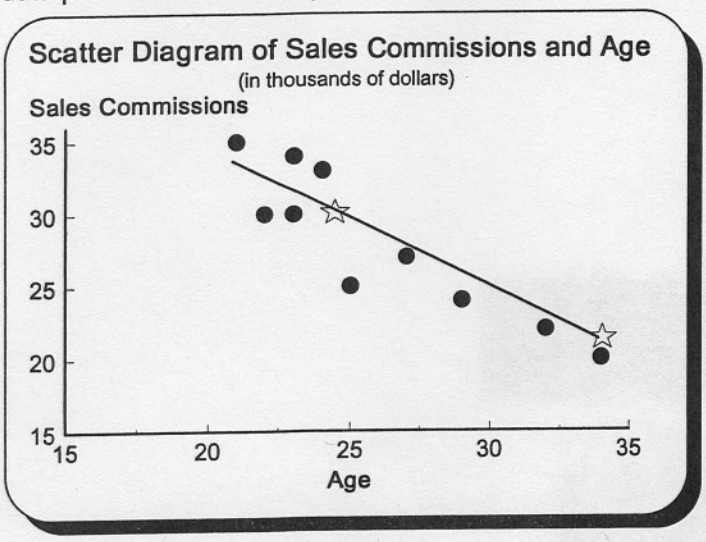
Practice Set 24 Simple Linear Regression Analysis

I. Having determined that age affects sales performance, Darin Jones wants to estimate sales commissions using the data presented in the chapter 23 practice set.



Age	Sales Commissions (000)	xy	x^2	y^2
23	30	690	529	900
25	25	625	625	625
34	20	680	1,156	400
29	24	696	841	576
21	35	735	441	1,225
32	22	704	1,024	484
23	34	782	529	1,156
24	33	792	576	1,089
27	27	729	729	729
22	30	660	484	900
260	280	7,093	6,934	8,084

A. Determine the regression equation to 3 significant digits.

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} = \frac{-1{,}870}{1{,}740} = -1.0747126$$

Note: Data for b was taken from the previous page.

$$a = \bar{Y} - b\bar{x} = \frac{\sum Y}{n} - b\frac{\sum X}{n} = \frac{280}{10} - (-1.0747126)\left(\frac{260}{10}\right) = 55.942527$$

$$\hat{y}_{\cdot x} = a + bx$$
$$\hat{y}_{\cdot x} = 55.9 - 1.07x$$

B. Estimate sales commissions for a group of 24-year-old salespeople.

$$\hat{y}_{\cdot 24} = 55.9 - .1.07x = 55.9 - 1.07(24) = 30.2$$

C. Graph the regression line.

$$\hat{y}_{\cdot x} = 55.9 - .1.07x = 55.9 - 1.07(34) = 19.5$$

Two points (x,y) will be used to draw a straight line. We will use the coordinates from questions B and C.

D. Determine the 99% confidence interval for the question B group.

$$S_{y.24} = \sqrt{\frac{\sum Y^2 - a(\sum Y) - b(\sum XY)}{n-2}} = \sqrt{\frac{8{,}084 - 55.942527(280) - (-1.0747126)(7{,}093)}{10-2}} = 2.319$$

$df = 10 - 2 = 8$

$\alpha/2 = .01/2 = .005 \rightarrow t = 3.355$

$$\bar{x} = \frac{\sum x}{n} = \frac{260}{10} = 26$$

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

$$\hat{y}_{\cdot 24} = 30.2 \pm 3.355(2.319)\sqrt{\frac{1}{10} + \frac{(24-26)^2}{6{,}934 - \frac{(260)^2}{10}}}$$

$$= 30.2 \pm 2.73$$

$$27.47 \leftrightarrow 32.93$$

E. What procedure should be followed if question D's range includes negative numbers?

The standard error of the estimate may be lowered with a larger sample. Here, the standard error is low and the range for y does not include zero.