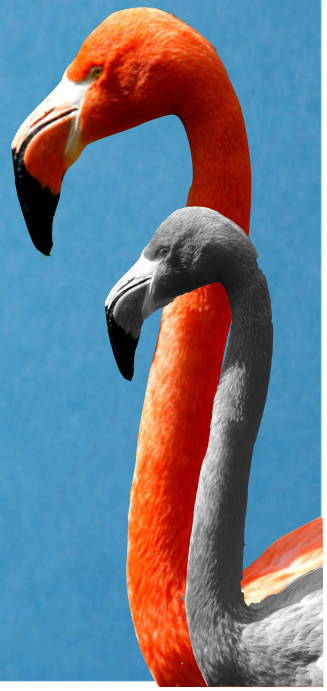
CHAPTER 2
Modeling Distributions of Data

2.1b
Describing Location in a Distribution



The Practice of Statistics, 5th Edition
Starnes, Tabor, Yates, Moore

Bedford Freeman Worth Publishers

Describing Location in a Distribution

Learning Objectives

After this section, you should be able to:

- ✓ FIND and INTERPRET the percentile of an individual value within a distribution of data.
- ✓ ESTIMATE percentiles and individual values using a cumulative relative frequency graph.
- ✓ FIND and INTERPRET the standardized score (z-score) of an individual value within a distribution of data.
- ✓ DESCRIBE the effect of adding, subtracting, multiplying by, or dividing by a constant on the shape, center, and spread of a distribution of data.

Transforming Data

Transforming converts the original observations from the original units of measurements to another scale. Transformations can affect the shape, center, and spread of a distribution.

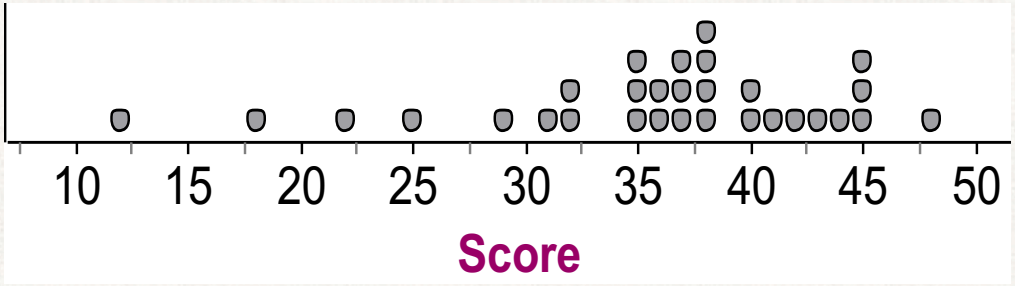
Effect of Adding (or Subtracting) a Constant

Adding the same number *a* to (subtracting *a* from) each observation:

- adds *a* to (subtracts *a* from) measures of center and location (mean, median, quartiles, percentiles), but
- Does not change the shape of the distribution or measures of spread (range, *IQR*, standard deviation).

Example: Test Scores

Here are a graph and table of summary statistics for a sample of 30 test scores. The maximum possible score on the test was 50 points. Describe the distribution.



	n	$\bar{x}$	s_x	Min	Q_1	Med	Q_3	Max	IQR	Range
Score	30	35.8	8.17	12	32	37	41	48	9	36

Describe the Distribution

Shape: The distribution of test scores appears skewed to the left with a peak at 38.

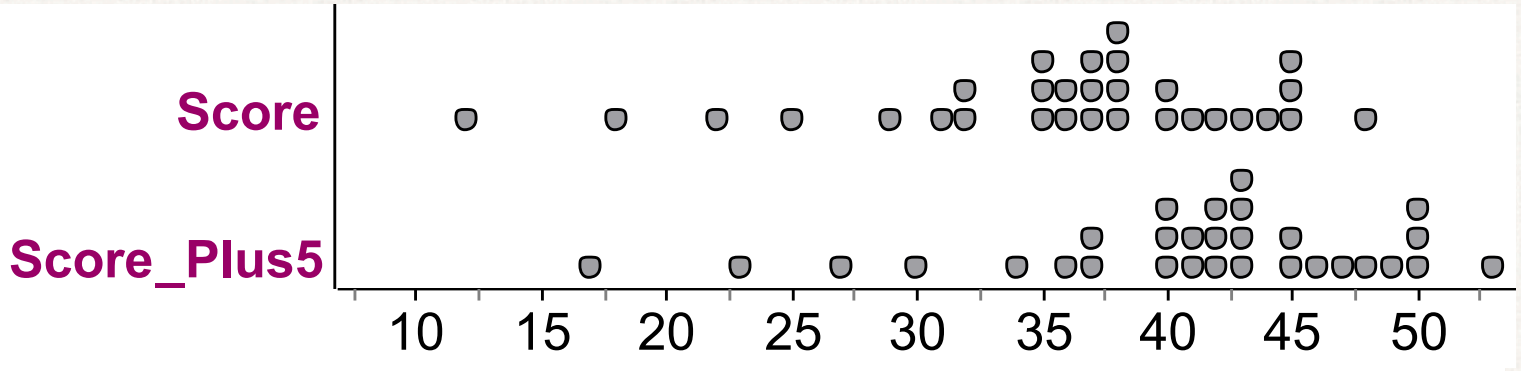
Center: The median score was 37 and the mean score was about 36. Due to the clear skewness and potential outliers, the median is a better choice for summarizing the "typical" score.

Spread: Because $Q_1 = 32$, about 25% of the students scored less than 32. The 75th percentile of the distribution is at about $Q_3 = 41$. The *IQR* of 9 points describes the spread of the middle 50% of students' scores. The standard deviation tells us that the typical distance of students' scores from the mean was about 8 points. Because standard deviation is not resistant to extreme values, we prefer the *IQR* to describe the variability of this distribution.

Outliers: By the $1.5 \times IQR$ rule, values greater than $41 + 1.5(9) = 53.5$ or less than $32 - 1.5(9) = 18.5$ are identified as outliers. So the two lowest scores—which are 12 and 18–are outliers.

Transforming Data – Test Scores

Suppose that the teacher was nice and added 5 points to each test score. How would this change the shape, center, and spread of the distribution?



	n	$\bar{x}$	s_x	Min	Q_1	Med	Q_3	Max	IQR	Range
Score	30	35.8	8.17	12	32	37	41	48	9	36
Score + 5	30	40.8	8.17	17	37	42	46	53	9	36

Transforming Data

Transforming converts the original observations from the original units of measurements to another scale. Transformations can affect the shape, center, and spread of a distribution.

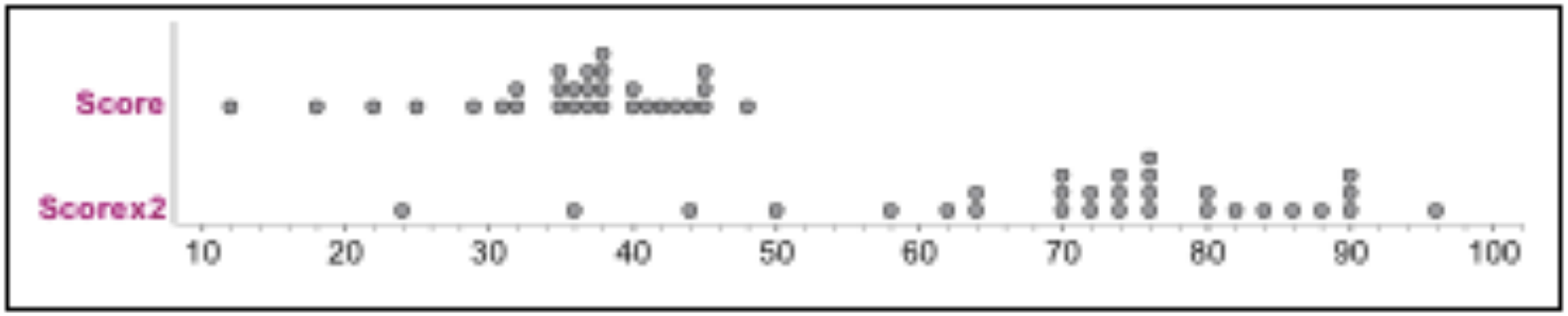
Effect of Multiplying (or Dividing) by a Constant

Multiplying (or dividing) each observation by the same number b:

- multiplies (divides) measures of center and location (mean, median, quartiles, percentiles) by b
- multiplies (divides) measures of spread (range, IQR, standard deviation) by $|b|$, but
- does not change the shape of the distribution

Example: Test Scores

Suppose that the teacher in the previous example wanted to convert the *original* test scores to percents. Because the test was out of 50 points, he should multiply each score by 2 to make them out of 100. How would this change the shape, center, and spread of the distribution?



	n	$\bar{x}$	s_x	Min	Q_1	Med	Q_3	Max	IQR	Range
Score	30	35.8	8.17	12	32	37	41	48	9	36
Score ×2	30	71.6	16.34	24	64	74	82	96	18	72

Example: Taxi Cabs

In 2010, taxi cabs in New York City charged an initial fee of $2.50 plus $2 per mile. In equation form, *fare* = 2.50 + 2(*miles*). At the end of a month, a businessman collects all his taxi cab receipts and calculates some numerical summaries. The mean fare he paid was $15.45, with a standard deviation of $10.20.

Problem: What are the mean and standard deviation of the lengths of his cab rides in miles?

Solve the equation for miles. $miles = \frac{fare - 2.50}{2}$ To find the mean number of miles, subtract 2.50 from the mean fare to get 12.95 and divide by 2 to get **6.475 miles**. To find the standard deviation, we divide the standard deviation of the fares by 2 to get **5.10 miles**. We don't subtract 2.50 because subtracting a constant from each observation does not change the spread of a distribution.

Describing Location in a Distribution

Section Summary

In this section, we learned how to…

- ✓ FIND and INTERPRET the percentile of an individual value within a distribution of data.
- ✓ ESTIMATE percentiles and individual values using a cumulative relative frequency graph.
- ✓ FIND and INTERPRET the standardized score (z-score) of an individual value within a distribution of data.
- ✓ DESCRIBE the effect of adding, subtracting, multiplying by, or dividing by a constant on the shape, center, and spread of a distribution of data.