A Company has authorized its marketing research department to make a study of customers who have been issued a credit card. The marketing research department hopes to be able to identify the significant variables that explain the variations in purchases. Once these variables are determined, the department intends to try to attract new customers who would be predicted to have a high volume of purchases.

Twenty-five customers were selected at random, and values for the following variables were recorded in the file Data.

y = Average monthly purchases (in dollars)
x_1 = Customer age
x_2 = Customer family income
x_3 = Family size

a.   A first step in regression analysis often involves developing a scatter plot of the data. Develop the scatter plots of all the possible pairs of variables, and with a brief statement indicate what each plot says about the relationship between the two variables.

b.   Compute the correlation matrix for these data. Develop the decision rule for testing the significance of each coefficient. Which, if any, correlations are not significant? Use an alpha level of 0.05.

c.   Use forward selection stepwise regression to develop the multiple regression model. The variable x_2, family income, was brought into the model. Discuss why this happened.

d.   Test the significance of the regression model at step 1 of the computer printout. Justify the significance level you have selected.

e.   Develop a 95% confidence level for the slope coefficient for the family income variable at step 1 of the model. Be sure to interpret this confidence interval.

f.   Describe the regression model at step 2 of the analysis. In your discussion, be sure to discuss the effect of adding a new variable on the standard error of the estimate and on R^2.

g.   Referring to Exercise (f) above, suppose the manager of the company's marketing department questions the appropriateness of adding a second variable. How would you respond to her question?

h.   Looking carefully at the stepwise regression model, you can see that the value of the slope coefficient for variable x_2, family income, changes each time a new variable is added to the regression model. Discuss why this change takes place.

i.   Analyze the regression model at step 3 and the intermediate results at step 1 and 2. Write a report to the marketing manager pointing out the strengths and weaknesses of the model. Be sure to comment on the department's goal of

being able to use the model to predict customers who will purchase high volumes from the company.

j.   Plot the residuals against the predicted value of y, and comment on what this plot means relative to the aptness of the model.

- Compute the standardized residuals and form these in a frequency histogram. What does this indicate about the normality assumption?

- Comment on the overall aptness of this model, and indicate what might be done to improve the model.