Two standard deviations above the mean is

$$\mu + 2\sigma = 7498 + 2(3639) = \$14{,}776$$

and two standard deviations below the mean is

$$\mu - 2\sigma = 7498 - 2(3639) = \$220.$$

Therefore, by Chebyshev's Theorem, we can say that at least 75% of the tuition and fees of colleges and universities in the United States is between \$220 and \$14,776 for 2019–2020.

Who Is King of the Hill?

In 1961, Wilt Chamberlain was the National Basketball Association (NBA) rebounding leader with 27 rebounds per game. In 1992, the colorful Dennis Rodman won the same honor with 18.7 rebounds per game. Common sense suggests that professional basketball in the 1990s is played at a much higher level than in the 1960s. So why has the rebounds per game for the rebounding leader fallen? Is it another case of "less is more"?

Researchers investigating this interesting puzzle considered two other variables: the number of rebounding opportunities (this had gone down since the field goal percentage has increased historically) and the average number of minutes played per game, which has also fallen.

Thus, when we adjust the actual rebounds obtained by the rebounding leaders to the number of minutes played and the total number of rebounding opportunities, we see a completely different picture. The adjusted rebound numbers for Chamberlain and Rodman are 35.42 and 51.06 respectively.

The Coefficient of Variation

Sometimes a data analyst wants to compare the variation of two or more data sets. The **coefficient of variation** is a unit-free statistical measure that enables the comparison of the variation in two or more data sets.

Formula

Coefficient of Variation

The coefficient of variation, another statistical measure, compares the variation in data sets.

For population data, the coefficient of variation is defined as

$$\text{CV} = \left(\frac{\sigma}{\mu} \cdot 100 \right)\%,$$

and for sample data,

$$\text{CV} = \left(\frac{s}{\bar{x}} \cdot 100 \right)\%.$$

When comparing the variation of data sets, many times the units of measure will be different. The coefficient of variation standardizes the variation measure by dividing it by the mean. The division has one interesting side effect: the unit of measure is removed from the statistic.

One of the primary focuses of quality control in manufacturing is the reduction in variation of the output of the process. A bolt manufacturer wants to compare the variability of two bolt manufacturing processes. One process creates bolts with a mean length of 2.5 cm and a standard deviation of 0.2 cm. Is this process more variable than one that produces a bolt that has a mean length of 1 inch and a standard deviation of 0.052 inches?

$$\text{CV}_{\text{Bolt 1}} = \frac{0.2}{2.5} \cdot 100 = 8.0\%$$

$$\text{CV}_{\text{Bolt 2}} = \frac{0.052}{1} \cdot 100 = 5.2\%$$

The coefficient of variation for Bolt 1 is 8.0% . This means that the variation is 8% of the mean value. The coefficient of variation for Bolt 2 is 5.2% of the mean. Therefore, the process used to make Bolt 2 is less variable than that for Bolt 1.

4.2 Exercises

Basic Concepts

1. Describe three measures of variation. Discuss the strengths and weaknesses of each.
2. What does the standard deviation measure?
3. Why are the variance and standard deviation more commonly used as measures of variability than the MAD?

4. Explain how the variance can be construed as an average.
5. True or false: The variance and standard deviation are resistant measures.
6. When is it appropriate to calculate the variance of a time series?
7. What is the empirical rule? When is it appropriate to use the empirical rule?
8. What is Chebyshev's Theorem?
9. Discuss the purpose of the coefficient of variation.
10. How is the coefficient of variation calculated?
11. Why is the coefficient of variation important?

Exercises

12. Find the missing age in the following set of four student ages.

Student Ages

Student	Age	Deviation from the Mean
A	19	–4
B	20	–3
C	?	+1
D	29	+6

13. Find the missing weight in the following data set.

Weights

Person	Weight	Deviation from the Mean
A	144	–20
B	156	–8
C	?	+1
D	176	+12

14. Consider the following time until failure for 10 randomly selected car batteries (measured in years).

Years Until Failure for Car Batteries

5	3	4	6	2	5	7	10	8	4

   a. Calculate the sample variance of the time until failure.
   b. Calculate the sample standard deviation of the time until failure.
   c. Calculate the range of the time until failure.
   d. What are some of the factors which might contribute to the variation in the observations?

15. Consider the following distances jumped (in feet) by 8 randomly selected long jumpers.

Jump Distances (Feet)

21	15	12	18	10	14	17	11

   a. Calculate the sample variance of the distances jumped.
   b. Calculate the sample standard deviation of the distances jumped.
   c. Calculate the range of the distances jumped.
   d. What are some of the factors which might contribute to the variation in the observations?

16. The interest rates on 30-year mortgages offered by seven randomly selected banks in a large metropolitan area are recorded in the following table.

Interest Rates (%)						
7.5	8.0	7.0	7.25	8.5	8.25	7.75

a. Calculate the sample variance of the interest rates.

b. Calculate the sample standard deviation of the interest rates.

c. Calculate the range of the interest rates.

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

17. A researcher has hypothesized that female college students are more disciplined than male college students.The researcher believes that a reasonable measure of discipline is performance on a statistics test in terms of both absolute scores and consistency of scores. Seven male statistics students and seven female statistics students are randomly selected and their scores on a statistics test are observed.

Test Scores							
Males	65	100	75	45	85	73	95
Females	75	80	95	85	82	72	49

a. Calculate the average test score for male students and female students separately.

b. Calculate the variance of the test scores for male students and female students separately.

c. Calculate the standard deviation of the test scores for male students and female students separately.

d. Do you think that the data tend to support the hypothesis that female college students are more disciplined than male college students based on the researcher's measurement?

e. What do you think about this particular measurement of discipline?

18. Consider the following market values of two portfolios of stocks at five randomly selected times during a year.

Market Values ($)					
Portfolio A	150,000	155,000	145,000	160,000	140,000
Portfolio B	130,000	175,000	100,000	150,000	195,000

a. What statistical criteria might you use to select the better portfolio? Justify your answer.

b. Calculate the statistics you proposed in part **a.**

c. Which portfolio has the least amount of risk? Why?

19. Add 20 to each of the following data values.

81	99	97	81	85	86
99	93	96	83	82	91

a. Compute the mean and standard deviation for both the original data and adjusted data.

b. Compare the mean and standard deviation of the adjusted data to the mean and standard deviation of the original data.

c. Describe the effect on the mean and standard deviation of adding a constant to a data set.

20. Adjust the following data values by subtracting 20 from each data value.

745	789	712	764	736
758	722	773	751	741

a. Calculate the mean and variance for the original and adjusted data.

b. Compare the mean and variance of the adjusted data to the mean and the variance of the original data.

c. Describe the effect of subtracting a constant value from each member of a data set on the mean and variance of the data.

21. The average score on a pre-employment test is 26 with a standard deviation of 7. Using Chebyshev's Theorem, state the range in which at least 88.89% of the data will reside.

22. The daily average number of phone calls to a call center is 972 with a standard deviation of 127. Using Chebyshev's Theorem, state the range in which at least 75% of the data will reside.

23. There is an annual chowder eating contest in a small New England town. The average amount of chowder eaten at the contest was 32 ounces with a variance of 64 ounces. Given that one hundred people participated in the contest, find:

a. The approximate number of people who ate between 24 and 40 ounces of chowder.

b. The approximate number of people who ate between 16 and 48 ounces of chowder.

c. What assumptions did you make about the amount of chowder eaten by each contestant in answering parts **a.** and **b.**?

24. The manager of a local diner has calculated his average daily sales to be $4500 with a standard deviation of $750.

a. In what range can the manager expect his daily sales to be 68% of the time?

b. In what range can the manager expect his daily sales to be 95% of the time?

c. In what range can the manager expect his daily sales to be 99.7% of the time?

d. What assumption did you make about daily sales when answering parts **a.**, **b.**, and **c.**?

25. A management consulting firm is evaluating the salary structure for a large insurance company. The goal of the study is to develop salary ranges for each of the possible job grades within the company. The company and the firm have agreed that a reasonable salary range for each job grade can be determined by finding the salary range in which 95% of the current salaries for that job grade fall.The average salary and the standard deviation of the salaries are listed in the following table for three of the job grades.

Salaries ($)			
Job Grade	**25**	**33**	**40**
$\bar{x}$	22,000	35,000	45,000
s	1500	2000	5000

a. Determine the appropriate salary ranges for the three job grades.

b. What assumption did you make about the salaries in each of the job grades in answering part **a.**?

26. A consumer interest group is interested in comparing two brands of vitamin C. One brand of vitamin C advertises that its tablets contain 500 mg of vitamin C. The other brand advertises that its tablets contain 250 mg of vitamin C. Tablets for each brand are randomly selected and the milligrams of vitamin C for each tablet are measured with the following results.

Vitamin C Content (mg)		
	Brand A (500 mg)	Brand B (250 mg)
$\bar{x}$	500	250
s	10	7

a. Calculate the coefficient of variation for Brand A.

b. Calculate the coefficient of variation for Brand B.

c. Which brand more consistently produces tablets as advertised? Explain.

27. A manufacturer of bolts has two different machines. One machine is used to produce $\frac{1}{4}$-inch bolts; the other machine is used to produce $\frac{1}{2}$-inch bolts. It is very important that the machines consistently produce bolts of the correct diameters, or the bolts will not fit on the corresponding nuts. In order to compare the two machines, management randomly selects bolts produced from each machine and computes the average diameter of the bolts and the standard deviation of the bolts. The results of the study are shown in the following table.

Bolt Diameter		
	Machine X $\left(\frac{1}{4}"\right)$	Machine Y $\left(\frac{1}{2}"\right)$
$\bar{x}$	0.25”	0.50”
s	0.03”	0.05”

a. Calculate the coefficient of variation for Machine X.

b. Calculate the coefficient of variation for Machine Y.

c. Which machine more consistently produces bolts of the correct diameter? Explain.

4.3 Measures of Relative Position

Suppose you want to know where an observation stands in relation to other values in a data set. For example, on many standardized tests such as the SAT, GMAT, and ACT, the test scores themselves are rather meaningless unless they are associated with some measure that tells you how well you did relative to others taking the same test. There are two principal methods of communicating relative position: **percentiles** and ***z*-scores**. Both of these methods are data transformations which change the scale of the data in some way.

Definition

P^{th} Percentile

Given a set of data $x_1, x_2, \ldots, x_n$, the **P^{th} percentile** is a value, say x, such that approximately P percent of the data is less than or equal to x and approximately $(100 - P)$ percent of the data is greater than or equal to x.

Percentiles

The most commonly used measure of relative position is the percentile. In fact, we have already discussed the 50th percentile; it is the median. For example, in data sets that do not contain significant quantities of identical data, the 30th percentile is a value such that about 30 percent of the values are below it, and about 70 percent are above it.