

standard units; the product of two negatives is positive. In the upper right quadrant, the product of two positives is positive. In the remaining two quadrants, the product of a positive and a negative is negative. The average of all these products is the correlation coefficient. If r is positive, then points in the two positive quadrants will predominate, as in figure 9b. If r is negative, points in the two negative quadrants will predominate, as in figure 9c.

Exercise Set D

1. For each of the data sets shown below, calculate r .

(a)	(b)	(c)
x y	x y	x y
1 6	1 2	1 7
2 7	2 1	2 6
3 5	3 4	3 5
4 4	4 3	4 4
5 3	5 7	5 3
6 1	6 5	6 2
7 2	7 6	7 1

2. Find the scatter diagram in figure 6 (p. 127) with a correlation of 0.95. In this diagram, the percentage of points where both variables are simultaneously above average is around

5% 25% 50% 75% 95%.

3. Repeat exercise 2, for a correlation of 0.00.
 4. Using figure 7, repeat exercise 2 for a correlation of -0.95 .

The answers to these exercises are on p. A57.

Technical note. There is another way to compute r , which is sometimes useful:⁸

$$r = \frac{\text{cov}(x, y)}{(\text{SD of } x) \times (\text{SD of } y)}$$

where

$$\text{cov}(x, y) = (\text{average of products } xy) - (\text{ave of } x) \times (\text{ave of } y).$$

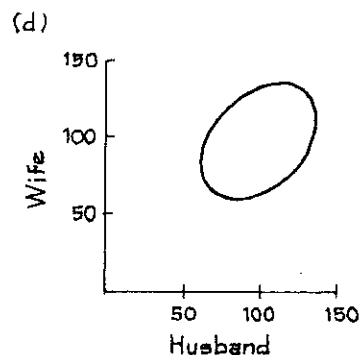
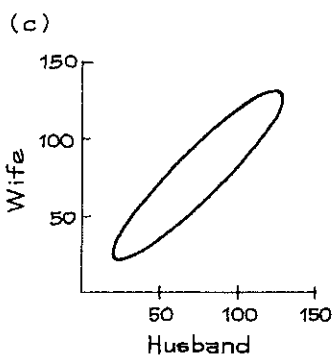
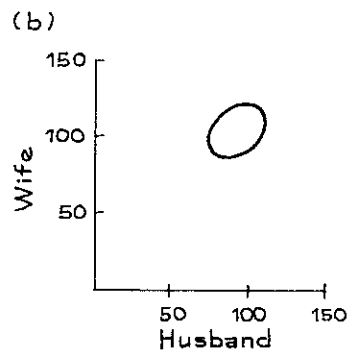
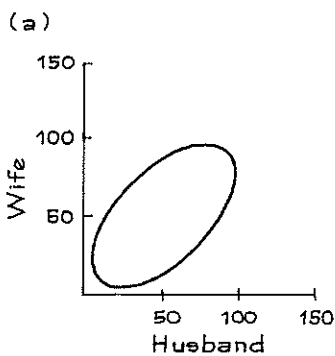
5. REVIEW EXERCISES

Review exercises may cover material from previous chapters.

1. A study of the IQs of husbands and wives obtained the following results:

$$\begin{array}{ll} \text{for husbands,} & \text{average IQ} = 100, \quad \text{SD} = 15 \\ \text{for wives,} & \text{average IQ} = 100, \quad \text{SD} = 15 \\ & r = 0.6 \end{array}$$

One of the following is a scatter diagram for the data. Which one? Say briefly why you reject the others.



2. (a) For a representative sample of cars, would the correlation between the age of the car and its gasoline economy (miles per gallon) be positive or negative?
 - (b) The correlation between gasoline economy and income of owner turns out to be positive.⁹ How do you account for this positive association?
3. Suppose men always married women who were exactly 8% shorter. What would the correlation between their heights be?
4. Is the correlation between the heights of husbands and wives in the U.S. around -0.9 , -0.3 , 0.3 , or 0.9 ? Explain briefly.
5. Three data sets are collected, and the correlation coefficient is computed in each case. The variables are
 - (i) grade point average in freshman year and in sophomore year
 - (ii) grade point average in freshman year and in senior year
 - (iii) length and weight of two-by-four boards

Possible values for correlation coefficients are

-0.50 0.0 0.30 0.60 0.95

Match the correlations with the data sets; two will be left over. Explain your choices.

3. In each case, say which correlation is higher, and explain briefly. (Data are from a longitudinal study of growth.)
- Height at age 4 and height at age 18, height at age 16 and height at age 18.
 - Height at age 4 and height at age 18, weight at age 4 and weight at age 18.
 - Height and weight at age 4, height and weight at age 18.
4. An investigator collected data on heights and weights of college students; results can be summarized as follows.

	Average	SD
Men's height	70 inches	3 inches
Men's weight	144 pounds	21 pounds
Women's height	64 inches	3 inches
Women's weight	120 pounds	21 pounds

The correlation coefficient between height and weight for the men was about 0.60; for the women, it was about the same. If you take the men and women together, the correlation between height and weight would be _____.

just about 0.60 somewhat lower somewhat higher

Choose one option, and explain briefly.

5. A number is missing in each of the data sets below. If possible, fill in the blank to make r equal to 1. If this is not possible, say why not.

(a)		(b)	
x	y	x	y
1	1	1	1
2	3	2	3
2	3	3	4
4	-	4	-

6. A computer program prints out r for the two data sets shown below. Is the program working correctly? Answer yes or no, and explain briefly.

(i)		(ii)	
x	y	x	y
1	2	1	5
2	1	2	4
3	4	3	7
4	3	4	6
5	7	5	10
6	5	6	8
7	6	7	9

$r = 0.8214$ $r = 0.7619$

7. In 1910, Hiram Johnson entered the California gubernatorial primaries. For each county, data are available to show the percentage of native-born Americans in that county, as well as the percentage of the vote for Johnson. A

political scientist calculated the correlation between these percentages.¹¹ It is 0.5. Is this a fair measure of the extent to which "Johnson received native, as opposed to immigrant, support?" Answer yes or no, and explain briefly.

8. For women age 25 and over in the U.S. in 2005, the relationship between age and educational level (years of schooling completed) can be summarized as follows:¹²

$$\begin{aligned} \text{average age} &\approx 50 \text{ years, } SD \approx 16 \text{ years} \\ \text{average ed. level} &\approx 13.2 \text{ years, } SD \approx 3.0 \text{ years, } r \approx -0.20 \end{aligned}$$

True or false, and explain: as you get older, you become less educated. If this statement is false, what accounts for the negative correlation?

9. At the University of California, Berkeley, Statistics 2 is a large lecture course with small discussion sections led by teaching assistants. As part of a study, at the second-to-last lecture one term, the students were asked to fill out anonymous questionnaires rating the effectiveness of their teaching assistants (by name), and the course, on the scale

1 2 3 4 5
poor fair good very good excellent

The following statistics were computed.

- The average rating of the assistant by the students in each section.
- The average rating of the course by the students in each section.
- The average score on the final for the students in each section.

Results are shown below (sections are identified by letter). Draw a scatter diagram for each pair of variables—there are three pairs—and find the correlations.

Section	Ave. rating of assistant	Ave. rating of course	Ave. score on final
A	3.3	3.5	70
B	2.9	3.2	64
C	4.1	3.1	47
D	3.3	3.3	63
E	2.7	2.8	69
F	3.4	3.5	69
G	2.8	3.6	69
H	2.1	2.8	63
I	3.7	2.8	53
J	3.2	3.3	65
K	2.4	3.3	64

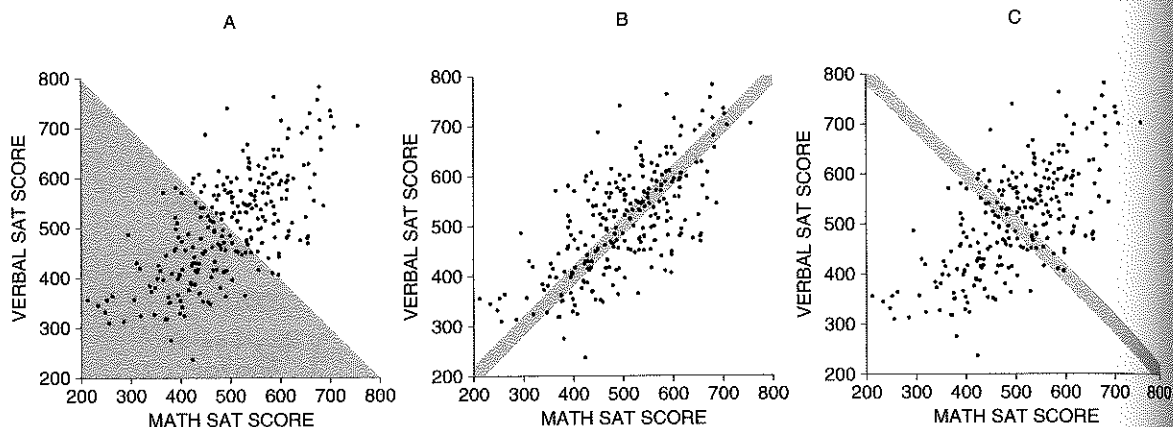
The data are section averages. Since the questionnaires were anonymous, it was not possible to link up student ratings with scores on an individual basis. Student ability may be a confounding factor. However, controlling for pre-test results turned out to make no difference in the analysis.¹³ Each assistant taught one section. True or false, and explain:

6. REVIEW EXERCISES

Review exercises may cover material from previous chapters.

1. Shown below is a scatter diagram for Math and Verbal SAT scores for graduating seniors at a certain high school. Three areas are shaded. Match the area with the description. (One description will be left over.)

- (i) Total score (Math + Verbal) is below 800.
- (ii) Total score (Math + Verbal) is around 800.
- (iii) Math score is about equal to Verbal score.
- (iv) Math score is less than Verbal score.



2. In a study of the stability of IQ scores, a large group of individuals is tested once at age 18 and again at age 35. The following results are obtained.

age 18: average score ≈ 100 , $SD \approx 15$

age 35: average score ≈ 100 , $SD \approx 15$, $r \approx 0.80$

- (a) Estimate the average score at age 35 for all the individuals who scored 115 at age 18.
 - (b) Predict the score at age 35 for an individual who scored 115 at age 18.
3. Pearson and Lee obtained the following results in a study of about 1,000 families:

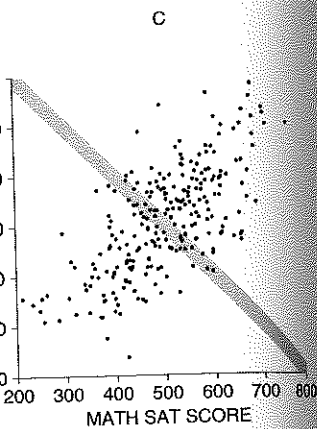
average height of husband ≈ 68 inches, $SD \approx 2.7$ inches

average height of wife ≈ 63 inches, $SD \approx 2.5$ inches, $r \approx 0.25$

Predict the height of a wife when the height of her husband is

- (a) 72 inches (b) 64 inches (c) 68 inches (d) unknown
4. In one study, the correlation between the educational level of husbands and wives in a certain town was about 0.50; both averaged 12 years of schooling completed, with an SD of 3 years.⁷

cores for gradu-
Match the area



Individuals is tested
e obtained.

≈ 0.80

Individuals who scored

red 115 at age 18.

ty of about 1,000

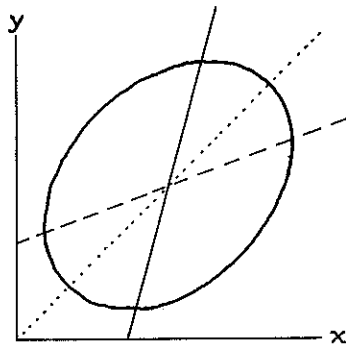
hes
hes, $r \approx 0.25$

d is

(d) unknown

el of husbands and
years of schooling

- (a) Predict the educational level of a woman whose husband has completed 18 years of schooling.
 - (b) Predict the educational level of a man whose wife has completed 15 years of schooling.
 - (c) Apparently, well-educated men marry women who are less well educated than themselves. But the women marry men with even less education. How is this possible?
5. An investigator measuring various characteristics of a large group of athletes found that the correlation between the weight of an athlete and the amount of weight that athlete could lift was 0.60. True or false, and explain:
- (a) On the average, an athlete can lift 60% of his body weight.
 - (b) If an athlete gains 10 pounds, he can expect to lift an additional 6 pounds.
 - (c) The more an athlete weighs, on the average the more he can lift.
 - (d) The more an athlete can lift, on the average the more he weighs.
 - (e) 60% of an athlete's lifting ability can be attributed to his weight alone.
6. Three lines are drawn across the scatter diagram below. One is the SD line, one is the regression line for y on x , and one is the regression line for x on y . Which is which? Why? (The "regression line for y on x " is used to predict y from x .)



7. A doctor is in the habit of measuring blood pressures twice. She notices that patients who are unusually high on the first reading tend to have somewhat lower second readings. She concludes that patients are more relaxed on the second reading. A colleague disagrees, pointing out that the patients who are unusually low on the first reading tend to have somewhat higher second readings, suggesting they get more nervous. Which doctor is right? Or perhaps both are wrong? Explain briefly.
8. A large study was made on the blood-pressure problem discussed in the previous exercise. It found that first readings average 130 mm, and second readings average 120 mm; both SDs were about 15 mm. Does this support either doctor's argument? Or is it the regression effect? Explain.