Question 1 *(20 points)*

Initial levels of advertising often bring a larger response in the market than later spending. In this question, advertising comes in the form of devoting more shelf space to the indicated product. The level of sales is the weekly total sales of this product at several outlets of a chain of markets. The display space gives the number of shelf feet used to display the item. The data include sales at 48 stores and are available on the CSV file *Display Space*.

a) Create a scatterplot for the level of sales on the number of shelf feet. Does the relationship appear linear? Do you think that it ought to be linear?

b) Fit a linear regression equation to the data, regressing sales on the number of shelf feet. Does this fitted model make substantive sense? What do the slope and intercept tell you?

c) If you see nonlinear relationship in part (a), then use Tukey's bulging rule and transform any variables. Create a scatterplot that shows the relationship between the new set of variables. Does the relationship seem more linear than in part (a)? Fit a linear regression equation to the transformed data. What do the slope and intercept tell you?

d) Compare the fit of the two models to the data. Use R^2, the model standard error, and draw any inferences about the slope to check. Which of the two models provide a better description of the pattern in the data?

Question 2 *(40 points)*

Profitability remains a challenge for banks and thrifts with less than \$2 billion of assets. The business problem facing a bank analyst relates to the factors that affect return on average assets (ROAA), an indicator of how profitable a company is relative to its total assets. Data collected on a sample of 199 community banks and stored on the CSV file *Community Banks* include the ROAA (%), the efficiency ratio (%) as a measure of bank productivity, and total risk-based capital (%) as a measure of capital adequacy.

a) State the multiple regression model.

b) Using the data, create the estimated multiple regression model.

c) Determine whether there is a significant relationship between ROAA and the two independent variables (efficiency ratio and total risk-based capital) at the 0.05 level of significance. In other words, run the Global F-test and comment on the overall model validity.

d) At the 0.05 level of significance, determine whether each independent variable makes a significant contribution to the regression model. On the basis of these results, indicate the independent variables to include in this model.

e) Interpret the partial slope coefficients.

f) Compute the coefficient of multiple determination, R^2 and interpret its value.

g) What is the value of adjusted R^2?

h) Construct a 95% interval estimate for the mean ROAA when the efficiency ratio is 60% and the total risk-based capital is 15%.

i) Construct a 95% interval for the ROAA for a particular community bank when the efficiency ratio is 60% and the total risk-based capital is 15%.

Question 3 *(40 points)*

Medical Company sells medical supplies to hospitals, clinics, and doctors' offices. The company currently markets in three regions of the United States: the South, the West, and the Midwest. These regions are divided into many smaller territories. Data are given on the CSV file *MedicalCo*.

Medical Company's management is concerned about the effectiveness of a new bonus program. The program is overseen by regional sales managers and provides bonuses to salespeople based on performance. Management wants to know if the bonuses paid were related to sales. In determining whether this relationship exists, they also want to take into account the effects of advertising. The variables used in the study include:

Y = Company's sales (in thousands of dollars) in each territory (SALES)
X_1 = the amount that Company spent on advertising in each territory (in hundreds of dollars) (ADV)
X_2 = the total amount of bonuses paid in each territory (BONUS)

Data for a random sample of 25 of the Company's sales territories are given on the CSV file. Management believes that, in addition to advertising and bonus, two other explanatory variables may be important in explaining the variation in sales. These variables are:

X_3 = market share currently held by the Company in each territory (MKTSHR)
X_4 = largest competitor's sales in each territory (COMPET)

Determine at 5% significance level if the variables X_3 and X_4 jointly have a statistically significant influence on Y. What proportion of variation in Sales (Y) is explained by these two variables that is not explained by X_1 and X_2 alone?