# Week 2 Assignment 3 – Experimental Design

By: Zach Adair

Regis University

MSDS – 650 – Data Analytics

## Factorial Design

Factorial design is used when researchers want to understand the effect of two or more independent variables upon a single dependent variable. For this R tutorial, it walks us through an example where a fast food franchise is testing 3 new menu items on the East and West Coasts and the test is designed to measure the popularity of these three menu items. By design, 12 restaurants are tested on each coast for a total of 24 total scores.

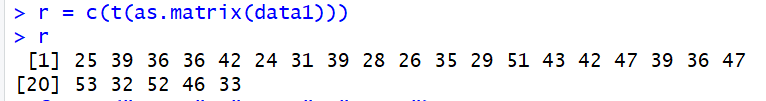
The problem faced here is we need to test if the mean sales volumes for the new items are all equal and then decide whether the mean sales volumes of the two coastal regions differ. From the construction of the problem it looks to be doing a hypothesis test, the null hypothesis being the mean sales volume of the new items are all equal.

To start we needed to read in the code through a csv, I created a .csv file with the information given to me by the tutorial and we want to read it into RStudio and assign the data to a data1 variable:



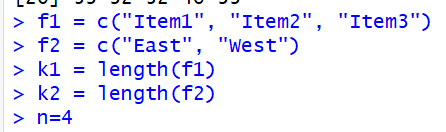
Now that I have my data read into RStudio, I can move on with the assignment.

Next, I’m going to concatenate my newly created data1 into a single vector named r:

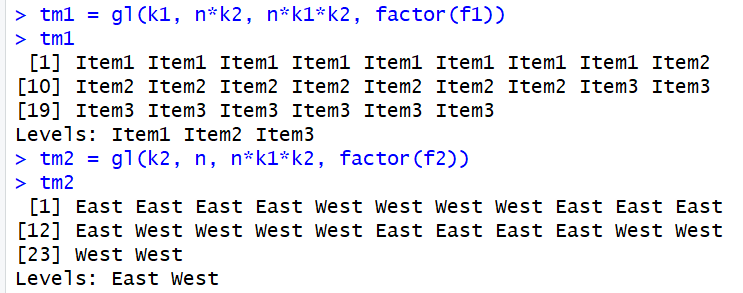


To create r, data1 needed to be converted into a matrix using the as.matrix command.

Now I will assign variables which will name treatment levels as well as number of records. The variables we are going to assign are f1, f2, k1, k2, and n.



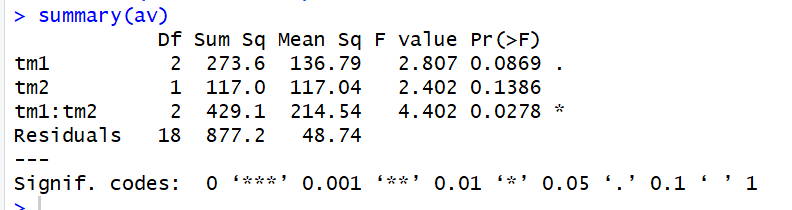
Now that we have created these variables, we are going to use these variables to construct the 1st and 2nd treatment level of response element by element:



After creating the tm1 and tm2 vectors it is time to use the aov formula to help describe the response of r by tm1 and tm2 with interaction.



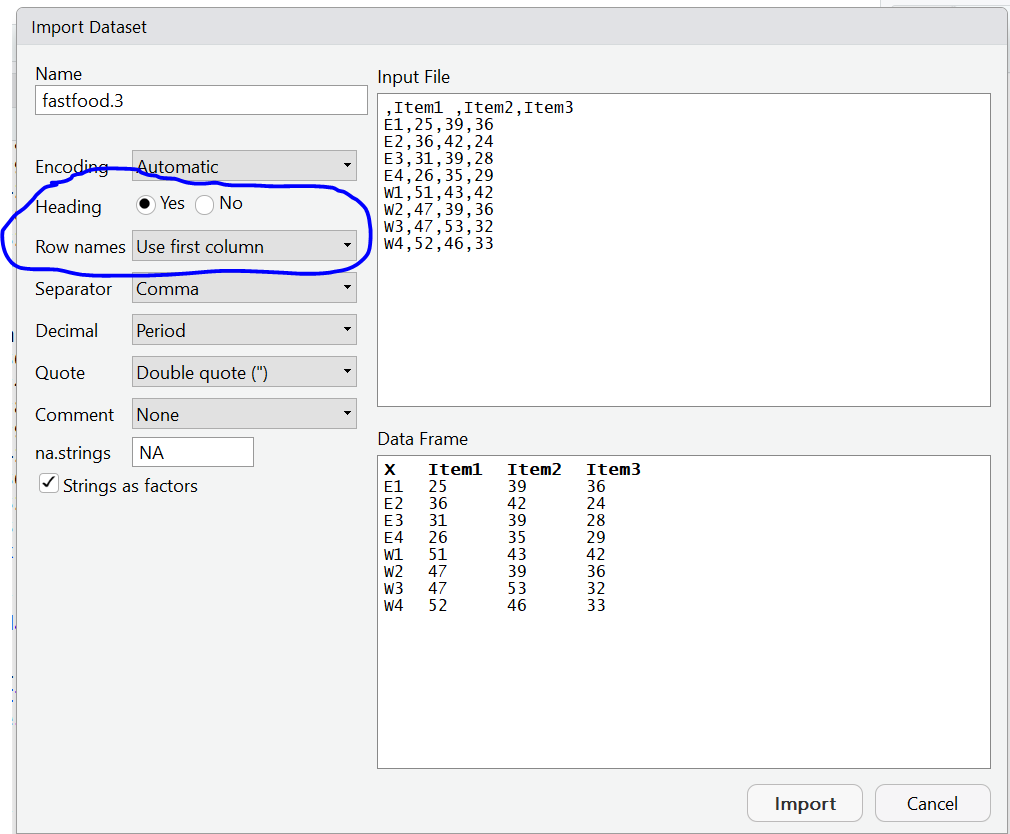
The formula av has been created and now what is left is to print out the ANOVA table summary:

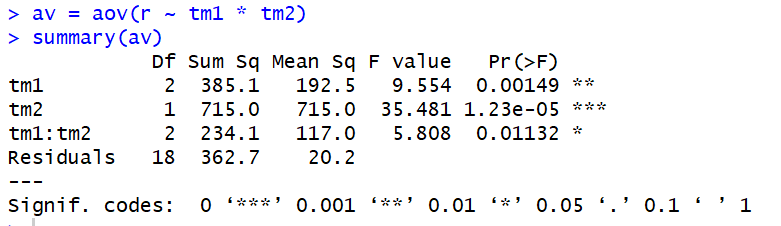


For some reason I went wrong here because the tutorial says I should have a tm1 p-value of 0.0015 which would cause me to reject the null hypothesis, but I got a 0.0869 which means I would accept the null hypothesis. This happens again for tm2, and tm1 : tm2.

From my data I would conclude that we do not reject the null hypothesis and state that the mean sales volume for the new items are all equal. I will do some further investigation though to try and find out how I got a different answer before moving forward on my assignment.

I resolved my issue and now have the same result as the tutorial! My issue was with how I got my data into RStudio. First it was reading my columns as headings, and 2nd I got rid of my row headings and that was somehow throwing off my data. So I use the import option and made sure I called out my headers in my columns and rows, ran through the process (data2 this time) and ended up with the same result as the tutorial.





From the expected result, with the p-values all being less than 0.05, we can deduce that we should reject the null hypothesis and that the mean sales volume of the new items are not all equal.

## Bonus Questions for the assignment

1. What does an Analysis of Variance tell you? What types of questions does it answer?

Analysis of Variance (ANOVA) tells the experimenter what are the potential differences between dependent variables having 2 or more categories. The types of questions that ANOVA can provide an answer to are ones only with nominal level variables yet have two or more categories. For example, in our test above we had 3 variables we were testing which were the three new menu items and we were trying to measure the difference in their mean sales volume between two variables East and West. ANOVA works for this kind of problem because to see if we can solve the null hypothesis of the mean sales are the same between East and West, we would need to figure out the difference in their scores for all 3 menu items. ANOVA can be used for more than just 2 or 3 categories as well, ANOVA is actually useful to the Nth degree because you can always measure the difference (variance) between the variables in question.

1. What then is the significance of experimental design?

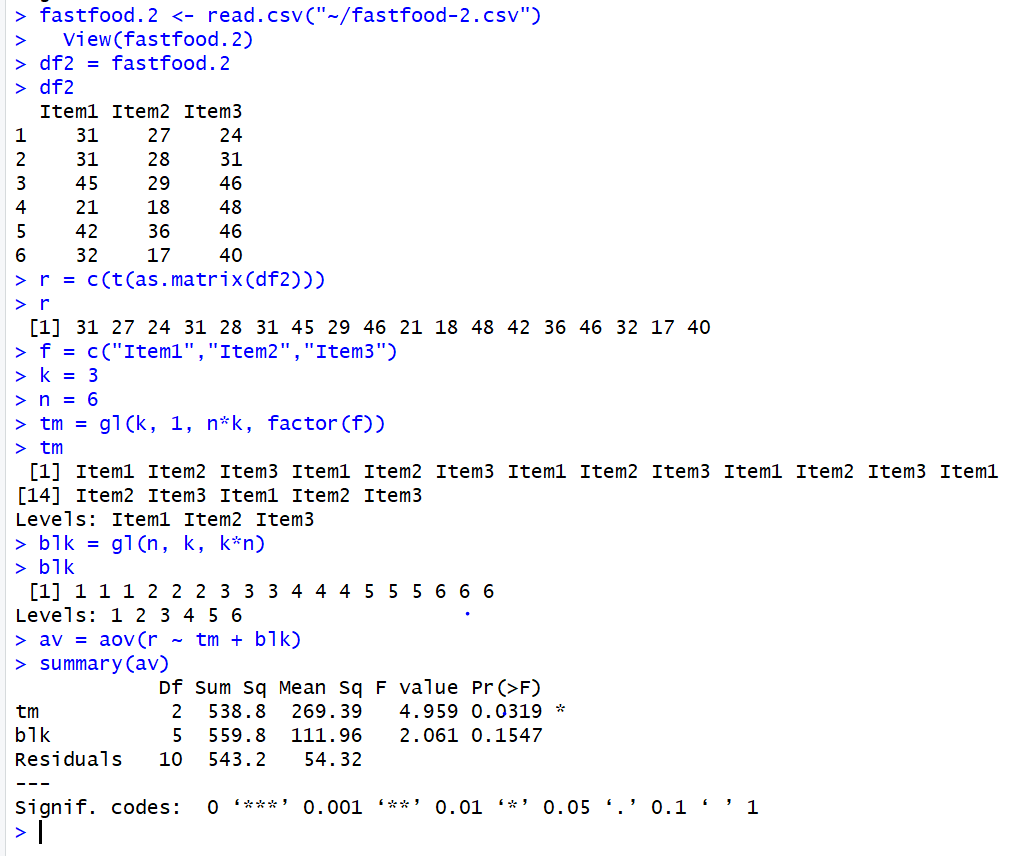
The importance of experimental design is that experimental design can help a researcher create a road map to hit whatever objectives they want an experiment to test and problem they want it to solve. Proper planning ensures the right data is used, the sample size is adequate, the proper research questions are being asked, the proper assumptions are being made, and the correct tests are looking at the data. Without being able to plan that then it makes it difficult to come out with reliable information which is also clear and done efficiently for the researchers own self and possible end user.

## Next Tutorials

I know this isn’t part of the assignment, but I wanted to just document the steps taken in the other two tutorials.

### Randomized Block Design

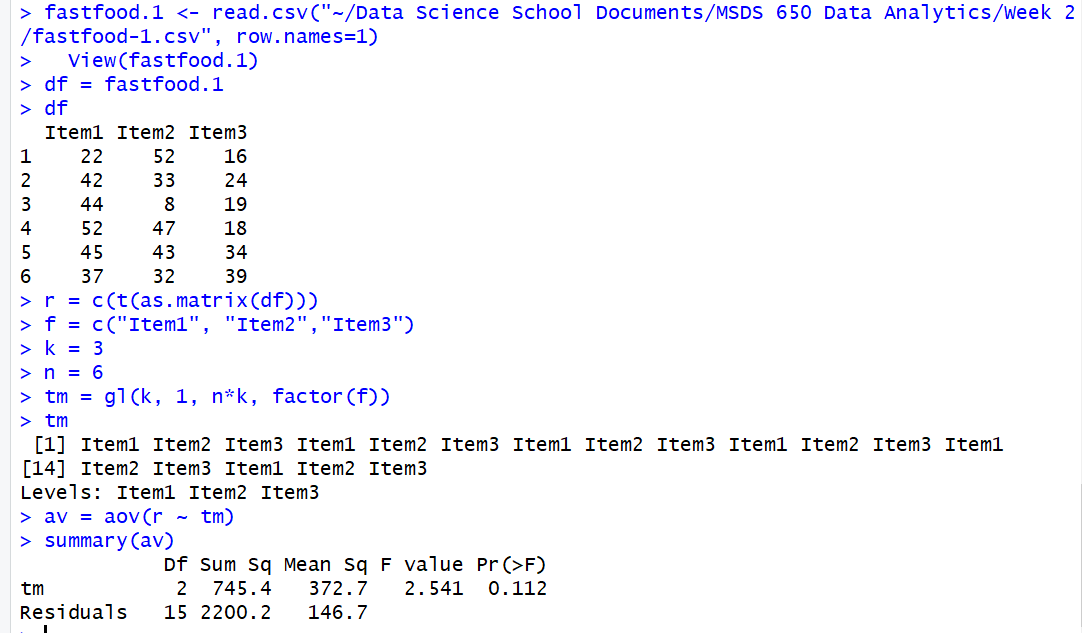
Problem: Each row in the following table represents the sales figures of the 3 new menu in a restraurant after a week of testing marketing. At 0.05 level of significance, test whether the mean sales volume for 3 new menu items are equal.



Conclusion: With the p-value being less than 0.05, we would reject the null hypothesis which stated that the mean sales volume for the 3 new menu items is equal.

## Completely Randomized Design

Problem: At 0.05 level of significance, test whether the mean sales volume for the 3 new menu items are all equal.



Conclusion: The p-value = 0.112 is greater than 0.05, thus we would not reject the null hypothesis stating the mean sales volume of the new menu items are all equal.