

## Teachout Math 140 Project: ANOVA Randomization Hypothesis Test

Plagiarism Warning: *Do not copy an online article or another math 140 student's work. This is plagiarism. Cheating will go on your permanent record. Further discipline will be determined by the COC dean of students. A plagiarism checker may be used if I suspect cheating. I have also had problems with students simply copying online ANOVA articles as their project. Do not do this either. This is also plagiarism. Your project answers should be intro level and use the given data and the simple straight forward language used during lectures and in the textbook. Online articles are usually written by people in advanced statistics and will not answer these questions appropriately.*

**Grading Rubric:** The project is worth 100 points. There are 25 things to put on your project listed below in bold as "Put these on the project report". Some of them are pictures found on StatKey. Others are answers to questions or sentences to write. Each of these is worth 4% of the project grade. (4 points each)

Key Note about Assumptions (Conditions): *This is a Census of students in the Fall 2015 semester. Even though it is not a random sample, we will be assuming that the sample data represents the population of COC statistics students. We will also be assuming that individual observations from this data are independent of each other, even though in reality this might not be the case.*

PART 1: Pick your two columns of data.

Open the Math 075/140 Combined Survey Data Fall 2015. [Data Link](#).

*This data was taken from statistics students (math 140) and pre-stat students (math 075) in the Fall 2015 semester.*

*There are 36 columns of data to choose from. Pick one categorical data set (column of words) and one quantitative data set (numerical measurement data) from the 075/140 combined survey data.*

**Put these on the Project Report:**

- **What categorical variable did you pick?** For example: The type of sandwich left out to spoil.
- **What quantitative variable did you pick?** For example: The number of ants

PART 2: Write the Null Hypothesis, Alternative Hypothesis and chose your Claim

- *Test the claim that there is a relationship between your categorical variable and your quantitative variable. Use the following null and alternative hypothesis. The number of options in your categorical data will determine the number of groups in your ANOVA test. If your categorical data has two options, then the null will be  $\mu_1 = \mu_2$ . If your data has three options, then the null will be  $\mu_1 = \mu_2 = \mu_3$ . If your categorical data has 6 options then your null will be  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$ . Here is an example null and alternative hypothesis. Your null and alternative hypothesis should have symbolic notation with " $\mu$ " and also the relationship statement implication. Do not just copy this null and alternative below. You should not say that the categorical and quantitative variables are related, you should say that the type of sandwich is related to the number of ants. If those were the variables you picked. Also, the claim could be either statement. Which do you think is true. Do you think the two columns of data you picked are related or not? That is the claim. Be sure to label which statement is the claim. This is a right tailed test. (Remember all ANOVA tests are right tailed.)*

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots$  (The type of sandwich is NOT RELATED to and the number of ants.)

$H_A : \text{at least one is } \neq$  (The type of sandwich is RELATED to and the number of ants.) CLAIM

**Put these on the Project Report:**

- **Write your null and alternative hypothesis as seen in the example above.** Be sure to include the symbolic notation with “ $\mu$ ” AND the relationship statement with the two variables you picked.
- **Is your claim that the categorical and quantitative data is related (Ha) or not related (Ho)?**
- **Is this a right-tailed, left-tailed, or two-tailed test?** Put this Answer: “ANOVA tests are all right-tailed.”

**PART 3: Paste your Data into StatKey and Find your F-test statistic**

Copy and paste your categorical column of data and the quantitative column of data next to each other in excel. The categorical column should be on the left. The quantitative column should be on the right. Now highlight both columns without the titles, right click and copy. Do NOT copy the titles.

Go to the “ANOVA for Difference in Means” under the “More Advanced Randomization Tests” menu. [StatKey Link](#).

Click on “Edit Data” and Copy and Paste the two columns of data (categorical on left and quantitative on the right) into StatKey. Do NOT paste the titles. If you do paste the data with the titles, delete the titles. Uncheck the box for “header row” and push OK.

**Put these on the Project Report:**

- **Copy and Paste the “Original Sample Statistics” ANOVA printout in the top right of the StatKey page into your Project report. It should show the F-test statistic from your data, and the sample size, mean and standard deviations for all your groups.** It should look like this but the numbers will be different. Do NOT copy and paste the “randomization sample” by mistake. Only copy the one that says “original sample”. The F from “original sample” is your one and only F test statistic for your data.

**Original Sample**

ANOVA Table

$n = 24, F = 5.627$

| Statistics         | Vegemite | Peanut Butter | Ham & Pickles | Overall |
|--------------------|----------|---------------|---------------|---------|
| Sample Size        | 8        | 8             | 8             | 24      |
| Mean               | 30.8     | 34.0          | 49.3          | 38.0    |
| Standard Deviation | 9.3      | 14.6          | 10.8          | 13.9    |

- **Assumption#1&2: Is the sample data either random or represents the population and are the individuals independent.** Put this Answer: “The data is not random, but we are assuming the data represents the population and that individuals are independent.”
- **Assumption#3: In the original sample printout, are all of the sample sizes at least 30?** “Yes, all of the sample sizes are 30 or higher.” OR “No. At least one of the sample sizes is below 30.” In the above Original Sample table, all of the sample sizes were 8 so the ant/sandwich data would not pass this assumption.
- **Assumption#4: In the original sample printout, there should not be any sample standard deviations that are more than twice as large as any other sample standard deviation. Is this the case for your data?** “Yes, the sample standard deviations are close.” OR “No, there is at least one sample standard deviation that is more than twice as large as the others.” In the above Original Sample printout, all of the sample standard deviations (9.3, 14.6, 10.8, 13.9) are all close. So this would pass the assumption.
- **What is the F-test statistic from your data?** (This is listed in the Original Sample printout. For example, in the above printout the F-test statistic was 5.627)
- **Write a sentence explaining the F-test statistic.** Here is an example: “The ratio of the variance between the groups to the variance within the groups is 5.627.”

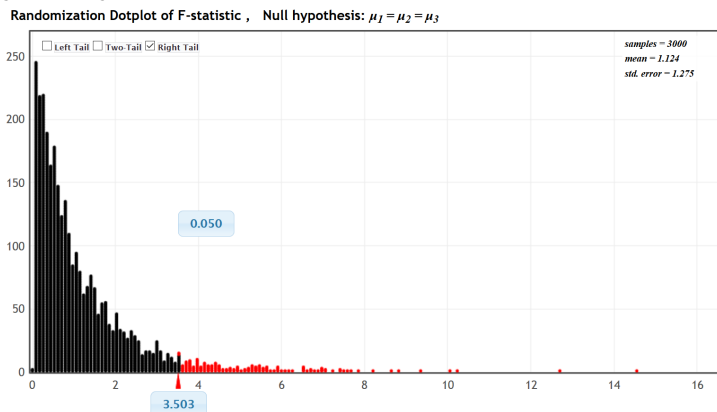
#### PART 4: Create a Randomized Simulation and Calculate the Critical Value

You will now be creating two simulated F-distributions. One to calculate the critical value and tail. Another to calculate the P-value. Click on "Generate 1000 Samples" a few times to create the simulated F-distribution.

Directions for calculating the Critical Value and tail. Click on "Right-Tail". Change the right-tail proportion (upper box in distribution) to 0.05. This corresponds to a 5% significance level. The number at the bottom is the Critical Value and the start of the right tail.

#### Put these on the Project Report:

- **Copy and paste the simulated F-distribution that has 0.05 in the right tail and the Critical Value into your Project report.** It should look like this but the critical value at the bottom will be different.



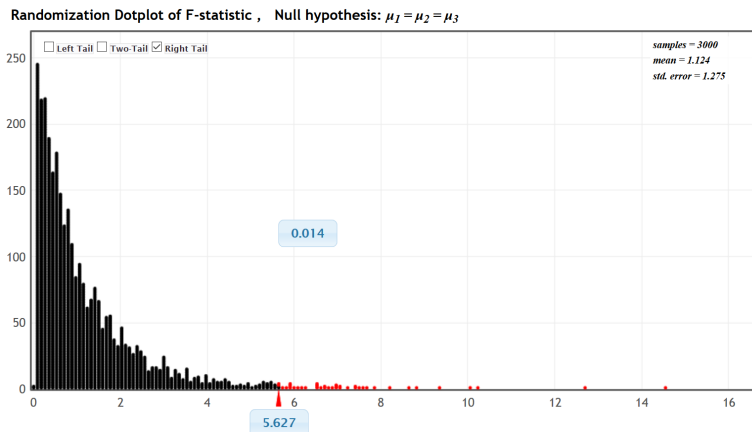
- **What is the critical value corresponding to 0.05 in the tail?** (In the example above the Critical Value is 3.503)
- **Does the real F-test statistic listed on "Original Sample Statistics" fall in the right tail determined by the Critical Value or does the F-test statistic NOT fall in the right tail?** In the ants/sandwich example, the test statistic was 5.627, so this would fall in the tail that starts at Critical Value of 3.503.
- **Does the sample data significantly disagree with the null hypothesis (F-test stat in tail) or does the sample data not significantly disagree with the null hypothesis (F-test stat not in tail)?**
- **Is the variance between the groups significantly higher than the variance within the groups (F-test stat in tail) or is the variance between not significantly different than the variance within (F-test stat not in tail)?**

## PART 5: Use the Randomized Simulation to Calculate the P-Value

Directions for calculating the P-value. Use the same distribution you have already created. Click on "Right-Tail". On the lower box where your Critical Value was, type in the actual F-test statistic listed under "Original Sample Statistics". Once you put in the F-test statistic into the bottom box in the right tail, the upper box is the P-value.

Put these on the Project Report:

- **Copy and paste the simulated F-distribution that has the F-test statistic in the lower box of the right tail and the P-value in the upper box of your right tail into your Project report.** It should look like this but the test stat and P-value will be different.



- **What is your P-value? Write it as a proportion and a percentage.** For example, in the simulation above the P-value is 0.014 or 1.4%.
- **Write a sentence to explain your P-value.** For the example above: "If the null hypothesis is true and the variables are not related, then there is a 1.4% probability of getting this sample data or more extreme by sampling variability."
- **Is your P-value lower or higher than your 5% significance level?** "The P-value is lower than the 5% significance level." OR "The P-value is higher than the 5% significance level". (Note: If a P-value was exactly 5%, that would be considered a low P-value.)
- **If the null hypothesis was true, could the sample data have occurred because of sampling variability (High P-value) or is it unlikely to be sampling variability (Low P-value)?**
- **Assuming the sample data met the assumptions, would your P-value be considered significant evidence (Low P-value) or not significant evidence (High P-value)?**

## PART 6: Interpretations and Final Conclusion

- **Should we reject the null hypothesis (Low P-value) or fail to reject the null hypothesis (High P-value).**
- **Write the standard conclusion sentence in context for your test addressing evidence and the claim that your categorical and quantitative variable are related or not related.** Here is an example though the variables will be different.

Low P-value with Passing ALL Assumptions: "There is significant evidence to (reject or support) the claim that the type of sandwich and the number of ants are (related or not related)."

Low P-value but One or More Assumptions did NOT pass: "There is NOT significant evidence to (reject or support) the claim that the type of sandwich and the number of ants are (related or not related)."

High P-value: "There is NOT significant evidence to (reject or support) the claim that the type of sandwich and the number of ants are (related or not related)."

- **Does the data indicate that your categorical data is related or NOT related?** Here are some examples though the variables will be different.

High P-value: "Data indicates that the type of sandwich is not related to the number of ants, however we do not have evidence."

Low P-value with Passing Assumptions: "The data gives evidence that the type of sandwich is related to the number of ants".

Low P-value but One or More Assumptions did NOT pass: "The data indicates that the type of sandwich is related to the number of ants. However, since this data did not pass the assumptions, we do not have evidence."