# Lab 9: ANOVA (Chapter 5)

## Objectives

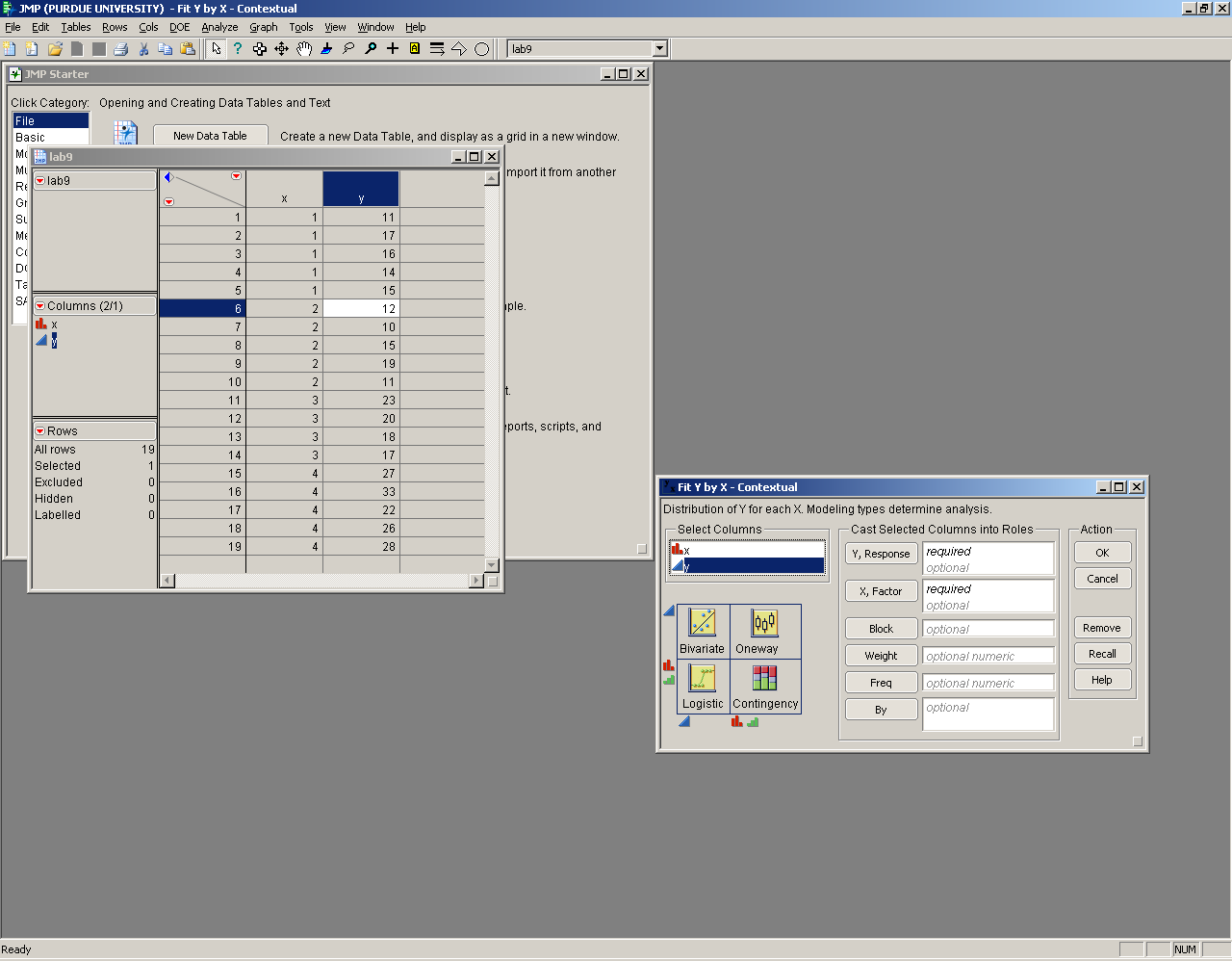
* Identify when ANOVA is used in statistics
* Use JMP to analyze data

## Analysis of Variance (ANOVA)

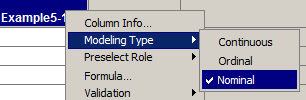
Now, we move onto analyzing data from more than two populations using Analysis of Variance (ANOVA). **We will consider multiple factors (or variables, or treatments) at different levels (settings).** For example, we may consider the environment in lab in terms of two different factors, like temperature and relative humidity, at different levels, like 25°C/30°C and 60%/70%. Normal hypothesis testing wouldn’t be sufficient in comparing all of these different scenarios, so we turn to ANOVA. **ANOVA essentially breaks down the observed variance into the different sources (for each treatment) and then compares the means of the different groups to see if any of there is any difference.** ANOVA can be done by hand (or Excel) using the equations in chapter 5, but today we’ll focus on using JMP for ANOVA to make it simpler.

## ANOVA in JMP

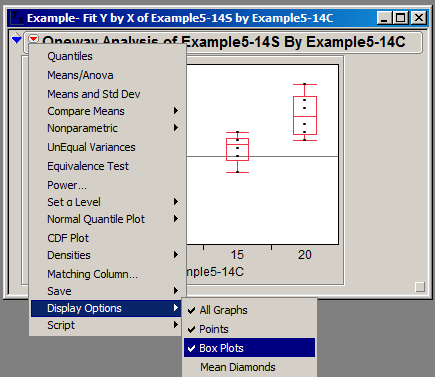
After loading your data in JMP, select Analyze > Fit Y by X.



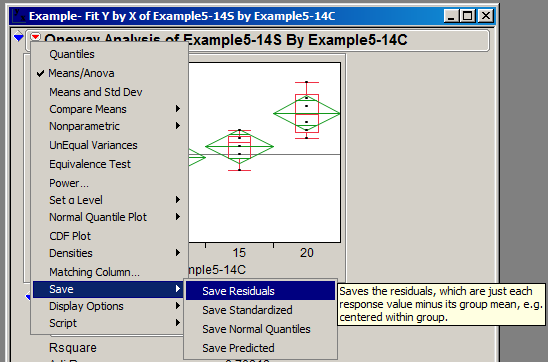
What we are doing is a one-way ANOVA analysis, which is on the upper right quadrant in the image in the bottom left. The symbols mean that our x variable must be discrete or nominal (denoted by the red or green icons) and the y variable must be continuous (denoted by the blue triangle). If you need to change your data to fit what analysis is being done, you can right click the column heading in the data and change the modeling type.



**Remember for JMP, you can always click the red triangle in the upper right to add more options. For ANOVA, you must click the second option (‘Means/ANOVA’). It’s also useful to add box plots (Display Options > Box Plots) to compare the means of the different factors you’re analyzing.**



**ANOVA assumes that data is normally distributed, that the variances for each factor are equal and that the factors are independent.** You can check these assumptions by looking at the residuals of the data, which are the differences between the observed value and predicted value based on your fitted model. You can save your residuals (Save > Save Residuals in the red triangle):



You can then analyze the distribution of your residuals (Analyze > Distribution) to check normality. You can also graph your residuals vs. X. This allows us to check that the variances are all equal (the residuals should distribute evenly vertically over the different factors). Lastly, if you graph your residuals vs. the predicted/fitted Ys (saved by selecting Save > Save Predicted, seen in the image above), you can check independence to make sure that the residuals do not change based on different expected values. You also could graph your residuals vs. the order that they were taken in or time to make sure the order of the measurements did not affect the data.

## Lab 9 Exercises

The Kenton Food Company wished to test four different package designs for a new breakfast cereal. Twenty stores, with approximately equal sales volumes, were selected as the experimental unit. Each store was randomly assigned one of the package designs, with each package design assigned to five stores. The stores were chosen to be comparable in location and sales volume. Other relevant conditions that could affect sales, such as price, amount and location of shelf space, and special promotional efforts were kept the same for all of the stores in the experiment. During the trial period, one of the stores had a fire and was closed for a day. That store's result was removed from the data set.

Using JMP, for the given data uploaded to Blackboard,

1. State H0 and H1 formally and describe what they mean in your own words.
2. State any assumptions about the data.
3. Run an ANOVA to determine if you can reject H0 or not.
4. Can you say anything about the individual cereal box designs from the JMP output?
5. Test your assumptions using the residuals.