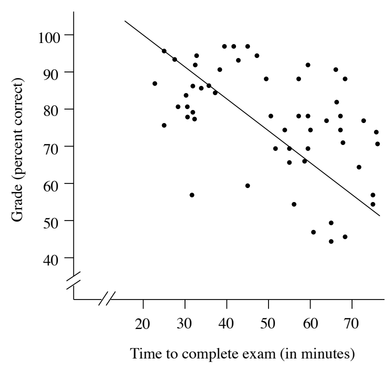
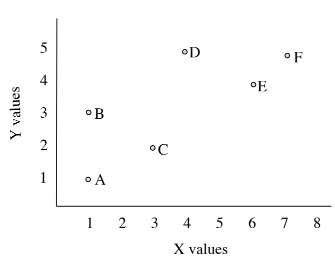
Chapter 16: Correlation



Person	X	Y
A	1	1
B	1	3
C	3	2
D	4	5
E	6	4
F	7	5

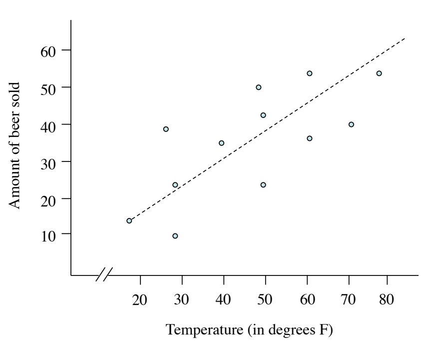


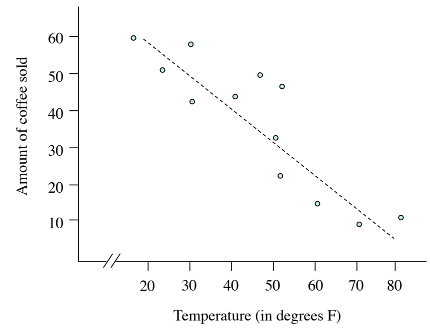
3 Characteristics of a Correlation:

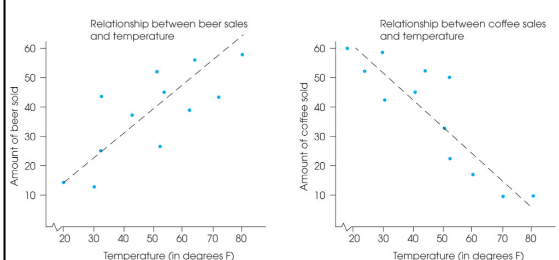
- Direction of relationship
- Form of the relation
- Degree of the relationship

Correlations: Measuring and Describing Relationships (cont.)

- The **direction** of the relationship is measured by the sign of the correlation (+ or -). A positive correlation means that the two variables tend to change in the same direction; as one increases, the other also tends to increase. A negative correlation means that the two variables tend to change in opposite directions; as one increases, the other tends to decrease.

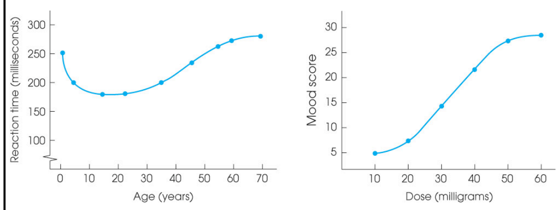


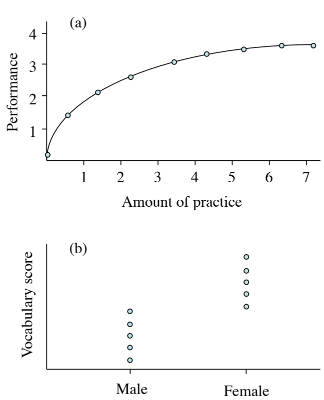




Correlations: Measuring and Describing Relationships (cont.)

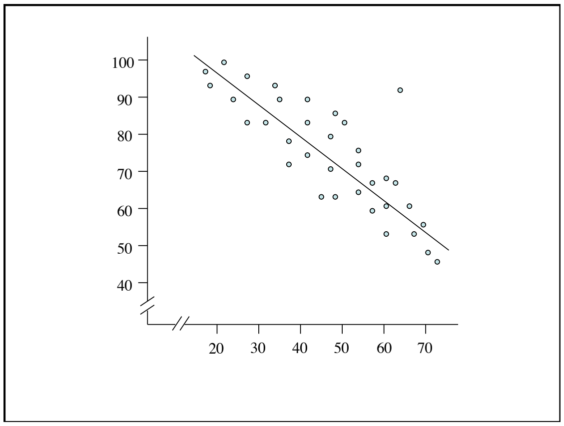
- The most common **form** of relationship is a straight line or linear relationship which is measured by the Pearson correlation.



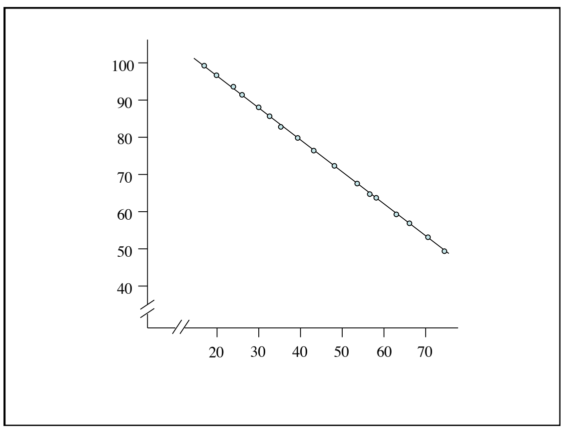


Correlations: Measuring and Describing Relationships (cont.)

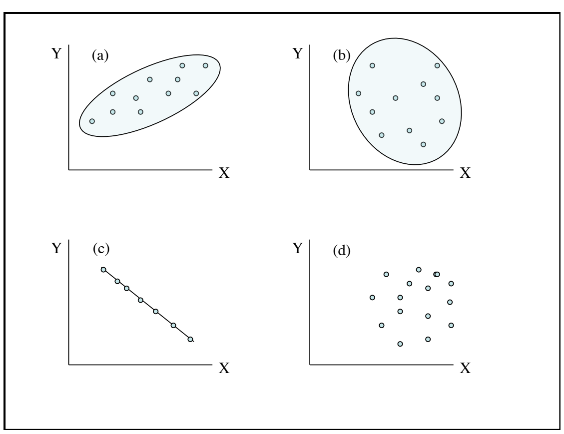
- The **degree** of relationship (the strength or consistency of the relationship) is measured by the numerical value of the correlation. A value of 1.00 indicates a perfect relationship and a value of zero indicates no relationship.



100
90
80
70
60
50
40
20
30
40
50
60
70



100
90
80
70
60
50
40
20
30
40
50
60
70



Y
(a)
X
Y
(b)
X
Y
(c)
X
Y
(d)
X

Where and Why Correlations are Used:

- Prediction
- Validity
- Reliability
- Theory Verification

Correlations: Measuring and Describing Relationships (cont.)

- To compute a correlation you need two scores, X and Y, for each individual in the sample.
- The Pearson correlation requires that the scores be numerical values from an interval or ratio scale of measurement.
- Other correlational methods exist for other scales of measurement.

The Pearson Correlation

- The **Pearson correlation** measures the direction and degree of linear (straight line) relationship between two variables.
- To compute the Pearson correlation, you first measure the variability of X and Y scores separately by computing SS for the scores of each variable (SS_X and SS_Y).
- Then, the covariability (tendency for X and Y to vary together) is measured by the sum of products (SP).
- The Pearson correlation is found by computing the ratio, $SP/\sqrt{(SS_X)(SS_Y)}$.

The Pearson Correlation (cont.)

- Thus the Pearson correlation is comparing the amount of covariability (variation from the relationship between X and Y) to the amount X and Y vary separately.
- The magnitude of the Pearson correlation ranges from 0 (indicating no linear relationship between X and Y) to 1.00 (indicating a perfect straight-line relationship between X and Y).
- The correlation can be either positive or negative depending on the direction of the relationship.

The Pearson Correlation

$$r = \frac{degree\ to\ which\ x\ and\ y\ vary\ together}{degree\ to\ which\ x\ and\ y\ vary\ separately}$$

$$r = \frac{covariability\ of\ x\ and\ y}{variability\ of\ x\ and\ y\ separately}$$

$$r = \frac{SP}{\sqrt{(SSx)(SSy)}} \qquad SP = \Sigma(x - \bar{x})(y - \bar{y})$$

$$SP = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$

Computational Examples

Understanding & Interpreting the Pearson Correlation

- Correlation is not causation
- Correlation greatly affected by the range of scores represented in the data
- One or two extreme data points (outliers) can dramatically affect the value of the correlation
- How accurately one variable predicts the other—the strength of a relation

The Spearman Correlation

- The **Spearman correlation** is used in two general situations:
  (1) It measures the relationship between two ordinal variables; that is, X and Y both consist of ranks.
  (2) It measures the consistency of direction of the relationship between two variables. In this case, the two variables must be converted to ranks before the Spearman correlation is computed.

The Spearman Correlation (cont.)

The calculation of the Spearman correlation requires:

1. Two variables are observed for each individual.
2. The observations for each variable are rank ordered. Note that the X values and the Y values are ranked separately.
3. After the variables have been ranked, the Spearman correlation is computed by either:
   a. Using the Pearson formula with the ranked data.
   b. Using the special Spearman formula (assuming there are few, if any, tied ranks).

The Point-Biserial Correlation and the Phi Coefficient

- The Pearson correlation formula can also be used to measure the relationship between two variables when one or both of the variables is dichotomous.
- A dichotomous variable is one for which there are exactly two categories: for example, men/women or succeed/fail.

The Point-Biserial Correlation and the Phi Coefficient (cont.)

With either one or two dichotomous variables the calculation of the correlation precedes as follows:

1. Assign numerical values to the two categories of the dichotomous variable(s). Traditionally, one category is assigned a value of 0 and the other is assigned a value of 1.
2. Use the regular Pearson correlation formula to calculate the correlation.

The Point-Biserial Correlation and the Phi Coefficient (cont.)

- In situations where one variable is dichotomous and the other consists of regular numerical scores (interval or ratio scale), the resulting correlation is called a **point-biserial correlation**.
- When both variables are dichotomous, the resulting correlation is called a **phi-coefficient**.

The Point-Biserial Correlation and the Phi Coefficient (cont.)

- The point-biserial correlation is closely related to the independent-measures t test introduced in Chapter 10.
- When the data consists of one dichotomous variable and one numerical variable, the dichotomous variable can also be used to separate the individuals into two groups.
- Then, it is possible to compute a sample mean for the numerical scores in each group.

The Point-Biserial Correlation and the Phi Coefficient (cont.)

- In this case, the independent-measures t test can be used to evaluate the mean difference between groups.
- If the effect size for the mean difference is measured by computing r^2 (the percentage of variance explained), the value of r^2 will be equal to the value obtained by squaring the point-biserial correlation.

TABLE 16.3

The same data are organized in two different formats. On the left-hand side, the data appear as two separate samples appropriate for an independent-measures *t* hypothesis test. On the right-hand side, the same data are shown as a single sample, with two scores for each individual: the original high school grade and a dichotomous score (*Y*) that identifies the condition (Seasame Street or not) in which the participant is located. The data on the right are appropriate for a point-biserial correlation.

Data for the Independent-Measures *t*. Two separate samples, each with $n = 10$ scores.

Average High School Grade			
Watched Seasame Street		Did Not Watch Sesame Street	
86	99	90	79
87	97	89	83
91	94	82	86
97	89	83	81
98	92	85	92
$n = 10$		$n = 10$	
$M = 93$		$M = 85$	
$SS = 200$		$SS = 160$	

Data for the Point-Biserial Correlation. Two scores, *X* and *Y* for each of the $n = 20$ participants.

Participant	Grade *X*	Condition *Y*
A	86	1
B	87	1
C	91	1
D	97	1
E	98	1
F	99	1
G	97	1
H	94	1
I	89	1
J	92	1
K	90	0
L	89	0
M	82	0
N	83	0
O	85	0
P	79	0
Q	83	0
R	86	0
S	81	0
T	92	0