Math 140 ANOVA Project
Categorical / Quantitative Analysis

Data Directions:

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.

Part 1: Categorical Data Analysis

Directions:

Go to "One Categorical Variable" under the "Descriptive Statistics and Graphs" menu in StatKey. StatKey Link.

Click on "Edit Data" and Copy and Paste the categorical data into StatKey. Click on "raw data" and "header row" and push ok.

On Project Report:

- **Copy and paste the Bar Chart and the Summary Statistics (counts and proportions) into your project report.** It should look like this but the numbers will be different.
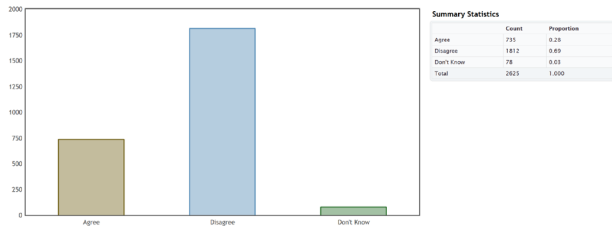
Part 2. One-Population Proportion Bootstrap Confidence Interval

Directions:

In your categorical summary statistics you have multiple counts and a total. Pick one categorical variable to make a proportion confidence interval for. Use just the count for that one variable and the total.

Go to "CI for Single Proportion" under the "Bootstrap Confidence Interval" menu in StatKey. StatKey Link.

Click on "Edit Data". Put in the count and total you selected. Push OK.

On Project Report:

- **Copy and Paste the "Original Sample" printout (count, sample size, proportion) on the top right of the page into your project report.** It should look like this. Do not use the one that says "randomization sample". Only give the one that says "original sample".

Original Sample

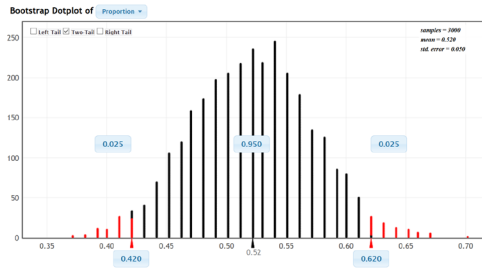| Count	Sample Size	Proportion
52	100	0.520

Click "Generate 1000 Samples" a few times to make the bootstrap distribution. Now click on "Two-Tail" to make the 95% confidence interval.

On Project Report:

- **Copy and Paste your 95% two-tailed Bootstrap distribution into your project report.** It should look like this, though the numbers at the bottom will be different.

- **Give the lower limit of the confidence interval on the lower box in the left tail of your bootstrap. Write it as both a proportion and percentage.**
- **Give the upper limit of the confidence interval on the lower box in the right tail of your bootstrap. Write it as both a proportion and percentage.**
- **Write a sentence in context explaining what the bootstrap distribution tells you about the population percentage for your variable.**


Part 3. *Quantitative Data Analysis*

Directions:

Go to "One Quantitative Variable" under the "Descriptive Statistics and Graphs" menu in StatKey. *StatKey Link.*

Click on "Edit Data" and Copy and Paste the quantitative data into StatKey. Do not paste the title. If you do paste the data with the title, delete the title. Uncheck the boxes for "identifier" and "header row" and push ok.

On Project Report:

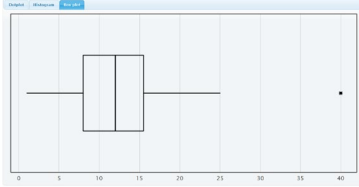- **Copy and paste the Histogram.** It should look like this but the shape may be different.
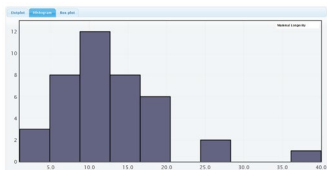
- **Copy and paste the Dot Plot.** It should look like this but the shape may be different.

- **Copy and paste the Box Plot.** It should look like this but the box and outliers will be different.

- **Copy and paste the Summary Statistics (Sample Size, Mean, Standard Deviation, Min, Q1, Median, Q3 and Max) on the right side of the page into your project report.** It should look like this but the numbers will be different.

Summary Statistics

Statistic	Value
Sample Size	40
Mean	13.150
Standard Deviation	7.245
Minimum	1
Q_1	8.000
Median	12.000
Q_3	15.500
Maximum	40

- **Calculate the Interquartile range (IQR) = Q3 – Q1.**
- **What is the shape of the data?**
- **Which average is most accurate, the mean or the median?**
- **Which spread is most accurate, the standard deviation or the IQR?**
- **List ALL of the outliers listed on the Box Plot.**

Part 4. *One-population mean or median bootstrap confidence interval*

Directions:

In your quantitative summary statistics you picked either the mean or the median as your best average. If you picked the mean as your average, then you will be making a bootstrap confidence interval for the mean average. If you picked the median as your average, then you will be making a bootstrap confidence interval for the median average.

Go to "CI for Single Mean, Median, Std Dev" under the "Bootstrap Confidence Interval" menu in StatKey. StatKey Link.

Click on "Edit Data" and Copy and Paste the quantitative data into StatKey. Do not paste the title. Or if you do paste the data with the title, delete the title. Uncheck the boxes for "identifier" and "header row" and push ok.

One Project Report:

- **Copy and paste the "Original Sample" printout (dot plot, sample size, mean, median, standard deviation) on the top right of the page into your project report.** It should look like this but your numbers will be different. Do not past the "bootstrap sample" by mistake. Only the one that says "original sample".

Original Sample
n = 25, mean = 15.98
median = 11.9, stdev = 11.114
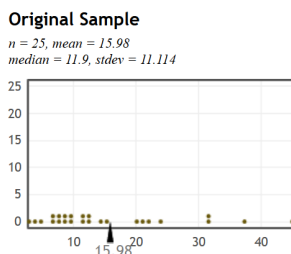
Under "Bootstrap Dot Plot of", click on "Mean" or "Median". Do <u>not</u> do both.

Click "Generate 1000 Samples" a few times to make the bootstrap distribution. Now click on "Two-Tail" to make the 95% confidence interval.

On Project Report:

- **Copy and Paste your 95% two-tailed Bootstrap distribution into your project report.** It should look like one of these but the numbers at the bottom will be different. Do not put both mean and median bootstraps. Only one.
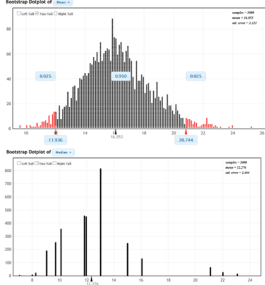
- **Give the lower limit of the confidence interval in the lower box in the left tail of your bootstrap distribution.**
- **Give the upper limit of the confidence interval in the lower box in the right tail of your bootstrap distribution.**
- **Write a sentence in context explaining what the bootstrap distribution tells you about the population mean or population median for your variable.**

Part 5: ANOVA Randomization Hypothesis Test

Directions:

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 $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \ldots$ *(The categorical and quantitative variables are not related.)*

$H_A :$ *at least one is ≠ (The categorical and quantitative variables are related.) CLAIM*

On Project Report:

- **Write your null and alternative hypothesis. Is your claim that the categorical and quantitative data is related (Ha) or not related (Ho).**

Copy and paste your categorical column of data and the quantitative column of data next to each other in either excel or pages. The categorical column should be on the left.

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. <u>Uncheck</u> the box for "header row" and push ok.

On Project Report:

- **Copy and Paste the "Original Sample Statistics" ANOVA printout in the top right of the 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 past the "randomization sample" by mistake. Only the one that says "original sample".

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

- **Check the following assumptions.**
  Random Sample or Representative? Since this data was a census from 1 semester, you can assume it is representative of all stat students at COC in all semesters.
  Sample sizes for each group at least 30? See the original sample statistics printout.
  Are individual stat students independent of each other?
  Are the sample standard deviations for each group close? See the original sample statistics printout. No standard deviation for 1 group should be more than twice as large as the standard deviations for any other group.
- **Give the F-test statistic from your data.**

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.

On Project Report:

- **Copy and past 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.
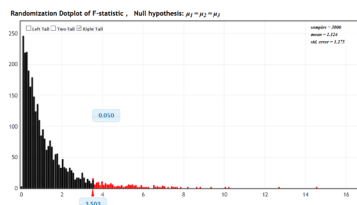
- **Does the real F-test statistic listed on "Original Sample Statistics" fall in the tail determined by the 5% significance level and the Critical Value?**
- **Does the sample data significantly disagree with the null hypothesis or not significantly disagree?**

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.

On Project Report:

- **Copy and past 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.
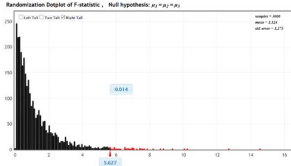
- **What is your P-value? Write it as a proportion and a percentage.**
- **Is your P-value lower or higher than your 5% significance level?**
- **If the null hypothesis was true, could the sample data have occurred because of sampling variability (random chance) or is it unlikely to be sampling variability?**
- **Should we reject the null hypothesis (low P-value) or fail to reject the null hypothesis (high 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)?**
- **Write the standard conclusion sentence in context for your test addressing evidence and the claim that your categorical and quantitative variable are related.**