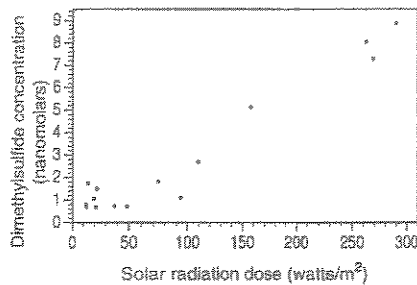


- 4.25. (a) The association is (weakly) positive. (b) The estimate is 4, which is an overestimate; that child had the lowest score on the test.

- 4.30. (a) SRD is the explanatory variable, so it should be on the horizontal axis.
 (b) The scatterplot shows a positive linear association. The correlation coefficient is $r \doteq 0.9685$, which is consistent with the strength of the association visible in the scatterplot.



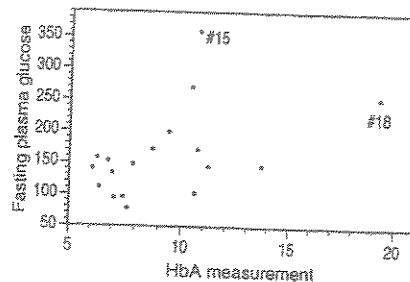
- 5.39. (a) The two unusual observations are marked on the scatterplot. (b) The correlations are

$$r_1 \doteq 0.4819 \text{ (all observations)}$$

$$r_2 \doteq 0.5684 \text{ (without Subject 15)}$$

$$r_3 \doteq 0.3837 \text{ (without Subject 18)}$$

Both outliers change the correlation. Removing Subject 15 increases r , because its presence makes the scatterplot less linear, while removing Subject 18 decreases r , because its presence decreases the relative scatter about the linear pattern.



- 5.41. The scatterplot from Exercise 5.39 is reproduced here with the regression lines added. The equations are

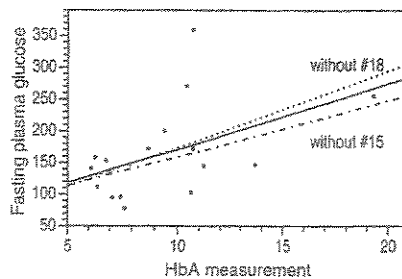
$$\hat{y} \doteq 66.4 + 10.4x \text{ (all observations)}$$

$$\hat{y} \doteq 69.5 + 8.92x \text{ (without #15)}$$

$$\hat{y} \doteq 52.3 + 12.1x \text{ (without #18)}$$

While the equation changes in response to removing either subject, one could argue that neither one is particularly influential, because the line moves very little over the range of x (HbA) values. Subject 15 is

an outlier in terms of its y value; such points are typically not influential. Subject 18 is an outlier in terms of its x value, but is not particularly influential because it is consistent with the linear pattern suggested by the other points.



- 8.29. (a) Assign labels 0001 through 1410. (b) Beginning at line 105, we choose plots 0769, 1315, 0094, 0720, and 0906.

- 8.39. Assign labels 001 through 290 to the men, and 001 through 110 to the women. Entering the table at line 130, first choose the men (174, 095, and 178), and then continue on to choose the women (019, 007, and 041). See note on page 54 about using Table B.

- 8.46. (a) The population is Ontario residents; the sample is the 61,239 people interviewed.
 (b) The sample size is very large, so if there were large numbers of both sexes in the sample—this is a safe assumption because we are told this is a “random sample”—these two numbers should be fairly accurate reflections of the values for the whole population.