might be to simply use the same regression line that we have already calculated (Figure 14-6), but such an inclination would lead us astray. Recall that the X-to-Y regression line is computed so as to minimize the squared error terms between the Y and Y's—in terms of Figure 14-7, we are concerned with minimizing the "vertical" errors. However, if we wish to predict X from Y, we want to minimize the errors between the observed and the predicted X's—that is, we are concerned with minimizing the *horizontal* distances between the observed X's and the regression line. It turns out that this criterion (usually) requires a different regression line.

Just as the formulas for obtaining the Y-from-X regression are

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

and

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

the analogous formulas for obtaining the X-from-Y regression equation are

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

and

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

The Pearson r^2, however, is identical whether predicting Y from X or predicting X from Y.

Given two variables, practical factors often dictate which should be the predictor and which should be the predicted variable. For example, suppose we are interested in the relationship between a person's IQ score as a teenager and the person's yearly income as an adult. Since teenage IQ occurs prior to adult income, it makes sense to want to predict the latter from the former, rather than vice versa. By convention, the predictor variable is usually labeled as X and the predicted variable, Y.

PROBLEMS

1. Draw scatterplots that would roughly correspond to the following situations (stated in terms of regression lines and Pearson r's).
   a. $Y = 0.65X + 2.5$ $\quad r = 0.25$
   b. $Y = -0.51X + 1.5$ $\quad r = 0.75$
   c. $Y = X + 3$ $\quad r = 1.00$

d. $Y = 4$ $\quad r = 0$
e. $Y = 3X + 14$ $\quad r = 0.95$
f. $Y = 1.6X + 2$ $\quad r = -0.05$
g. $Y = 1.6X + 3$ $\quad r = -0.99$

2. A sports psychologist is interested in the relationship between players' ages and their batting averages. Twelve players, chosen at random, yielded the following data:

	Player											
	1	2	3	4	5	6	7	8	9	10	11	12
Age	18	17	31	25	22	24	28	21	21	18	35	41
Average	0.225	0.350	0.150	0.275	0.269	0.200	0.320	0.315	0.195	0.200	0.310	0.275

a. Compute the regression line predicting average from age.
b. From this regression line compute each player's predicted average and error (difference between predicted and actual average).
c. Compute the variance of the actual averages, predicted averages, and errors. How do these variances relate to one another?
d. What is the ratio of the variance of the predicted averages to the variance of the actual averages?
e. Compute a Pearson r^2 for these data. How does the value of the r^2 compare to your answer to question d?

3. The same sports psychologist is interested in the relationship between weight and discus-throwing ability. He selects a random sample of 15 people from the population and measures both the person's weight and the distance they are able to throw the discus. The data are as follows:

Person	Weight (pounds)	Discus distance (feet)
Maggie	120	125
Fred	165	215
Elaine	105	145
Suzie	128	129
Robert	220	175
Geoff	170	209
Beth	115	141
Earl	156	223
Linda	125	130
Tom	190	200
Betsy	160	132
George	130	250
John	200	180
Alinda	100	150
Sylvia	130	135

a. Draw a scatterplot of these data.
b. Compute the regression line predicting discus distance from weight.
c. Compute the Pearson r^2 relating these two variables. Is this Pearson r^2 significantly different from 0?
d. Now repeat questions b and c for men and women separately.
e. What conclusion do you draw about one potential danger of computing correlations?

4. A sociologist is interested in the relationship between the size of a city and its per capita murder rate. She chooses 10 cities in the United States. The sizes and per capita murder rates are shown below:

	City									
	1	2	3	4	5	6	7	8	9	10
Size (in 1000s)	150	990	75	520	610	304	190	100	60	120
Murder rate (per 100,000 per year)	1.2	11.4	3.1	4.0	3.1	4.2	2.0	0.3	1.1	0.9

a. Draw a scatterplot depicting these data.
b. Compute regression equations to predict per capita murder rate from city size, and vice versa.
c. What is the correlation (Pearson r) between the two variables?
d. Is the correlation significantly different from zero?

5. Consider the data of Chapter 5, problem 8.
a. Find the regression equation to predict the IQ of the high-SES twin from the IQ of the low-SES twin.
b. Find the regression equation to predict the IQ of the low-SES twin from the IQ of the high-SES twin.
c. What is the Pearson r^2 between IQ scores of high- and low-SES twins?
d. Is this correlation significantly different from zero?

6. A developmental psychologist is interested in whether there is a relationship between sex and handedness. He selects a random sample of 24 children. To each child, he assigns a 1 or a 0 corresponding to whether the child is male or female and also a 1 or a 0 corresponding to whether the child is right- or left-handed. The data are as follows:

Child	Sex	Handed	Child	Sex	Handed	Child	Sex	Handed
1	1	1	9	0	1	17	1	0
2	0	1	10	1	1	18	1	0
3	0	0	11	0	1	19	0	1
4	1	0	12	1	1	20	1	0
5	0	1	13	0	0	21	0	1
6	1	1	14	0	0	22	1	1
7	1	0	15	1	1	23	1	1
8	1	1	16	1	0	24	0	0

Compute the correlation (Pearson r^2) between sex and handedness. Is this correlation significant?

7. A study is done by the telephone company to determine whether there is any relationship between sex (male or female) and type of job (operator or supervisor). Ten phone company employees are selected at random and assigned a 1 or a 0 corresponding to whether they are female or male and another 1 or 0 corresponding to whether they are an operator or a supervisor, respectively. The data are as follows:

Person	**Female/Male (x)**	**Operator/Supervisor (y)**
1	1	1
2	0	1
3	1	0
4	1	0
5	0	0
6	1	1
7	0	0
8	1	1
9	1	1
10	0	0

a. Determine r^2 for these data.
b. Test the significance of this r^2.

8. A dental researcher is interested in the relationship between the number of times teeth are brushed per day and number of cavities. The researcher asks six people how many times per day (one, two, or three) they brush their teeth and how many cavities they have. The data are as follows, showing the number of cavities for people who brush one, two, or three times a day:

Number of Brushings (x)			
1	2	3	
8	1	2	Number of cavities
6	5		
7			

a. Compute the regression line relating number of brushings and number of cavities.
b. Compute r^2 for these data.
c. Test whether a significant nonzero relationship exists between the two variables.