**AP STATISTICS PROJECT #2:**

Scatter Plots and Regressions

**PART 1: Find or Produce Data**

Find or produce some quantitative data for two variables that you want to determine if there is a correlation between. For example, you could ask Moses Brown students how many minutes it takes them to get to school versus how many DSP’s they had last year. Or you could go online and find a data set where data has already been produced for you (check out my resources online or just do some research on Google). It is pretty easy to find some data on a topic of your choice. For example you could look at local sports data on the Providence Journal website. As a soccer coach, I might look at the box scores from this season’s MB games. Does the number of shots a team takes correlate to the number of goals they score in a game? You could even look at the 2d:4d ratio of participants and compare it to a measure of their self professed competitiveness?

Note\*\* if you decided to collect data with a survey, you don’t need to go too crazy with how many participants you collect data from (around 10-20 is sufficient). We will be doing plenty of projects later where the method by which you collect your data and the number of participants will be important. But for now, simply collect data and you can always comment on the bias of your sample in your conclusion. For this project you only need to ask to look at ONE set of data to make a scatter plot of.

**PART 2: Analyze your Data/Make graphs**

Using the tools that we have learned from Chapter 3 on scatter plots and regressions, analyze your data. Make a scatter plot with a regression line to display your data. Choose graphs appropriately that display your data effectively (i.e. do you have a categorical variable that you want to display on the scatter plot with colors?)



Also calculate a numerical summary like the mean and standard deviation for each of the variables separately. Make sure you also find the equation of the regression line and the r-squared value.

You will be using Minitab or some other statistical software to make these plots, which will be part of your final product. For this project, excel typically works just fine too.

Residual Plot:Also make a plot of the residuals, which you will include in your summary.



**PART 3: WRITING AND INTERPRETING YOUR FINDINGS**

Write a paper where you present your findings. Your paper should be concise (as you are only looking at one scatter plot) and should include the following:

Introduction: (1 paragraph) Explain what question you are examining and why you are interested in that particular question. Provide any relevant background information that you think could be relevant to a reader and things that could “hook” the reader into why your question is relevant. Also include your hypothesis. What did you expect to find before conducting the study?

Method: Explain briefly how (or from where) you collected your data.

NOTE: Your intro and method should address the Who? What? Why? Where? When? How? By Whom?

Results:Display the scatter plot of your findings (with the regression line) and any numerical summaries that you conducted. Be sure to label axes and title graphs. You don’t need to interpret your results in this section as you will do so in your conclusion. PLEASE INCLUDE YOUR RESIDUAL PLOT EVEN THOUGH WE TYPICALLY WOULDN’T IN A NORMAL STUDY.

Conclusion:Write an analysis of your findings. Describe the form, strength and direction of the data that you collected? What stood out to you? Were there any problems in the way that you collected the data? What are the implications of your findings? In addition, use your trend line to extrapolate or interpolate (i.e. pick a value of the explanatory variable?

*NOTE: I will also provide you with a grading rubric to help you with the writing of this paper.*