VI. Answer the following questions using the data on the preceding page.

A. Use the method of least squares to determine a regression equation.

Data from page T160

$$\hat{y}._x = a + bx$$

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

$$= \frac{1,210}{185}$$

$$= 6.5405405$$

$$a = \bar{Y} - b\bar{x}$$

$$= \frac{\sum Y}{n} - b\frac{\sum X}{n}$$

$$= \frac{310}{6} - 6.5405405\left(\frac{35}{6}\right)$$

$$= 13.513515$$

$$\hat{y}._x = 13.5 + 6.54x$$

When using a regression equation, values for x should be limited to the actual data range. Here, 3 to 10.

B. Calculate the estimated profit for next year when R & D will be $8,000,000.

$$\hat{y}._x = 13.5 + 6.54x$$

$$\hat{y}._8 = 13.5 + 6.54(8)$$

$$= 13.5 + 52.32$$

$$= \$65.82 \text{ million}$$

C. Draw the regression line on the page 160 scatter diagram.

Two points (x,y) may be used to draw a straight line. Here, 8 and 65.82 from question B, and the y-intercept (0,13.5) are used.

D. Calculate the 99% confidence interval for question B.

$$S_{y.x} = \sqrt{\frac{\sum Y^2 - a(\sum Y) - b(\sum XY)}{n-2}} = \sqrt{\frac{17,700 - 13.513515(310) - (6.5405405)(2,010)}{6-2}} = 9.54$$

$$df = 6 - 2 = 4$$

$$\alpha/2 = .01/2 = .005 \rightarrow t = 4.604$$

$$\bar{x} = \frac{\sum x}{n} = \frac{35}{6} = 5.83$$

Note the use of +1 under the radical. This is necessary because the question concerns a particular value (the next value) of y, and not the mean value of y. Predicting a particular value of y increases the confidence interval.

$$\hat{y}._x \pm ts_{y.x}\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$\hat{y}._8 = 65.82 \pm 4.604(9.54)\sqrt{1 + \frac{1}{6} + \frac{(8 - 5.83333)^2}{235 - \frac{(35)^2}{6}}}$$

$$= 65.82 \pm 50.44$$

$$15.38 \leftrightarrow 116.26$$

E. What procedure should be followed if the range for the answer to question D includes zero or a negative number?

If the range expresses the possibility of a negative number, the confidence level may be lowered with a larger sample. This happens because a larger sample lowers t. Here, if only 3 million dollars is invested in R & D, the average value y is only 13.5 + 6.54(3) = 33.12. For the 99% confidence level, acceptable error is approximately 52.47 (calculations for this number are not shown). This means profit could be negative (33.12 - 52.47). However, because profits can be negative, a larger sample is not required. But, this range is very large and may not be useful.