Example 4.3.4

Determining *z*-Scores for Biology and Psychology Test Scores

Suppose you scored an 86 on your biology test and a 94 on your psychology test. The mean and standard deviations of the two tests are given in the following table.

Test Scores		
Course	**Mean**	**Standard Deviation**
Biology	74	10
Psychology	82	11

What are the z-scores for your two tests? On which of the tests did you perform relatively better?

Solution

The z-score for the biology test is $z = \dfrac{86 - 74}{10} = 1.20$.

The z-score for the psychology test is $z = \dfrac{94 - 82}{11} \approx 1.09$.

On the biology test you scored 1.2 standard deviations above the mean, compared to only 1.09 standard deviations above the mean for the psychology test. Even though the raw score on the psychology test is larger than the raw score on the biology test, relative to the means and variability in the data sets, the performance on the biology test was slightly better. Once again, changing the scale of the data (to standard deviation units) has beneficial effects. It enables the comparison of two measurements that are drawn from different populations.

Properties of a *z*-Score

- If a z-score is negative, the corresponding data value is less than the mean.
- Conversely, if a z-score is positive, the corresponding data value is greater than the mean.
- The z-score is a unit-free measure. That is, regardless of the original units of measurement (centimeters, meters, or kilometers), an observation's z-score will be the same.

PROPERTIES

4.3 Exercises

Basic Concepts

1. What are two methods for describing relative position?
2. If a data value is determined to be the 72nd percentile, what does this mean?
3. Describe how to find the percentile of a particular value.
4. What are quartiles? Are they equivalent to percentiles? If so, how?
5. What is the interquartile range? What does it measure?
6. What are the advantages of using a box plot to display a data set?
7. What are the key calculations needed in order to construct a box plot?

8. What is an outlier? How can outliers be identified?
9. What is a z-score? Why is it useful?

Exercises

10. According to the U.S. Constitution in Article 2, the minimum age to be elected president is 35 years old. To date, the youngest president was Theodore Roosevelt who was elected in 1900 at the age of 42. The oldest president is Joe Biden at the age of 78. 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. Use the data shown to determine the median age for the presidents elected from 1900 to the present.
   b. Determine the 80th percentile for the data. Interpret this value in the context of the presidents' ages.
   c. When Barack Obama was elected in 2008, he was 47 years old. Determine the percentile for his age. What percentage of most recent presidents were younger than him?
   d. Ronald Reagan was the oldest president to date at his inauguration in 1980. Determine the percentile for his age, 69 years. What percentage of most recent presidents were older than him?

11. The *safety score* variable in the OECD Better Life Index 2022 data set is a measure of how safe a person in this country feels walking alone at night. Using the *safety score* variable, answer the following questions.

Data
stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > OECD Better Life Index 2022

   a. What level of measurement does the data possess?
   b. Determine the 20th percentile.
   c. Determine the 95th percentile.
   d. Interpret the meaning of each of these percentiles.

12. Use the *safety score* variable from the OECD Better Life Index data set from the previous problem to determine the following. Round your answers to the nearest whole number.
   a. Determine the percentile rank for Ireland's safety score.
   b. Determine the percentile rank for Chile's safety score.
   c. Based on the OECD Better Life Index, in which of these two countries would you feel safer to walk alone at night?

13. Using the *Adult.smoking* variable in the US County Data, answer the following questions.

Data
stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > US County Data

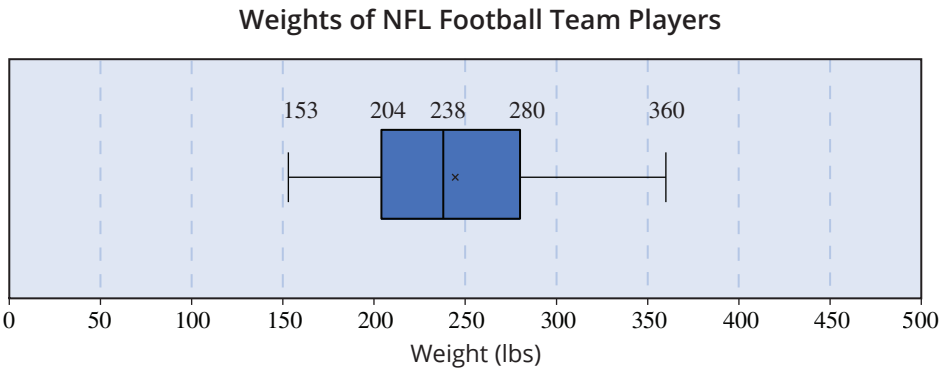
   a. What level of measurement does the data possess?
   b. Determine the 20th percentile. Round your answer to three decimal places.
   c. Determine the 95th percentile. Round your answer to three decimal places.
   d. Interpret the meaning of each of these percentiles.

14. Use the US County Data from the previous problem to find the following.

a. Determine the percentile rank for Lee County, Kentucky's *Adult.smoking* percentage. Round to the nearest whole number.

b. Determine the percentile rank for Ozaukee County, Wisconsin's *Adult.smoking* percentage. Round to the nearest whole number.

15. A team in the National Football League (NFL) has 55 players on the active roster. The 5-number summary of the players' weights is given in the box plot.



a. What is the median weight of the players?

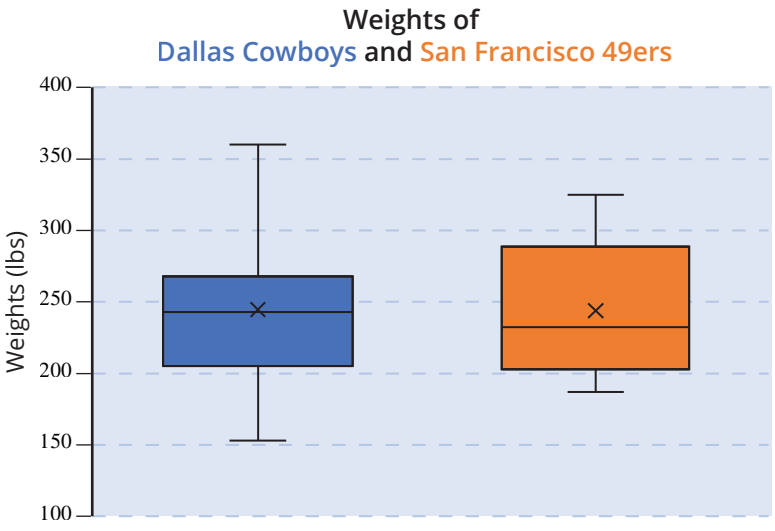
b. What is the interquartile range of the players' weights?

c. Approximately, what is the percentage of players weighing greater than or equal to 204 pounds?

d. How many players weigh between 238 and 280 pounds?

16. The following is a 5-number summary and box plot of the player weights of the Dallas Cowboys and the San Francisco 49ers.[13] Compare the weights of the two teams and note any differences that you observe.

Football Players' Weights (lbs)		
	Cowboys	49ers
Min	153	187
Q1	206	204.25
Median	243	232.5
Q3	268	288.75
Max	360	325
IQR	62	84.5
Mean	244.6032	243.9667
Std Dev.	46.03296	43.30049



17. Subjects in a marketing study were shown a product video and at the end of the video were given a test to measure their recall. The scores are listed below.

26	27	28	31	31	35	38	40	45	49	57
57	58	61	61	61	62	64	68	72	78	81
84	85	86	87	92	95	97	100			

a. What level of measurement does the data possess?

b. Determine Q_1, the first quartile.

c. Determine Q_2, the second quartile.

d. Determine Q_3, the third quartile.

e. Explain the meaning of these percentiles in the context of the marketing study.

f. Determine the interquartile range.

g. Construct a box plot for the test scores. Are there any outliers?

h. Compute the z-score for a test score of 81.

i. Compute the z-score for a test score of 62.

j. Explain what the z-scores in parts **h.** and **i.** are measuring.

18. Use the marketing study data from the previous problem to find the following.

a. Determine the percentile rank for the subject who scored 49.

b. Determine the percentile rank for the subject who scored 95.

19. Using the on-base percentage *(OBP)* variable from the Moneyball data set, answer the following questions.

Data

stat.hawkeslearning.com
Discovering Statistics and Data, Fourth Edition > Data Sets > Moneyball

a. What level of measurement does the data possess?

b. Determine Q_1, the first quartile. Round your answer to three decimal places.

c. Determine Q_2, the second quartile. Round your answer to three decimal places.

d. Determine Q_3, the third quartile. Round your answer to three decimal places.

e. Explain the meaning of these percentiles in the context of the on-base percentages.

f. Determine the interquartile range.

g. Construct a box plot for the on-base percentages. Are there any outliers?

h. Compute the z-score and percentile for an on-base percentage of 0.280. Round the z-score to three decimal places and the percentile to the nearest whole number.

i. Compute the z-score and percentile for an on-base percentage of 0.355. Round the z-score to three decimal places and the percentile to the nearest whole number.

j. Explain what the z-scores in parts **h.** and **i.** are measuring.

20. Using the on-base percentage (*OBP*) variable from the previous problem, find the following.

a. Determine the percentile rank for the Chicago Cubs in the year 2012. Round to the nearest whole number.

b. Determine the percentile rank for the New York Yankees in the year 2012. Round to the nearest whole number.

21. Consider a set of data in which the sample mean is 64 and the sample standard deviation is 21. For the following specific values, compute the z-score and interpret the results.

a. $x = 80$ **b.** $x = 64$ **c.** $x = 40$

22. A statistics student scored a 78 on the first exam of the semester and an 87 on the second exam of the semester. The average score and standard deviation of scores for the two exams are given in the following table. On which exam did the student perform relatively better?

Test Scores		
Statistic	**First Exam**	**Second Exam**
μ	74	85
σ	10	6

23. A hospital measures babies' lengths when they are born in both inches and centimeters. Eight babies are randomly selected and the following lengths are recorded in both inches and centimeters.

Newborn Lengths								
Baby	1	2	3	4	5	6	7	8
Inches	17.75	18.50	19.25	19.75	20.25	20.50	20.50	20.75
Centimeters	45.09	46.99	48.90	50.17	51.44	52.07	52.07	52.71

a. Determine the mean length in inches and centimeters for the babies.

b. Determine the standard deviation of the lengths of the babies in both inches and centimeters.

c. Prior to calculating the z-score for the length of a particular baby in inches and the z-score for the length of that same baby in centimeters, what does your intuition tell you about the values of the z-scores?

d. Determine the z-score for the length of Baby 3 measured in inches.

e. For Baby 3, determine the z-score for its length measured in centimeters.

f. Consider the z-scores determined in parts **d.** and **e.** Are the z-scores as you expected them to be? Explain.