10.2: The Linear Correlation Coefficient

Learning Objectives

To learn what the linear correlation coefficient is, how to compute it, and what it tells us about the relationship between two variables x and y

Figure $10.2.1$ illustrates linear relationships between two variables x and y of varying strengths. It is visually apparent that in the situation in panel (a), x could serve as a useful predictor of y, it would be less useful in the situation illustrated in panel (b), and in the situation of panel (c) the linear relationship is so weak as to be practically nonexistent. The *linear correlation coefficient* is a number computed directly from the data that measures the strength of the linear relationship between the two variables x and y.

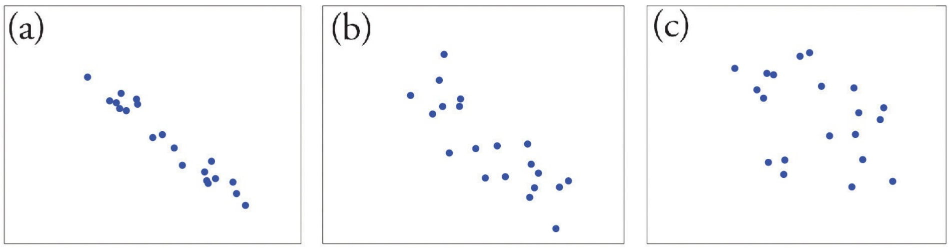

Figure 10.2.1: Linear Relationships of Varying Strengths

Definition: linear correlation coefficient

The *linear correlation coefficient* for a collection of n pairs x of numbers in a sample is the number r given by the formula

The linear correlation coefficient has the following properties, illustrated in Figure $10.2.2$

1. The value of r lies between -1 and 1, inclusive.
2. The sign of r indicates the direction of the linear relationship between x and y:
3. The size of $|r|$ indicates the strength of the linear relationship between x and y:
   1. If $|r|$ is near 1 (that is, if r is near either 1 or -1), then the linear relationship between x and y is strong.
   2. If $|r|$ is near 0 (that is, if r is near 0 and of either sign). then the linear relationship between x and y is weak.

so that

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{2.44.583}{\sqrt{(46.916)(1690.916)}} = 0.868$$

The number quantifies what is visually apparent from Figure $10.2.2$ weights tends to increase linearly with height (r is positive) and although the relationship is not perfect, it is reasonably strong (r is near 1).

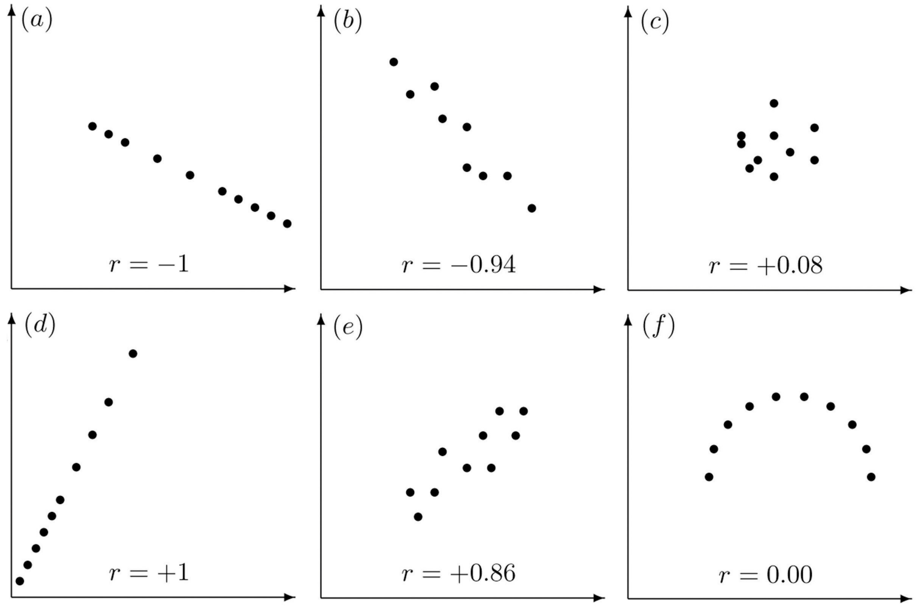

Figure 10.2.2: Linear Correlation Coefficient r. Pay particular attention to panel (f), which shows a perfectly deterministic relationship between x and y, but $f=0$ because the relationship is not linear. (In this particular case the points lie on the top half of a circle.)

✔ Example 10.2.1

Compute the linear correlation coefficient for the height and weight pairs plotted in Figure $10.2.2$

Solution:

Even for small data sets like this one computations are too long to do completely by hand. In actual practice the data are entered into a calculator or computer and a statistics program is used. In order to clarify the meaning of the formulas we will display the data and related quantities in tabular form. For each

	x	y	x^2	y^2
68	151	4624	10268	22801
69	146	4761	10074	21316
70	157	4900	10990	24649
70	164	4900	11480	26896
71	171	5041	12141	29241
72	160	5184	11520	25600
72	163	5184	11736	26569
72	180	5184	12960	32400
73	170	5329	12410	28900
73	175	5329	12775	30625
74	178	5476	13172	31684
75	188	5625	14100	35344
859	2003	61537	143626	336025

Key Takeaway

- The linear correlation coefficient measures the strength and direction of the linear relationship between two variables x and y.
- The sign of the linear correlation coefficient indicates the direction of the linear relationship between x and y.
- When r is near 1 or -1 the linear relationship is strong; when it is near 0 the linear relationship is weak.