

**1.15 Fast-Twitch Muscles and Race** [Example 1.5](#) studied a variant of the gene *ACTN3* which inhibits fast-twitch muscles and seems to be less prevalent in sprinters. A separate study<sup>9</sup> indicated ethnic differences: Approximately 20% of a sample of Caucasians, approximately 25% of a sample of Asians, and approximately 1% of a sample of Africans had the gene variant. What are the variables in this study? Classify each as categorical or quantitative.

**1.16 Rowing Solo Across the Atlantic Ocean** On January 14, 2012, Andrew Brown of Great Britain set the world record time (40 days) for rowing solo across the northern Atlantic Ocean. On March 14, 2010, Katie Spotz of the United States became the youngest person to ever row solo across the Atlantic when she completed it in 70 days at the age of 22 years old. [Table 1.3](#) shows times for males and females who rowed solo across the Atlantic Ocean in the last few years.<sup>10</sup>

<b>TABLE 1.3</b>	
<i>Number of days to row alone across the Atlantic Ocean</i>	
Male times:	40, 87, 78, 106, 67
Female times:	70, 153, 81

- How many cases are there in this dataset? How many variables are there and what are they? Is each categorical or quantitative?
- Display the information in [Table 1.3](#) as a dataset with cases as rows and variables as columns.

**1.22 Don't Text While Studying!** For the 2015 Intel Science Fair, two brothers in high school recruited 47 of their classmates to take part in a two-stage study. Participants had to read two different passages and then answer questions on them, and each person's score was recorded for each of the two tests. There were no distractions for one of the passages, but participants received text messages while they read the other passage. Participants scored significantly worse when distracted by incoming texts. Participants were also asked if they thought they were good at multitasking (yes or no) but "even students who were confident of their abilities did just as poorly on the test while texting."<sup>45</sup>

- What are the cases?
- What are the variables? Is each variable categorical or quantitative?
- If we create a dataset of the information with cases as rows and variables as columns, how many rows and how many columns would the dataset have?

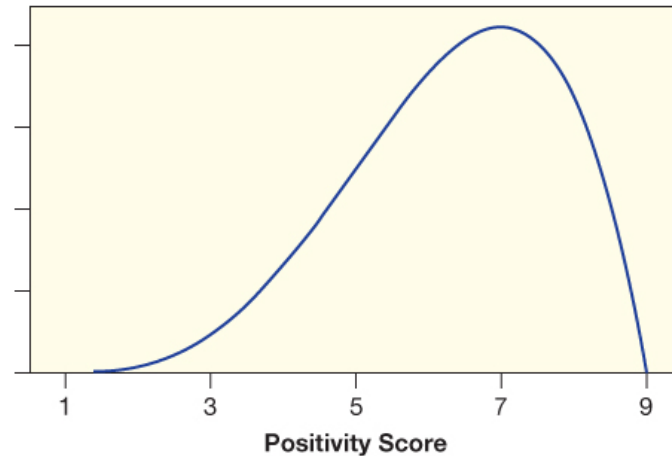
**2.20 Culture and Mental Illness** A recent study<sup>43</sup> examining the link between schizophrenia and culture interviewed 60 people who had been diagnosed with schizophrenia and who heard voices in their heads. The participants were evenly split between the US, India, and Ghana, and each was interviewed to determine whether the voices were mostly negative, mostly neutral, or mostly positive. The results are shown in [Table 2.11](#). "Learned cultural expectations about the nature of mind and self" appear to influence how the voices are perceived.

<b>TABLE 2.11</b>				
<i>How do people with schizophrenia perceive the voices?</i>				
	US	India	Ghana	Total
Negative	14	4	2	20
Neutral	6	3	2	11
Positive	0	13	16	29
Total	20	20	20	60

- What proportion of all the participants felt that the voices are mostly negative?
- What proportion of all US participants felt that the voices are mostly negative?
- What proportion of non-US participants felt that the voices are mostly negative?
- What proportion of participants hearing positive voices are from the US?
- Does culture appear to be associated with how voices are perceived by people with schizophrenia?

**2.66 Is Language Biased toward Happiness?** “Are natural languages neutrally, positively, or negatively biased?” That is the question a recent study<sup>39</sup> set out to answer. They found the top 5000 words used in English in each of four different places: Twitter, books on the Google Book Project, *The New York Times*, and music lyrics. The resulting complete list was 10,222 unique words in the English language. Each word was then evaluated independently by 50 different people, each giving a rating on how the word made them feel on a 1 to 9 scale, where 1 = least happy, 5 = neutral, and 9 = most happy. (The highest rated word was “laughter,” while the lowest was “terrorist.”) The distributions of the ratings for all 10,222 words for each of the four media sources were surprisingly similar, and all had approximately the shape shown in [Figure 2.14](#).

[Figure 2.14](#) Distribution of ratings of words where 9 = most positive

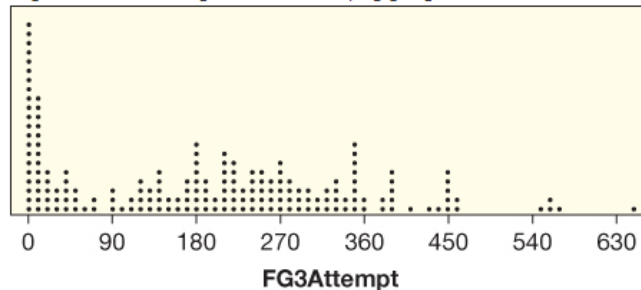


- Describe the shape of the distribution.
- Which of the following values is closest to the median of the distribution:  
3.5   5   6.5   7   7.5   8
- Will the mean be smaller or larger than the value you gave for the median in [part \(b\)](#)?

Exercises 2.116 to 2.118 refer to the dataset **NBAPlayers2015**, which contains information on many variables for players in the NBA (National Basketball Association) during the 2014–2015 season. The dataset includes information for all players who averaged more than 24 minutes per game, and includes  $n = 182$  players and 25 variables.

**2.116 Distribution of Three-Point Attempts in the NBA** In basketball, a basket is awarded three points (rather than the usual two) if it is shot from farther away. Some players attempt lots of three-point shots and quite a few attempt none, as we see in the distribution of number of three-point attempts by players in the NBA in [Figure 2.30](#). The data are available in **NBAPlayers2015** under the variable name *FG3Attempt*. Is it appropriate to use the 95% rule with this dataset? Why or why not?

[Figure 2.30](#) Number of three-point shot attempts in the NBA, by player



**2.117 Distribution of Blocked Shots in the NBA** The variable *Blocks* in the dataset **NBAPlayers2015** includes information on the number of blocked shots during the season for each of the 182 players in the dataset.

- a. Use technology to find the mean and the standard deviation of the number of blocked shots.
- b. Use technology to find the five number summary for the same variable.
- c. Which set of summary statistics, those from [part \(a\)](#) or [part \(b\)](#), is more resistant to outliers and more appropriate if the data are heavily skewed?
- d. Use technology to create a graph of the data in *Blocks* and describe the shape of the distribution.
- e. Is it appropriate to use the 95% rule with these data? Why or why not?