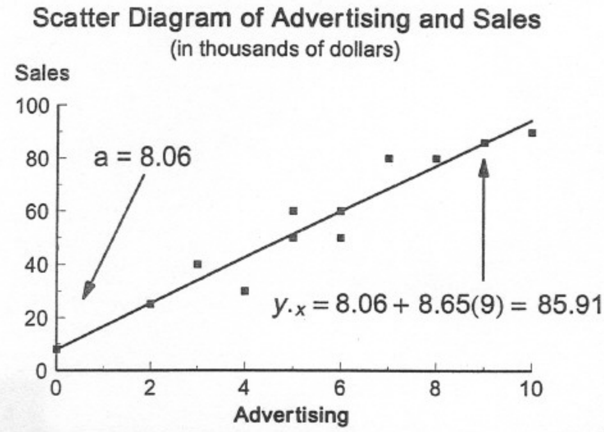
III. Drawing a regression line

A. Two points (x,y) may be used to draw a straight line.
B. The y-intercept (0, $8,060) will be one point.
C. The estimated value of y for x of $9,000 is $85,910. It will be the second point (see page 152).



IV. The standard error of the estimate

A. The standard error of the estimate measures the dispersion of the scatter (plots) around the regression line.
B. It is the standard deviation of y given some value of x.
C.

$$S_{y.x} = \sqrt{\frac{\sum\left(Y-\bar{Y}\right)^2}{n-2}} = \sqrt{\frac{\sum Y^2 - a(\sum Y) - b(\sum XY)}{n-2}}$$

$$S_{y.x} = \sqrt{\frac{36{,}225 - 8.055556(565) - 8.6507936(3{,}600)}{10-2}} = 8.145$$

V. An interval estimate for the conditional mean of y for some given value of x

A. A confidence interval will be determined using the small sample t distribution.
B.

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

Note: The correction factor following the standard error of the estimate is needed because the sample is small and the scatter of sales data might not be normal.

C. Linda Smith wants to determine the 95% confidence interval for expected sales for months when advertising expenditures are $9,000.

Basic Assumptions Concerning Linear Regression Analysis

1. There are a number of y values for each value of x.
2. The conditional distributions of y given x are normal.
3. The variance of the conditional distributions are equal.
4. Predictions of y are limited to the existing range for x.
Note: Predicting an individual value (next month's sales) rather than the mean of Y (sales) requires inserting a +1 under the radical.

Problem Notes

$\hat{y}._x = 8.06 + 8.65(x) = \$85,910$ when $x = 9$. See page 152.		
Degrees of freedom for t will be n - 2 because both a and b were estimated in determining $\hat{y}._x$. df = n - 2 = 10 - 2 = 8		
$\alpha/2 = .05/2 = .025 \rightarrow 2.306$ for t		
$\bar{x} = \frac{\sum x}{n} = \frac{56}{10} = 5.6$	$S_{y.x} = 7.89$	n = 10

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

$$85.91 \pm 2.306(8.145)\sqrt{\frac{1}{10} + \frac{(9-5.6)^2}{364 - \frac{56^2}{10}}}$$

$$85.91 \pm 10.779$$

$$75.131 \leftrightarrow 96.689$$

D. For regression analysis to be valid, the range for variables a and b must consist of realistic values. Here, the y-intercept cannot be negative because negative sales are not possible. But, determining the 95% confidence interval for the y-intercept (0,8.06) by recalculating acceptable error (E) results in a negative lower limit (8.06 - 15.96 = -7.90). This concern might be solved by lowering the standard error of the estimate with a larger sample. In addition, procedures exist for determining a confidence interval for the slope. The possibility of a negative slope would cause people to question the relationship between advertising and sales. A larger sample might also solve the problem of a negative slope.