

## Stats Project #2

### Due Thursday, April 19

Spring 2018

For Project #2, you are to choose *one* of the following two options. As before, you can work alone or in teams of two or three people (from either of my two stats classes). Your report should be typed and professionally organized (*i.e.*, should look nice), and you should use Minitab to do the statistical analysis required.

#### Option #1. Bivariate data analysis

The purpose of this project is to test for a linear correlation between two sets of quantitative data. Your team needs to obtain bivariate quantitative data from at least 30 experimental units. Be sure that you are taking numerical data from larger populations, not actually comparing populations (e.g., looking for a correlation between winning percentage and year in the history of the Cleveland Browns). Let me know if you need ideas for a topic. Your team must devise a sampling plan, carry out the plan, and use a statistical program to carry out the analysis. In your report, you must include the following.

- (a) What are the two variables that you are studying? Which is the explanatory variable and which is the response variable? Why are you interested in this topic? Before looking at the data, what would you guess the strength and direction of the correlation are?
- (b) What was your sampling plan? Talk about the pros and cons of your plan. Did you find it difficult to carry out?
- (c) Using whatever graphs/tables you see fit, summarize your data. You should have at least one graph for both variables measured and a scatterplot. Be sure to talk about shape, center, spread, and any unusual observations.
- (d) Use technology to find the equation of the regression line and the correlation coefficient. Interpret both the correlation coefficient and the slope of your regression line. Be specific!
- (e) Use the regression line to make a specific prediction; e.g., an  $x$ -value of \_\_\_\_ is predicted to have a  $y$ -value of \_\_\_\_\_. Find the residual for one of your data points.
- (f) In the final paragraph, summarize your results in everyday language.

## Option #2. Comparing two population means

The purpose of this project is to compare two population means using the appropriate  $t$ -test. Your team needs to obtain a representative sample of at least 15 data points from both populations that are being compared. Be sure that you are taking samples from a larger population, not actually comparing populations (e.g. average score/game for the 2016 Browns versus the 2015 Browns). Let me know if you need ideas for a topic. Your team must devise a sampling plan, carry out the plan, and use a statistical program to carry out the analysis. In your report, you must include the following.

- (a) What two population means are you comparing? Why are you interested in this topic?
- (b) What was your sampling plan? Talk about the pros and cons of your plan. Did you find it difficult to carry out?
- (c) Using whatever graphs/tables you see fit, summarize your data. You should have at least one graph for both variables measured and a graph comparing both variables side-by-side. Be sure to talk about shape, center, spread, and any unusual observations. Based on these graphs, would you guess that your two population means are different?
- (d) Use the appropriate  $t$ -test (*i.e.*, paired or unpaired) to test for a difference in your two population means. Be sure to include the null and alternative hypotheses and the  $p$ -value. Be specific in your conclusion! What assumptions are you making when using a  $t$ -distribution here?
- (e) Use Minitab to find a 95% confidence interval for the difference in your two population means. Interpret this interval in the context of your problem. (**Hint:** Your interval should agree with the conclusion in part (d).)
- (f) In the final paragraph, summarize your results in everyday language.