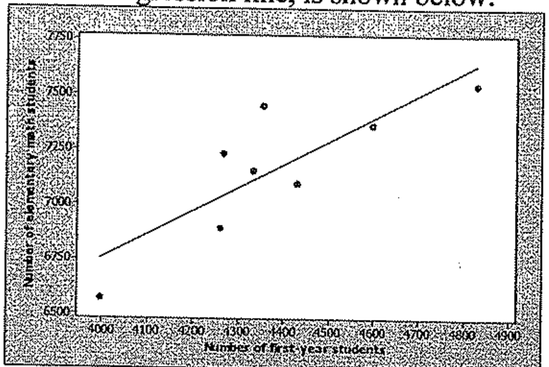
HW pp 242-243 ; pp 244-247

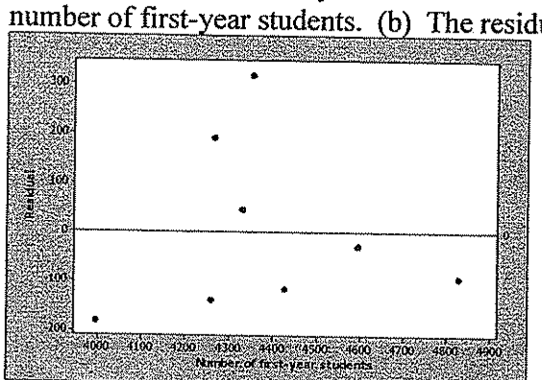
3.63 Higher income can cause better health: higher income means more money to pay for medical care, drugs and better nutrition, which in turn results in better health. Better health can cause higher income: if workers enjoy better health, they are better able to get and hold a job, which can increase their income.

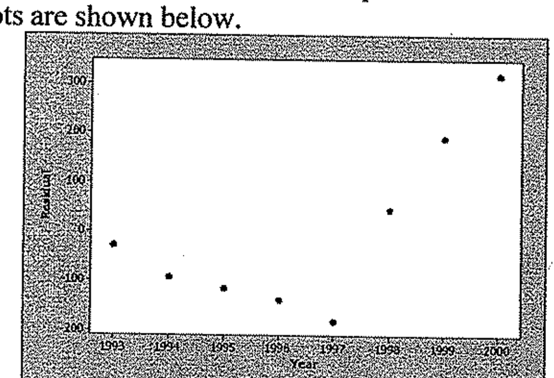
3.64 No, you cannot shorten your stay by choosing a smaller hospital. The positive correlation does not imply a cause and effect relationship. Larger hospitals tend to see more patients in poor condition, which means that the patients will tend to require a longer stay.

3.66 (a) *Who?* The individuals are students at a large state university. *What?* The variables are the number of first-year students and the number of students who enroll in elementary mathematics courses. Both variables are quantitative and take on integer values from several hundred to several thousand, depending on the size of the university. *Why?* The data were collected to try to predict the number of students who will enroll in elementary mathematics courses. *When, where, how, and by whom?* Faculty members in the mathematics department at a large state university obtained the enrollment data and class sizes from 1993 to 2000. These data were probably extracted from a historical data base in the Registrar's office. A scatterplot, with the regression line, is shown below.



The regression line appears to provide a reasonable fit. About 69.4% of the variation in enrollments for elementary math classes is accounted for by the linear relationship with the number of first-year students. (b) The residual plots are shown below.



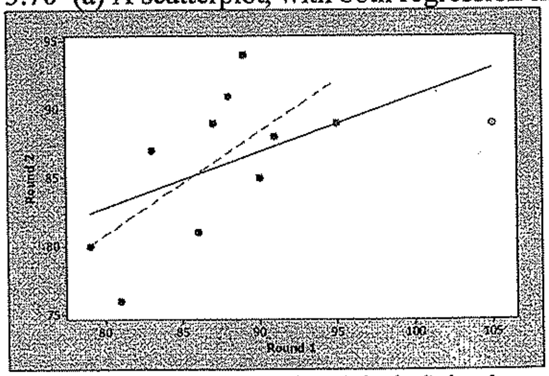


The plot of the residuals against x shows that a somewhat different line would fit the five lower points well. The three points above the regression line represent a different relation between the number of first-year students and mathematics enrollments. The plot of the residuals against year clearly illustrates that the five negative residuals are from the years 1993 to 1997, and the three positive residuals are from 1998, 1999, and 2000. (c) The change in requirements was not visible on the scatterplot in part (a) or the plot of the residuals against x. However, the change is clearly illustrated (negative residuals before 1998 and positive residuals after 1998) on the plot of the residuals against year.

3.67 The correlation for individual stocks would be lower. Individual stock performances will be more variable weakening the relationship.

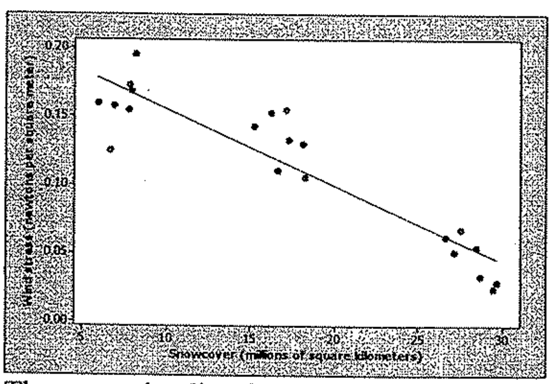
3.69 (a) Yes, but the relationship is not very strong. (b) The mortality rate is extremely variable for those hospitals that treat few heart attacks. As the number of patients treated increases the variability decreases and the mortality rate appears to decrease giving the appearance of an exponentially decreasing pattern of points in the plot. The nonlinearity strengthens the conclusion that heart attack patients should avoid hospitals that treat few heart attacks.

3.70 (a) A scatterplot, with both regression lines, is below.



The influential observation (circled) is observation 7, (105, 89). (b) The line with the larger slope is the line that omits the influential observation (105, 89). The influential point pulls the regression line with all of the points downward in order to minimize the overall prediction error.

3.73 (a) *Who?* The individuals are land masses. *What?* The two quantitative variables are the amount of snow cover (in millions of square kilometers) and summer wind stress (in newtons per square meter). *Why?* The data were collected to explore a possible effect of global warming. *When, where, how, and by whom?* The data from Europe and Asia appear to be collected over a 7 year period during the months of May, June, and July. The amount of snow cover may have been estimated from arial photographs or satellite images and the summer wind stress measurements may have been collected by meteorologists. The scatterplot below suggests a negative linear association, with correlation $r = -0.9179$.



The regression line for predicting y = wind stress from x = snow cover is $\hat{y} \doteq 0.212 - 0.0056x$; $r^2 = 0.843$. The linear relationship explains 84.3% of the variation in wind stress. We have good evidence that decreasing snow cover is strongly associated with increasing wind stress. (b) The graph shows 3 clusters of 7 points.