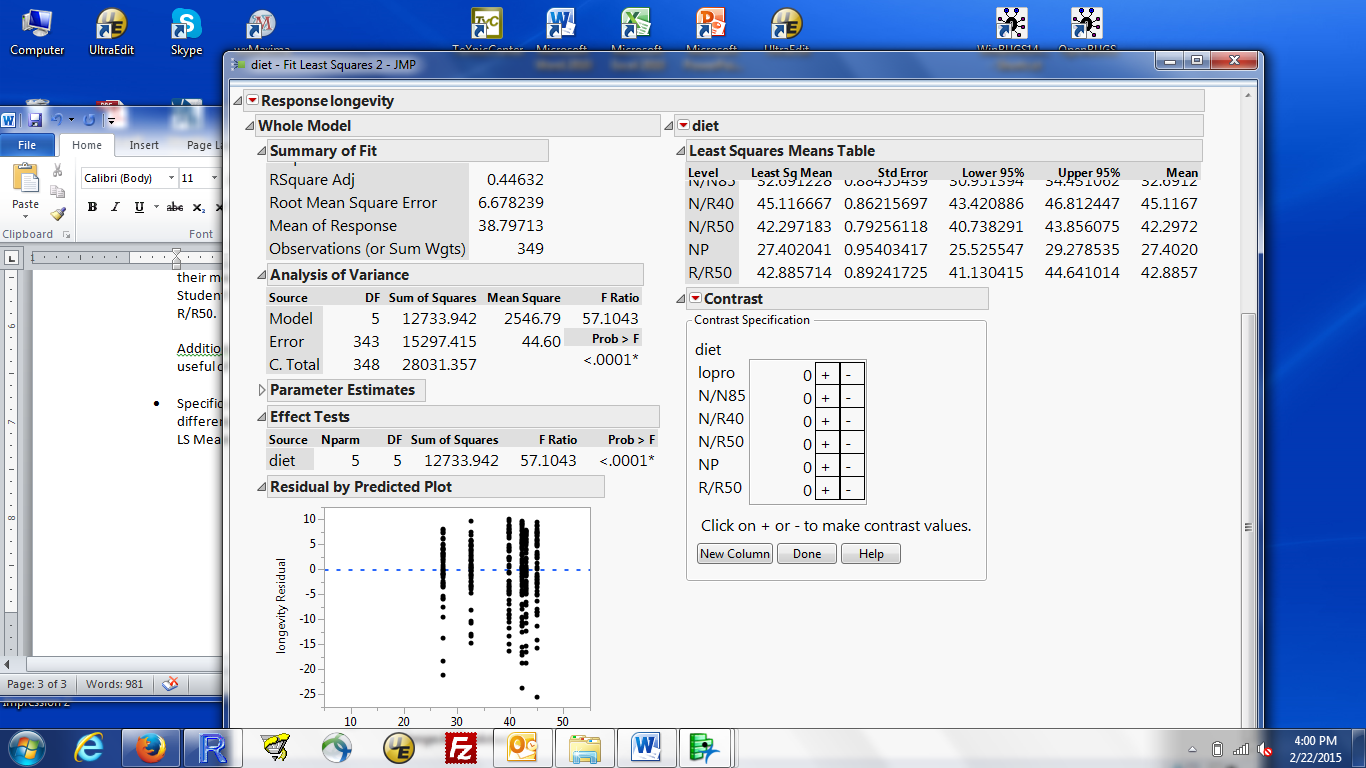
* 95% confidence intervals for group means: right click on the column names in the Least Squares Means Table (i.e. Level, Least Sq Mean, …), mouse over Columns and select Lower 95% and Upper 95% to add those columns to the results.
* Pairwise comparisons among group means: there are a couple of ways to get these. I think the easiest is to click the red triangle by the name of the grouping factor (diet). Two choices give you all pairs of comparisons: LS Means student’s T and LS Means Tukey’s HSD. The Student’s T version gives you tests without multiple comparisons adjust; the Tukey’s HSD gives you tests with Tukey’s adjustment.

The output is a compact summary with one box for each combination of two groups. The key is in the top left: the four numbers in each box, going from top to bottom, are: the estimated difference, its se, and the 95% confidence interval. Red text indicates comparisons that are significantly different from 0; black comparisons that are not.

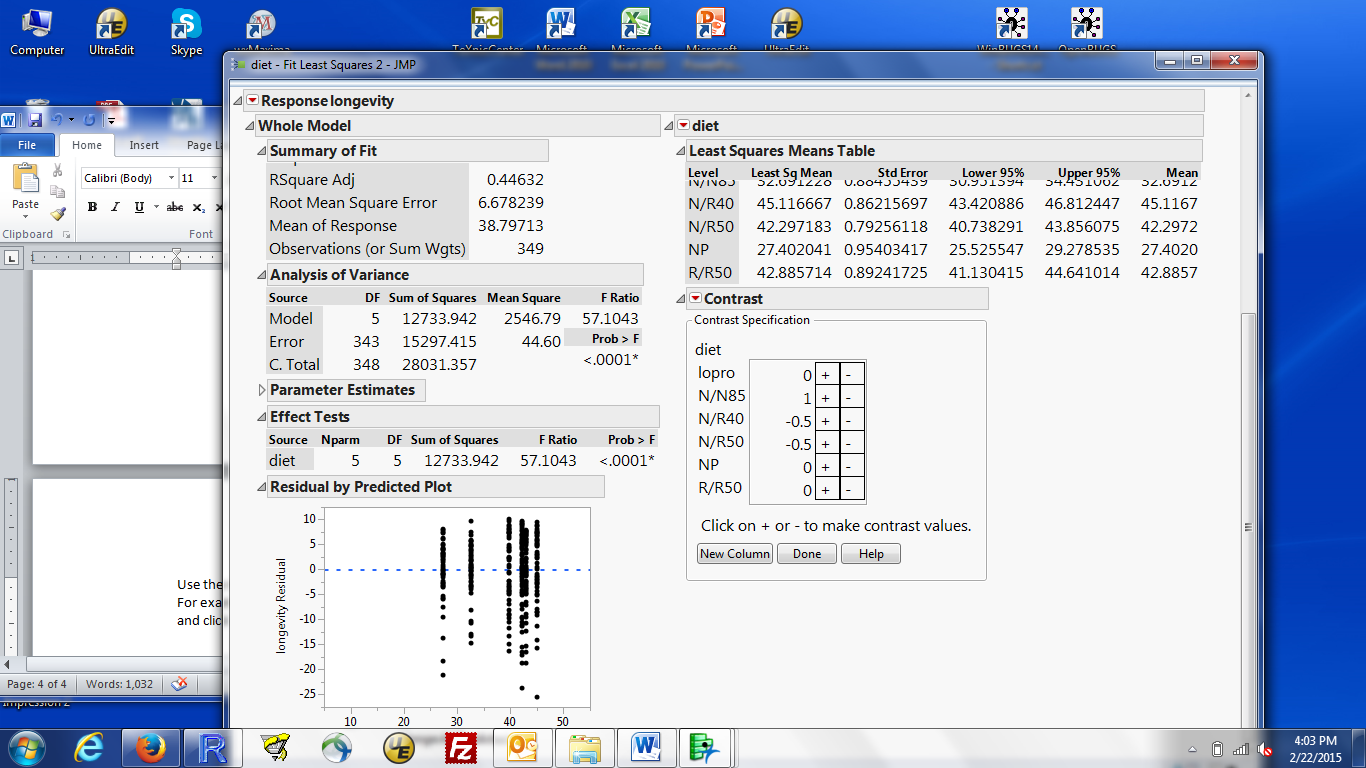
The details for each pair are followed by a letters table. Groups are listed in descending order of their mean. Groups that share the same letter are not significantly different. So (using Student’s T): N/R40 is significantly different from N/R50 but not significantly different from R/R50.

Additional forms of output is obtained by clicking the red triangle by LSMeans. I find the most useful of these is the ordered differences report.

* Specific comparisons among group means: To get a comparison, other than the simple difference between two groups, click the red triangle by the group name (e.g. diet) and choose LS Means Contrast. You will get a box that looks like:



Use the mouse to left click either the + or the – by groups that are involved in the comparison. For example, to compare N/N85 to the average of N/R40 and N/R50, click the + by N/N85 **once** and click the – by N/R40 and N/R50. The contrast box will now look like:



Note that after you click the – by N/R40 and before you click that by N/R50, the coefficients will be 1 and -1. They change to -0.5 after you click the – by two groups.

If you then click the – by R/R50, the three negative coefficients will change to -0.333. To undo this, click the + by R/R50 and the coefficients will revert to what is above.

If one of the values (e.g. for N/R40) is -0.667, you clicked – twice for this group. Click + once and the coefficients will go back to -0.5 and -0.5.

When you have the coefficients you want, click Done and you’ll get the test of this contrast = 0. The default output has some numbers associated with the general theory of contrasts and the p-value for the test of coefficient = 0.

To see the estimate, click the small grey triangle by Test Detail. That box will expand. You will see a summary of the test you requested (do check it was the one you wanted), followed by:

Estimate: The estimated value of this contrast

Std Error: the se of this contrast

t Ratio: the T statistic for this contrast

Prob > |t| the p-value for the test of contrast = 0. Matches (perhaps with less rounding) the p-value from the default output.

Unfortunately, I don't know any way to get the confidence interval for a contrast.