STATISTICS
INFORMED DECISIONS USING DATA
Fifth Edition

Chapter 3
Numerically Summarizing Data

Pearson

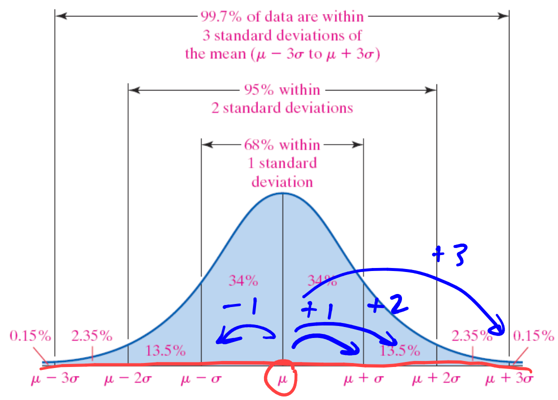
3.4 Measures of Position and Outliers

Learning Objectives

1. Determine and interpret ***z*-scores**
2. Interpret **percentiles**
3. Determine and interpret **quartiles**
4. Determine and interpret the **interquartile range (IQR)**
5. Check a set of data for outliers

3.4 Measures of Position and Outliers
3.4.1 Determine and Interpret z-Scores (1 of 3)

The **z-score** represents the distance that a data value is from the mean in terms of the number of standard deviations. I.e. the # of st. dev away from mean.



Population z-score

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

Ex: $\mu = 75$
$\sigma = 11$
$x = 86$

$$z = \frac{86 - 75}{11} = 1.$$

Sample z-score

$$\boxed{z = \frac{x - \bar{x}}{s}}$$

The z-score is unitless. It has mean 0 and standard deviation 1.

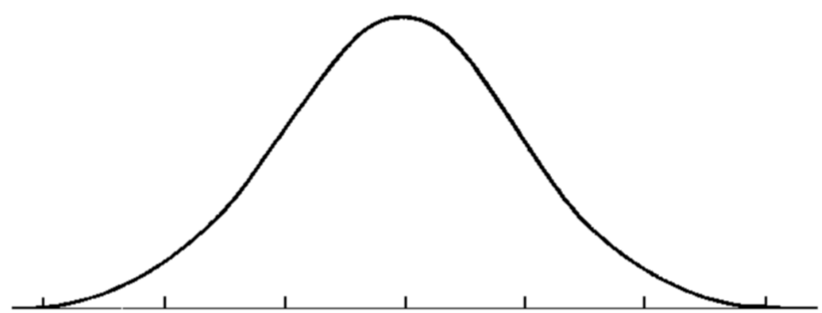
3.4 Measures of Position and Outliers
3.4.1 Determine and Interpret *z*-Scores (1 of 3)

Why are z-scores useful?

Allows us to compare two different normally distributed groups. Quick example: two sections of Math 136 with different instructors. You want to compare how you did with your friend's score. You can compare z-scores to see which had the higher z-score.

Population z-score

$$z = \frac{x - \mu}{\sigma}$$



The *z*-score is unitless. It has mean 0 and standard deviation 1.

3.4 Measures of Position and Outliers

3.4.1 Determine and Interpret *z*-Scores (1 of 3)

Remember:

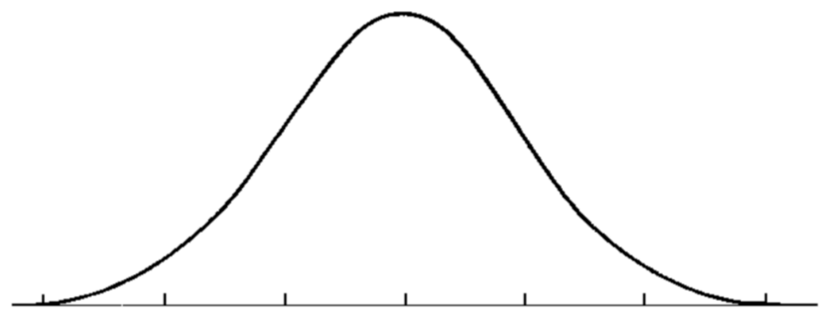
z-scores don't have units!

Rounding Rule for z-scores:

round all z-scores to two decimal places

Population *z*-score

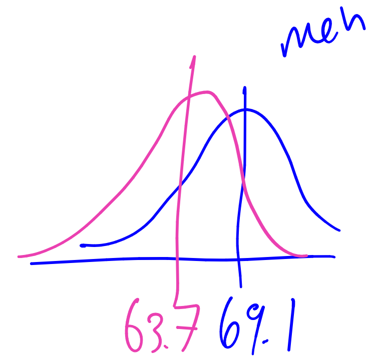
$$z = \frac{x - \mu}{\sigma}$$



The *z*-score is unitless. It has mean 0 and standard deviation 1.

3.4 Measures of Position and Outliers

3.4.1 Determine and Interpret *z*-Scores (2 of 3)



EXAMPLE Using Z-Scores

The mean height of males 20 years or older is 69.1 inches with a standard deviation of 2.8 inches. μ σ

The mean height of females 20 years or older is 63.7 inches with a standard deviation of 2.7 inches. μ σ

Data is based on information obtained from National Health and Examination Survey. **Who is relatively taller?**

relatively taller

Kevin Garnett whose height is 83 inches

$z = \frac{83 - 69.1}{2.8} = \boxed{4.96}$ KG

or

Candace Parker whose height is 76 inches

CP $z = \frac{76 - 63.7}{2.7}$

$z = \boxed{4.56}$

3.4 Measures of Position and Outliers

3.4.1 Determine and Interpret *z*-Scores (3 of 3)

$$z_{kg} = \frac{83 - 69.1}{2.8} = 4.96$$

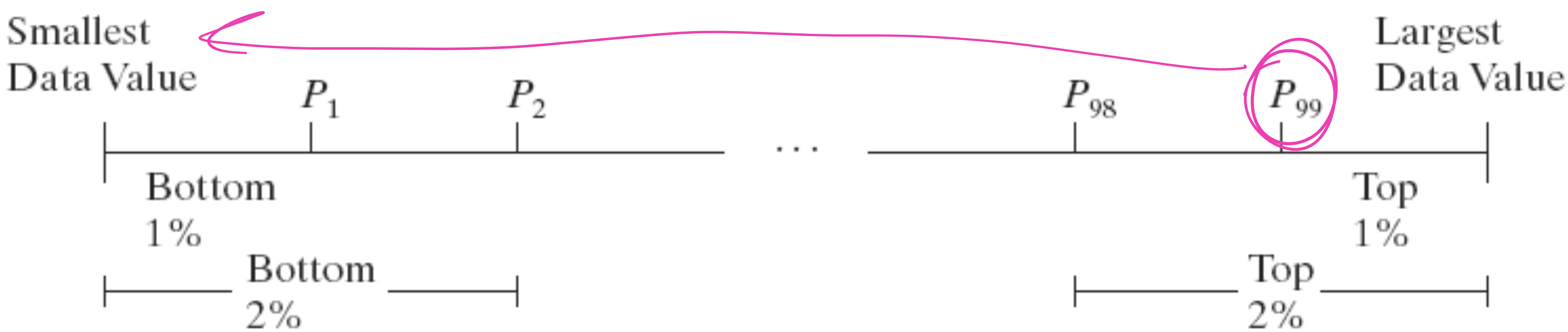
$$z_{cp} = \frac{76 - 63.7}{2.7} = 4.56$$

Kevin Garnett’s height is 4.96 standard deviations above the mean. Candace Parker’s height is 4.56 standard deviations above the mean. Kevin Garnett is relatively taller.

3.4 Measures of Position and Outliers

3.4.2 Interpret Percentiles (1 of 3)

The ***k*th percentile**, denoted, P_k, of a set of data is a value such that k percent of the observations are less than or equal to the value.



99 percentile = means 99 % of values are less.

3.4 Measures of Position and Outliers

3.4.2 Interpret Percentiles (2 of 3)

EXAMPLE Interpret a Percentile

The Graduate Record Examination (GRE) is a test required for admission to many U.S. graduate schools. The University of Pittsburgh Graduate School of Public Health requires a GRE score no less than the 70th percentile for admission into their Human Genetics MPH or MS program.

(Source: http://www.publichealth.pitt.edu/interior.php?pageID=101.)

Interpret this admissions requirement.

3.4 Measures of Position and Outliers

3.4.2 Interpret Percentiles (3 of 3)

EXAMPLE Interpret a Percentile

In general, the 70^{th} percentile is the score such that 70% of the individuals who took the exam scored worse, and 30% of the individuals scores better.

In order to be admitted to this program, an applicant must score as high or higher than 70% of the people who take the GRE.

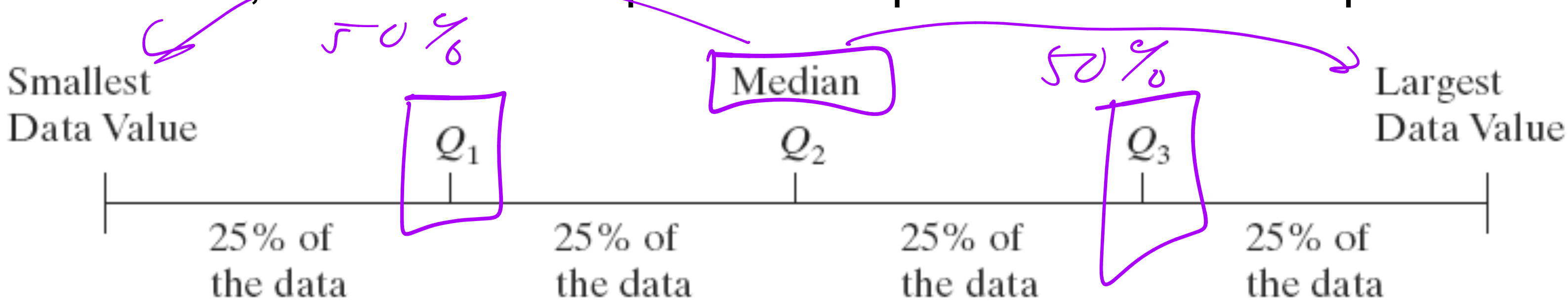
Put another way, the individual's score must be in the top 30%.

3.4 Measures of Position and Outliers

3.4.3 Determine and Interpret Quartiles (1 of 5)

Quartiles divide data sets into fourths, or four equal parts.

- The 1st quartile, denoted Q_1, divides the bottom 25% the data from the top 75%. Therefore, the 1st quartile is equivalent to the 25th percentile.
- The 2nd quartile divides the bottom 50% of the data from the top 50% of the data, so that the 2nd quartile is equivalent to the 50th percentile, which is equivalent to the median.
- The 3rd quartile divides the bottom 75% of the data from the top 25% of the data, so that the 3rd quartile is equivalent to the 75th percentile.



Pearson

3.4 Measures of Position and Outliers

3.4.3 Determine and Interpret Quartiles (2 of 5)

Finding Quartiles

Step 1: Arrange the data in ascending order.

Step 2: Determine the median, *Med*, or second quartile, Q_2 .

Step 3: Divide the data set into halves: the observations below (to the left of) *Med* and the observations above *Med*. The first quartile, Q_1, is the median of the bottom half, and the third quartile, Q_3, is the median of the top half.

Calculator: Enter data into a list. Compute "1-Var Stats".

It will tell you Q1, Med(=Q2), & Q3

3.4 Measures of Position and Outliers

3.4.3 Determine and Interpret Quartiles (3 of 5)

EXAMPLE Finding and Interpreting Quartiles

A group of Brigham Young University—Idaho students (Matthew Herring, Nathan Spencer, Mark Walker, and Mark Steiner) collected data on the speed of vehicles traveling through a construction zone on a state highway, where the posted speed was 25 mph. The recorded speed of 14 randomly selected vehicles is given below:

20, 24, 27, 28, 29, 30, 32, 33, 34, 36, 38, 39, 40, 40

Find and interpret the quartiles for speed in the construction zone.

min Q1 Med Q3 max

20 28 32.5 38 40

IQR = 38 − 28 = 10

3.4 Measures of Position and Outliers

3.4.3 Determine and Interpret Quartiles (4 of 5)

EXAMPLE Finding and Interpreting Quartiles

Step 1: The data is already in ascending order.

Step 2: There are n = 14 observations, so the median, or second quartile, Q_2, is the mean of the 7th and 8th observations. Therefore, *Med* = 32.5.

Step 3: The median of the bottom half of the data is the first quartile, Q_1.

20, 24, 27, 28, 29, 30, 32

The median of these seven observations is 28. Therefore, $Q_1 = 28$. The median of the top half of the data is the third quartile, Q_3. Therefore, $Q_3 = 38$.

3.4 Measures of Position and Outliers

3.4.3 Determine and Interpret Quartiles (5 of 5)

Interpretation:

- 25% of the speeds are <u>less than or equal to</u> the first quartile, 28 miles per hour, and 75% of the speeds are greater than 28 miles per hour.
- 50% of the speeds are <u>less than or equal to</u> the second quartile, 32.5 miles per hour, and 50% of the speeds are greater than 32.5 miles per hour.
- 75% of the speeds are <u>less than or equal to</u> the third quartile, 38 miles per hour, and 25% of the speeds are greater than 38 miles per hour.

3.4 Measures of Position and Outliers

3.4.4 Determine and Interpret the Interquartile Range (1 of 3)

The **interquartile range, IQR**, is the range of the middle 50% of the observations in a data set.

That is, the IQR is the difference between the third and first quartiles and is found using the formula

$$\boxed{\mathbf{IQR} = Q_3 - Q_1}$$

3.4 Measures of Position and Outliers

3.4.4 Determine and Interpret the Interquartile Range (2 of 3)

EXAMPLE Determining and Interpreting the Interquartile Range

Determine and interpret the interquartile range of the speed data.

$IQR = 38 - 28$

$= \boxed{10 \text{ mph}}$

3.4 Measures of Position and Outliers

3.4.4 Determine and Interpret the Interquartile Range (2 of 3)

EXAMPLE Determining and Interpreting the Interquartile Range

Determine and interpret the interquartile range of the speed data.

$Q_1 = 28 \qquad Q_3 = 38$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 38 - 28 \\ &= 10 \end{aligned}$$

The difference from the slowest to fastest speeds of the middle 50% of the cars traveling through the construction zone is 10 miles per hour.

3.4 Measures of Position and Outliers

3.4.4 Determine and Interpret the Interquartile Range (3 of 3)

Suppose a 15^{th} car travels through the construction zone at 100 miles per hour.

How does this value impact the mean, median, standard deviation, and interquartile range?

3.4 Measures of Position and Outliers

3.4.4 Determine and Interpret the Interquartile Range (3 of 3)

Suppose a 15th car travels through the construction zone at 100 miles per hour. How does this value impact the mean, median, standard deviation, and interquartile range?

	Without 15th car	With 15th car
Mean	32.1 mph	36.7 mph
Median	32.5 mph	33 mph
Standard deviation	6.2 mph	18.5 mph
IQR	10 mph	11 mph

Summary: Which Measures to Report

Shape of Distribution	**Measures of Central Tendency**	**Measures of Dispersion**
Symmetric	Mean	Standard deviation
Skewed left or skewed right	Median	Interquartile range

3.4 Measures of Position and Outliers

3.4.5 Check a Set of Data for Outliers (1 of 2)

Checking for Outliers by Using Quartiles

Step 1 Determine the first and third quartiles of the data.

Step 2 Compute the interquartile range.

Step 3 Determine the fences. **Fences** serve as cutoff points for determining outliers.

$$\textbf{Lower Fence} = Q_1 - 1.5(\text{IQR})$$

$$\textbf{Upper Fence} = Q_3 + 1.5(\text{IQR})$$

Step 4 If a data value is less than the lower fence or greater than the upper fence, it is considered an **outlier**.

3.4 Measures of Position and Outliers

3.4.5 Check a Set of Data for Outliers (2 of 2)

EXAMPLE Determining and Interpreting the Interquartile Range

Check the speed data for outliers.

3.4 Measures of Position and Outliers

3.4.5 Check a Set of Data for Outliers (2 of 2)

EXAMPLE Determining and Interpreting the Interquartile Range

Check the speed data for outliers.

Step 1: The first and third quartiles are Q_1 = 28 mph and Q_3 = 38 mph.

Step 2: The interquartile range is 10 mph.

Step 3: The fences are

Lower Fence = $Q_1 - 1.5(\text{IQR}) = 28 - 1.5(10) = 13$ mph

Upper Fence = $Q_3 + 1.5(\text{IQR}) = 38 + 1.5(10) = 53$ mph

Step 4: There are no values less than 13 mph or greater than 53 mph. Therefore, there are no outliers.