

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. Why is a “bell-shaped” distribution associated with the empirical rule?
8. What is the empirical rule? When is it appropriate to use? What statistical techniques would be useful in establishing appropriateness?
9. What is Chebyshev’s Theorem?
10. Suppose you had a data set with 100,000 SAT scores and you wished to describe an equidistant interval around the mean with 75% of the scores. Aside from the empirical rule or Chebyshev’s theorem, how could you define the interval? Hint: Think empirically.

Exercises

11. Since Super Bowl football games are a sample of all NFL football games, use the Super Bowl data set from the companion website to determine the following.
 - a. Determine the sample variance of the variable *Winner_Rush Attempts*. Round your answer to three decimal places.
 - b. Determine the sample standard deviation of the variable *Winner_Rush Attempts*. Round your answer to three decimal places.
 - c. Determine the range of the variable *Winner_Rush Attempts*.
 - d. What are some of the factors which might contribute to the variation in the observations?
12. Since Super Bowl football games are a sample of all NFL football games, use the Super Bowl data set from the companion website to determine the following.
 - a. Determine the sample variance of the variable *Loser_Rush Attempts*. Round your answer to three decimal places.
 - b. Determine the sample standard deviation of the variable *Loser_Rush Attempts*. Round your answer to three decimal places.
 - c. Determine the range of the variable *Loser_Rush Attempts*.
 - d. What are some of the factors which might contribute to the variation in the observations? Try to use different factors than the ones used in the previous exercise.

Data

stat.hawkeslearning.com

Discovering Statistics and Data,
Fourth Edition > Data Sets > Super
Bowl

13. The interest rates on 30 year mortgages offered by seven randomly selected banks in a large metropolitan area are recorded below.

7.5% 8.0% 7.0% 7.25% 8.5% 8.25% 7.75%

- Determine the sample variance of the interest rates.
 - Determine the sample standard deviation of the interest rates.
 - Determine the range of the interest rates.
 - What are some of the factors which might contribute to the variation in the observations?
14. A researcher has hypothesized that sophomore college students are more disciplined than freshman college students. The researcher believes that a reasonable measure of discipline is the performance on a statistics test in terms of both absolute score and consistency of scores. Seven freshman statistics students and seven sophomore statistics students are randomly selected and their scores on a statistics test are observed.

| Statistics Test Scores | | | | | | | |
|------------------------|----|-----|----|----|----|----|----|
| Freshman | 65 | 100 | 75 | 45 | 85 | 73 | 95 |
| Sophomore | 75 | 80 | 95 | 85 | 82 | 72 | 49 |

- Determine the average test score for freshman students and sophomore students separately.
 - Determine the variance of the test scores for freshman students and sophomore students separately.
 - Determine the standard deviation of the test scores for freshman students and sophomore students separately.
 - Do you think that the data tend to support the hypothesis that sophomore college students are more disciplined than freshman college students based on the researcher's measurement?
 - What do you think about this particular measurement of discipline?
15. An initial investment of \$135,000 was placed in each of 10 stocks and the stocks were divided into two portfolios. The market values of the stocks in each portfolio at the end of year one of the investment are presented below.

| 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 |

- What level of measurement does the data possess?
- What statistical criteria might you use to select the better portfolio? Justify your answer.
- Compute the statistics you proposed in part **b**.
- Intuitively, which portfolio has the least amount of risk? Why?

16. Given the data set shown below.

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

- Compute the mean and standard deviation for the original data.
 - Add 20 to each of the data points in the original data. Compute the mean and standard deviation for this adjusted data set and compare them to the mean and standard deviation of the original data. What do you notice?
 - Subtract 10 from each of the data points in the original data. Compute the mean and standard deviation for this adjusted data set and compare them to the mean and standard deviation of the original data. What do you notice?
 - Make an empirical generalization about the effect of adding or subtracting a constant value to each member of a data set on the mean and standard deviation of the data.
17. Given the data set shown below.

8 14 6 10 20 12 4 10 18 2

- Compute the mean and standard deviation for the original data.
 - Multiply each of the data points by 10. Compute the mean and the standard deviation for the adjusted data set and compare to the mean and the standard deviation of the original data. What do you notice?
 - Divide each of the data points by 2. Compute the mean and standard deviation for the adjusted data set to the mean and the standard deviation for the original data. What do you notice?
 - Make an empirical generalization about the effect of multiplying or dividing a constant value from each member of a data set on the mean and standard deviation of the data.
18. 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.
19. 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.
20. The length of a full-term human pregnancy is often defined as 40 weeks. From empirical data the length of time from conception to birth is on average 266 days with a standard deviation of 16 days. Use the empirical rule to determine the following ranges for the length of a human pregnancy.
- What time interval (in days) will contain the lengths of approximately 68% of human pregnancies?
 - What time interval (in days) will contain the lengths of approximately 95% of human pregnancies?
 - What time interval (in days) will contain the lengths of approximately 99.7% of human pregnancies?
 - What assumption did you make in order to use the empirical rule in this scenario? ©HAWKES LEARNING

21. For people with diabetes, it is important to determine the level of glucose in the bloodstream, avoiding amounts that are too high or too low, to prevent negative health consequences. The measurements of glucose levels in the blood can be obtained through a finger stick with a lancet or by wearing a device with a subcutaneous probe attached to the skin. Both methods result in a tremendous amount of data used by medical professionals to monitor the person's health and prescribe treatment. Over a 4-week span of measurements, the blood glucose levels of one patient had a bell-shaped distribution with a mean of 114 mg/dl and a standard deviation of 24 mg/dl. Use the empirical rule to determine the following.
- The percentage of measurements that the patient had blood glucose levels "in range", which is the desired target interval between 90 and 138 mg/dl.
 - The percentage of measurements that the patient had blood glucose levels less than 66 mg/dl, which is seriously low, categorized as hypoglycemia.
 - The percentage of measurements that the patient had blood glucose levels greater than 138 mg/dl, which is considered elevated and should be treated with an insulin dosage.
22. 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 table below for three of the job grades.

| Salary (\$) | | | |
|-------------|--------|--------|--------|
| Job Grade | 25 | 33 | 40 |
| \bar{x} | 40,000 | 55,000 | 70,000 |
| s | 3,000 | 2,000 | 5,000 |

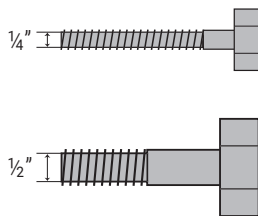
- Determine the appropriate salary ranges for the three job grades.
 - What assumption did you make about the salaries in each of the job grades in answering part a.?
23. A consumer interest group is interested in comparing two brands of vitamin C. One brand of vitamin C advertises that its tablets contain 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 |

- Compute the coefficient of variation for Brand A.
- Compute the coefficient of variation for Brand B.
- Which brand more consistently produces tablets as advertised? Explain.



24. A manufacturer of bolts has two different machines. One machine is used to produce 1/4 inch bolts; the other machine is used to produce 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 table below.



| 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" |

- Compute the coefficient of variation for Machine X.
- Compute the coefficient of variation for Machine Y.
- Relative to their mean diameters, which machine more consistently produces bolts of the correct diameter? Explain.

4.3 Measures of Relative Position, Box Plots, and Outliers

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.

Percentiles

The P^{th} percentile is a data transformation. It transforms a data value into its relative position in the data set. In fact, we have already discussed the 50th percentile; it is the median. 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 around 70 percent are above it.

P^{th} Percentile

Given a set of data x_1, x_2, \dots, 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 .

DEFINITION

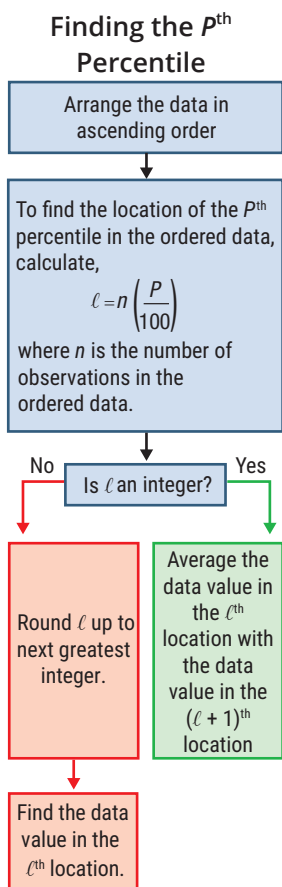


Figure 4.3.1