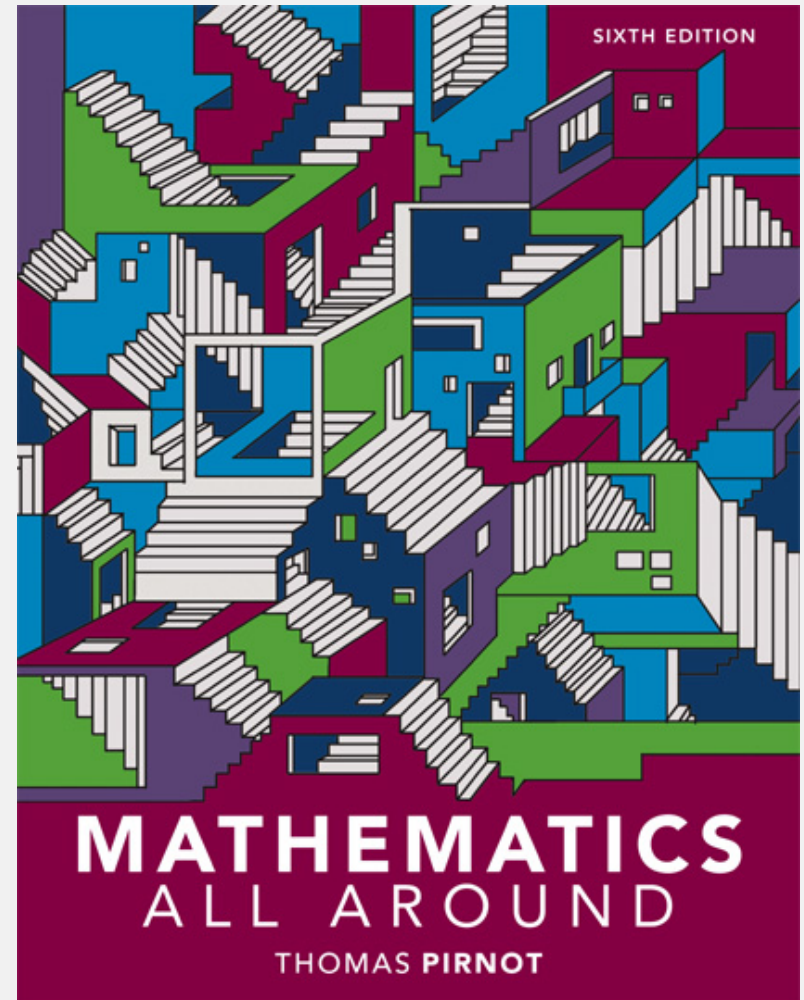


14.1

Descriptive Statistics



14.1 Organizing and Visualizing Data

- Explain the difference between a sample and a population.
- Organize data in a frequency table.
- Use a variety of methods to represent data visually.

Populations and Samples

Statistics is an area of mathematics in which we are interested in gathering, organizing, analyzing, and making predictions from numerical information called *data*.

The set of all items under consideration is called the population. Often only a *sample* or subset of the population is considered.

We will describe a sample as *biased* if it does not accurately reflect the population as a whole with regard to the data that we are gathering.

Populations and Samples

Bias could occur because of the way in which we decide how to choose the people to participate in the survey. This is called selection bias.

Another issue that can affect the reliability of a survey is the way we ask the questions, which is called leading-question bias.

Frequency Tables

We refer to a collection of information as **data**. Data can be **qualitative** or **quantitative**.

- **Qualitative data** consists of information about a characteristic such as eye color or gender; e.g. Marriage Patterns of U.S. adults.
- **Quantitative data** gives a meaningful numerical measurement – e.g. Gasoline Prices for the years 2015, 2016, 2017, and 2018.

Frequency Tables

Frequency table is one method used to organize data. To construct a frequency distribution, we determine the count, or **frequency (f)**, with which each value occurs.

We show the percent of the time that each item occurs in a frequency distribution using a ***relative frequency distribution (rf)***.

We often present a frequency distribution as a ***frequency table*** where we list the values in one column and the frequencies of the values in another column.

Example: Using Tables to Summarize TV Program Evaluations

Assume that 25 viewers were surveyed to evaluate a preview of an episode of *NCIS*.

The possible evaluations are (E)xcellent, (A)bove average, a(V)erage, (B)elow average, (P)oor.

After the show, the 25 evaluations were as follows:

A, V, V, B, P, E, A, E, V, V, A, E, P, B, V, V, A, A, A, E, B, V, A, B, V.

Construct a frequency table and a relative frequency table for this list of evaluations.

Example: Drawing a Bar Graph of the Viewer Evaluation Data

- a) Draw a bar graph of the frequency distribution of the *NCIS* viewers' responses from the table.

Evaluation	Frequency
E	4
A	7
V	8
B	4
P	2
Total	25

Example: Using Tables to Summarize TV Program Evaluations (cont)

Solution

We organize the data in the frequency table shown below.

Evaluation	Relative Frequency
E	$\frac{4}{25} = 0.16$
A	$\frac{7}{25} = 0.28$
V	$\frac{8}{25} = 0.32$
B	$\frac{4}{25} = 0.16$
P	$\frac{2}{25} = 0.08$
Total	1.00

Example: Using Tables to Summarize TV Program Evaluations (cont)

We construct a relative frequency distribution for these data by dividing each frequency in the table by 25. For example, the relative frequency of E is $\frac{4}{25} = 0.16$.

Evaluation	Relative Frequency
E	$\frac{4}{25} = 0.16$
A	$\frac{7}{25} = 0.28$
V	$\frac{8}{25} = 0.32$
B	$\frac{4}{25} = 0.16$
P	$\frac{2}{25} = 0.08$
Total	1.00

Example: Grouping Data Values into Classes

Suppose 40 health care workers take an AIDS awareness test and earn the following scores:

79, 62, 87, 84, 53, 76, 67, 73, 82, 68,
82, 79, 61, 51, 66, 77, 78, 66, 86, 70,
76, 64, 87, 82, 61, 59, 77, 88, 80, 58,
56, 64, 83, 71, 74, 79, 67, 79, 84, 68

Construct a frequency table and a relative frequency table for these data.

Example: Grouping Data Values into Classes (cont)

Solution

Because there are so many different scores in this list, the data is grouped in ranges.

Range of Scores on Depression Awareness Test	Frequency
50–54	2
55–59	3
60–64	5
65–69	6
70–74	4
75–79	9
80–84	7
85–89	4
Total	40

Example: Grouping Data Values into Classes (cont)

We divide each count in the frequency column by 40. For example, in the row labeled 55–59, we divide 3 by 40 to get 0.075 in the third column.

Range of Scores on Depression Awareness Test	Frequency	Relative Frequency
50–54	2	0.05
55–59	3	$\frac{3}{40} = 0.075$
60–64	5	0.125
65–69	6	0.15
70–74	4	0.10
75–79	9	0.225
80–84	7	0.175
85–89	4	0.10
Total	40	1.00

Representing Data Visually

A bar graph is one way to visualize a frequency distribution. In drawing a bar graph, we specify the **classes on the horizontal axis** and the **frequencies on the vertical axis**.

If we are graphing a relative frequency distribution, then the heights of the bars correspond to the size of the relative frequencies. Graphing the relative frequencies, rather than the actual values in data sets, allows us to compare the distributions.

Example: Drawing a Bar Graph of the Viewer Evaluation Data

- a) Draw a bar graph of the frequency distribution of the *NCIS* viewers' responses from the table.

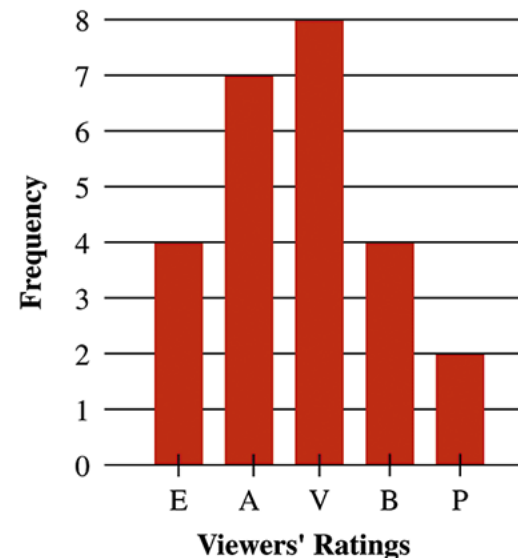
Evaluation	Frequency
E	4
A	7
V	8
B	4
P	2
Total	25

Example: Drawing a Bar Graph of the Viewer Evaluation Data (cont)

Solution

a) Because the highest frequency is 8, we labeled the vertical axis from 0 to 8. Next we drew five bars of heights 4, 7, 8, 4, and 2 to indicate the frequencies of the evaluations E, A, V, B, and P, as shown.

Evaluation	Frequency
E	4
A	7
V	8
B	4
P	2
Total	25



Example: Drawing a Bar Graph of the Viewer Evaluation Data (cont)

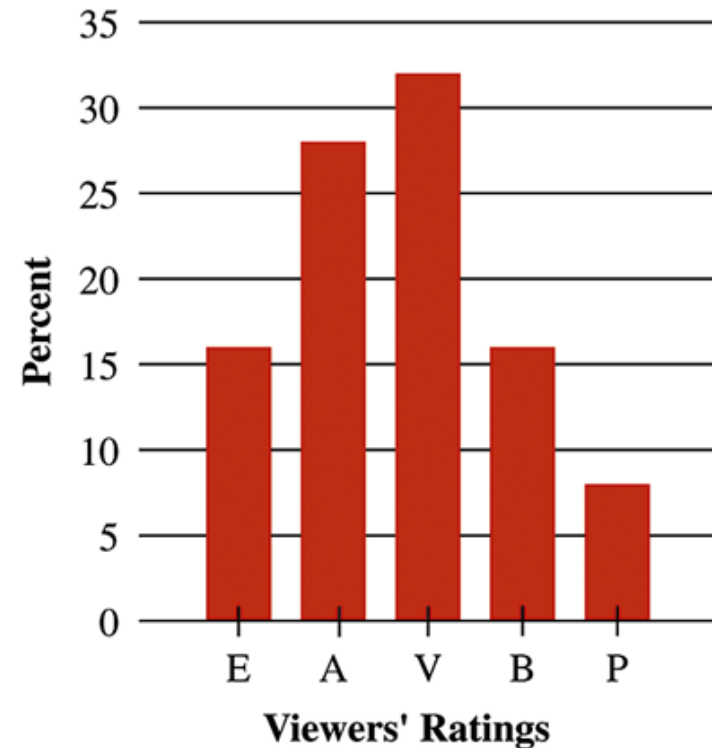
- b) Draw a bar graph of the relative frequency distribution of the *NC/S* viewers' responses summarized in the table.

Evaluation	Relative Frequency
E	$\frac{4}{25} = 0.16$
A	$\frac{7}{25} = 0.28$
V	$\frac{8}{25} = 0.32$
B	$\frac{4}{25} = 0.16$
P	$\frac{2}{25} = 0.08$
Total	1.00

Example: Drawing a Bar Graph of the Viewer Evaluation Data (cont)

b)

Evaluation	Relative Frequency
E	$\frac{4}{25} = 0.16$
A	$\frac{7}{25} = 0.28$
V	$\frac{8}{25} = 0.32$
B	$\frac{4}{25} = 0.16$
P	$\frac{2}{25} = 0.08$
Total	1.00



Representing Data Visually

A variable quantity that cannot take on arbitrary values is called **discrete**. Other quantities, called *continuous variables*, can take on arbitrary values.

The number of children in a family is an example of a discrete variable. Weight is an example of a **continuous** variable.

We use a special type of bar graph called a **histogram** to graph a frequency distribution when we are dealing with a continuous variable quantity or a variable quantity that is discrete, but has a very large number of different possible values.

Example: Drawing a Histogram to Represent an Increase in Exam Scores

A professor has collected data regarding the increase in test scores by her students after a ban on the use of all social media during class. The data, which includes some fractional values, is organized in the table. Compute the relative frequencies and then use them to draw a histogram.

Points Gained	Frequency
0 to 10	14
10+ to 20	23
20+ to 30	17
30+ to 40	8
40+ to 50	3
Total	65

Example: Drawing a Histogram to Represent an Increase in Exam Scores (cont)

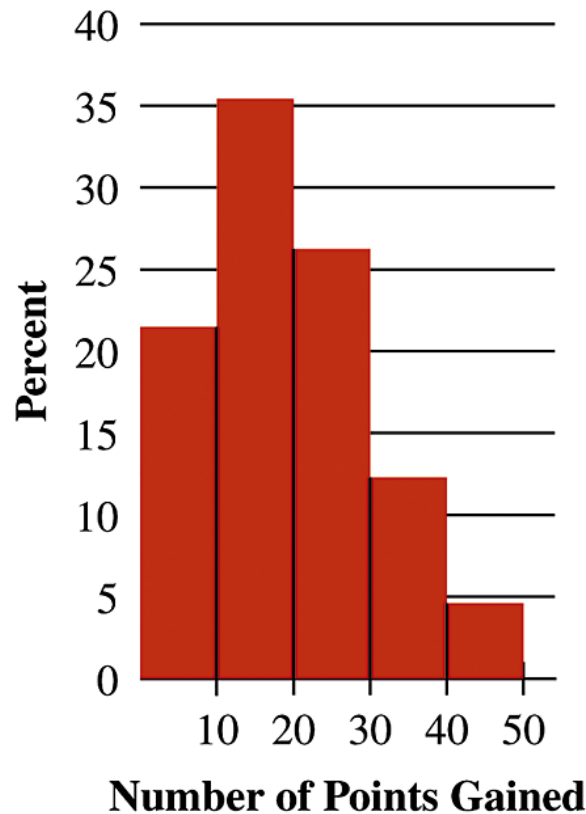
Solution

We first find the relative frequency distribution.

Points Gained	Frequency	Relative Frequency
0 to 10	14	0.215
10+ to 20	23	0.354
20+ to 30	17	0.262
30+ to 40	8	0.123
40+ to 50	3	0.046
Total	65	1.00

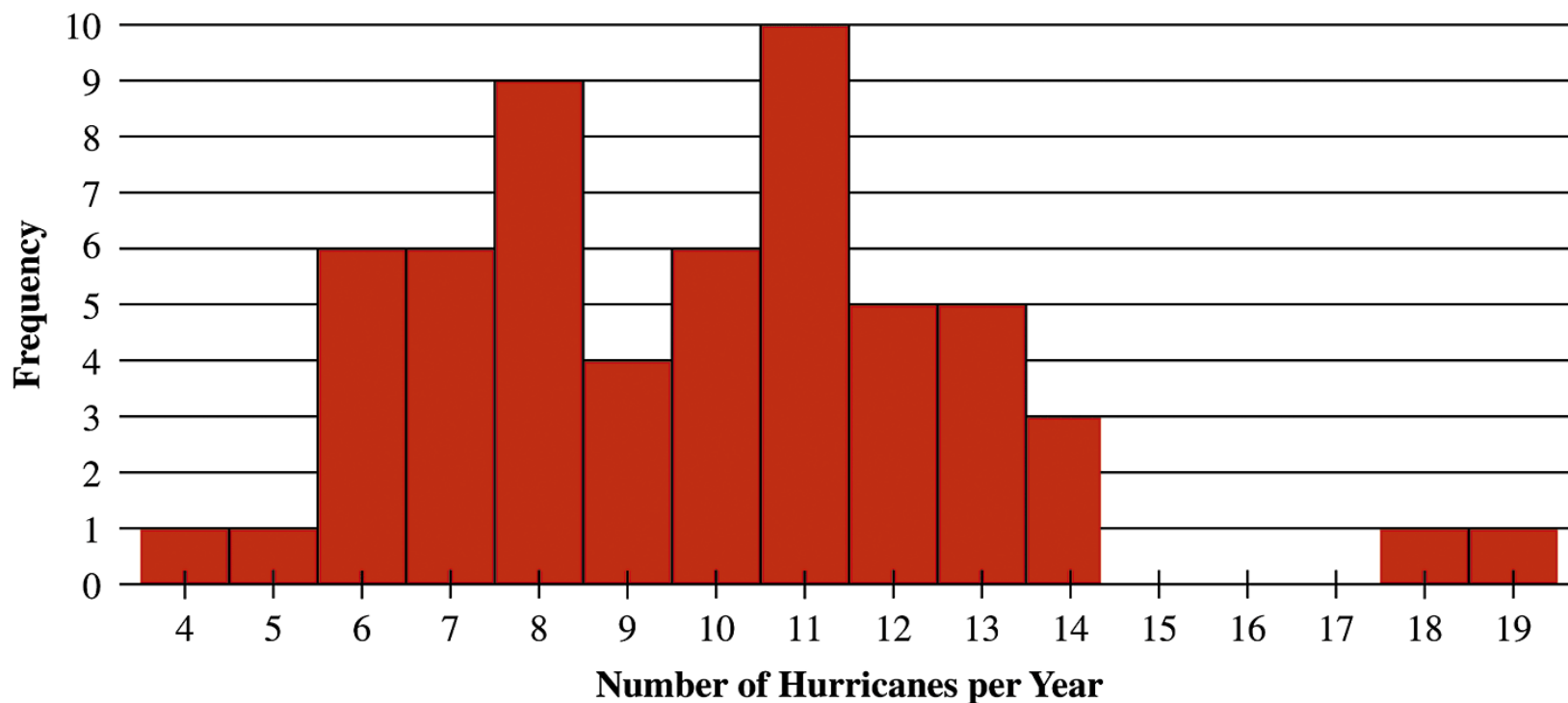
Example: Drawing a Histogram to Represent an Increase in Exam Scores (cont)

Draw the histogram exactly like a bar graph except that we do not allow spaces between the bars.



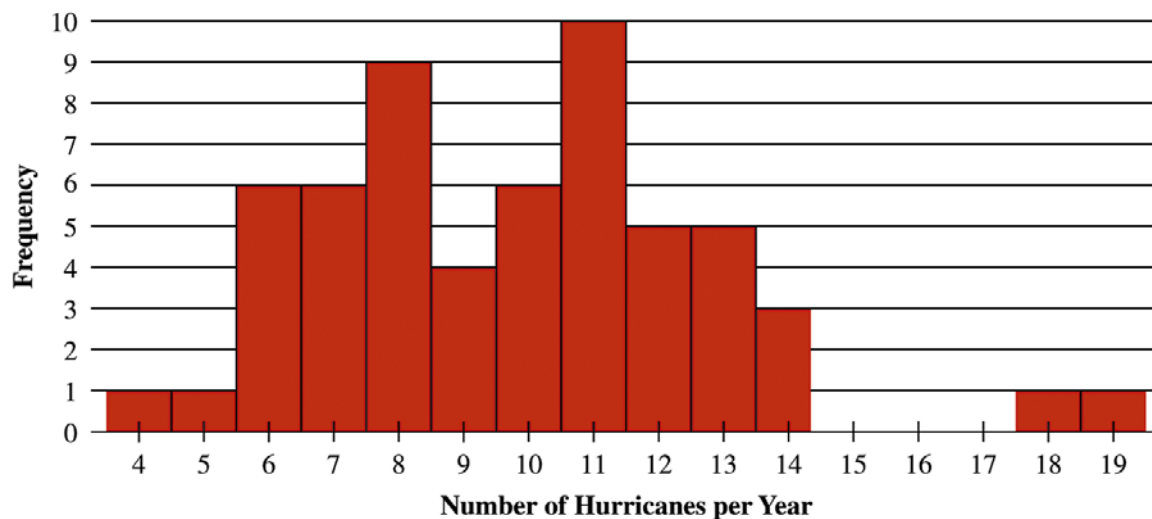
Example: Determining Information from a Graph

The bar graph shows the number of Atlantic hurricanes over a period of years. Answer the questions on the slides that follow.



Example: Determining Information from a Graph (cont)

- a) What was the smallest number of hurricanes in a year during this period? What was the largest?

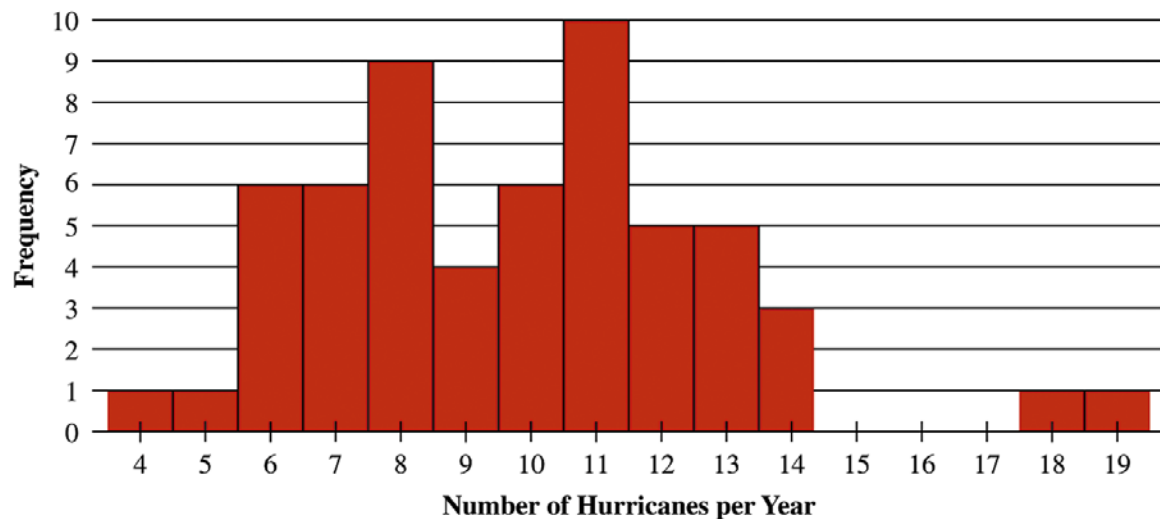


Solution

- a) The smallest number of hurricanes in any year during this time period was 4. The largest number was 19.

Example: Determining Information from a Graph (cont)

b) What number of hurricanes per year occurred most frequently?

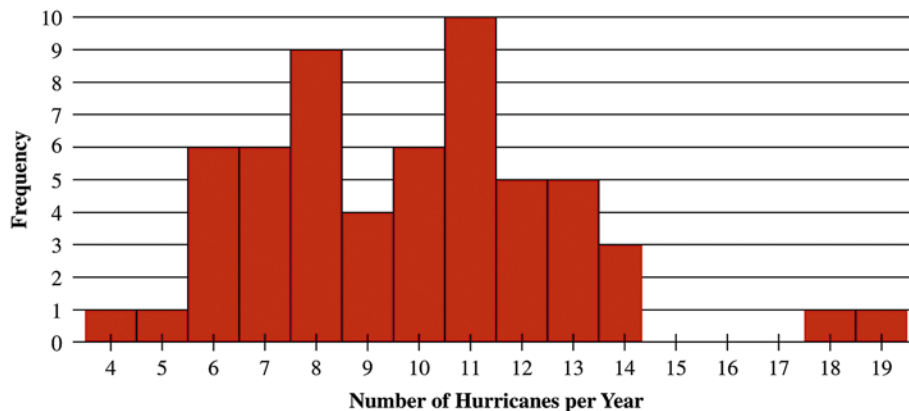


Solution

b) The number of hurricanes per year that occurred most frequently corresponds to the tallest bar, which appears over the number 11. Therefore, 11 hurricanes occurred in 10 different years.

Example: Determining Information from a Graph (cont)

c) How many years were the hurricanes counted?



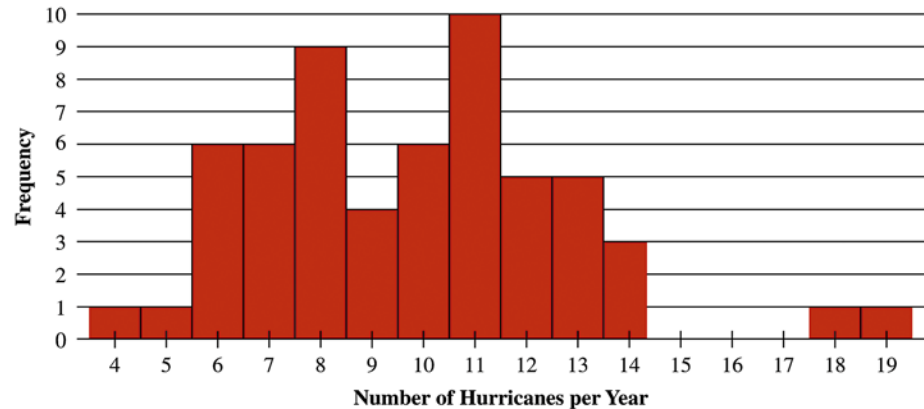
Solution

c) To find the total number of years for which the data were gathered, we add the heights of all of the bars to get

$$1 + 1 + 6 + 6 + 9 + 4 + 6 + 10 + 5 + 5 + 3 + 1 + 1 = 58 \text{ years.}$$

Example: Determining Information from a Graph (cont)

d) In what percentage of the years were there more than 10 hurricanes?



Solution

d) First, we count the number of years in which there were more than 10 hurricanes. If we add the heights of the bars above these values, we get $10 + 5 + 5 + 3 + 1 + 1 = 25$. Because there are 58 years of data, we calculate $25/58 \approx 0.431$, which is approximately 43%.