PROJECT-SET
Statistics Education for Teachers

Regression Culminating Activity

In this project, you will perform a hands-on investigation and explore relationships between two variables. You will use real data that you will download from the Census and School project website. Note that this is real data and thus it is not cleaned in any way. Therefore, you will have to examine it to make sure that everything makes sense. If it does not, you will then have to decide the best way to deal with the issue.

On the Census at School website, examine the questionnaire that is given to participants. Pose a question about the relationship between two quantitative variables that you would be interested in investigating.

To answer the question you choose, you are going to download and work with a real data set. To download the data set, go to the following website:

http://www.amstat.org/censusatschool/

To download the data, do the following:

1. Click on Random Sampler

2. Accept the Terms & Conditions

3. Select a sample size of 100 from All States and 9, 10, 11, and 12 grade levels. Include All Genders and All Years of data collection.

4. Download the data into Excel.

5. Open the data in Excel. You will see a large number of variables (labeled in each column).

6. Delete the columns except for the ones you are interested in. Bring your data into Fathom (or another technology that will allow you to estimate the regression line).

7. Plot a scatterplot to visually see the data. Does the relationship appear to be linear? Are there any outliers? What are some possible explanations for why there could be outliers? Should you eliminate the outliers in your data set? Why or why not? Is there any data clean up that needs to take place? Explain.

8. Write the model (population equation) of the relationship between your variables. This is the relationship you are going to estimate.

9. Estimate the least squares regression line for your downloaded data.

10. Interpret the slope (β) in the context of the question.

11. Interpret the y-intercept (α) in the context of the question.

12. Interpret the correlation coefficient, *r*, in the context of the question.

13. What conclusions can you draw about the population regression model given the sample estimated regression equation that you found in this project? How can you use what you have learned about sampling variability to make inferences about the population on the basis of your sample regression equation?

Regression Wrap-Up

14. Has your participation in the course changed the way that you think about the teaching of any aspects of regression?

a) If "yes" – What has changed in your thinking? How will that influence your teaching practice? What will you do differently as a result?
b) If "no" – How did you approach the teaching of regression previously? How is that aligned with what we did in the class and read for the course?