Describing Data:
Numerical Measures

Chapter 3

The Tony Fernandes School of Business UC-Cambodia

Course MTH120: Fundamental of Statistic, International Program

Instructor: Heng Serey Rathana, BA,BS,MA,MBA,MCE,Ph.d Candidate

Academic Year: 2023-2024

Learning Objectives

LO1 Explain the concept of central tendency. ការសិក្សាចំនុចកណ្តាល

LO2 Identify and compute the arithmetic mean. ការកណត់និងគណនាមធ្យមនព្វន្ត

LO3 Compute and interpret the weighted mean. គណនានិងបកស្រាយមធ្យមរួម

LO4 Determine the median. ការកំណត់នៅមេដ្យាន

LO5 Identify the mode. ការកណត់ម៉ូត

LO6 Explain and apply measures of dispersion. ពន្យល់និងអនុវត្តនការវាស់វែងបម្រែបម្រួល

LO7 Compute and interpret the standard deviation. គណនានិងបកស្រាយនៅគលាតស្តង់ដា

LO8 Explain Chebyshev's Theorem and the Empirical Rule. ពន្យល់អំពីទ្រឹស្តីបទ Chebysheve និងទ្រឹស្តី Empirical

LO9 Compute the mean and standard deviation of grouped data. គណនាមធ្យមនិងគំលាតស្តង់ដាសម្រាប់ទិន្នន័យជាក្រុម

Describing Data Numerically

- Describing Data Numerically
  - Central Tendency
    - Arithmetic Mean
    - Median
    - Mode
  - Variation
    - Range
    - Interquartile Range
    - Variance
    - Standard Deviation
    - Coefficient of Variation

Measures of Central Tendency

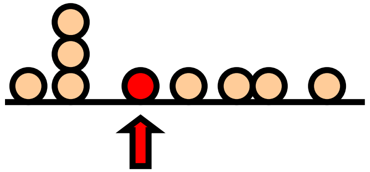
Central Tendency

Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

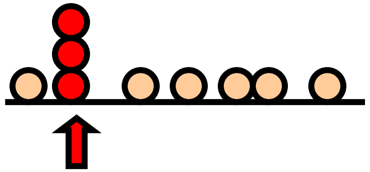
Arithmetic average

Median



Midpoint of ranked values

Mode

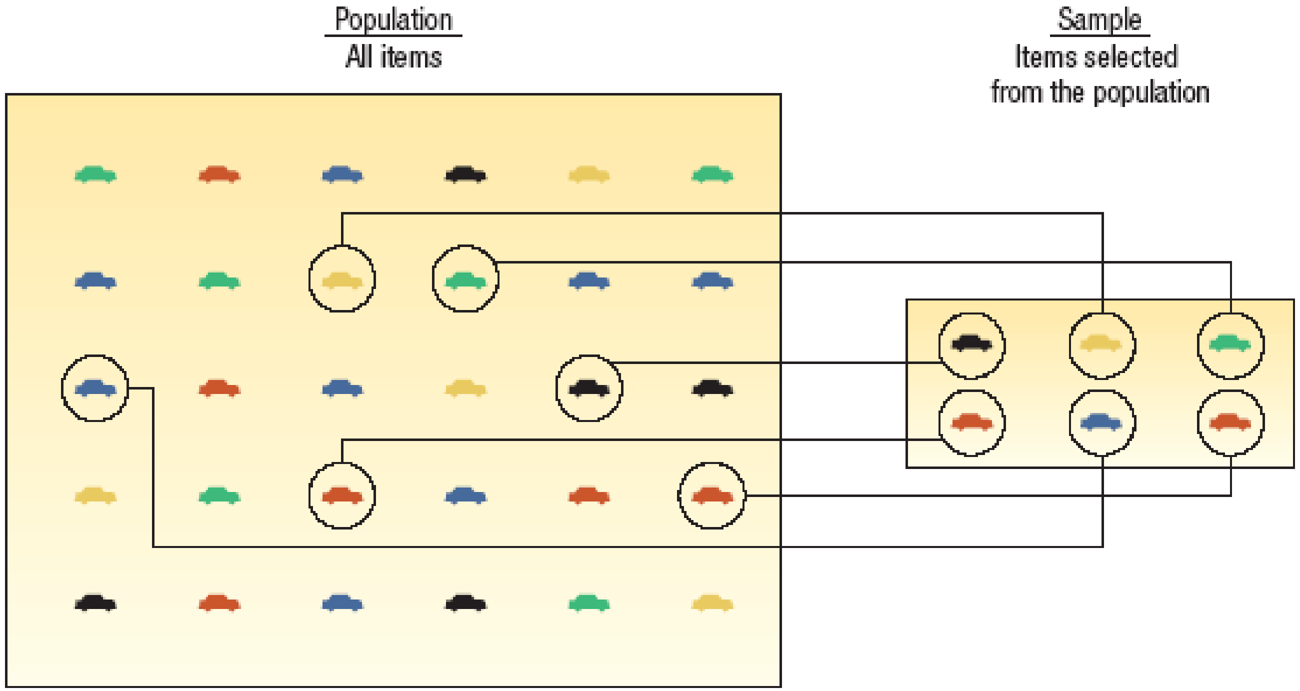


Most frequently observed value

Parameter Versus Statistics

PARAMETER **A measurable characteristic of a *population*.**

STATISTIC **A measurable characteristic of a *sample*.**



Population Mean

For ungrouped data, the population mean is the sum of all the population values divided by the total number of population values. The sample mean is the sum of all the sample values divided by the total number of sample values.

POPULATION MEAN $$\mu = \frac{\Sigma X}{N} \qquad [3\text{–}1]$$

$$\text{Population mean} = \frac{\text{Sum of all the values in the population}}{\text{Number of values in the population}}$$

where:

- μ represents the population mean. It is the Greek lowercase letter "mu."
- N is the number of values in the population.
- X represents any particular value.
- Σ is the Greek capital letter "sigma" and indicates the operation of adding.
- ΣX is the sum of the X values in the population.

Any measurable characteristic of a population is called a **parameter.** The mean of a population is an example of a parameter.

PARAMETER A characteristic of a population.

Example There are 42 exits on I-75 through the state of Kentucky. Listed below are the distances between exits (in miles).

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

Why is this information a population? What is the mean number of miles between exits?

This is a population because we are considering all the exits in Kentucky. We add the distances between each of the 42 exits. The total distance is 192 miles. To find the arithmetic mean, we divide this total by 42. So the arithmetic mean is 4.57 miles, found by 192/42. From formula (3–1):

$$\mu = \frac{\Sigma X}{N} = \frac{11 + 4 + 10 + \cdots + 1}{42} = \frac{192}{42} = 4.57$$

How do we interpret the value of 4.57? It is the typical number of miles between exits. Because we considered all the exits in Kentucky, this value is a population parameter.

The Sample Mean

SAMPLE MEAN $\bar{X} = \frac{\Sigma X}{n}$ **[3–2]**

$$\text{Sample mean} = \frac{\text{Sum of all the values in the sample}}{\text{Number of values in the sample}}$$

where:

- $\bar{X}$ represents the sample mean. It is read "X bar."
- n is the number of values in the sample.
- X represents any particular value.
- Σ is the Greek capital letter "sigma" and indicates the operation of adding.
- ΣX is the sum of the X values in the sample.

The mean of a sample, or any other measure based on sample data, is called a **statistic.** If the mean weight of a sample of 10 jars of Smucker's strawberry jam is 41 ounces, this is an example of a statistic.

STATISTIC A characteristic of a sample.

Example

SunCom is studying the number of minutes used by clients in a particular cell phone rate plan. A random sample of 12 clients showed the following number of minutes used last month.

90	77	94	89	119	112
91	110	92	100	113	83

What is the arithmetic mean number of minutes used?

Using formula (3–2), the sample mean is:

$$\text{Sample mean} = \frac{\text{Sum of all values in the sample}}{\text{Number of values in the sample}}$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{90 + 77 + \cdots + 83}{12} = \frac{1170}{12} = 97.5$$

The arithmetic mean number of minutes used last month by the sample of cell phone users is 97.5 minutes.

PROPERTIES OF THE ARITHMTIC MEAN

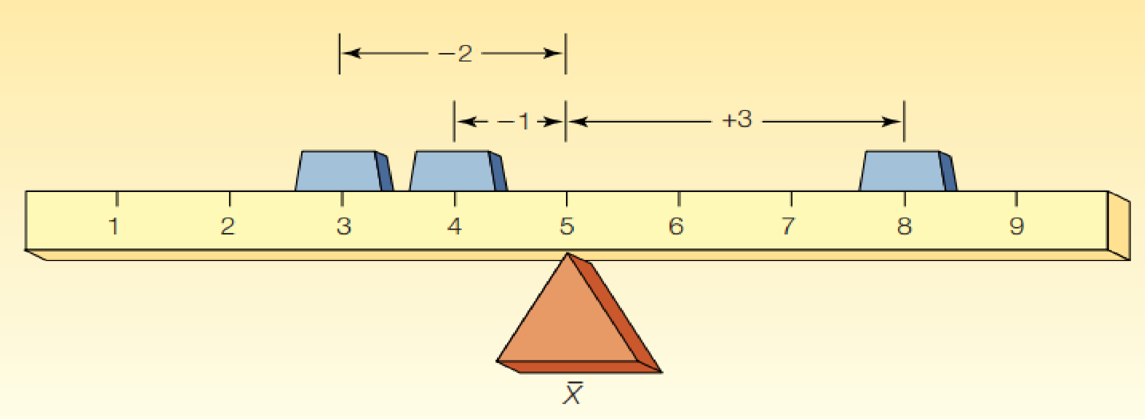
1. **Every set of interval- or ratio-level data has a mean.** Example: ratio-level data include such data as ages, incomes, and weights, with the distance between numbers being constant.
2. **All the values are included in computing the mean.**
3. **The mean is unique.** That is, there is only one mean in a set of data
4. **The sum of the deviations of each value from the mean is zero.**

Expressed symbolically:

$$\Sigma(X - \bar{X}) = 0$$

As an example, the mean of 3, 8, and 4 is 5. Then:

$$\begin{aligned}\Sigma(X - \bar{X}) &= (3 - 5) + (8 - 5) + (4 - 5)\\ &= -2 + 3 - 1\\ &= 0\end{aligned}$$



- Mean as a balance point
- Mean unduly affected by unusually large or small values

Try This

Self-Review

1. The annual incomes of a sample of middle-management employees at Westinghouse are: $62,900, $69,100, $58,300, and $76,800.
   (a) Give the formula for the sample mean.
   (b) Find the sample mean.
   (c) Is the mean you computed in (b) a statistic or a parameter? Why?
   (d) What is your best estimate of the population mean?
2. All the students in advanced Computer Science 411 are a population. Their course grades are 92, 96, 61, 86, 79, and 84.
   (a) Give the formula for the population mean.
   (b) Compute the mean course grade.
   (c) Is the mean you computed in (b) a statistic or a parameter? Why?

1. Compute the mean of the following population values: 6, 3, 5, 7, 6.
2. Compute the mean of the following population values: 7, 5, 7, 3, 7, 4.
3. **a.** Compute the mean of the following sample values: 5, 9, 4, 10.
b. Show that $\Sigma(X - \overline{X}) = 0$.
4. **a.** Compute the mean of the following sample values: 1.3, 7.0, 3.6, 4.1, 5.0.
b. Show that $\Sigma(X - \overline{X}) = 0$.
5. Compute the mean of the following sample values: 16.25, 12.91, 14.58.
6. Suppose you go to the grocery store and spend $61.85 for the purchase of 14 items. What is the mean price per item?

The Weighted Mean

The weighted mean is a special case of the arithmetic mean. It occurs when there are several observations of the same value.

WEIGHTED MEAN
$$\bar{X}_w = \frac{w_1X_1 + w_2X_2 + w_3X_3 + \cdots + w_nX_n}{w_1 + w_2 + w_3 + \cdots + w_n} \qquad [3\text{–}3]$$

This may be shortened to:

$$\boxed{\bar{X}_w = \frac{\Sigma(wX)}{\Sigma w}}$$

Example

The Carter Construction Company pays its hourly employees \$16.50, \$19.00, or \$25.00 per hour. There are 26 hourly employees, 14 of which are paid at the \$16.50 rate, 10 at the \$19.00 rate, and 2 at the \$25.00 rate. What is the mean hourly rate paid the 26 employees?

To find the mean hourly rate, we multiply each of the hourly rates by the number of employees earning that rate. From formula (3–3), the mean hourly rate is

$$\bar{X}_w = \frac{14(\$16.50) + 10(\$19.00) + 2(\$25.00)}{14 + 10 + 2} = \frac{\$471.00}{26} = \$18.1154$$

The weighted mean hourly wage is rounded to \$18.12.

Try This

Springers sold 95 Antonelli men's suits for the regular price of \$400. For the spring sale, the suits were reduced to \$200 and 126 were sold. At the final clearance, the price was reduced to \$100 and the remaining 79 suits were sold.

(a) What was the weighted mean price of an Antonelli suit?

(b) Springers paid \$200 a suit for the 300 suits. Comment on the store's profit per suit if a salesperson receives a \$25 commission for each one sold.

The Median

MEDIAN The midpoint of the values after they have been ordered from the smallest to the largest, or the largest to the smallest.

PROPERTIES OF THE MEDIAN

1. There is a unique median for each data set.
2. It is not affected by extremely large or small values. Therefore, the median is a valuable measure of location when such values do occur.
3. It can be computed for ordinal-level data or higher.

Prices Ordered from Low to High		Prices Ordered from High to Low
$ 60,000		$275,000
65,000		80,000
70,000	← Median →	70,000
80,000		65,000
275,000		60,000

The Median

Finding the Median

- The location of the median:

$$\text{Median position} = \frac{n+1}{2} \text{ position in the ordered data}$$

  - If the number of values is odd, the median is the middle number
  - If the number of values is even, the median is the average of the two middle numbers

- Note that $\frac{n+1}{2}$ is not the *value* of the median, only the *position* of the median in the ranked data

The Median

Median less affected by extreme values

The ages for a sample of five college students are:

21, 25, 19, 20, 22

Arranging the data in ascending order gives:

19, 20, 21, 22, 25.

Thus the median is 21.

The heights of four basketball players, in inches, are:

76, 73, 80, 75

Arranging the data in ascending order gives:

73, 75, 76, 80.

Thus the median is **(75+76) /2 =75.5**

Example

Facebook is a popular social networking website. Users can add friends and send them messages, and update their personal profiles to notify friends about themselves and their activities. A sample of 10 adults revealed they spent the following number of hours last month using Facebook.

3	5	7	5	9	1	3	9	17	10

Find the median number of hours.

Note that the number of adults sampled is even (10). The first step, as before, is to order the hours using Facebook from low to high. Then identify the two middle times. The arithmetic mean of the two middle observations gives us the median hours. Arranging the values from low to high:

1	3	3	5	5	7	9	9	10	17

The median is found by averaging the two middle values. The middle values are 5 hours and 7 hours, and the mean of these two values is 6. We conclude that the typical Facebook user spends 6 hours per month at the website. Notice that the median is not one of the values. Also, half of the times are below the median and half are above it.

The Mode

MODE The value of the observation that appears most frequently.

- A measure of central tendency
- Value that occurs most often
- Not affected by extreme values
- Used for either numerical or categorical data
- There may maybe *no mode*
- There may be *several modes*

The annual salaries of quality-control managers in selected states are shown below. What is the modal annual salary?

State	Salary	State	Salary	State	Salary
Arizona	$35,000	Illinois	$58,000	Ohio	$50,000
California	49,100	Louisiana	60,000	Tennessee	60,000
Colorado	60,000	Maryland	60,000	Texas	71,400
Florida	60,000	Massachusetts	40,000	West Virginia	60,000
Idaho	40,000	New Jersey	65,000	Wyoming	55,000

A perusal of the salaries reveals that the annual salary of $60,000 appears more often (six times) than any other salary. The mode is, therefore, $60,000.

The mode is especially useful in summarizing **nominal-level data**

The Mode

Example

Recall the data regarding the distance in miles between exits on I-75 through Kentucky. The information is repeated below.

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

What is the modal distance?

The first step is to organize the distances into a frequency table. This will help us determine the distance that occurs most frequently.

Distance in Miles between Exits	Frequency
1	8
2	7
3	7
4	3
5	4
6	1
7	3
8	2
9	1
10	4
11	1
14	1
Total	42

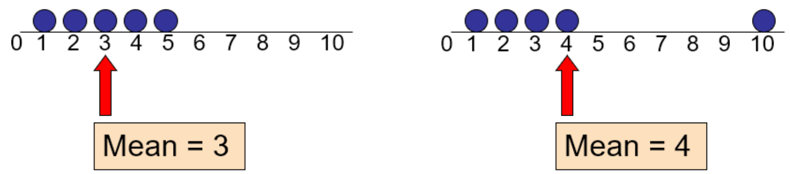
The distance that occurs most often is one mile. This happens eight times—that is, there are eight exits that are one mile apart. So the modal distance between exits is one mile.

Which of the three measures of location (mean, median, or mode) best represents the central location of this data? Is the mode the best measure of location to represent the Kentucky data? No. The mode assumes only the nominal scale of measurement and the variable miles is measured using the ratio scale. We calculated the mean to be 4.57 miles. See page 59. Is the mean the best measure of location to represent this data? Probably not. There are several cases in which the distance between exits is large. These values are affecting the mean, making it too large and not representative of the distances between exits. What about the median? The median distance is 3 miles. That is, half of the distances between exits are 3 miles or less. In this case, the median of 3 miles between exits is probably a more representative measure of the distance between exits.

We can determine the mode for all levels of data—nominal, ordinal, interval, and ratio. The mode also has the advantage of not being affected by extremely high or low values.

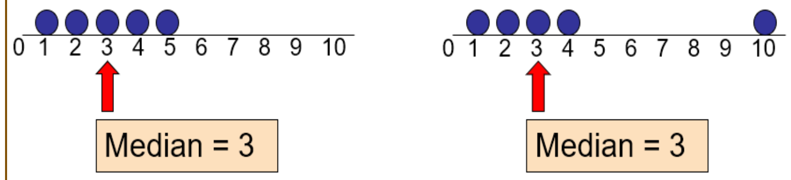
Summary Central Tendency

- The most common measure of central tendency
- Mean = sum of values divided by the number of values
- Affected by extreme values (outliers)



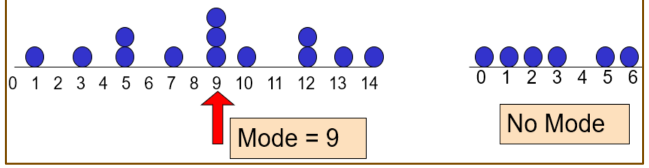
$$\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3 \qquad \frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$

- In an ordered list, the median is the "middle" number (50% above, 50% below)



- Not affected by extreme values

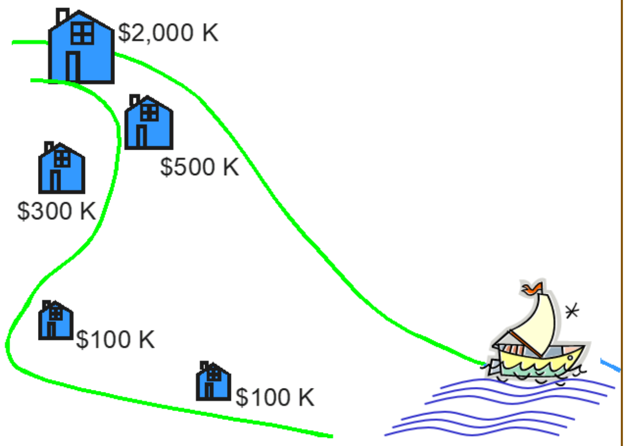
- A measure of central tendency
- Value that occurs most often
- Not affected by extreme values
- Used for either numerical or categorical data
- There may maybe no mode
- There may be several modes



- Five houses on a hill by the beach

House Prices:

$2,000,000
500,000
300,000
100,000
100,000



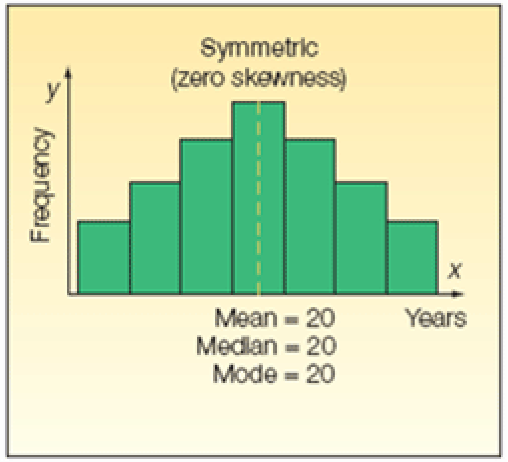
House Prices:

$2,000,000
500,000
300,000
100,000
100,000

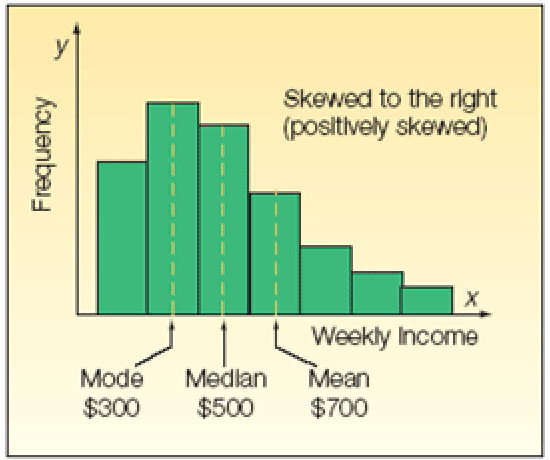
Sum 3,000,000

- **Mean:** ($3,000,000/5) = **$600,000**
- **Median:** middle value of ranked data = **$300,000**
- **Mode:** most frequent value = **$100,000**

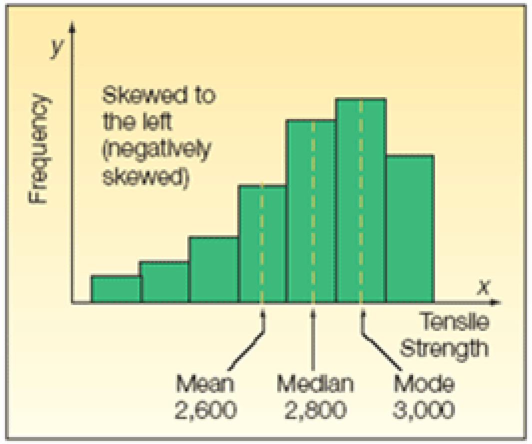
The Relative Positions of the Mean, Median and the Mode



zero skewness
mode = median = mean



positive skewness
mode < median < mean

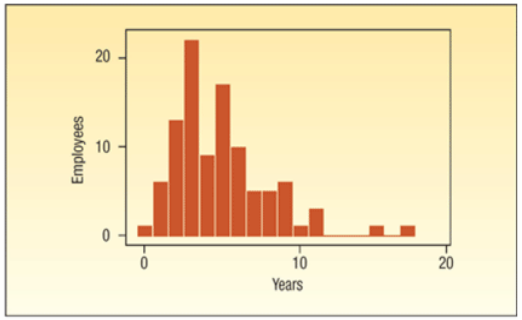


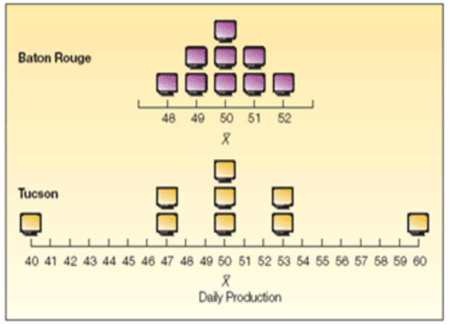
negative skewness
mode > median > mean

Measures of Dispersion

- A measure of location, such as the mean or the median, only describes the center of the data. It is valuable from that standpoint, but it does not tell us anything about the *spread* of the data.
- For example, if your nature guide told you that the river ahead averaged 3 feet in depth, would you want to wade across on foot without additional information? Probably not. You would want to know something about the variation in the depth.
- A second reason for studying the dispersion in a set of data is to compare the spread in two or more distributions.

RANGE Range = Largest value − Smallest value **[3–6]**





MEAN DEVIATION $MD = \frac{\Sigma|X - \bar{X}|}{n}$ **[3–7]**

POPULATION VARIANCE $\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$ **[3–8]**

POPULATION STANDARD DEVIATION $\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$ **[3–9]**

EXAMPLE – Mean Deviation

MEAN DEVIATION

$$MD = \frac{\Sigma|X - \overline{X}|}{n} \quad [3–7]$$

EXAMPLE:

The number of cappucci Airport between 4 and 7 and 80. Determine the m

where:
X is the value of each observation.
$\overline{X}$ is the arithmetic mean of the values.
n is the number of observations in the sample.
$||$ indicates the absolute value.

Step 1: Compute the mean

Step 2: Subtract the mean (50) is negative

Step 3: Sum the absolute differences found in step 2 then divide by the number of observations

MEAN DEVIATION The arithmetic mean of the absolute values of the deviations from the arithmetic mean.

Number of Cappuccinos Sold Daily	$(X - \overline{X})$	Absolute Deviation
20	(20 − 50) = −30	30
40	(40 − 50) = −10	10
50	(50 − 50) = 0	0
60	(60 − 50) = 10	10
80	(80 − 50) = 30	30
	Total	80

$$MD = \frac{\Sigma|X - \overline{X}|}{n} = \frac{80}{5} = 16$$

Try This

Self-Review

$$MD = \frac{\Sigma|X - \bar{X}|}{n}$$

where:

X is the value of each observation.
$\bar{X}$ is the arithmetic mean of the values.
n is the number of observations in the sample.
$||$ indicates the absolute value.

The weights of containers being shipped to Ireland are (in thousands of pounds):

95	103	105	110	104	105	112	90

(a) What is the range of the weights?
(b) Compute the arithmetic mean weight.
(c) Compute the mean deviation of the weights.

Variance and Standard Deviation

- The variance and standard deviations are nonnegative and are zero only if all observations are the same.
- For populations whose values are *near the mean*, the variance and standard deviation will be small.
- For populations whose values are *dispersed from the mean*, the population variance and standard deviation will be large.

VARIANCE The arithmetic mean of the squared deviations from the mean.

STANDARD DEVIATION The square root of the variance.

Variance and Standard Deviation

POPULATION VARIANCE $$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$$ **[3–8]**

where:

- σ^2 is the population variance (σ is the lowercase Greek letter sigma). It is read as "sigma squared."
- X is the value of an observation in the population.
- μ is the arithmetic mean of the population.
- N is the number of observations in the population.

POPULATION STANDARD DEVIATION $$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$$ **[3–9]**

Note the process of computing the variance.

1. Begin by finding the mean.
2. Find the difference between each observation and the mean, and square that difference.
3. Sum all the squared differences.
4. Divide the sum of the squared differences by the number of items in the population.

Variance and Standard Deviation

Example

The number of traffic citations issued last year by month in Beaufort County, South Carolina, is reported below.

Month	January	February	March	April	May	June	July	August	September	October	November	December
Citations	19	17	22	18	28	34	45	39	38	44	34	10

Determine the population variance.

$$\mu = \frac{\Sigma X}{N}$$

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N} = \frac{1488}{12} = 124$$

Month	**Citations (X)**	$X - \mu$	$(X - \mu)^2$
January	19	−10	100
February	17	−12	144
March	22	−7	49
April	18	−11	121
May	28	−1	1
June	34	5	25
July	45	16	256
August	39	10	100
September	38	9	81
October	44	15	225
November	34	5	25
December	10	−19	361
Total	348	0	1,488

Try This

Self-Review

$$\mu = \frac{\Sigma X}{N}$$

$$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$$

The Philadelphia office of Price Waterhouse Coopers LLP hired five accounting trainees this year. Their monthly starting salaries were: $3,536; $3,173; $3,448; $3,121; and $3,622.

(a) Compute the population mean.
(b) Compute the population variance.
(c) Compute the population standard deviation.
(d) The Pittsburgh office hired six trainees. Their mean monthly salary was $3,550, and the standard deviation was $250. Compare the two groups.

The annual incomes of the five vice presidents of TMV Industries are: $125,000; $128,000; $122,000; $133,000; and $140,000. Consider this a population.

a. What is the range?
b. What is the arithmetic mean income?
c. What is the population variance? The standard deviation?
d. The annual incomes of officers of another firm similar to TMV Industries were also studied. The mean was $129,000 and the standard deviation $8,612. Compare the means and dispersions in the two firms.

Variance and Standard Deviation

SAMPLE VARIANCE $$s^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1}$$ **[3–10]**

where:

s^2 is the sample variance.
X is the value of each observation in the sample.
$\overline{X}$ is the mean of the sample.
n is the number of observations in the sample.

SAMPLE STANDARD DEVIATION $$s = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n - 1}}$$ **[3–11]**

Variance and Standard Deviation

Example

The hourly wages for a sample of part-time employees at Home Depot are: \$12, \$20, \$16, \$18, and \$19. What is the sample variance?

The sample variance is computed by using formula (3–10).

$$\bar{X} = \frac{\Sigma X}{n} = \frac{\$85}{5} = \$17$$

Hourly Wage (X)	$X - \bar{X}$	$(X - \bar{X})^2$
\$12	−\$5	25
20	3	9
16	−1	1
18	1	1
19	2	4
\$85	0	40

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{40}{5 - 1}$$

$$= 10 \text{ in dollars squared}$$

Try This

Self-Review

$$s^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1}$$

$$s = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n - 1}}$$

The years of service for a sample of seven employees at a State Farm Insurance claims office in Cleveland, Ohio, are: 4, 2, 5, 4, 5, 2, and 6. What is the sample variance? Compute the sample standard deviation.

The Houston, Texas, Motel Owner Association conducted a survey regarding weekday motel rates in the area. Listed below is the room rate for business-class guests for a sample of 10 motels.

\$101	\$97	\$103	\$110	\$78	\$87	\$101	\$80	\$106	\$88

a. Compute the sample variance.
b. Determine the sample standard deviation.

Interpretation and Uses of the Standard Deviation

Chebyshev's Theorem

- ✓ The *Russian mathematician P. L. Chebyshev (1821–1894)* developed a theorem that allows us to determine the minimum proportion of the values that lie within a specified number of standard deviations of the mean.

CHEBYSHEV'S THEOREM For any set of observations (sample or population), the proportion of the values that lie within k standard deviations of the mean is at least $1 - 1/k^2$, where k is any constant greater than 1.

Interpretation and Uses of the Standard Deviation

Chebyshev's Rule

For any quantitative data set and any real number k greater than or equal to 1, at least $1 - 1/k^2$ of the observations lie within k standard deviations to either side of the mean, that is, between $\bar{x} - k \cdot s$ and $\bar{x} + k \cdot s$.

Two specific cases of Chebyshev's rule are applied frequently, namely, when $k = 2$ and $k = 3$. These specific cases can be stated as follows.

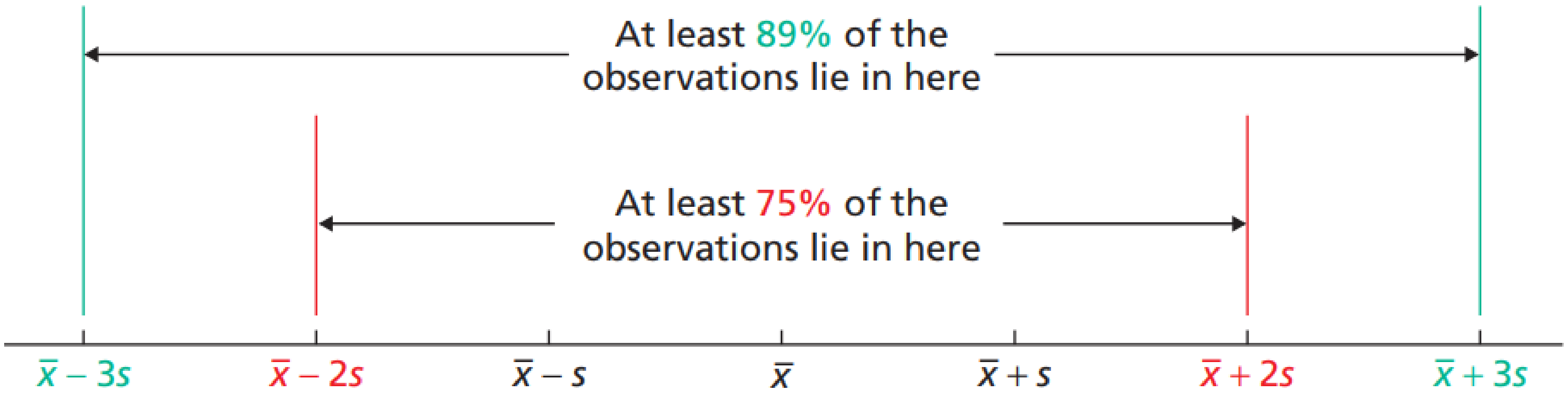
- $k = 2$: At least 75% of the observations in any data set lie within two standard deviations to either side of the mean, that is, between $\bar{x} - 2s$ and $\bar{x} + 2s$.
- $k = 3$: At least 89% of the observations in any data set lie within three standard deviations to either side of the mean, that is, between $\bar{x} - 3s$ and $\bar{x} + 3s$.

Interpretation and Uses of the Standard Deviation

Let us illustrate how to get these two specific cases of Chebyshev's rule by considering the specific case $k = 2$. Then

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0.75,$$

or 75%. Thus, by Chebyshev's rule with $k = 2$, at least 75% of the observations lie within two standard deviations to either side of the mean.



Chebyshev's rule with $k = 2$ and $k = 3$

Interpretation and Uses of the Standard Deviation

EXAMPLE

Chebyshev's Rule

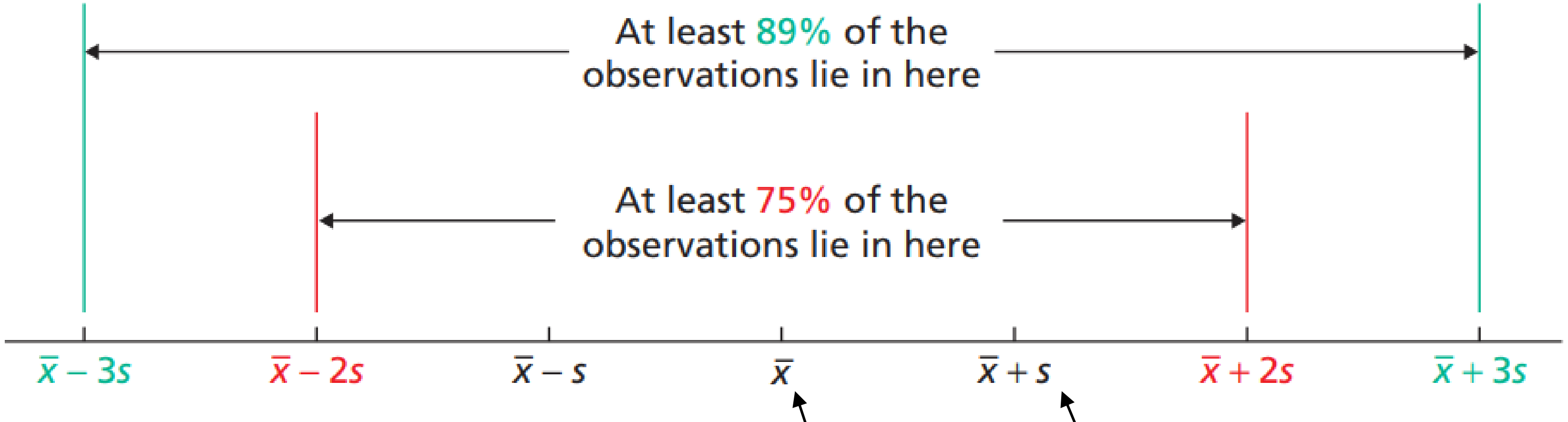
Forearm Length In 1903, K. Pearson and A. Lee published the paper "On the Laws of Inheritance in Man. I. Inheritance of Physical Characters" (*Biometrika*, Vol. 2, pp. 357–462). The article examined data on forearm length, in inches, for a sample of 140 men. The mean and standard deviation of the forearm lengths are 18.8 in. and 1.12 in., respectively.

a. Apply Chebyshev's rule with $k = 2$ to make pertinent statements about the forearm lengths of the men in the sample.

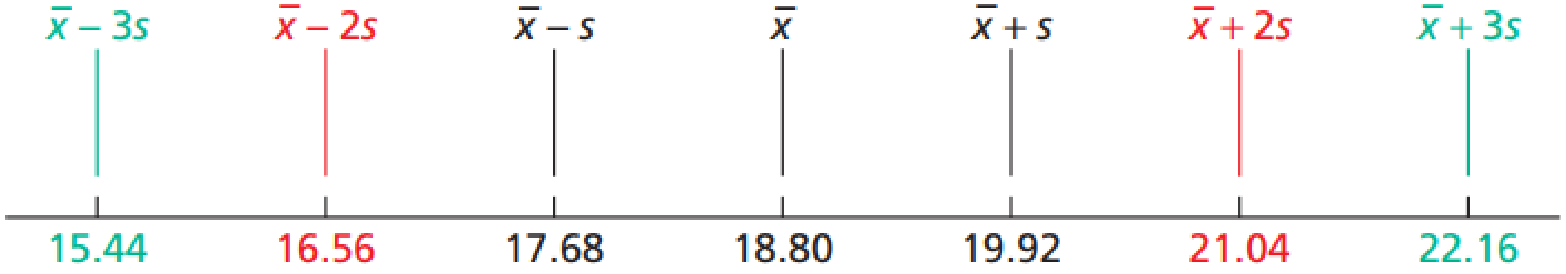
b. Repeat part (a) with $k = 3$.

Interpretation and Uses of the Standard Deviation

The mean and one, two, and three standard deviations to either side of the mean for the forearm-length data

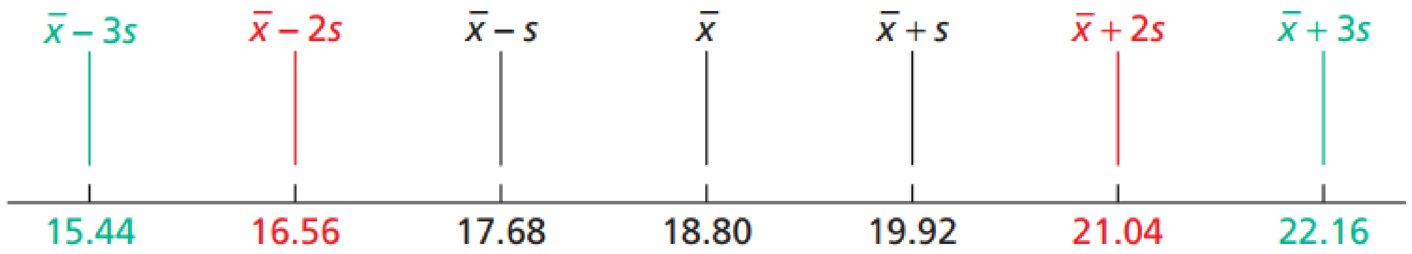


Solution Based on the fact that $\boxed{\bar{x} = 18.8}$ and $\boxed{s = 1.12,}$ we constructed
Notice, for instance, that $\bar{x} - 2s = 18.8 - 2 \cdot 1.12 = 16.56$.



Interpretation and Uses of the Standard Deviation

FIGURE 3.8
The mean and one, two, and three standard deviations to either side of the mean for the forearm-length data



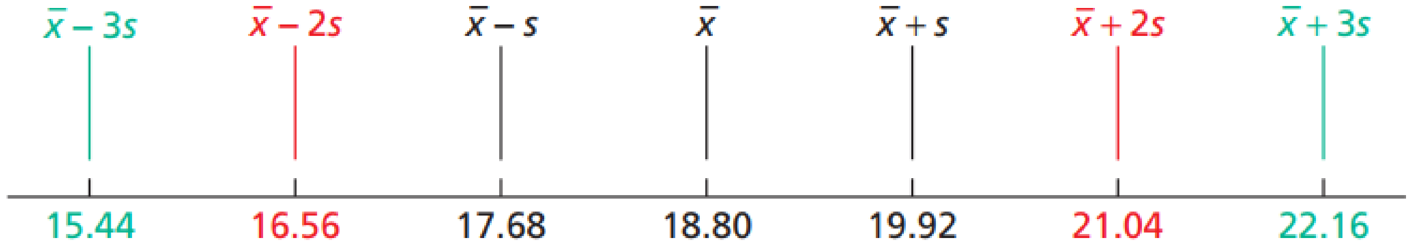
a. By Chebyshev's rule with $k = 2$, at least 75% of the men in the sample have forearm lengths within two standard deviations to either side of the mean. Now, 75% of 140 is 105, and two standard deviations to either side of the mean is from 16.56 to 21.04, as we see from Fig. 3.8.

Interpretation At least 105 of the 140 men in the sample have forearm lengths between 16.56 in. and 21.04 in.

Interpretation and Uses of the Standard Deviation

FIGURE 3.8

The mean and one, two, and three standard deviations to either side of the mean for the forearm-length data



b. By Chebyshev's rule with $k = 3$, at least 89% of the men in the sample have forearm lengths within three standard deviations to either side of the mean. Now, 89% of 140 is 124.6, and three standard deviations to either side of the mean is from 15.44 to 22.16, as we see from Fig. 3.8.

Interpretation At least 125 (124.6 rounded up) of the 140 men in the sample have forearm lengths between 15.44 in. and 22.16 in.

Interpretation and Uses of the Standard Deviation

The Empirical Rule

For data sets with approximately bell-shaped distributions, we can improve on the estimates given by Chebyshev's rule by applying the **empirical rule.**

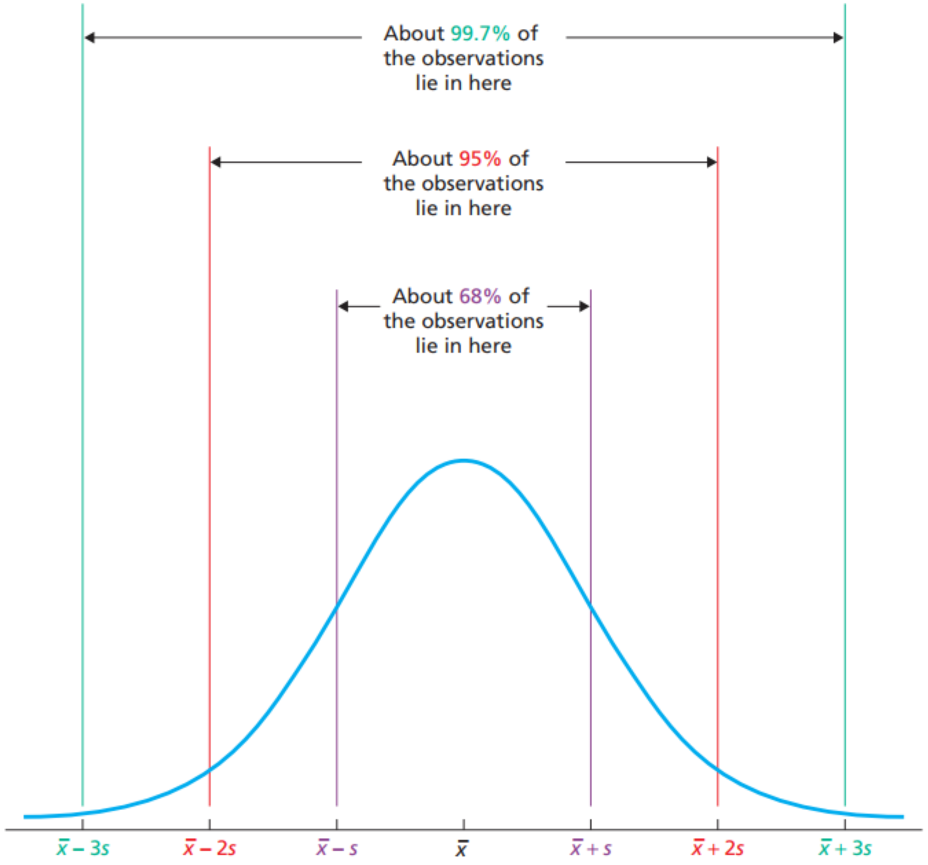
Empirical Rule

For any quantitative data set with roughly a bell-shaped distribution, the following properties hold.

Property 1: Approximately 68% of the observations lie within one standard deviation to either side of the mean, that is, between $\bar{x} - s$ and $\bar{x} + s$.

Property 2: Approximately 95% of the observations lie within two standard deviations to either side of the mean, that is, between $\bar{x} - 2s$ and $\bar{x} + 2s$.

Property 3: Approximately 99.7% of the observations lie within three standard deviations to either side of the mean, that is, between $\bar{x} - 3s$ and $\bar{x} + 3s$.



Note the following:

- The empirical rule is also known as the **68-95-99.7 rule.**
- The percentage approximations given by the empirical rule can be somewhat rough.
- The percentage approximations given by the empirical rule tend to be accurate when the distribution of the data set is close to bell shaped.

Interpretation and Uses of the Standard Deviation

EXAMPLE The Empirical Rule

Forearm Length From Example 3.16 (page 140), recall that the mean and standard deviation of forearm lengths for a sample of 140 men are 18.8 in. and 1.12 in., respectively.

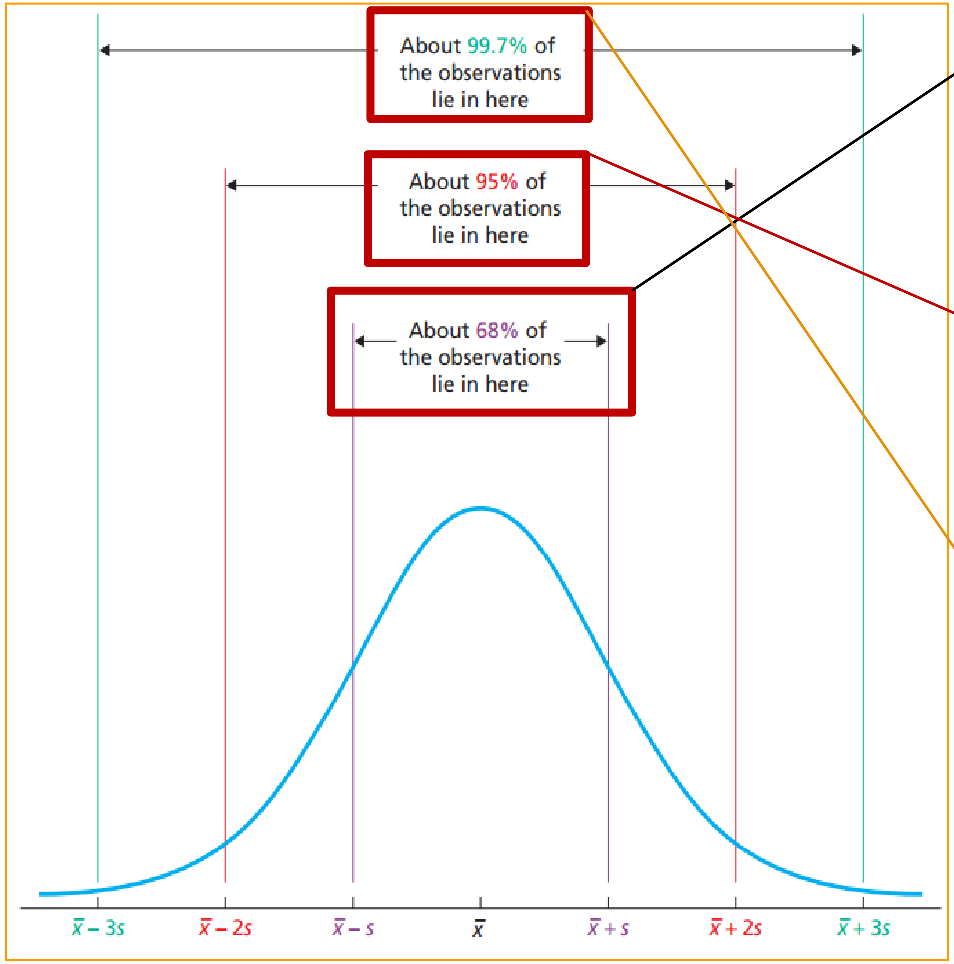
In Example 3.16, we used Chebyshev's rule to make pertinent statements about the forearm lengths of the men in the sample. Presuming that the distribution of forearm lengths of the men in the sample is roughly bell shaped (which, in fact, it actually is), we can use the empirical rule to get more accurate estimates.

a. Apply Property 1 of the empirical rule to make pertinent statements about the forearm lengths of the men in the sample.

b. Repeat part (a) for Property 2 of the empirical rule.

c. Repeat part (a) for Property 3 of the empirical rule.

Interpretation and Uses of the Standard Deviation



a. By Property 1 of the empirical rule, approximately 68% of the men in the sample have forearm lengths within one standard deviation to either side of the mean. Now, 68% of 140 is 95.2, and one standard deviation to either side of the mean is from 17.68 to 19.92, as we see from Fig. 3.8.

Interpretation Approximately 95 (95.2 rounded) of the 140 men in the sample have forearm lengths between 17.68 in. and 19.92 in.

b. By Property 2 of the empirical rule, approximately 95% of the men in the sample have forearm lengths within two standard deviations to either side of the mean. Now, 95% of 140 is 133, and two standard deviations to either side of the mean is from 16.56 to 21.04, as we see from Fig. 3.8.

Interpretation Approximately 133 of the 140 men in the sample have forearm lengths between 16.56 in. and 21.04 in.

c. By Property 3 of the empirical rule, approximately 99.7% of the men in the sample have forearm lengths within three standard deviations to either side of the mean. Now, 99.7% of 140 is 139.58, and three standard deviations to either side of the mean is from 15.44 to 22.16, as we see from Fig. 3.8.

Interpretation Approximately 140 (139.58 rounded), that is, approximately all of the 140 men in the sample have forearm lengths between 15.44 in. and 22.16 in.

Try This

Self-Review

Exam Scores. Consider the following sample of exam scores, arranged in increasing order.

28	57	58	64	69	74
79	80	83	85	85	87
87	89	89	90	92	93
94	94	95	96	96	97
97	97	97	98	100	100

The sample mean and sample standard deviation of these exam scores are 85 and 16.1, respectively.

a. Compare the percentage of the observations that actually lie within two standard deviations to either side of the mean with that given by Chebyshev's rule with $k = 2$.

b. Repeat part (a) with $k = 3$.

c. Interpret your results from parts (a) and (b).

The Mean and Standard Deviation of Grouped Data

ARITHMETIC MEAN OF GROUPED DATA $$\bar{X} = \frac{\Sigma fM}{n}$$ **[3–12]**

where:

- $\bar{X}$ is the designation for the sample mean.
- M is the midpoint of each class.
- f is the frequency in each class.
- fM is the frequency in each class times the midpoint of the class.
- ΣfM is the sum of these products.
- n is the total number of frequencies.

The Mean and Standard Deviation of Grouped Data

Example

The computations for the arithmetic mean of data grouped into a frequency distribution will be shown based on the Applewood Auto Group profit data. Recall in Chapter 2, in Table 2–7 on page 33, we constructed a frequency distribution for the vehicle profit. The information is repeated below. Determine the arithmetic mean profit per vehicle.

Profit	Frequency
\$ 200 up to \$ 600	8
600 up to 1,000	11
1,000 up to 1,400	23
1,400 up to 1,800	38
1,800 up to 2,200	45
2,200 up to 2,600	32
2,600 up to 3,000	19
3,000 up to 3,400	4
Total	180

The Mean and Standard Deviation of Grouped Data

Solution

Profit	Frequency (f)	Midpoint (M)	fM
\$ 200 up to \$ 600	8	\$ 400	\$ 3,200
600 up to 1,000	11	800	8,800
1,000 up to 1,400	23	1,200	27,600
1,400 up to 1,800	38	1,600	60,800
1,800 up to 2,200	45	2,000	90,000
2,200 up to 2,600	32	2,400	76,800
2,600 up to 3,000	19	2,800	53,200
3,000 up to 3,400	4	3,200	12,800
Total	180		\$333,200

Solving for the arithmetic mean using formula (3–12), we get:

$$\bar{X} = \frac{\Sigma fM}{n} = \frac{\$333{,}200}{180} = \$1{,}851.11$$

We conclude that the mean profit per vehicle is about \$1,851.

The Mean and Standard Deviation of Grouped Data

Standard Deviation

To calculate the standard deviation of data grouped into a frequency distribution, we need to adjust formula (3–11) slightly. We weight each of the squared differences by the number of frequencies in each class. The formula is:

STANDARD DEVIATION, GROUPED DATA

$$s = \sqrt{\frac{\Sigma f(M - \overline{X})^2}{n - 1}} \qquad \textbf{[3–13]}$$

where:

- s is the symbol for the sample standard deviation.
- M is the midpoint of the class.
- f is the class frequency.
- n is the number of observations in the sample.
- $\overline{X}$ is the designation for the sample mean.

The Mean and Standard Deviation of Grouped Data

Example

Refer to the frequency distribution for the Applewood Auto Group profit data reported in Table 3–1. Compute the standard deviation of the vehicle selling prices.

Following the same practice used earlier for computing the mean of data grouped into a frequency distribution, *f* is the class frequency, *M* the class midpoint, and *n* the number of observations.

Profit	Frequency (f)	Midpoint (M)	fM	$(M - \bar{X})$	$(M - \bar{X})^2$	$f(M - \bar{X})^2$
\$ 200 up to \$ 600	8	400	3,200	−1,451	2,105,401	16,843,208
600 up to 1,000	11	800	8,800	−1,051	1,104,601	12,150,611
1,000 up to 1,400	23	1,200	27,600	−651	423,801	9,747,423
1,400 up to 1,800	38	1,600	60,800	−251	63,001	2,394,038
1,800 up to 2,200	45	2,000	90,000	149	22,201	999,045
2,200 up to 2,600	32	2,400	76,800	549	301,401	9,644,832
2,600 up to 3,000	19	2,800	53,200	949	900,601	17,111,419
3,000 up to 3,400	4	3,200	12,800	1,349	1,819,801	7,279,204
Total	180		333,200			76,169,780

The Mean and Standard Deviation of Grouped Data

Example

To find the standard deviation:

Step 1: Subtract the mean from the class midpoint. That is, find $(M - \overline{X}) = (\$400 - \$1{,}851 = -\$1{,}451)$ for the first class, for the second class $(\$800 - \$1{,}851 = -\$1{,}051)$, and so on.

Step 2: Square the difference between the class midpoint and the mean. For the first class, it would be $(\$400 - \$1{,}851)^2 = 2{,}105{,}401$ for the second class $(\$800 - \$1{,}851)^2 = 1{,}104{,}601$, and so on.

Step 3: Multiply the squared difference between the class midpoint and the mean by the class frequency. For the first class, the value is $8(\$400 - \$1{,}851)^2 = 16{,}843{,}208$; for the second, $11(\$800 - \$1{,}851)^2 = 12{,}150{,}611$, and so on.

Step 4: Sum the $f(M - \overline{X})^2$. The total is 76,169,920. To find the standard deviation, we insert these values in formula (3–13).

$$s = \sqrt{\frac{\Sigma f(M - \overline{X})^2}{n - 1}} = \sqrt{\frac{76{,}169{,}780}{180 - 1}} = 652.33$$

Try This

Self-Review

The net incomes of a sample of large importers of antiques were organized into the following table:

Net Income ($ millions)	Number of Importers	Net Income ($ millions)	Number of Importers
2 up to 6	1	14 up to 18	3
6 up to 10	4	18 up to 22	2
10 up to 14	10		

(a) What is the table called?
(b) Based on the distribution, what is the estimate of the arithmetic mean net income?
(c) Based on the distribution, what is the estimate of the standard deviation?

Determine the mean and the standard deviation of the following frequency distribution.

Class	Frequency
0 up to 5	2
5 up to 10	7
10 up to 15	12
15 up to 20	6
20 up to 25	3

Homework-03

1 Owens Orchards sells apples in a large bag by weight. A sample of seven bags contained the following numbers of apples: 23, 19, 26, 17, 21, 24, 22.

a. Compute the mean and median number of apples in a bag.

b. Verify that $\Sigma(X - \overline{X}) = 0$.

2 A sample of households that subscribe to United Bell Phone Company for land line phone service revealed the following number of calls received per household last week. Determine the mean and the median number of calls received.

52	43	30	38	30	42	12	46	39	37
34	46	32	18	41	5				

Homework-03

3 A recent study of the laundry habits of Americans included the time in minutes of the wash cycle. A sample of 40 observations follows. Determine the mean and the median of a typical wash cycle.

35	37	28	37	33	38	37	32	28	29
39	33	32	37	33	35	36	44	36	34
40	38	46	39	37	39	34	39	31	33
37	35	39	38	37	32	43	31	31	35

4 Trudy Green works for the True-Green Lawn Company. Her job is to solicit lawn-care business via the telephone. Listed below is the number of appointments she made in each of the last 25 hours of calling. What is the arithmetic mean number of appointments she made per hour? What is the median number of appointments per hour? Write a brief report summarizing the findings.

9	5	2	6	5	6	4	4	7	2	3	6	3
4	4	7	8	4	4	5	5	4	8	3	3	

Homework-03

5 The weights (in pounds) of a sample of five boxes being sent by UPS are: 12, 6, 7, 3, and 10.

a. Compute the range.
b. Compute the mean deviation.
c. Compute the standard deviation.

6 The enrollments of the 13 public universities in the state of Ohio are listed below.

College	Enrollment
University of Akron	25,942
Bowling Green State University	18,989
Central State University	1,820
University of Cincinnati	36,415
Cleveland State University	15,664
Kent State University	34,056
Miami University	17,161
Ohio State University	59,091
Ohio University	20,437
Shawnee State University	4,300
University of Toledo	20,775
Wright State University	18,786
Youngstown State University	14,682

a. Is this a sample or a population?
b. What is the mean enrollment?
c. What is the median enrollment?
d. What is the range of the enrollments?
e. Compute the standard deviation.

Homework-03

7

Health issues are a concern of managers, especially as they evaluate the cost of medical insurance. A recent survey of 150 executives at Elvers Industries, a large insurance and financial firm located in the Southwest, reported the number of pounds by which the executives were overweight. Compute the mean and the standard deviation.

Pounds Overweight	Frequency
0 up to 6	14
6 up to 12	42
12 up to 18	58
18 up to 24	28
24 up to 30	8

8

Bidwell Electronics Inc. recently surveyed a sample of employees to determine how far they lived from corporate headquarters. The results are shown below. Compute the mean and the standard deviation.

Distance (miles)	Frequency	*M*
0 up to 5	4	2.5
5 up to 10	15	7.5
10 up to 15	27	12.5
15 up to 20	18	17.5
20 up to 25	6	22.5

Homework-03

9 The following frequency distribution reports the electricity cost for a sample of 50 two-bedroom apartments in Albuquerque, New Mexico, during the month of May last year.

Electricity Cost	Frequency
\$ 80 up to \$100	3
100 up to 120	8
120 up to 140	12
140 up to 160	16
160 up to 180	7
180 up to 200	4
Total	50

a. Estimate the mean cost.
b. Estimate the standard deviation.
c. Use the Empirical Rule to estimate the proportion of costs within two standard deviations of the mean. What are these limits?