Describing Data: Numerical Measures

Chapter 3

McGraw-Hill/Irwin

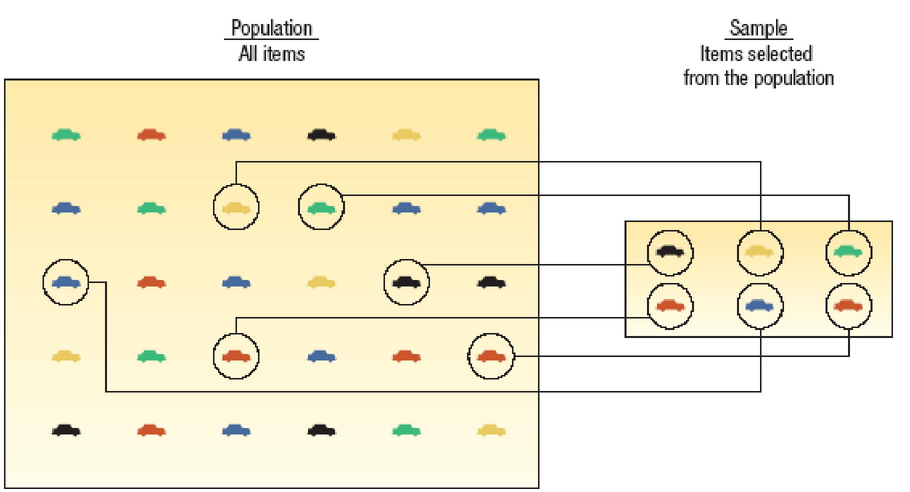
Learning Objectives

LO1 Explain the concept of central tendency.
LO2 Identify and compute the arithmetic mean.
LO3 Compute and interpret the weighted mean.
LO4 Determine the median.
LO5 Identify the mode.
LO6 Calculate the geometric mean.
LO7 Explain and apply measures of dispersion.
LO8 Compute and interpret the standard deviation.
LO9 Explain Chebyshev's Theorem and the Empirical Rule.
LO10 Compute the mean and standard deviation of grouped data.

Parameter Versus Statistics

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

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



Population Mean

LO1 Explain the concept of central tendency

LO2 Identify and compute the arithmetic 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 $\mu = \frac{\Sigma X}{N}$ **[3–1]**

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.

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

where:

$\overline{X}$ is the sample mean. It is read "X bar."
n is the number of values in the sample.

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?

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

SunCom is studying the number of minutes used monthly 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?

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

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

The Weighted Mean

WEIGHTED MEAN

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

- 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.

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 and is therefore a valuable measure of central tendency when such values occur.
3. It can be computed for ratio-level, interval-level, and ordinal-level data.
4. It can be computed for an open-ended frequency distribution if the median does not lie in an open-ended class.

EXAMPLES:

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:

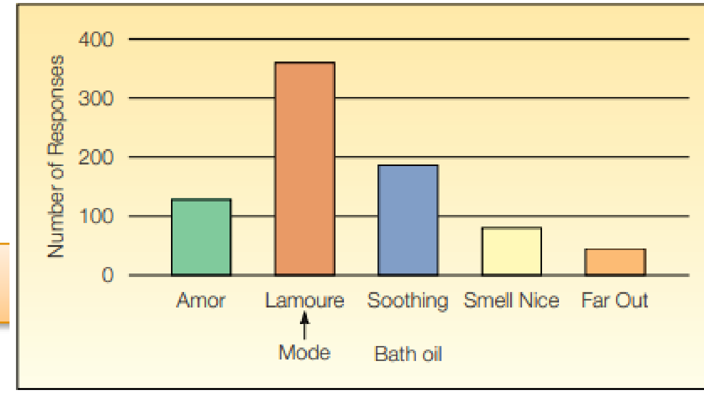
76, 73, 80, 75

Arranging the data in ascending order gives:

73, 75, 76, 80.

Thus the median is 75.5

The Mode



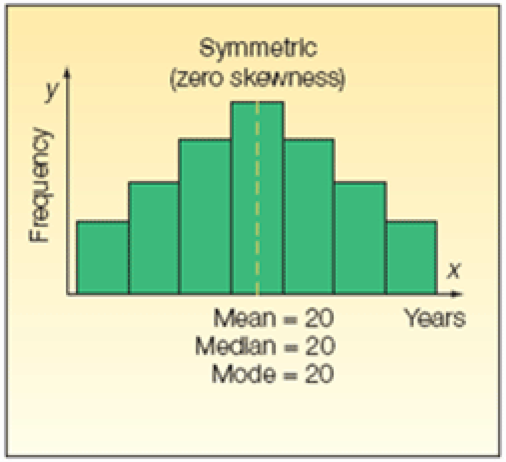
MODE The value of the observation that appears most frequently.

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

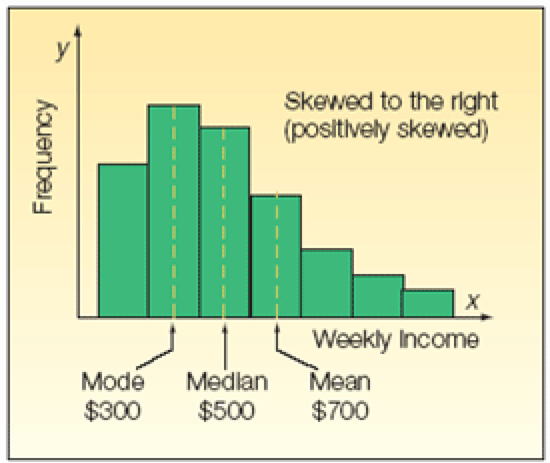
State	Salary	State	Salary	State	Salary
Arizona	$35,000	Illinois	$58,000	Ohio	$50,000
California	49,100	Louisiana	60,000	Tennessee	60,000
Colorado	60,000	Maryland	60,000	Texas	71,400
Florida	60,000	Massachusetts	40,000	West Virginia	60,000
Idaho	40,000	New Jersey	65,000	Wyoming	55,000

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.

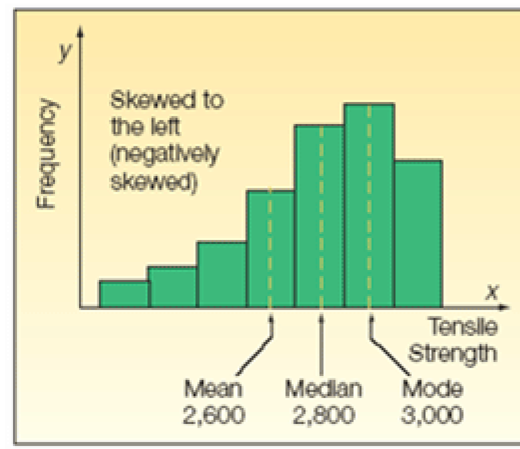
The Relative Positions of the Mean, Median and the Mode



zero skewness
mode = median = mean



positive skewness
mode < median < mean



negative skewness
mode > median > mean

The Geometric Mean

GEOMETRIC MEAN $$GM = \sqrt[n]{(X_1)(X_2) \cdots (X_n)} \quad [3\text{–}4]$$

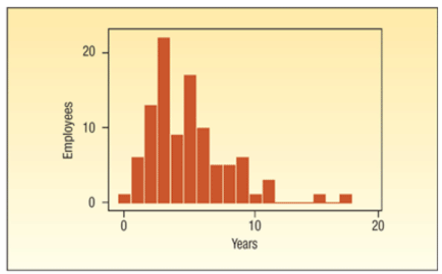
- Useful in finding the average change of percentages, ratios, indexes, or growth rates over time.
- It has a wide application in business and economics because we are often interested in finding the percentage changes in sales, salaries, or economic figures, such as the GDP, which compound or build on each other.
- The geometric mean will always be less than or equal to the arithmetic mean.
- The formula for the geometric mean is written:

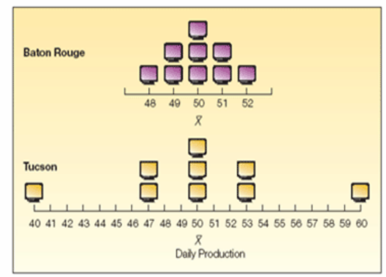
EXAMPLE:
The return on investment earned by Atkins Construction Company for four successive years was: 30 percent, 20 percent, -40 percent, and 200 percent. What is the geometric mean rate of return on investment?

$$GM = \sqrt[n]{(X_1)(X_2) \cdots (X_n)} = \sqrt[4]{(1.3)(1.2)(0.6)(3.0)} = \sqrt[4]{2.808} = 1.294$$

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

- MEAN DEVIATION

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

- VARIANCE AND STANDARD DEVIATION

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]**

LO7

EXAMPLE – Mean Deviation

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

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.

EXAMPLE:

The number of cappuccinos sold at the Starbucks location in the Orange Country Airport between 4 and 7 p.m. for a sample of 5 days last year were 20, 40, 50, 60, and 80. Determine the mean deviation for the number of cappuccinos sold.

Step 1: Compute the mean

$$\bar{x} = \frac{\sum x}{n} = \frac{20+40+50+60+80}{5} = 50$$

Step 2: Subtract the mean (50) from each of the observations, convert to positive if difference is negative

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

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

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

Variance and Standard Deviation

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

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

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

STANDARD DEVIATION The square root of the variance.

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.

- 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.
- The variance overcomes the weakness of the range by using all the values in the population

LO7

EXAMPLE – Population Variance and Population Standard Deviation

The number of traffic citations issued during the last five months 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

What is the population variance?

Step 1: Find the mean.

$$\mu = \frac{\sum x}{N} = \frac{19+17+\ldots+34+10}{12} = \frac{348}{12} = 29$$

Step 2: Find the difference between each observation and the mean, and square that difference.

Step 3: Sum all the squared differences found in step 3

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

Step 4: Divide the sum of the squared differences by the number of items in the population.

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N} = \frac{1{,}488}{12} = 124$$

Sample Variance and Standard Deviation

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

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

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

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?

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

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

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

Chebyshev's Theorem and Empirical Rule

The arithmetic mean biweekly amount contributed by the Dupree Paint employees to the company's profit-sharing plan is \$51.54, and the standard deviation is \$7.51. At least what percent of the contributions lie within plus 3.5 standard deviations and minus 3.5 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.

$$1 - \frac{1}{k^2} = 1 - \frac{1}{(3.5)^2} = 1 - \frac{1}{12.25} = 0.92$$

EMPIRICAL RULE For a symmetrical, bell-shaped frequency distribution, approximately 68 percent of the observations will lie within plus and minus one standard deviation of the mean; about 95 percent of the observations will lie within plus and minus two standard deviations of the mean; and practically all (99.7 percent) will lie within plus and minus three standard deviations of the mean.

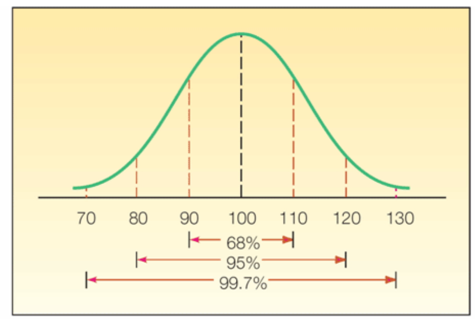


CHART 3–7 A Symmetrical, Bell-Shaped Curve Showing the Relationships between the Standard Deviation and the Observations

LO10 Compute the mean and standard deviation of grouped data.

The Arithmetic 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.

EXAMPLE:

Determine the arithmetic mean vehicle profit given in the frequency table below.

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:

$$\bar{X} = \frac{\Sigma fM}{n} = \frac{\$333{,}200}{180} = \$1{,}851$$

STANDARD DEVIATION, GROUPED DATA $s = \sqrt{\frac{\Sigma f(M - \bar{X})^2}{n - 1}}$ **[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.
- $\bar{X}$ is the designation for the sample mean.

EXAMPLE

Compute the standard deviation of the vehicle profit in the frequency table below.

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

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