Additional Exercises

1. The following data shows the age (in years) at inauguration of all United States presidents from 1900 to 2022.

42 43 46 47 51 51 51 52 54 54
55 55 56 56 60 61 62 64 69 70 78

a. Determine the mean, median, and mode for the ages of the U.S. presidents data.

b. Determine the 5-number summary of the data.

c. Determine the interquartile range.

d. Does the data set contain any outliers?

2. The maximum heart rates achieved while performing a particular aerobic exercise are measured (in beats per minute) for 9 randomly selected individuals.

Maximum Heart Rates (BPM)								
145	155	130	185	170	165	150	160	125

a. Determine the sample variance of the maximum heart rate achieved.

b. Determine the sample standard deviation of the maximum heart rate achieved.

c. Determine the range of the maximum heart rate achieved.

d. What are some of the factors which might contribute to the variation in the observations?

e. Create a box plot of the data.

f. What is the percentile associated with a BPM of 145?

3. A sample of executives from small businesses were asked how many vacation days they used during the previous calendar year. The frequency distribution table summarizes the results.

Number of Vacation Days Last Year			
Number of Days	Frequency	Number of Days	Frequency
0	13	6	0
1	18	7	3
2	11	8	3
3	7	9	0
4	4	10	2
5	3		

a. What proportion of the executives used at least 3 vacation days in the previous year?

b. Find the sample mean and sample standard deviation of the number of vacation days using the formulas for grouped data.

c. Compute the interval one standard deviation about the mean.

d. Find the percent of data falling in the interval one standard deviation about the mean.

e. Is the percent of the data falling in the interval one standard deviation about the mean close to what the Empirical Rule predicts? If there is a discrepancy, what is a likely reason for the discrepancy?

4. A high school math teacher summarized the 35 math SAT scores for the students in her calculus class. The mean for the class was 521 and the median was 535. The range of the scores was 235 and the highest score in the entire class was 675. Approximately 40% of the class scored higher than 562. State whether each of the following is true or false.

a. The 45^{th} percentile exceeds 540.

b. The lowest score in the class was 440.

c. The z-score for a score of 510 is a negative number.

d. The third quartile exceeds 562.

e. The percentile rank of 562 is 40.

5. Consider the following number of defective circuit boards produced by two different machines on seven randomly selected days.

Defective Circuit Boards							
Machine A	2	3	7	4	5	1	0
Machine B	2	3	4	3	4	2	4

a. Determine the average number of defective circuit boards produced by each machine.

b. Determine the variance of the number of defective circuit boards produced by each machine.

c. Determine the standard deviation of the number of defective circuit boards produced by each machine.

d. If both machines produce approximately the same number of circuit boards each day, which machine do you think is better? Why?

e. Is there another way of approaching this problem?

6. A basketball coach has one remaining scholarship to offer and has narrowed his choice to two players. Listed in the following table are the points scored per game over the last season for each player.

Points Scored		
Game Number	**Braudrick**	**Douglas**
1	27	35
2	34	21
3	29	50
4	25	28
5	28	missed
6	35	32

Points Scored		
Game Number	**Braudrick**	**Douglas**
7	31	29
8	33	missed
9	33	23
10	25	35
11	28	31
12	32	36
TOTAL	**360**	**320**

a. What level of measurement do the data possess?

b. What statistical criteria might you use to select the better player? Justify your answer.

c. Compute the statistics you proposed in **b.**

d. Which player is more consistent? Why?

e. What biases or errors might be present in the data?

f. Is there another way of approaching this problem?

7. Consider the literacy data from recent years that is compiled in the following table.[18]

Literacy Rates			
Country	**Literacy Rate (%)**	**Country**	**Literacy Rate (%)**
Argentina	99.00	**Ecuador**	93.60
Chile	96.40	**Suriname**	94.40
Brazil	93.20	**Mexico**	95.20
Colombia	95.60	**Guatemala**	80.80
Uruguay	98.80	**Costa Rica**	97.90
Paraguay	94.50	**El Salvador**	89.10
Bolivia	92.50	**Honduras**	88.50
Peru	94.50	**Nicaragua**	82.60
Venezuela	97.10	**Panama**	95.70
Guyana	88.50		

a. What is the mean literacy rate for these selected countries?

b. What is the standard deviation of these literacy rates?

c. Using the empirical rule, how many countries in this group would we expect to have literacy rates between one standard deviation below the mean and one standard deviation above the mean?

d. How many countries in this group actually have literacy rates between one standard deviation below the mean and one standard deviation above the mean?

e. What assumption did you make in answering part **c.** above?

f. Using the Literacy Rate data from the companion website, compare the literacy rates for Asia to those in South America.

Data

Discovering Statistics and Data, Fourth Edition > Data Sets > Literacy Rates 2022

8. A manufacturer considers her production process to be "in control" if the proportion of defective items is less than 3%. She randomly selects 200 items and determines that 9 of the items are defective.

a. Determine the sample proportion of defective items.

b. Based on the sample, do you think it is reasonable for the manufacturer to conclude that the production process is "out of control"? Why or why not?

9. Late in the summer of 1996, Tiger Woods became a professional golfer. This highly publicized event followed a sensational college career at Stanford University, where Tiger won three United States Amateur championships. Tiger was not a professional very long before he had his first win on the pro tour, the Las Vegas Invitational. He received a total of $297,000 for his accomplishment. The prize money (in thousands) for the top 40 finishers in the tournament are given in the data set Las Vegas Golf Invitational and shown below.[19]

Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > Las Vegas Golf Invitational

Las Vegas Invitational Prize Money (Thousands of Dollars)							
297.0	60.2	46.2	31.3	21.4	14.5	10.7	8.5
178.2	60.2	31.3	31.3	14.5	14.5	10.7	8.5
95.7	46.2	31.3	24.7	14.5	14.5	8.5	8.5
95.7	46.2	31.3	21.4	14.5	14.5	8.5	8.5
60.2	46.2	31.3	21.4	14.5	10.7	8.5	8.5

a. Find the mean.

b. Find the median.

c. Find the mode.

d. Find the 10% trimmed mean and compare it to the mean and the median.

e. Comment on the skewness of the distribution.

10. A pharmacist is interested in studying the relationship between the amount of a particular drug in the bloodstream (in mg) and reaction time (in seconds) of subjects taking the drug. Ten subjects are randomly selected and administered various doses of the drug. The reaction times (in seconds) are measured 15 minutes after the drug is administered with the following results.

Reaction Times			
Amount of Drug (mg)	**Reaction Time (Seconds)**	**Amount of Drug (mg)**	**Reaction Time (Seconds)**
1	0.5	6	0.8
2	0.7	7	0.9
3	0.6	8	0.6
4	0.7	9	0.9
5	0.8	10	1.0

Analyze the data collected for the study by answering the following questions:

a. Do the variables selected for measurement seem appropriate for answering the question the pharmacist is interested in?

b. What biases or errors might be present in the data?

c. What level of measurement (nominal, ordinal, interval, ratio) do the data possess?

11. Our galaxy, the Milky Way, is 120,000 light years across and contains more than 100 billion stars. In the 1930s it was unknown whether the galaxies were moving away from one another (i.e., the universe was expanding) or moving toward one another (contracting). The astronomer Edwin Hubble attempted to measure the rate at which the galaxies were moving and discovered they were moving apart. He originally estimated they were moving apart at the rate of 500 km/s/Mpc or about 160 km/sec per million-light-years from Earth, which was called Hubble's constant. As new telescopes have come online and new methods and technologies have become available, Hubble's constant has been reestimated. The current contenders are a team at the European Space Agency which estimates the constant at 67.8 km/s/Mpc, while another team has estimated the constant at 73.24 km/s/Mpc. A recent *Science* magazine article said, "the two values are separated by a gulf of 3.4 sigma."[20]

a. What is the value of sigma in this case?

b. What questions would you like to ask about this conclusion?