

Problem 7. Mark is shopping for a diamond and is interested in learning more about how these gems are priced. He has heard about the four C's: carat, color, cut, clarity. Now he wants to see if there is any relationship between these diamond characteristics and the price. (SAS problem)

Mark has a dataset from AwesomeGems.com (July, 2005) $n = 351$.

- a. Plot TotalPrice versus Carat for diamonds and describe the graph.
- b. Run a Quadratic Regression Model with TotalPrice as response and Carat as explanatory.
- c. Describe the residuals.
- d. Run a Cubic Regression Model with TotalPrice as response and Carat as explanatory.
- e. Describe the residuals.
- f. Make a decision between these models by looking at RMSE, Adjusted R^2 and residuals and p-values for individual tests.
- g. Run a quadratic model using Depth.
- h. Describe Residuals, Interpret RMSE.
- i. Run a two predictor model using Carat and Depth.
- j. Describe Residuals, Interpret RMSE.
- k. Run a three predictor model that adds interaction for Carat and Depth.
- l. Describe Residuals, Interpret RMSE.
- m. Run a complete second order model using Carat and Depth. Note: A complete 2nd order model has two predictors with linear terms, quadratic terms and an interaction term.
- n. From a to m which model would you recommend using for predicting TotalPrice.
- o. I'm aware that we have a problem with residuals. The variability of the residuals tends to increase for the more expensive diamonds. Using the model you picked on part n transform the response variable to logPrice. Checked the residuals.
- p. What average total price does the quadratic model predict for a 0.5 carat diamond? (with logPrice as response, remember to use the inverse of the natural log).
- q. Compute 95% confidence intervals and 95% prediction intervals for question p. Write conclusions.