Chapter 6 – Describing Associations of Two Variables Graphically
Section 2 – Determining the Four Characteristics of an Association

Objectives
1. Determine the shape of a distribution.
2. Determine the strength of a distribution.
3. Compute and interpret the correlation coefficient.
4. Determine whether there are any outliers and what to do with them.
5. Determine the four characteristics of an association in the correct order.
6. Explain why strong (or weak) association does not guarantee causation.

Vocabulary
1. linear/nonlinear association
2. shape of the association
3. exact/strong/weak association
4. positive/negative association
5. correlation coefficient
6. lurking variable

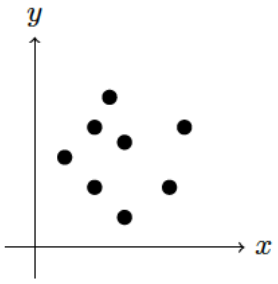
Lesson/Activity

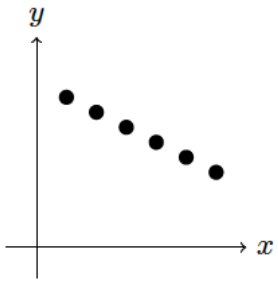
OBJECTIVE 1 – Determine the shape of a distribution.

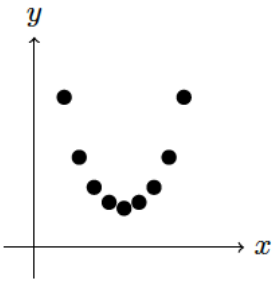
- If the points of a scatterplot lie close to (or on) a line, we say the variables are **linearly associated** and that there is a **linear association**. [Draw a figure.]
- If the points of a scatterplot lie close to (or on) a curve that is not a line, we say there is a **nonlinear association**. [Draw a figure.]
- If no curve comes close to all the points of a scatterplot, we say there is **no association**. [Draw a figure.]

For any pair of numerical variables, there is either a linear association, a nonlinear association, or no association. We will refer to the **shape** of an association as being one of these three types.

1. For each of the following scatterplots, determine whether there is a linear association, nonlinear association, or no association.







OBJECTIVE 2 – Determine the strength of a distribution.

- If a curve passes through all the points of a scatterplot, we say there is an **exact association** with respect to the curve. [Draw a figure.]
- If a curve comes close to all the points, we say there is a **strong association** with respect to the curve. [Draw a figure.]
- If a curve comes somewhat close to all the points, we say there is a **weak association** with respect to the curve. [Draw a figure.]

2. The carbohydrates, calories, and fat for 29 pizzas made by six of the leading pizza companies are shown in the following table.

Carbohydrates (g)	Calories	Fat (g)	Carbohydrates (g)	Calories	Fat (g)
39	368	17	35	342	16
36	308	13	37	359	17
20	259	15	26	297	16
39	348	15	32	305	13
36	311	13	53	499	21
20	249	14	64	549	21
40	357	16	42	410	19
27	279	13	63	579	24
28	246	9	43	395	17
30	276	11	29	260	9
37	338	15	37	277	9
41	367	16	30	269	11
35	320	14	29	286	14
36	329	15	33	309	14
26	266	13			

Source: Domino's, Little Caesar's, Papa John's, Pizza Hut, Doiron Frozen, Kashi Frozen

a. Construct a scatterplot that compares carbohydrates and calories.
b. Construct a scatterplot that compares carbohydrates and fat.
c. For each of the scatterplots you constructed, describe the shape of the association.
d. Compare the strengths of the two associations.
e. For each of the scatterplots you constructed, identify whether the association is positive, negative, or neither.
f. Use an appropriate scatterplot to estimate the caloric contents of the pizzas with 20 g of carbohydrates.
g. Use an appropriate scatterplot to estimate the fat content of the pizzas with 20 g of carbohydrates.

OBJECTIVE 3 − Compute and interpret the correlation coefficient.
We know that we can measure the **spread** of a distribution for a single numerical variable using the range, standard deviation, or IQR. We now learn that we can measure the strength of the association between two numerical variables using the **correlation coefficient**, **r**.

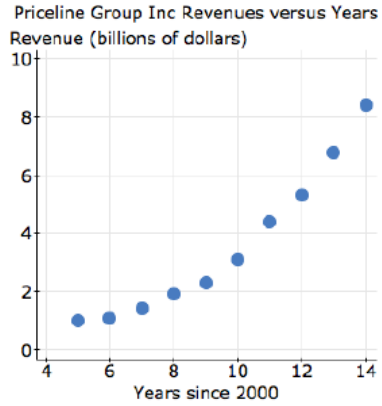
Properties of the Linear Correlation Coefficient

Assume **r** is the linear correlation coefficient for the association between two numerical variables. Then

- The values of **r** fall between −1 and 1, inclusive.
- If **r** is positive, then the variables are positively associated.
- If **r** is negative, then the variables are negatively associated.
- If r = 0, there is no *linear* association.
- The larger the value of |r|, the stronger the linear association will be.
- If r = 1, then the points lie exactly on a line and the association is positive.
- If r = −1, then the points lie exactly on a line and the association is negative.

[Draw figures to illustrate all of the above concepts.]

- If r = 0, then there is no linear association but there may be a *nonlinear* association.
- To determine the shape and strength of an association, we should inspect a scatterplot of the data as well as compute r. For example, even though r ≈ 0.96 for the data described by the following scatterplot, the points lie much closer to a curve than they do to a line.

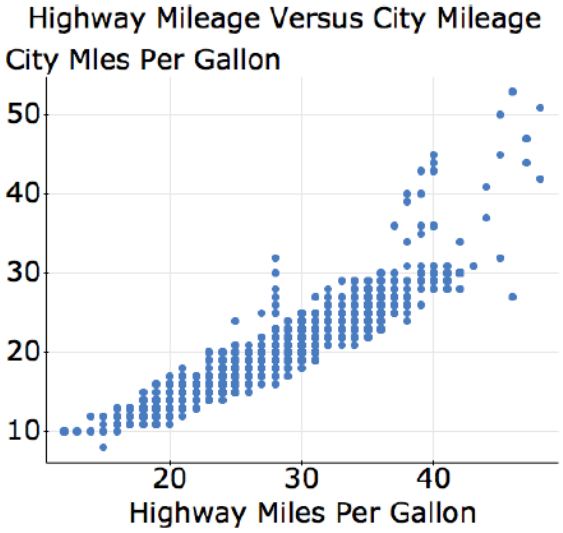


OBJECTIVE 4 – Determine whether there are any outliers and what to do with them.

- For an association between two numerical variables, a point is an outlier if it does not fit the overall pattern of the points in the scatterplot. [Draw a figure.]

- The correlation coefficient is very sensitive to outliers. [Draw scatterplots to show how r can change from 0 to approximately 0.95 by adding an outlier. See Figs. 43 and 44 on page 387 of the textbook.]

3. A scatterplot comparing the highway mileages and city mileages of 1195 types of cars is displayed in the following figure.



   a. Identify the outliers. Should they be removed? If yes, for what purpose?

   b. Assume the outliers have been removed. For the data that remain,
      i. Determine the shape of the association.
      ii. Determine the strength of the association.
      iii. Determine whether the association is positive, negative, or neither.

OBJECTIVE 5 – Determine the four characteristics of an association in the correct order.

Order of Determining the Four Characteristics of an Association

We determine the four characteristics of an association in the following order:

1. Identify all **outliers**.
   a. For outliers that stem from errors in measurement or recording, correct the errors if possible. If the errors cannot be corrected, remove the outliers.
   b. For other outliers, determine whether they should be analyzed in a separate study.
2. Determine the **shape** of the association.
3. If the shape is linear, then based on **r** and the scatterplot, determine the **strength**.
   If the shape is nonlinear, then based on the scatterplot, determine the **strength**.
4. Determine the **direction**. In other words, determine whether the association is positive, negative, or neither.

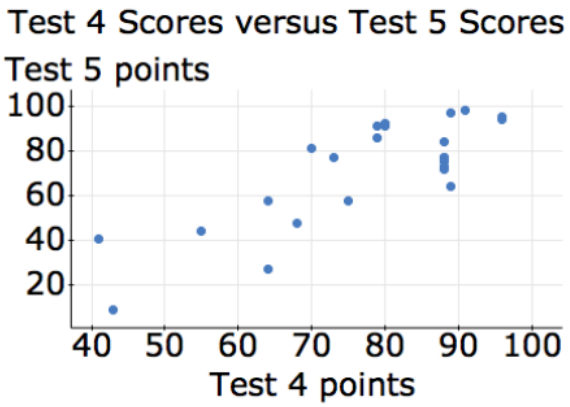
OBJECTIVE 6 – Explain why strong (or weak) association does not guarantee causation.

Definition: Lurking variable

A **lurking variable** is a variable that causes both the explanatory and response variables to change during the study.

A strong association between two variables does not guarantee that a change in the explanatory variable will cause a change in the response variable.

4. The scatterplot in the following figure compares the scores of Test 4 and Test 5 for a calculus course taught by the author.



   a. Describe the four characteristics of the association.
   b. Does a higher score on Test 4 cause a higher score on Test 5? If no, describe at least one possible lurking variable.
   c. Based on the fairly strong, positive association between Test 4 and Test 5 scores, a student concludes that there must have been a fairly strong, positive association between any pair of tests for the course. What would you tell the student?

Homework/Assessment

1, 3, 7, 9, 11, 15, 17, 19, 21, 23, 29, 33, 35, 37, 44, 47