**Project 1**

**This will be due as indicated on the syllabus.**

Regression Project:

You are hired by Rochester Regional Health to research the attributes contributing to Obesity in America. They have provided you with a dataset that they sourced from the USDA which gives obesity rates and measures of factors that may contribute to obesity. They will use your work to identify partners and regions to connect with throughout the US. The dataset provided includes information about ALL counties in the United States.

The data file and file with the field definitions are in my courses.

Initial approach:

1. Review the project requirements and discuss the approach and review the data with your team.
2. Do some research on Obesity in America for background information (Know your data)
3. Do some initial analysis on the data (graphs, histograms, etc.) (Bottom up approach to explore fields and learn about the data). Use your summary analysis to identify variables and relationships that you want to explore further with your regression analysis.
4. Use what you’ve learned in your regression analysis.

Regression:

Analyze the data file and determine which independent variables you want to include in your model. Create an actionable regression model for Obesity rate. It should explain a lot of the variability but only identify the most important factors. Incorporate what you’ve learned in your initial approach before you start using regression. This will provide ideas on what to consider and can help validate your findings.

Hand in a report that contains:

* An executive summary of Obesity in America, *which can be read by non-statistical nurses, doctors or administrators*. This should be a short executive summary, a page or two long.
* Identify any highly correlated predictors that you have in your data and decide if you will include them in your analysis. Include your reasoning why you made the choice(s) in your analysis.
* The main body of your report should summarize the types of analysis you did to pick your independent variables and arrive at your conclusions. In this section of the report you should also show that your model is valid and that you’ve tested for common problems with regression and how you’ve done that. You can use statistical references and details here.
* Determine the best way you could convey your findings to your audience.
* Include a summary explaining your regression model means and how well it predicts.
* Explain your findings!
* Create a power point (or any other type of presentation) that one or more of your team members will present. Usually teams have everyone present a part, but it’s up to you. This should be 15 – 20 minutes. Some teams have gone longer if they want to get deeper into explanations.
  + Remember to focus your presentation as if you are addressing your client.
  + All other groups can ask questions and ask why you chose the methods and visualizations that you used.
  + We’ll discuss this more in class. Most groups prepare a power point when they do their presentations.

Answer these questions in the body of your report too:

1. Which states of the country are most prone to obesity?
2. Conduct hypothesis tests to determine if the following lead to **more** Obesity (Column BM). Use alpha of 5%. For each, you will need to break the data for these variables into two sets (HIGH/LOW) at the mean. Minitab’s recode function can help with this or you can do it in excel.
   1. Fast food restaurants per 1000 people (Column W)
   2. Smoking (%smokers) (Column G)
   3. # low income & >1 mile to store (Column F)
3. Create a 4 level variable for exercise (% who do not exercise Column BN) and conduct an ANOVA comparing the four groups. Use excel or recode in Minitab to break the exercise variable into 4 parts (quartiles) and compare the obesity rates of the different groups.
4. Conduct a correlation analysis to see which factors are most highly correlated with Obesity Rate (and which are correlated with each other!).
5. Include the summary output from Minitab, or whatever software you use to analyze your data, explain what the key points from that output tell you.

If you have a problem with any of your team members you can confidentially let me know. You will also have the opportunity to rate each of your team members at the end of the projects.

Remember too, learn what skills each of your teammates has and leverage those to the team’s benefit when working on the project. Normally teams have every team member present some material during the presenatation.