

CHAPTER 2: TWO-VARIABLE REGRESSION ANALYSIS: SOME BASIC IDEAS

- 2.1 It tells how the mean or average response of the sub-populations of Y varies with the fixed values of the explanatory variable (s).
- 2.2 The distinction between the sample regression function and the population regression function is important, for the former is an estimator of the latter; in most situations we have a sample of observations from a given population and we try to learn something about the population from the given sample.
- 2.3 A regression model can never be a completely accurate description of reality. Therefore, there is bound to be some difference between the actual values of the regressand and its values estimated from the chosen model. This difference is simply the stochastic error term, whose various forms are discussed in the chapter. The residual is the sample counterpart of the stochastic error term.
- 2.4 Although we can certainly use the mean value, standard deviation and other summary measures to describe the behavior of the regressand, we are often interested in finding out if there are any causal forces that affect the regressand. If so, we will be able to better predict the mean value of the regressand. Also, remember that econometric models are often developed to test one or more economic theories.
- 2.5 A model that is linear in the parameters; it may or may not be linear in the variables.
- 2.6 Models (a), (b), (c) and (e) are linear (in the parameter) regression models. If we let $\alpha = \ln \beta_1$, then model (d) is also linear.
- 2.7 (a) Taking the natural log, we find that $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$, which becomes a linear regression model.
(b) The following transformation, known as the **logit** transformation, makes this model a linear regression model:
$$\ln [(1 - Y_i)/Y_i] = \beta_1 + \beta_2 X_i + u_i$$

(c) A linear regression model
(d) A nonlinear regression model
(e) A nonlinear regression model, as β_2 is raised to the third power.
- 2.8 A model that can be made linear in the parameters is called an intrinsically linear regression model, as model (a) above. If β_2 is 0.8 in model (d) of Question 2.7, it becomes a linear regression

model, as $e^{-0.8(X_i - 2)}$ can be easily computed.

- 2.9** (a) Transforming the model as $(1/Y_i) = \beta_1 + \beta_2 X_i$ makes it a linear regression model.
 (b) Writing the model as $(X_i/Y_i) = \beta_1 + \beta_2 X_i$ makes it a linear regression model.
 (c) The transformation $\ln[(1 - Y_i)/Y_i] = -\beta_1 - \beta_2 X_i$ makes it a linear regression model.
Note: Thus the original models are intrinsically linear models.

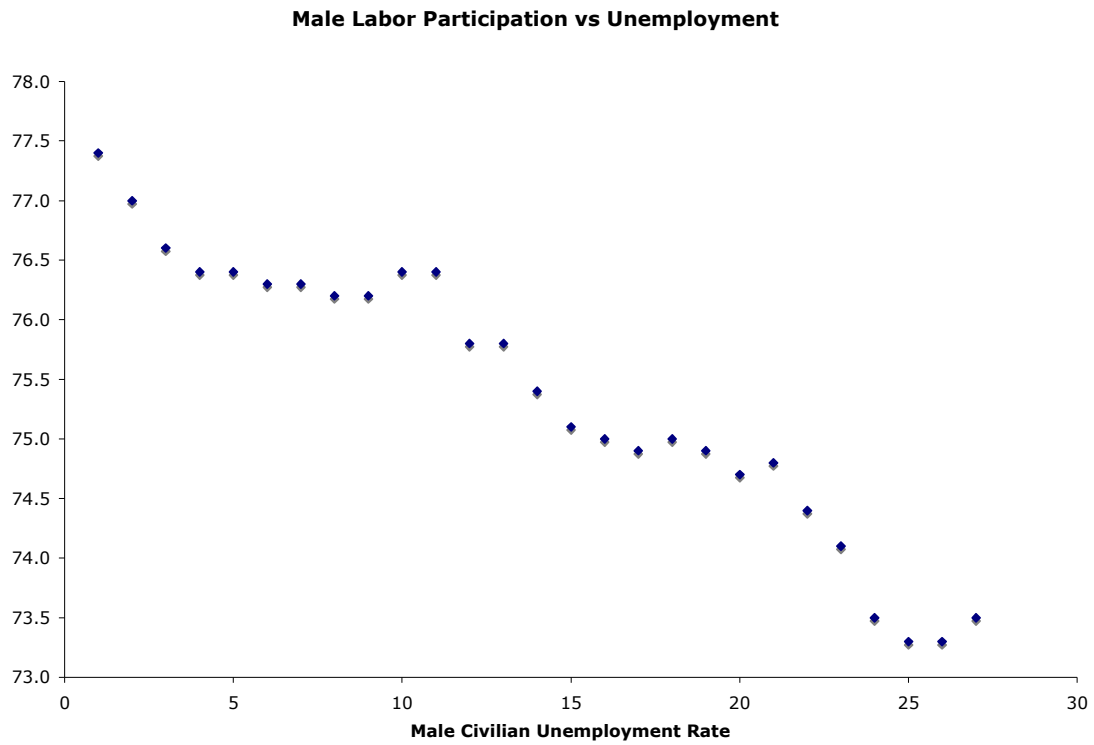
- 2.10** This scattergram shows that more export-oriented countries on average have more growth in real wages than less export oriented countries. That is why many developing countries have followed an export-led growth policy. The regression line sketched in the diagram is a sample regression line, as it is based on a sample of 50 developing countries.

- 2.11** According to the well-known Heckscher-Ohlin model of trade, countries tend to export goods whose production makes intensive use of their more abundant factors of production. In other words, this model emphasizes the relation between factor endowments and comparative advantage.

- 2.12** This figure shows that the higher is the minimum wage, the lower is per head GNP, thus suggesting that minimum wage laws may not be good for developing countries. But this topic is controversial. The effect of minimum wages may depend on their effect on employment, the nature of the industry where it is imposed, and how strongly the government enforces it.

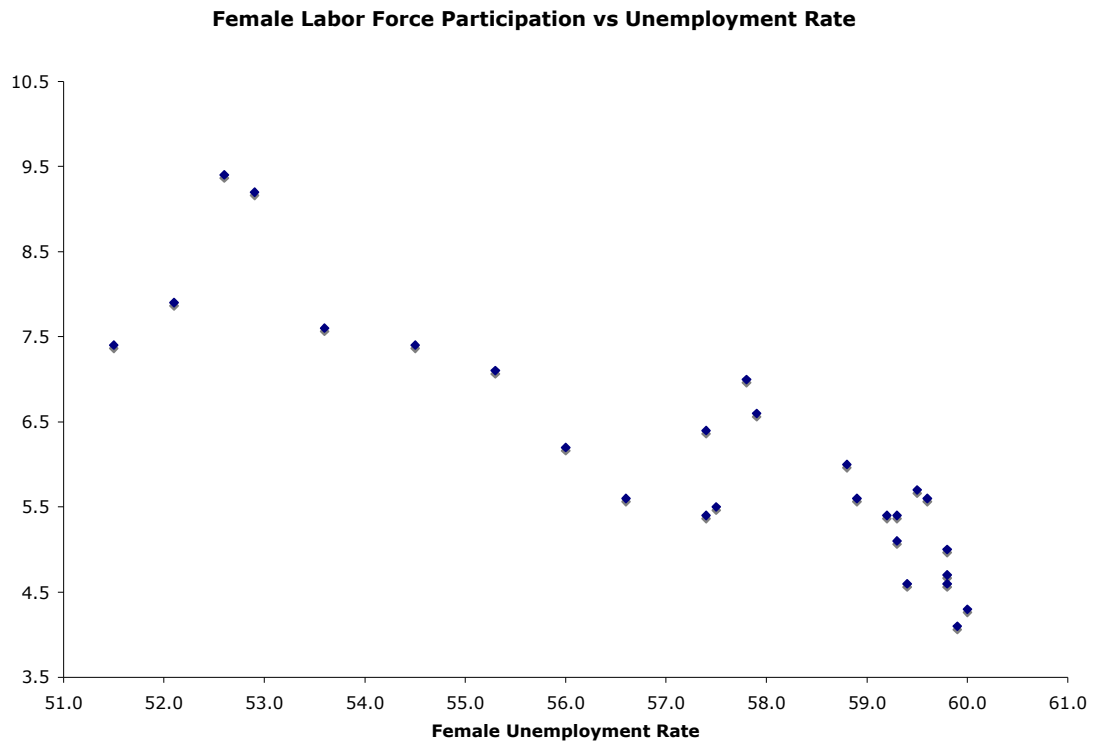
- 2.13** It is a sample regression line because it is based on a sample of 15 years of observations. The scatter points around the regression line are the actual data points. The difference between the actual consumption expenditure and that estimated from the regression line represents the (sample) residual. Besides GDP, factors such as wealth, interest rate, etc. might also affect consumption expenditure.

2.14 (a) The scattergram is as follows:



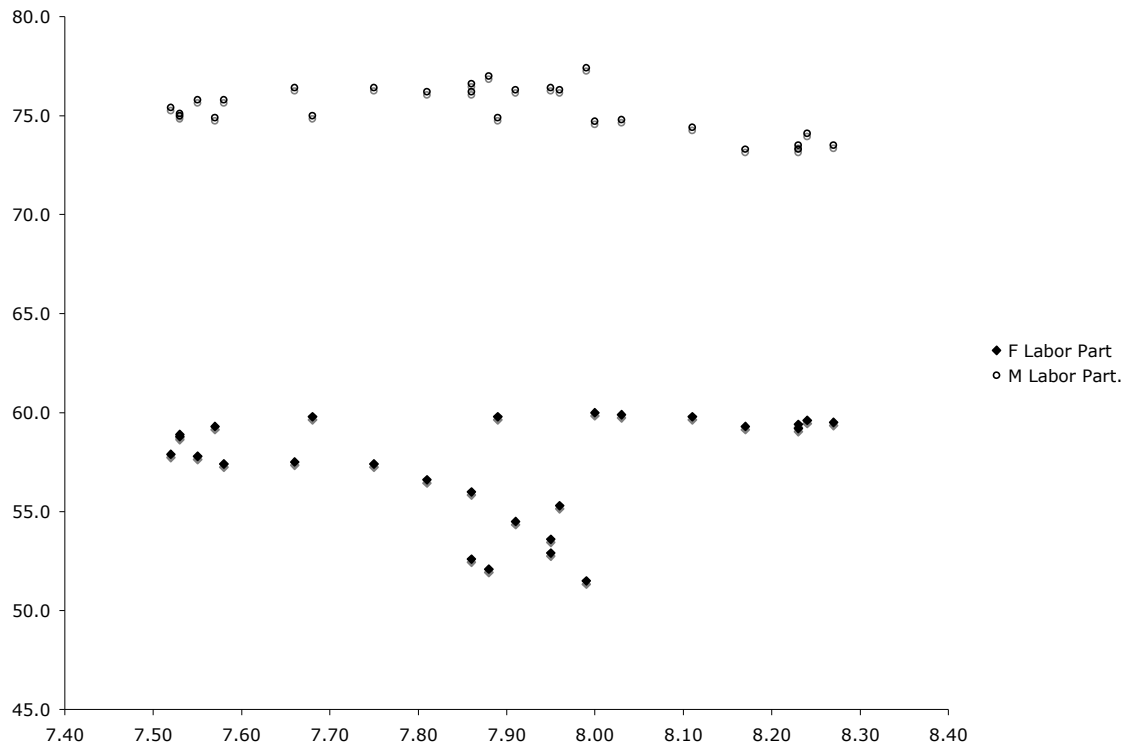
The negative relationship between the two variables seems seems relatively reasonable. As the unemployment rate increases, the labor force participation rate decreases, although there are several minor peaks and valleys in the graph.

(b) The scattergram is as follows:



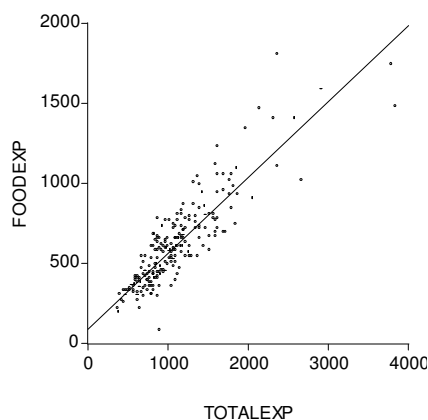
Here the **discouraged worker hypothesis** of labor economics seems to be at work: unemployment discourages female workers from participating in the labor force because they fear that there are no job opportunities.

(c) The plot of Male and Female Labor Force Participation against AH82 shows the following:



There is a similar relationship between the two variables for males and females, although the Male Labor Participation Rate is always significantly higher than that of the Females. Also, there is quite a bit more variability among the Female Rates. To validate these statements, the average Male Rate is 75.4 %, whereas the Female average is only 57.3 %. With respect to variability, the Male Rate standard deviation is only 1.17 %, but the Female standard deviation is 2.73 %, more than double that of the Male Rate. Keep in mind that we are doing simple bivariate regressions here. When we study multiple regression analysis, we may have some different conclusions.

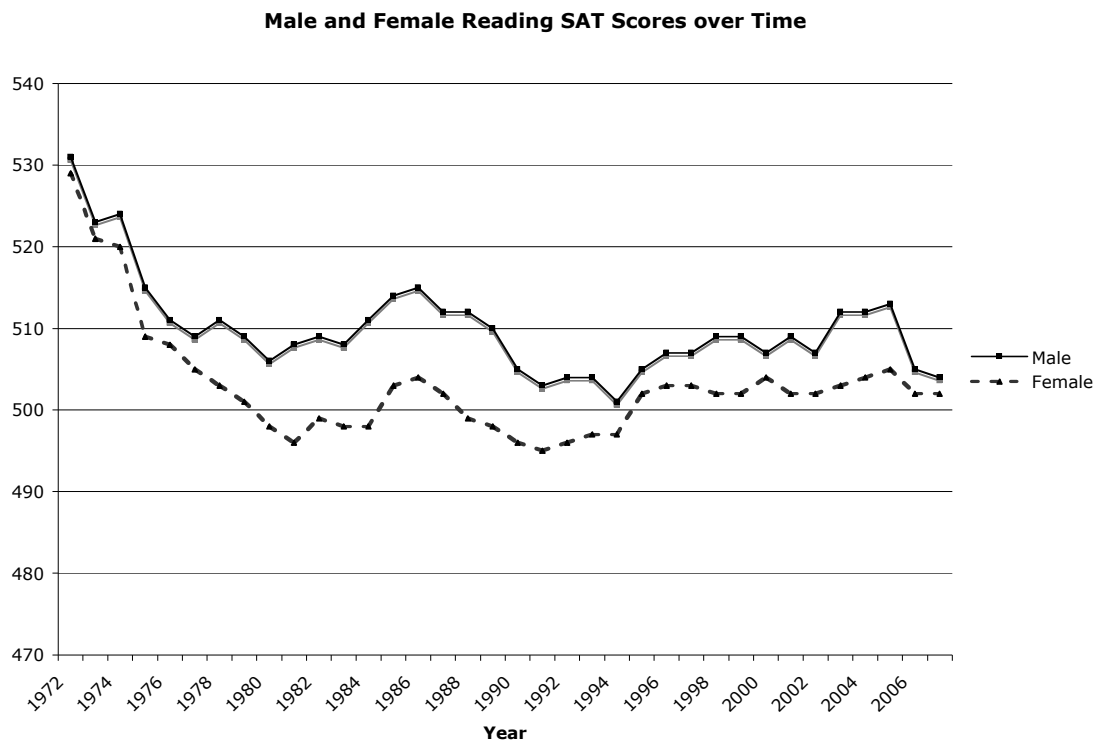
2.15 (a) The scattergram and the regression line look as follows:



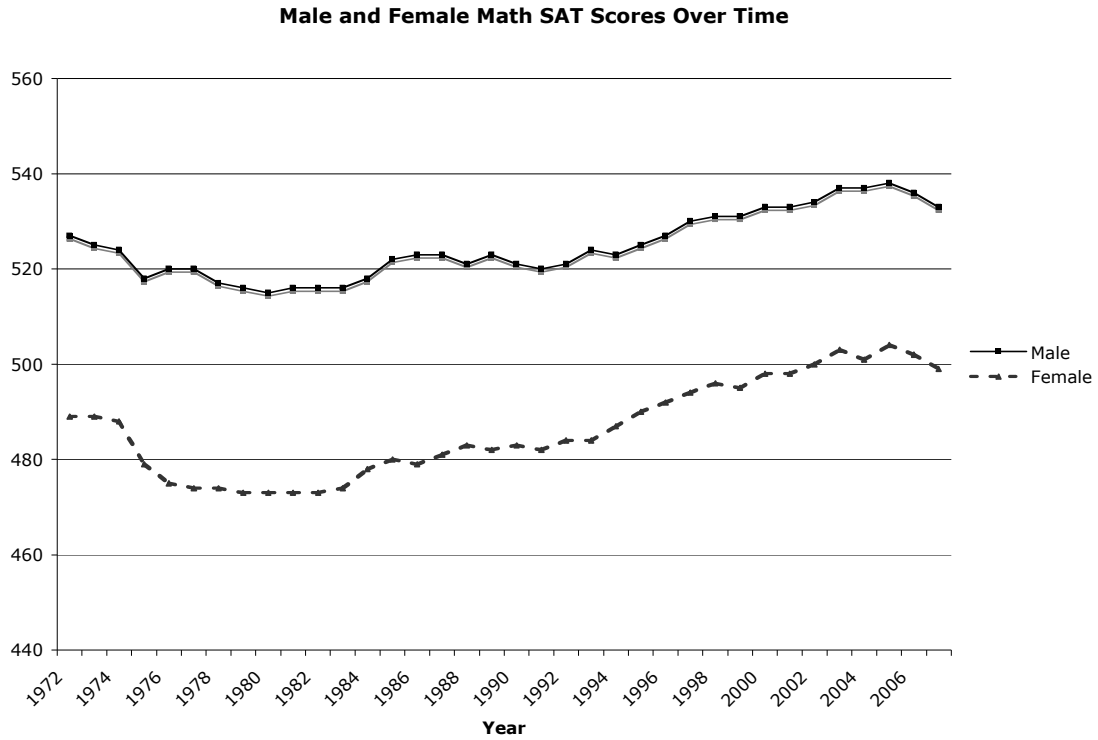
(b) As total expenditure increases, on the average, expenditure on food also increases. But there is greater variability between the two after the total expenditure exceeds the level of Rs. 2000.

(c) We would not expect the expenditure on food to increase linearly (i.e., in a straight line fashion) for ever. Once basic needs are satisfied, people will spend relatively less on food as their income increases. That is, at higher levels of income consumers will have more discretionary income. There is some evidence of this from the scattergram shown in (a): At the income level beyond Rs. 2000, expenditure on food shows much more variability.

2.16 (a) The scatter plot for male and female verbal scores is as follows:



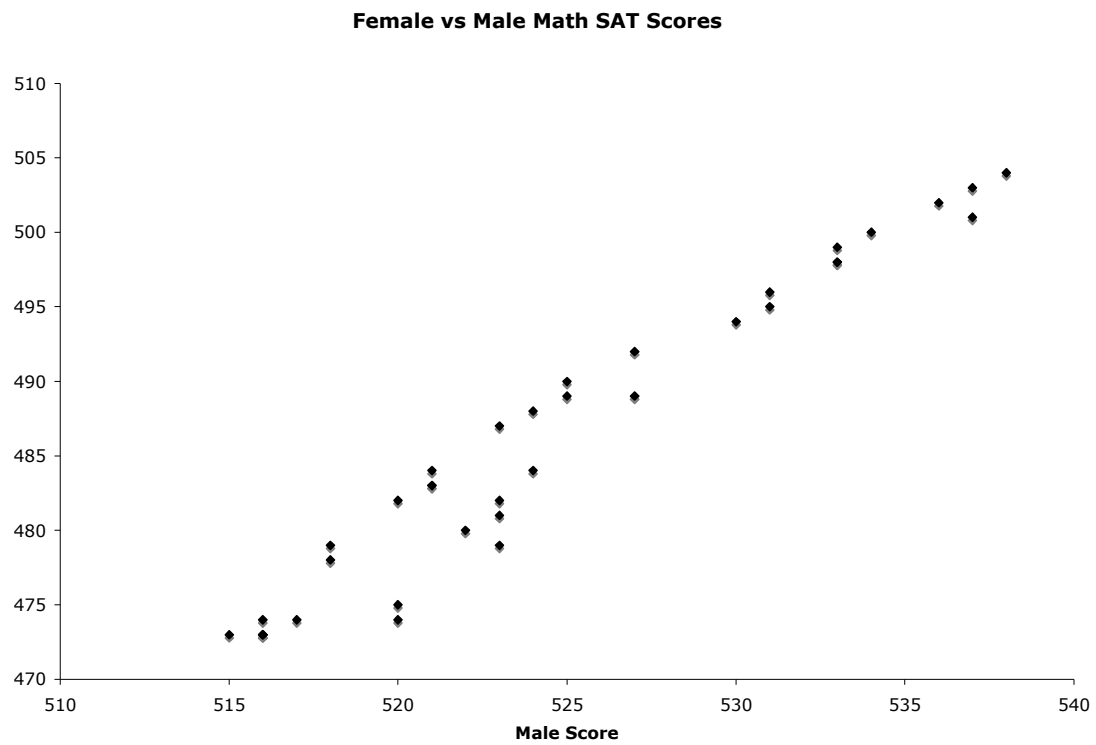
And the corresponding plot for male and female math score is as follows:



(b) Over the years, the male and female reading scores show a slight downward trend, although there seems to be a leveling in the mid-1990's. The math scores, however, show a slight increasing trend, especially starting in the early 1990's. In both graphs it seems the male scores are generally higher than the female scores, of course with year-to-year variation.

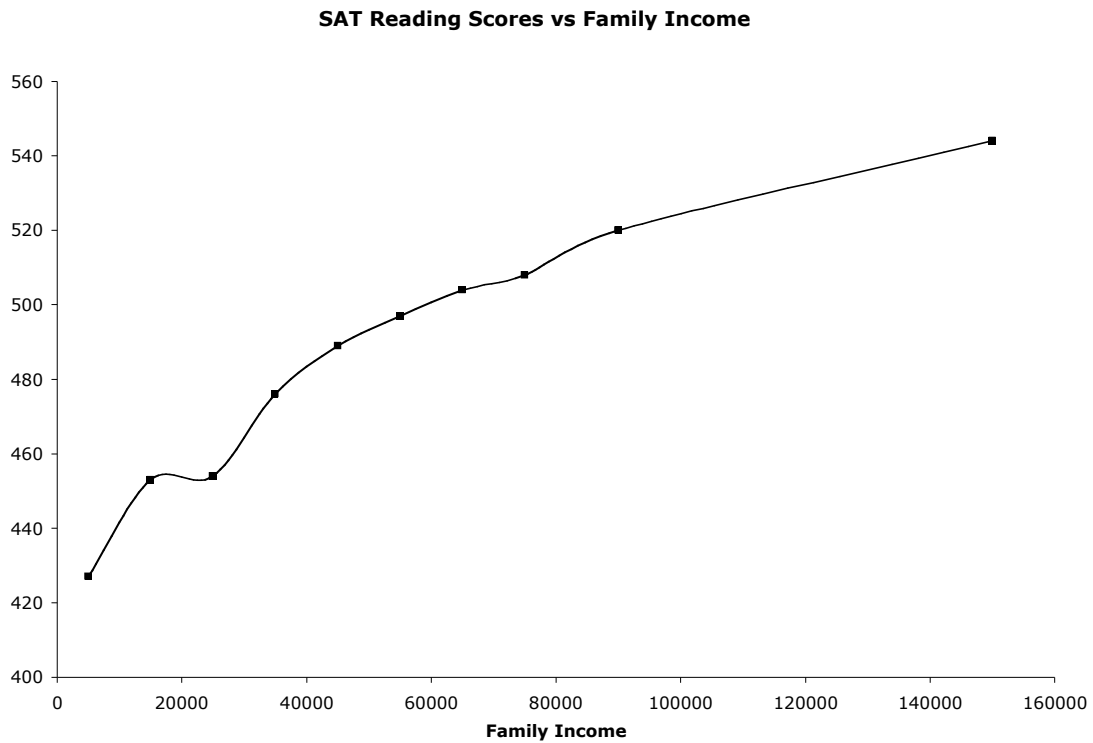
(c) We can develop a simple regression model regressing the math score on the verbal score for both sexes.

(d) The plot is as follows:



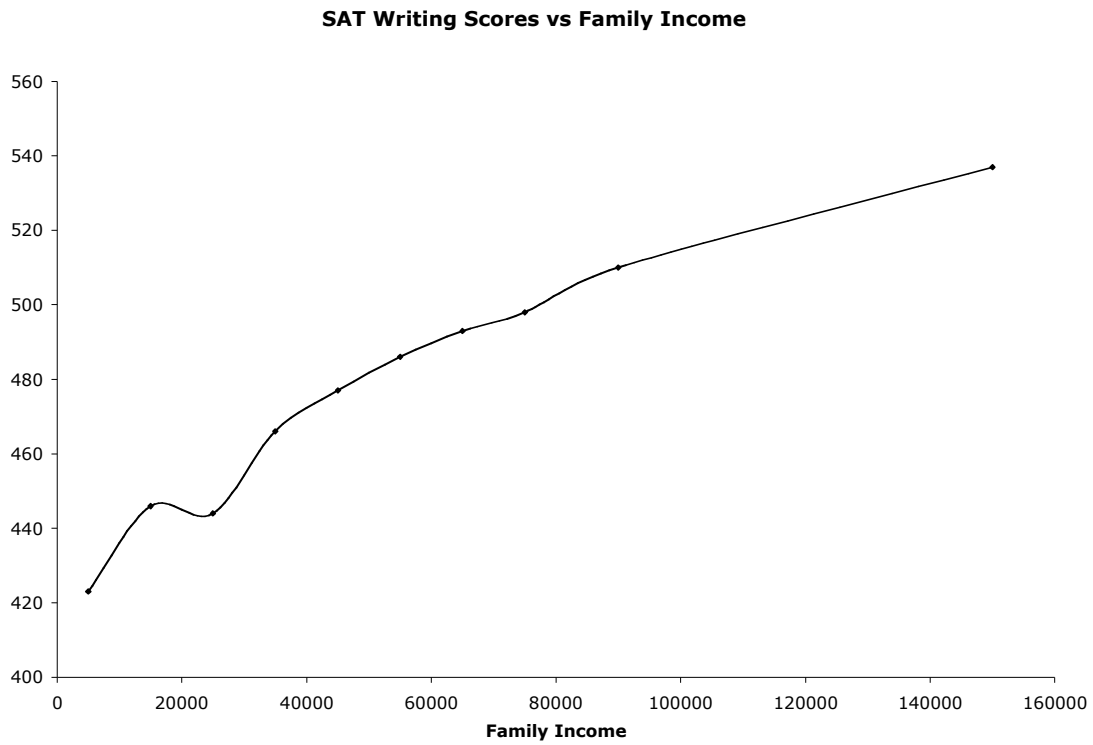
As the graph shows, the two genders seem to move together, although the male scores are always higher than the female scores.

2.17 (a) The scatter plot for male and female verbal scores is as follows:



The results are somewhat similar to those from Figure 2.7. It seems that the average Reading Score increases as the average Family Income increases.

b)



This graph looks almost identical to the previous ones, especially the Reading Score graph.

c) Apparently, there seems to be a positive relationship between average Family Income and SAT scores.