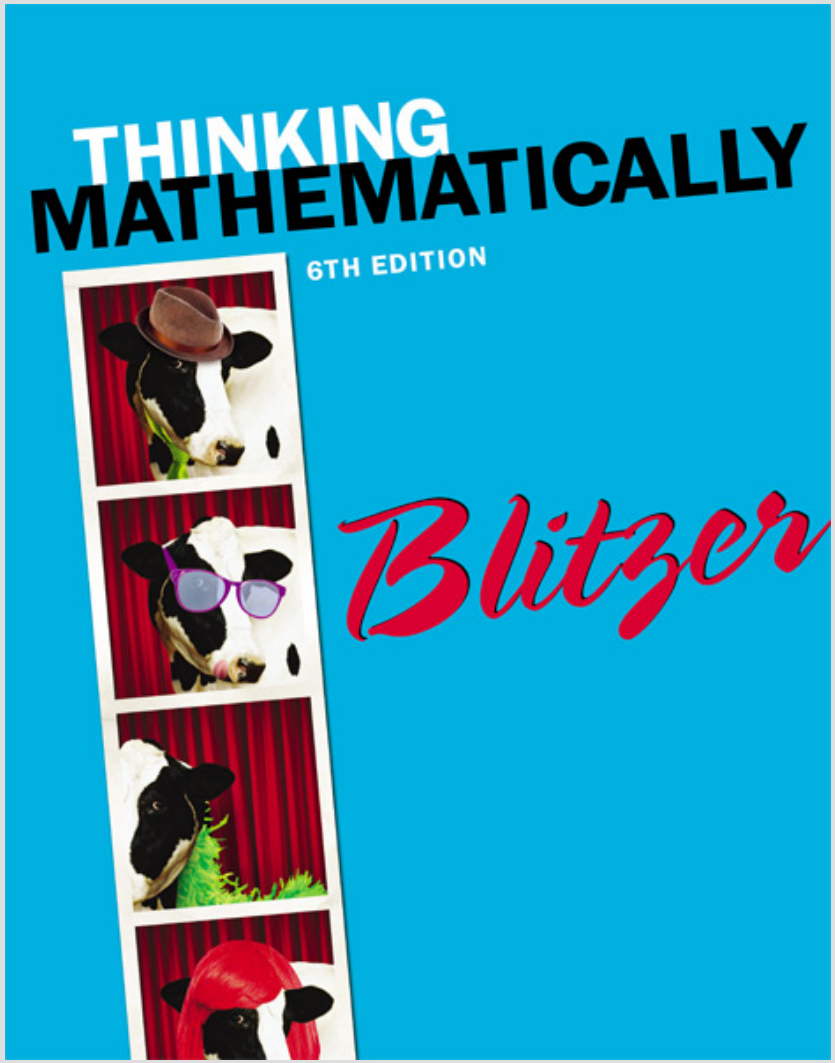
CHAPTER 12

Statistics



12.6

Scatter Plots, Correlation, and Regression Lines

Objectives

1. Make a scatter plot for a table of data items.
2. Interpret information given in a scatter plot.
3. Compute the correlation coefficient.
4. Write the equation of the regression line.
5. Use a sample's correlation coefficient to determine whether there is a correlation in the population.

Scatter Plots and Correlation

A **scatter plot** is a collection of data points, one data point per person or object.

Can be used to determine whether two quantities are related.

Correlation

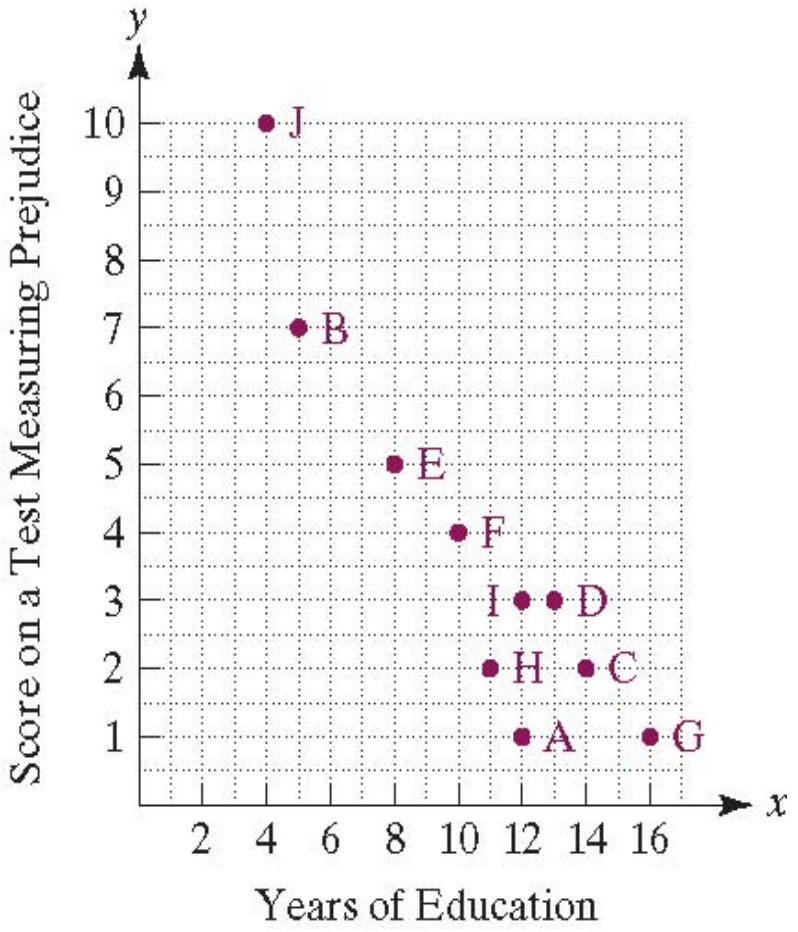
a clear relationship between two quantities.

used to determine if there is a relationship between two variables and, if so, the strength and direction of that relationship.

Scatter Plots and Correlation

The scatter plot shows a downward trend among the data points, with some exceptions.

People with increased education tend to have a lower score on the test measuring prejudice.



Correlation and Causal Connections

Although the scatter plot shows a correlation between education and prejudice, we cannot conclude that increased education causes a person's level of prejudice to decrease.

1. The correlation could be simply a coincidence.
2. Education usually involves classrooms with a variety of different kinds of people. Increased exposure to diversity in the classroom might be an underlying cause.
3. Education requires people to look at new ideas and see things in different ways. Thus, education causes one to be more tolerant and less prejudiced.

Regression Lines

Regression line is a line that best fits the data points in a scatter plot.

Correlation Coefficients

Correlation coefficient, designated by ***r***, is a measure that is used to describe the strength and direction of a relationship between variables whose data points lie on or near a line. The relationship is:

Positively correlated if they tend to increase or decrease together.

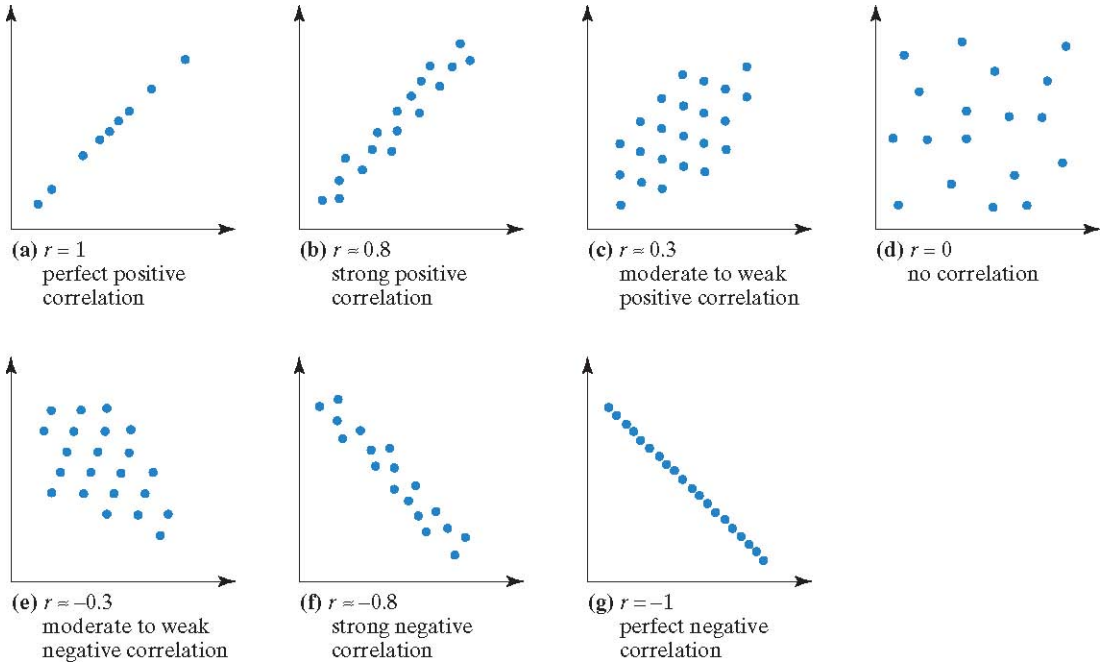
Negatively correlated if one variable tends to decrease while the other increases.

Correlation Coefficients

Perfect positive correlation in which all points lie precisely on the regression line that rises from left to right.

Perfect negative correlation in which all points in the scatter point lie precisely on the regression line that falls from left to right.

Scatter Plots and Correlation Coefficients



Computing the Correlation Coefficient

The following formula is used to calculate the correlation coefficient, r:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \cdot \sqrt{n(\sum y^2) - (\sum y)^2}}$$

Computing the Correlation Coefficient

In the formula:

n = the number of data points, (x, y)

Σx = the sum of the x-values

Σy = the sum of the y-values

Σxy = the sum of the product of x and y in each pair

Σx^2 = the sum of the squares of the x-values

Σy^2 = the sum of the squares of the y-values

$(\Sigma x)^2$ = the square of the sum of the x-values

$(\Sigma y)^2$ = the square of the sum of the y-values

Example: Computing the Correlation Coefficient

Shown below are the data involving the number of years of school, *x*, completed by 10 randomly selected people and their scores on a test measuring prejudice, *y*. Determine the correlation coefficient between years of education and scores on a prejudice test.

Respondent	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Years of education (x)	12	5	14	13	8	10	16	11	12	4
Score on prejudice test (y)	1	7	2	3	5	4	1	2	3	10

Example continued

x	y	xy	x^2	y^2
12	1	12	144	1
5	7	35	25	49
14	2	28	196	4
13	3	39	169	9
8	5	40	64	25
10	4	40	100	16
16	1	16	256	1
11	2	22	121	4
12	3	36	144	9
4	10	40	16	100
$\Sigma x = 105$	$\Sigma y = 38$	$\Sigma xy = 308$	$\Sigma x^2 = 1235$	$\Sigma y^2 = 218$

Example continued

$(\Sigma x)^2 = 11{,}025 \qquad (\Sigma y)^2 = 1444$

Calculating r:

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

$$= \frac{10(308) - 105(38)}{\sqrt{10(1235) - 11{,}025}\sqrt{10(218) - 1444}}$$

$$= \frac{-910}{\sqrt{1325}\sqrt{736}}$$

$$\approx -0.92$$

Formula

WRITING THE EQUATION OF THE REGRESSION LINE BY HAND

The equation of the regression line is

$$y = mx + b,$$

where

$$m = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \quad \text{and} \quad b = \frac{\Sigma y - m(\Sigma x)}{n}.$$

Example: Writing the Equation of the Regression Line

To find the regression line for the previous data:

$\Sigma x = 105 \quad \Sigma y = 38 \quad \Sigma xy = 308 \quad \Sigma x^2 = 1235 \quad \Sigma y^2 = 218$

$$m = \frac{10(308) - 105(38)}{10(1235) - (105)^2} = \frac{-910}{1325} \approx -0.69$$

$$b = \frac{38 - (-0.69)(105)}{10} = \frac{110.45}{10} \approx 11.05$$

The equation of the regression line is $y = -0.69x + 11.05$