Because learning changes everything.®

Chapter 3

Describing Data: Numerical Measures

Learning Objectives

LO3-1 Compute and interpret the mean, the median, and the mode.

LO3-2 Compute a weighted mean.

LO3-3 Compute and interpret the geometric mean.

LO3-4 Compute and interpret the range, variance, and standard deviation.

LO3-5 Explain and apply Chebyshev’s theorem and the Empirical Rule.

Measures of Location

A measure of location is a value used to describe the central tendency of a set of data.

Common measures of location.

- Mean.
- Median.
- Mode.

The arithmetic mean is the most widely reported measure of location.

The mean is both a population parameter and sample statistic.

Population Mean [1]

Many studies involve all the individuals in a population.

When the values are not summarized in a frequency distribution, $\mu = \frac{\sum x}{N}$

- μ is the mean.
- *N* is the number of values in the population.
- *x* is any particular value.
- $\sum$ is "sigma" and indicates the operation is adding.
- $\sum x$ is the sum of the values.

Parameter A characteristic of a population.

The mean, μ, of a population is a parameter.

Population Mean 2

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

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

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

Population Mean 3

- Example continued

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

- This is a population because we are considering all the exits on I-75 and the distances between them

- $\mu = \frac{\sum x}{N} = \frac{11+4+10+...+1}{46} = \frac{192}{46} = 4.17 \text{ miles}$

Sample Mean [1]

We select a sample from the population to estimate a specific characteristic.

The sample mean is $\bar{x} = \frac{\sum x}{n}$

- $\bar{x}$ is read "X bar"
- *n* is the number of values in the sample.
- *x* is any particular value.
- $\sum$ is "sigma" and indicates the operation is adding.
- $\sum x$ is the sum of the values.

Statistic A characteristic of a sample.

The sample mean is a statistic.

Sample Mean 2

Example: The number of hours per day that Verizon customer use their mobile phones.

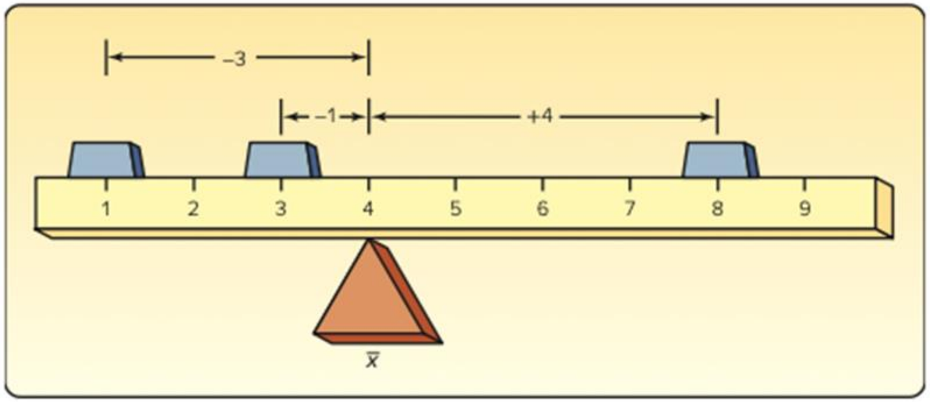
4.1	3.7	4.3	4.2	5.5	51
4.2	5.1	4.2	4.6	5.2	3.8

What is the arithmetic mean number of hours per day used?

- $$\bar{x} = \frac{\sum x}{n} = \frac{4.7 + 3.7 + ... + 3.8}{12} = \frac{54.0}{12} = 4.5 \text{ hours}$$

Sample Mean [3]

- Interval or ratio scale of measurement is required.
- All the data values are used in the calculation.
- The mean is unique.
- The sum of the deviations from the mean equals zero.
- A weakness of the mean is that it is affected by extreme values (large or small).



Access the text alternative for slide images.

The Median [1]

- For data containing extreme values, the mean may not fairly represent the central location.

  Median The midpoint of the values after they have been ordered from the minimum to the maximum values.

- Example: The median price of housing units.

Price Ordered from Minimum to Maximum		Price Ordered from Maximum to Minimum
$60,000		$275,000
65,000		80,000
70,000	← Median →	70,000
80,000		65,000
275,000		60,000

The Median 2

- The median is the value in the middle of a set of ordered data.
- At least the ordinal scale of measurement is required.
- It is not influenced by extreme values.
- Fifty percent of the observations are larger than the median.
- Fifty percent of the observations are smaller than the median.
- It is unique to a set of data.

The Median [3]

Odd numbered data: The median is the middle value

Even numbered data: The median is the average of the two middle values.

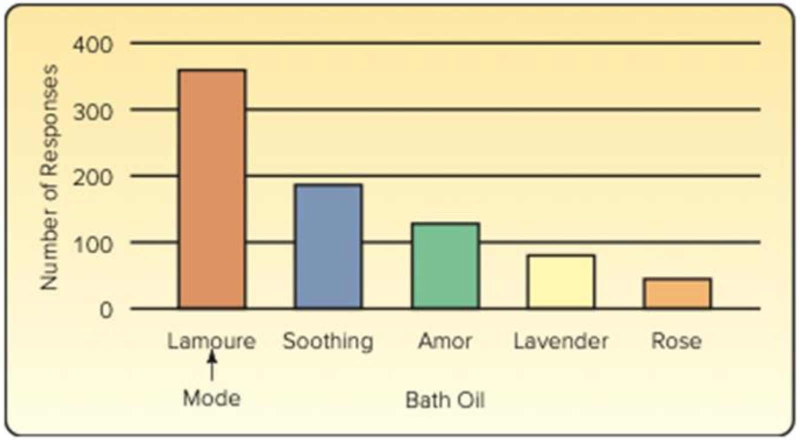
Example: The number of hours a sample of 10 adults used Facebook last month.

- Unsorted data: 3 5 7 5 9 1 3 9 17 10
- Sorted data: 1 3 3 5 **5** **7** 9 9 10 17
- The median is 6

The Mode

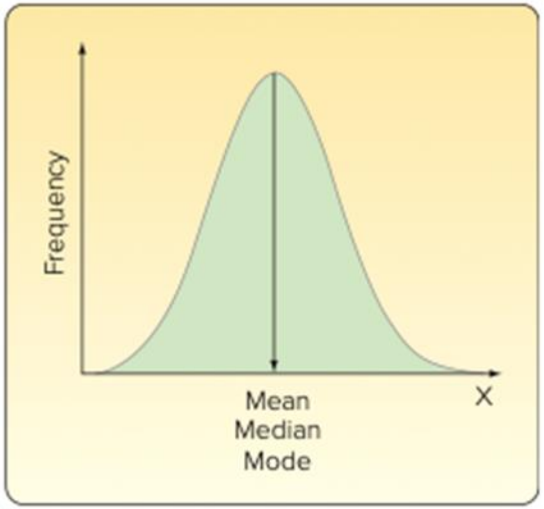
MODE The value of the observation that occurs most frequently.

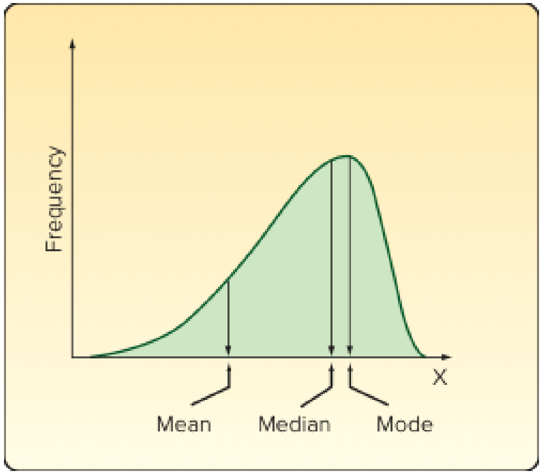
- The mode can be found for nominal level data.
- A set of data can have more than one mode.
- A set of data could have no mode.
- Example: The number of respondents that favor bath oils.

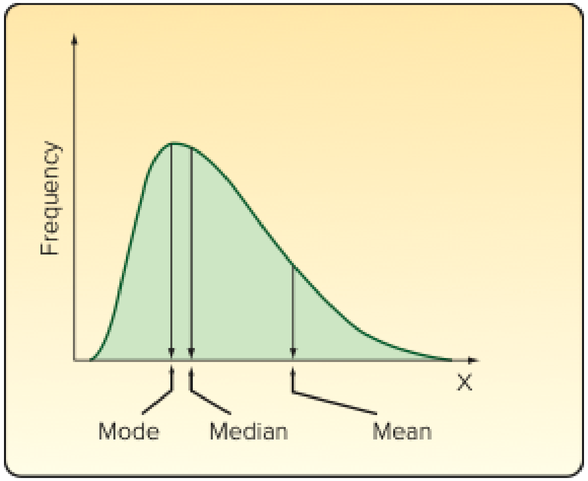


Access the text alternative for slide images.

Relative Positions of Mean, Median, and Mode







Access the text alternative for slide images.

The Weighted Mean [1]

The weighted mean is found by multiplying each observation by its corresponding weight.

A convenient way to compute the mean when there are several observations with the same value.

$$\bar{x} = \frac{w_1 + x_1 + w_2 + x_2 + \ldots + w_n x_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum(wx)}{\sum w}$$

- $\bar{x}_w$ is read "× bar sub w"
- $x_1, x_2, \ldots, x_n$ is the set of numbers
- $w_1, w_2, \ldots, w_n$ are the corresponding weights

The denominator is always the sum of the weights.

The Weighted Mean 2

- Example: The Carter Construction Company pays its hourly employees \$16.50, \$19.00, or \$25.00 per hour.
- There are 26 hourly employees: 14 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 for the 26 employees?

$$\bar{x}_w = \frac{14(\$16.50) + 10(\$19.00) + 2(\$5)}{14 + 10 + 2}$$

$$= \frac{\$471.00}{26} = \$18.1154$$

The Geometric Mean 1

- The geometric mean is useful in finding average rates of change over time.
- The rates can be expressed as percentages (or ratios).
- A percentage change is always 1.0 + change
- Wide applications in business and economics (example GDP).
- The geometric mean will always be no more than the arithmetic mean.
- The data values must be positive.
- $$GM = \sqrt[n]{(x_1)(x_2)...(x_n)}$$

The Geometric Mean 2

- Example: The RIO earned by a company for four consecutive years: 30%, 20%, −40% and 200%.
- $$GM = \sqrt[4]{(1.30)(1.20)(0.60)(3.00)} = 1.294$$
- The RIO is 29.4%.
- Note if you compute the arithmetic mean it would be 52.5%.
- This would overstate the true rate of return.

The Geometric Mean [3]

- A second application of the geometric mean is based on the beginning and ending values over a specified time period.
- $$\text{GM} = \sqrt[n]{\frac{\text{Value at end of period}}{\text{Value at start of period}}} - 1$$
- Example: You earned $45,000 in 2015 and $100,000 in 2022.
- $$GM = \sqrt[12]{\frac{\$100,00}{\$45,000}} - 1 = 1.0688 - 1 = 6.88\%$$

Why Study Dispersion? 1

Measures of location only describe the center.

Do not describe the spread or variation.

The dispersion is the variation or spread in a set of data.

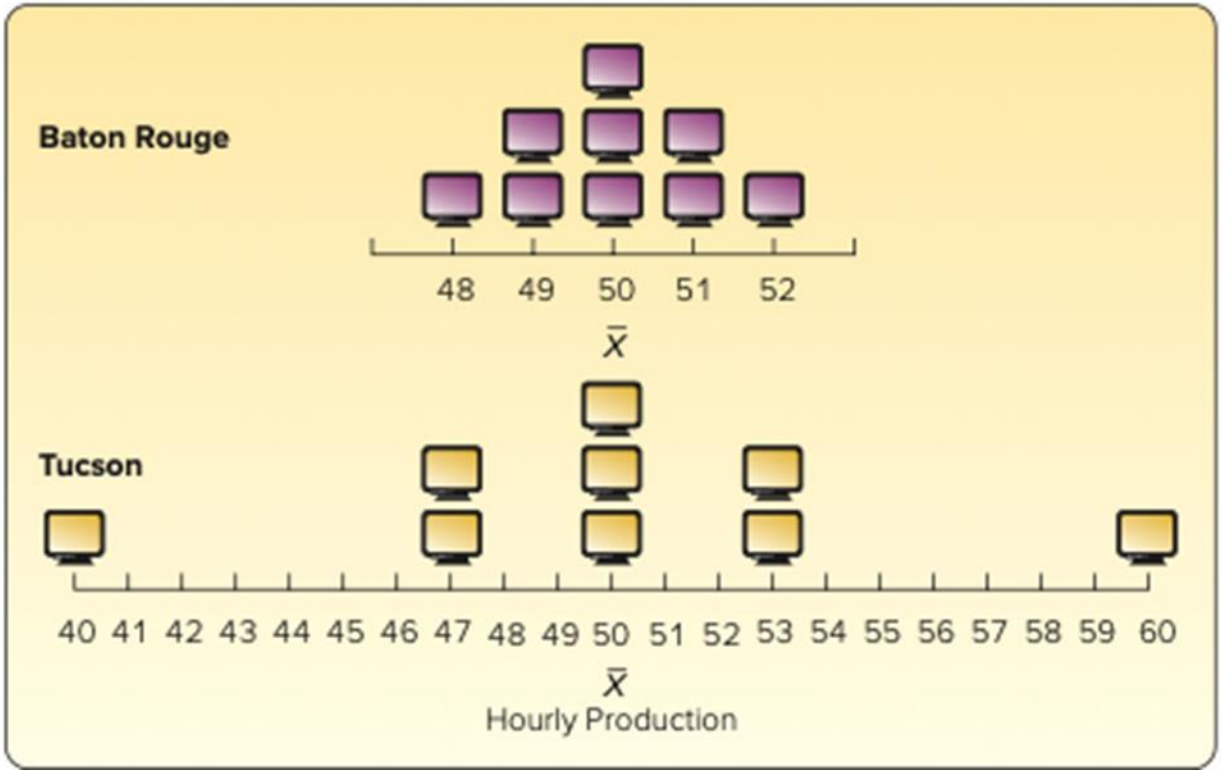
Study dispersion to compare the spread in two or more distributions.

Measures of dispersion include.

- Range.
- Variance.
- Standard Deviation.

Why Study Dispersion? 2

- Example: Hourly production at two computer monitor plants.



Access the text alternative for slide images.

Range

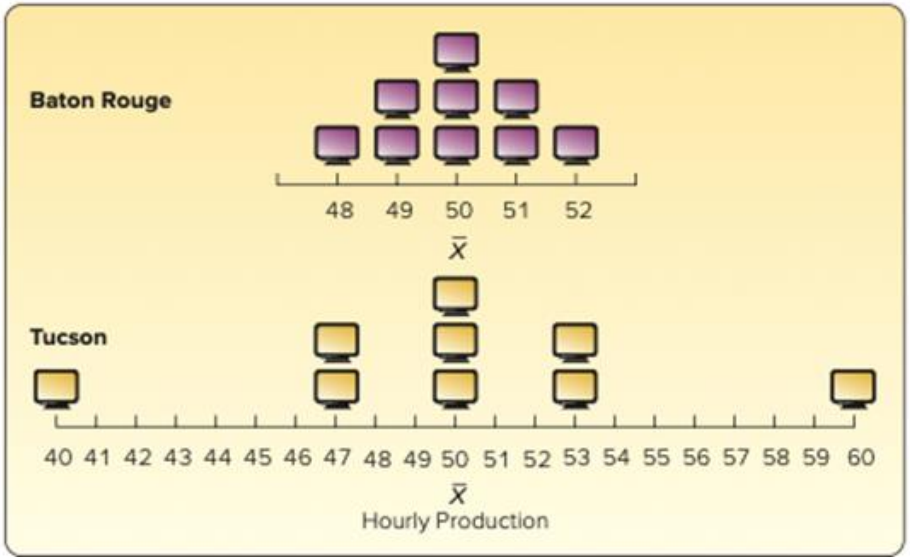
The simplest measure of dispersion is range.

Range = Maximum value − Minimum value.

- Only two values are used in its calculation.
- It is influenced by extreme values.
- It is easy to compute and to understand.

Example: Hourly production at two computer monitor plants.

- Baton Rouge: 52 − 48 = 4.
- Tucson: 60 − 40 = 20.



Access the text alternative for slide images.

Population Variance [1]

The range is based on only two numbers.

The variance measures how much the values vary from their mean.

- The mean squared deviation from the mean.
- Units are the units of measurement squared.

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

- σ^2 is lower case Greek "sigma squared".
- x is the value of a particular observation.
- μ is the population mean.
- N is number of observations in the population.

Population Variance 2

The process for computing the mean is implied by the formula.

1. Begin by finding the mean.
2. Find the difference between each observation and the mean, 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.

Population Variance [3]

- Example: The number of traffic citations.

Citations by Month

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

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

Population Variance 4

- $$\sigma^2 = \frac{\sum(x-\mu)^2}{N} = \frac{1,488}{12} = 124$$
- Units are citations squared!

Citations

Month	(x)	(x−μ)	$(x-\mu)^2$
January	19	−10	100
February	17	−12	144
March	22	−7	49
April	18	−11	121
May	28	−1	1
June	34	5	25
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

Standard Deviation

The units of the variance is the units of measure squared.

The variance is difficult to interpret.

The standard deviation is the square root of the variance.

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

Example: The number of traffic citations.

- The variance is 124 citations^2
- The standard deviation is 11.14 citations.

Sample Variance and Standard Deviation [1]

The sample variance and standard deviation uses the sample.

mean, $\bar{x} = \frac{\sum x}{n}$

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1} \text{ and } s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

- Using *n* in the denominator tends to underestimate the population variance.
- Using *n* − 1 provides an appropriate correction.

The units of the variance is the units of measurement squared.

Sample Variance and Standard Deviation 2

- Example: The hourly wages for part-time employees are \$12, \$20, \$16, \$18 and \$19.
- $$\bar{x} = \frac{\sum x}{n} = \frac{\$85}{5} = \$17$$

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

- $$s^2 = \frac{\sum (x - \bar{x})^2}{n-1} = \frac{40}{5-1} = 10$$ dollars squared
- s = 3.16 dollars.

Interpretations and Uses of the Standard Deviation 1

The standard deviation is to compare the spread of two or more sets of observations.

- Small: The values are close to the mean.
- Large: The values are widely scattered about the mean.

Chebyshev’s theorem defines the dispersion of data around the mean.

- For any set of observations, the proportion of values within k standard deviations is at least $1-\frac{1}{k^2}$ for $k>1$.
- k = 2 then $1-\frac{1}{2^2}=0.75$, at least 75% of the observation are within 1 standard deviation of the mean.
- k = 3 then $1-\frac{1}{3^2}=.889$ or at least 88.9%.

Interpretations and Uses of the Standard Deviation 2

Example: Employees at a company contribute a mean of $51.54 to the company profit-sharing plan every 2 weeks.

The standard deviation is $7.51.

At least what proportion of employees make a contribution within 3.5 standard deviations of the mean?

3.5 standard deviations of the mean is a range of values.

- 51.54 − 2×7.51 = $25.26.
- 51.54 + 2×7.51 = $77.83.

k = 3.5 then $1 - \frac{1}{3.5^2} = 0.92$

At least 92% of the employees contribute between $25.26 and $77.83.

Interpretations and Uses of the Standard Deviation 3

Chebyshev's theorem is for any set of values regardless of the shape of the distribution.

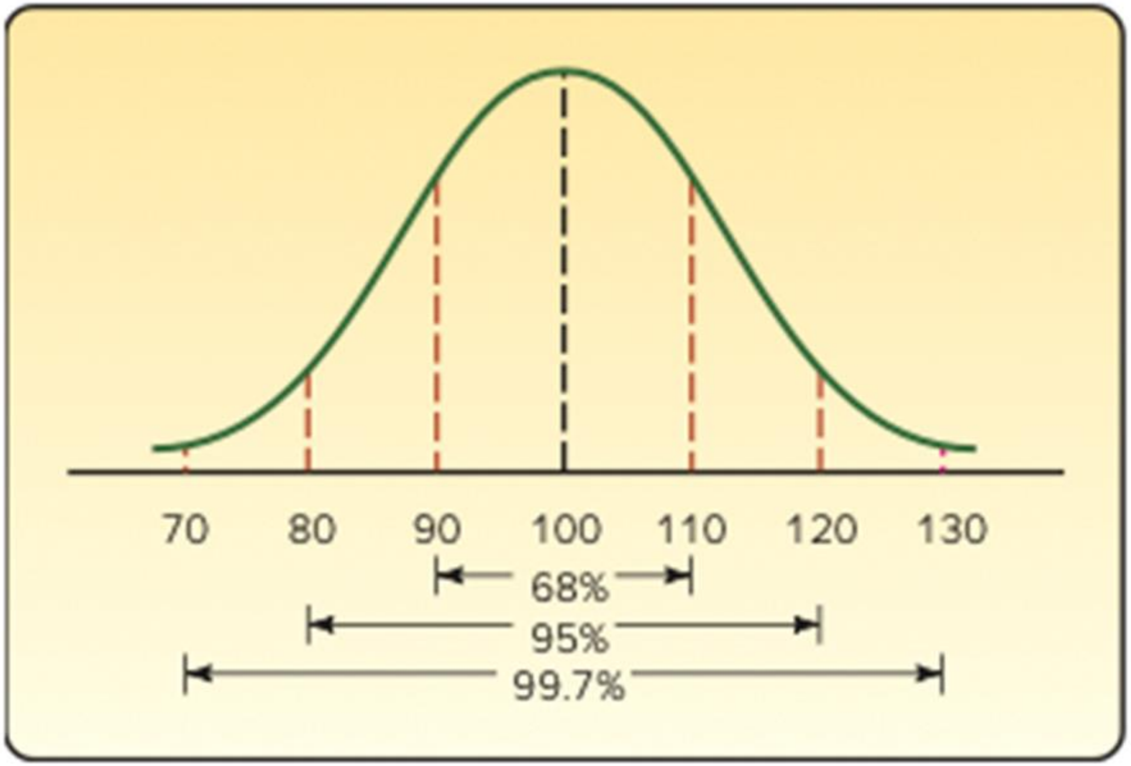
If the data have a symmetrical and bell-shaped distribution, we can be more precise without the "at least".

The Empirical Rule or Normal Rule provides an approximation.

- 1 standard deviation of the mean: about 68% of values.
- 2 standard deviations of the mean: about 95% of values.
- 3 standard deviations of the mean: about 99% of values.

Interpretations and Uses of the Standard Deviation 4

- Example: A mean of 100 and a standard deviation of 10.



Access the text alternative for slide images.

Interpretations and Uses of the Standard Deviation 5

- Example: Rental rates are approximately symmetric and bell shaped. The mean is $500 and the standard deviation is $20.
- About 68%: $500 \pm 1 \times 20$ = $480 to $520.
- About 95%: $500 \pm 2 \times 20$ = $460 to $540.
- About 99%: $500 \pm 3 \times 20$ = $440 to $560.

Ethics and Reporting Results

- Useful to know the advantages and disadvantages of mean, median, and mode as we report statistics and as we use statistics to make decisions.
- Similarly for measures of dispersion.
- Important to maintain an independent and principled point of view.
- Statistical reporting requires objective and honest communication of any results.

Chapter 3 Practice Problems

Question 19

LO3-1

The accounting firm of Rowatti and Koppel specializes in income tax returns for self-employed professionals, such as physicians, dentists, architects, and lawyers. The firm employs 11 accountants who prepare the returns. For last year, the number of returns prepared by each accountant was:

58	75	31	58	46	65	60	71	45	58	80

Find the mean, median, and mode for the number of returns prepared by each accountant. If you could report only one, which measure of location would you recommend reporting?

Question 25

LO3-2

The Loris Healthcare System employs 200 persons on the nursing staff. Fifty are nurse’s aides, 50 are practical nurses, and 100 are registered nurses. Nurse’s aides receive $12 an hour, practical nurses $20 an hour, and registered nurses $29 an hour. What is the weighted mean hourly wage?

Question 27

LO3-3

Compute the geometric mean of the following monthly percent increases: 8, 12, 14, 26, and 5.

Question 35

LO3-3

In 2010 there were 232.2 million cell phone subscribers in the United States. By 2020 the number of subscribers increased to 276.7 million.

a. What is the geometric mean annual percent increase for the period?

b. Further, the number of subscribers is forecast to increase to 276.7 million by 2020.

c. What is the rate of increase from 2010 to 2020?

d. Is the rate of increase expected to slow?

Question 39

LO3-1,4

Dave’s Automatic Door installs automatic garage door openers. The following list indicates the number of minutes needed to install 10 door openers: 28, 32, 24, 46, 44, 40, 54, 38, 32, and 42.

Calculate the following:

a. Range

b. Mean

c. Variance

Question 47

LO3-1,4

Plywood Inc. reported these returns on stockholder equity for the past 5 years: 4.3, 4.9, 7.2, 6.7, and 11.6. Consider these as population values.

Compute the following:

a. Range.

b. Arithmetic mean.

c. Variance.

d. Standard deviation.

Question 51

LO3-4

Dave’s Automatic Door installs automatic garage door openers. Based on a sample, following are the times, in minutes, required to install 10 door openers: 28, 32, 24, 46, 44, 40, 54, 38, 32, and 42.

a. Compute the sample variance.

b. Determine the sample standard deviation.

Question 55

LO3-5

According to Chebyshev’s theorem, at least what percent of any set of observations will be within 1.8 standard deviations of the mean?

Question 57

LO3-5

The distribution of the weights of a sample of 1,400 cargo containers is symmetric and bell-shaped. According to the Empirical Rule, what percent of the weights will lie:

- Between $\bar{x} - 2s$ and $\bar{x} + 2s$?
- Between $\bar{x}$ and $\bar{x} + 2s$? Above $\bar{x} + 2s$?

Because learning changes everything.®

www.mheducation.com