Describing Data: Frequency Tables, Frequency Distributions, and Graphic Presentation

Chapter 2

Learning Objectives

LO2-1 Summarize qualitative variables with frequency and relative frequency tables

LO2-2 Display a frequency table using a bar or pie chart

LO2-3 Summarize quantitative variables with frequency and relative frequency distributions

LO2-4 Display a frequency distribution using a histogram or frequency polygon

Constructing Frequency Tables

FREQUENCY TABLE A grouping of qualitative data into mutually exclusive and collectively exhaustive classes showing the number of observations in each class.

- Mutually exclusive means the data fit in just one class
- Collectively exhaustive means there is a class for each value

TABLE 2–1 Frequency Table for Vehicles Sold Last Month at Applewood Auto Group by Location

Location	Number of Cars
Kane	52
Olean	40
Sheffield	45
Tionesta	43
Total	180

Constructing Frequency Tables

- To construct a frequency table
  - First sort the data into classes
  - Count the number in each class and report as the class frequency
- Convert each frequency to a relative frequency
  - Each of the class frequencies is divided by the total number of observations
  - Shows the fraction of the total number observations in each class

Relative Class Frequencies

- In the vehicle sales example, we may want to know the percentage of total cars sold at each of the four location.

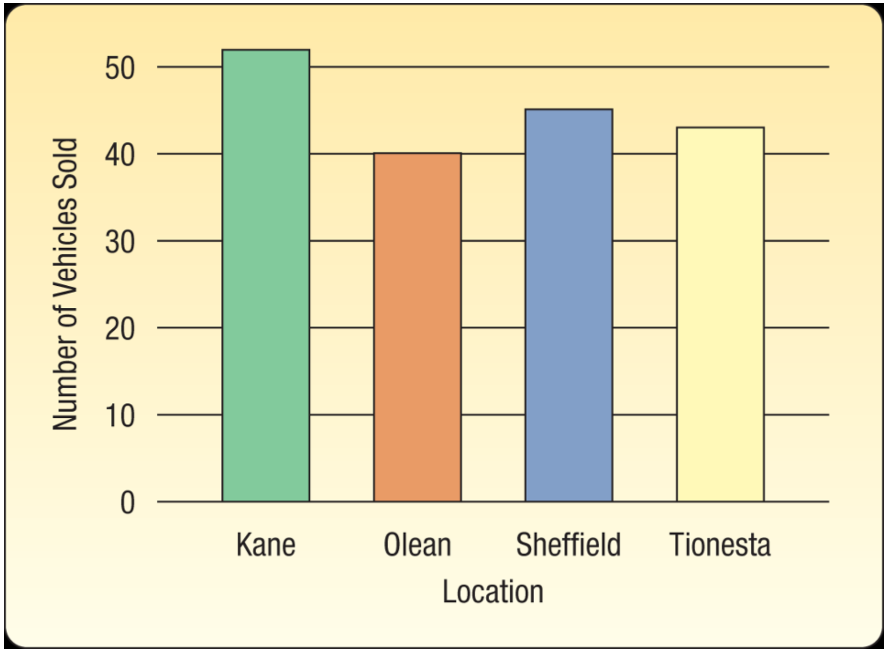
TABLE 2–2 RELATIVE FREQUENCY TABLE OF VEHICLES SOLD BY LOCATION LAST MONTH AT APPLEWOOD AUTO GROUP

Location	Number of Cars	Relative Frequency	Found by
Kane	52	.289	52/180
Olean	40	.222	40/180
Sheffield	45	.250	45/180
Tionesta	43	.239	43/180
Total	180	1.000	

Graphic Presentation of Qualitative Data

BAR CHART A graph that shows the qualitative classes on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are proportional to the heights of the bars.

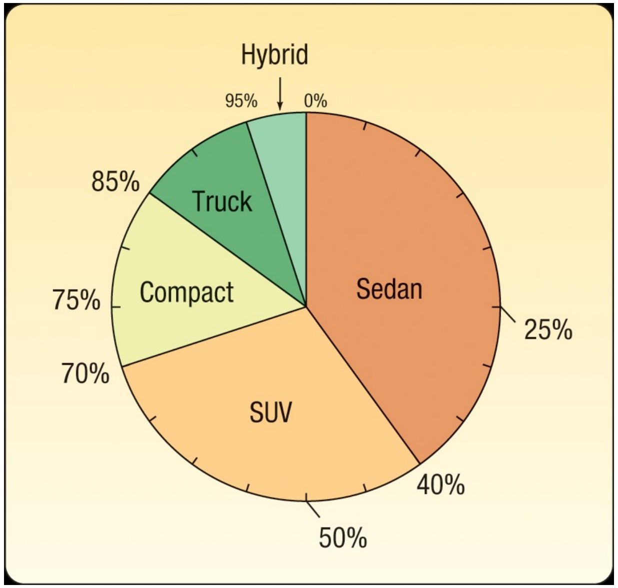
Use a bar chart when you wish to compare the number of observations for each class of a qualitative variable.



Graphic Presentation of Qualitative Data

PIE CHART A chart that shows the proportion or percentage that each class represents of the total number of frequencies.

Use a pie chart when you wish to compare relative differences in the percentage of observations for each class of a qualitative variable.



Example

EXAMPLE

SkiLodges.com is test-marketing its new website and is interested in how easy its website design is to navigate. It randomly selected 200 regular Internet users and asked them to perform a search task on the website. Each person was asked to rate the relative ease of navigation as poor, good, excellent, or awesome. The results are shown in the following table:

Awesome	102
Excellent	58
Good	30
Poor	10

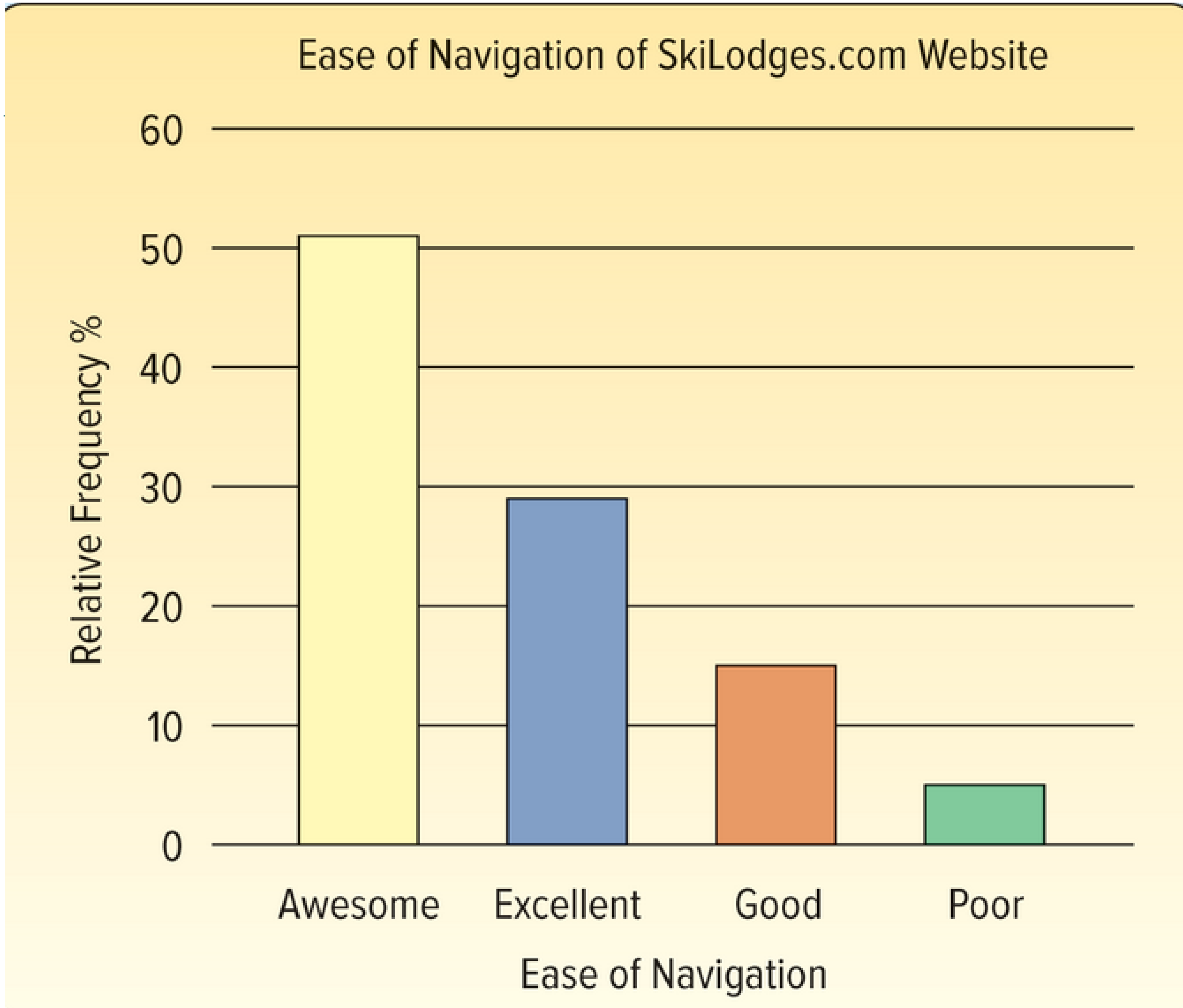
1. What type of measurement scale is used for ease of navigation?
2. Draw a bar chart for the survey results.
3. Draw a pie chart for the survey results.

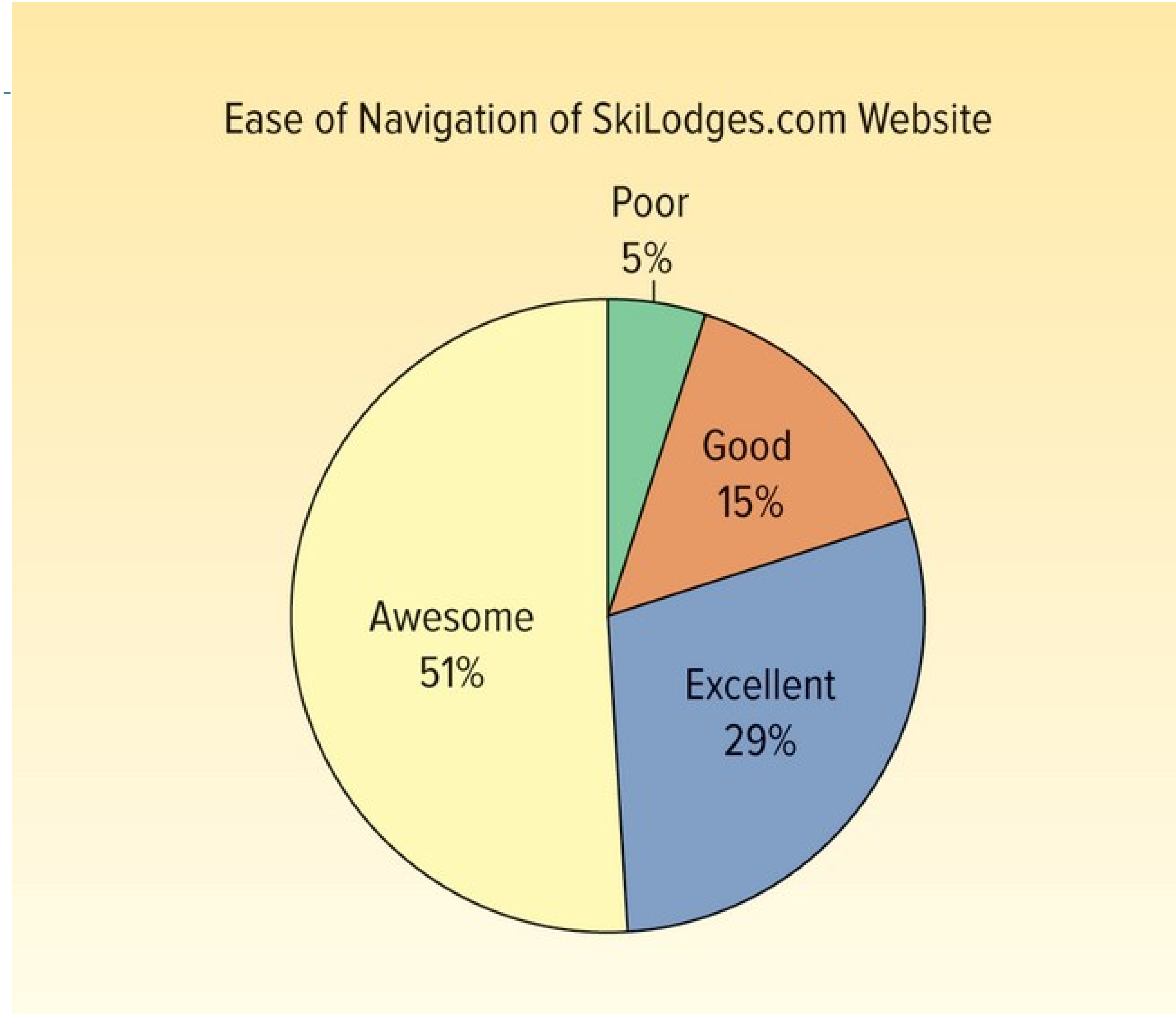
Solution

SOLUTION

The data are measured on an ordinal scale. That is, the scale is ranked in relative ease of navigation when moving from "awesome" to "poor." The interval between each rating is unknown so it is impossible, for example, to conclude that a rating of good is twice the value of a poor rating.

We can use a bar chart to graph the data. The vertical scale shows the relative frequency and the horizontal scale shows the values of the ease-of-navigation variable.


Ease of Navigation of SkiLodges.com Website
60
50
40
30
20
10
0
Relative Frequency %
Awesome
Excellent
Good
Poor
Ease of Navigation


Ease of Navigation of SkiLodges.com Website
Poor
5%
Good
15%
Awesome
51%
Excellent
29%

Constructing Frequency Distributions

FREQUENCY DISTRIBUTION A grouping of quantitative data into mutually exclusive and collectively exhaustive classes showing the number of observations in each class.

- This is a four-step process

1. Decide on the number of classes
2. Determine the class interval
3. Set the individual class limits
4. Tally the data into classes and determine the number of observations in each class

Frequency Distributions

- Step 1 Decide on the number of classes
- Use the $2^k > n$ rule, here n=180
  - k is the number of classes
  - n is the number of values in the data set
  - $2^k > 180$, let k = 8
  - So use 8 classes

TABLE 2–4 Profit on Vehicles Sold Last Month by the Applewood Auto Group

Maximum

$1,387	$2,148	$2,201	$ 963	$ 820	$2,230	$3,043	$2,584	$2,370
1,754	2,207	996	1,298	1,266	2,341	1,059	2,666	2,637
1,817	2,252	2,813	1,410	1,741	3,292	1,674	2,991	1,426
1,040	1,428	323	1,553	1,772	1,108	1,807	934	2,944
1,273	1,889	352	1,648	1,932	1,295	2,056	2,063	2,147
1,529	1,166	482	2,071	2,350	1,344	2,236	2,083	1,973
3,082	1,320	1,144	2,116	2,422	1,906	2,928	2,856	2,502
1,951	2,265	1,485	1,500	2,446	1,952	1,269	2,989	783
2,692	1,323	1,509	1,549	369	2,070	1,717	910	1,538
1,206	1,760	1,638	2,348	978	2,454	1,797	1,536	2,339
1,342	1,919	1,961	2,498	1,238	1,606	1,955	1,957	2,700
443	2,357	2,127	294	1,818	1,680	2,199	2,240	2,222
754	2,866	2,430	1,115	1,824	1,827	2,482	2,695	2,597
1,621	732	1,704	1,124	1,907	1,915	2,701	1,325	2,742
870	1,464	1,876	1,532	1,938	2,084	3,210	2,250	1,837
1,174	1,626	2,010	1,688	1,940	2,639	377	2,279	2,842
1,412	1,762	2,165	1,822	2,197	842	1,220	2,626	2,434
1,809	1,915	2,231	1,897	2,646	1,963	1,401	1,501	1,640
2,415	2,119	2,389	2,445	1,461	2,059	2,175	1,752	1,821
1,546	1,766	335	2,886	1,731	2,338	1,118	2,058	2,487

Minimum

Frequency Distributions

- Step 2 Determine the class interval, i
  - i ≥ (highest value – lowest value)/k
  - Round up to some convenient number

$$i \geq \frac{H - L}{k} = \frac{\$3{,}292 - \$294}{8} = \$374.75$$

  - So, decide to use an interval of $400
  - The interval is also referred to as the class width

Frequency Distributions

- Step 3 Set the individual class limits
  - Lower limits should be rounded to an easy to read number when possible

Classes
$ 200 up to $ 600
600 up to 1,000
1,000 up to 1,400
1,400 up to 1,800
1,800 up to 2,200
2,200 up to 2,600
2,600 up to 3,000
3,000 up to 3,400

Frequency Distributions

- Step 4 Tally the individual data into the classes and determine the number of observations in each class
  - The number of observations is the class frequency

Profit	Frequency
\$ 200 up to \$ 600	~~IIII~~ III
600 up to 1,000	~~IIII~~ ~~IIII~~ I
1,000 up to 1,400	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ III
1,400 up to 1,800	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ III
1,800 up to 2,200	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~
2,200 up to 2,600	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ II
2,600 up to 3,000	~~IIII~~ ~~IIII~~ ~~IIII~~ IIII
3,000 up to 3,400	IIII

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

Relative Frequency Distributions

- To find the relative frequencies, simply take the class frequency and divide by the total number of observations

TABLE 2–7 Relative Frequency Distribution of Profit for Vehicles Sold Last Month at Applewood Auto Group

Profit	Frequency	Relative Frequency	Found by
\$ 200 up to \$ 600	8	.044	8/180
600 up to 1,000	11	.061	11/180
1,000 up to 1,400	23	.128	23/180
1,400 up to 1,800	38	.211	38/180
1,800 up to 2,200	45	.250	45/180
2,200 up to 2,600	32	.178	32/180
2,600 up to 3,000	19	.106	19/180
3,000 up to 3,400	4	.022	4/180
Total	180	1.000	

Exercise

11. **FILE** Wachesaw Manufacturing Inc. produced the following number of units in the last 16 days.

27	27	27	28	27	25	25	28
26	28	26	28	31	30	26	26

The information is to be organized into a frequency distribution.

a. How many classes would you recommend?

b. What class interval would you suggest?

c. What lower limit would you recommend for the first class?

d. Organize the information into a frequency distribution and determine the relative frequency distribution.

e. Comment on the shape of the distribution.

Answer

Units	*f*	Relative Frequency
24.0 up to 25.5	2	0.125
25.5 up to 27.0	4	0.250
27.0 up to 28.5	8	0.500
28.5 up to 30.0	0	0.000
30.0 up to 31.5	2	0.125
Total	16	1.000

Graphic Presentation of a Frequency Distribution

HISTOGRAM A graph in which the classes are marked on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are represented by the heights of the bars, and the bars are drawn adjacent to each other.

- A histogram shows the shape of a distribution
- Each class is depicted as a rectangle, with the height of the bar representing the number in each class

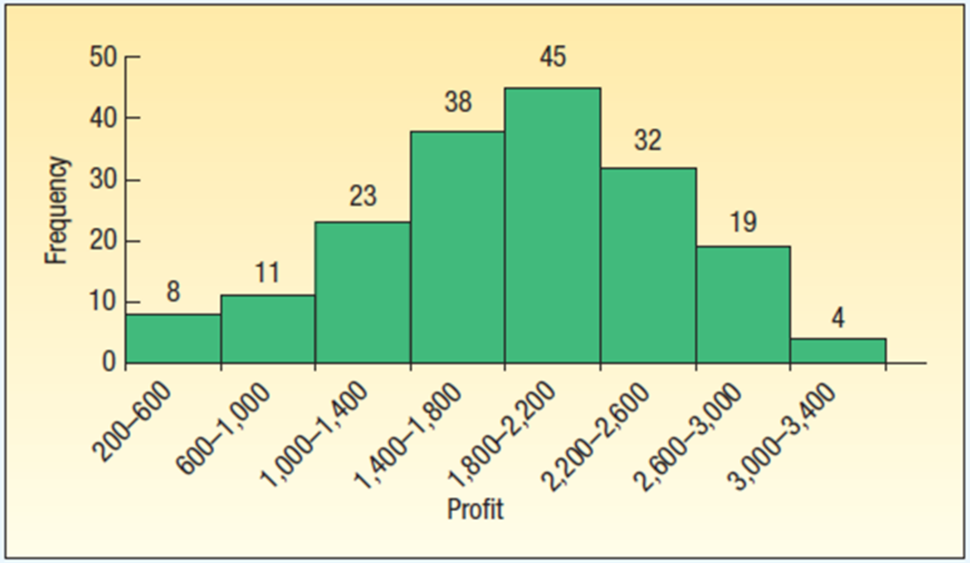
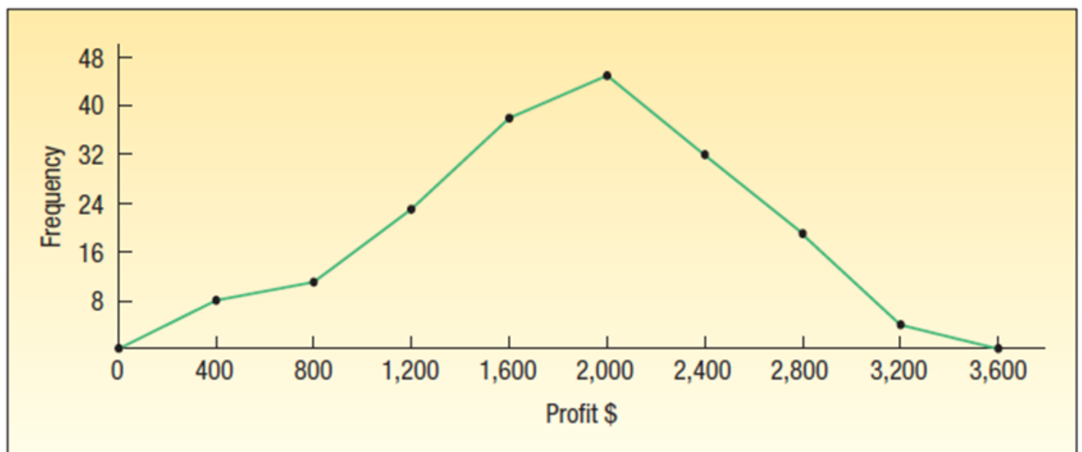


CHART 2–4 Histogram of the Profit on 180 Vehicles Sold at the Applewood Auto Group

Graphical Presentation of a Frequency Distribution

- A frequency polygon, similar to a histogram, also shows the shape of a distribution
- These are good to use when comparing two or more distributions

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



Cumulative Frequency Distributions

Profit	Cumulative Frequency	Found by
Less than $ 600	8	8
Less than 1,000	19	8 + 11
Less than 1,400	42	8 + 11 + 23
Less than 1,800	80	8 + 11 + 23 + 38
Less than 2,200	125	8 + 11 + 23 + 38 + 45
Less than 2,600	157	8 + 11 + 23 + 38 + 45 + 32
Less than 3,000	176	8 + 11 + 23 + 38 + 45 + 32 + 19
Less than 3,400	180	8 + 11 + 23 + 38 + 45 + 32 + 19 + 4

Cumulative Frequency Polygon

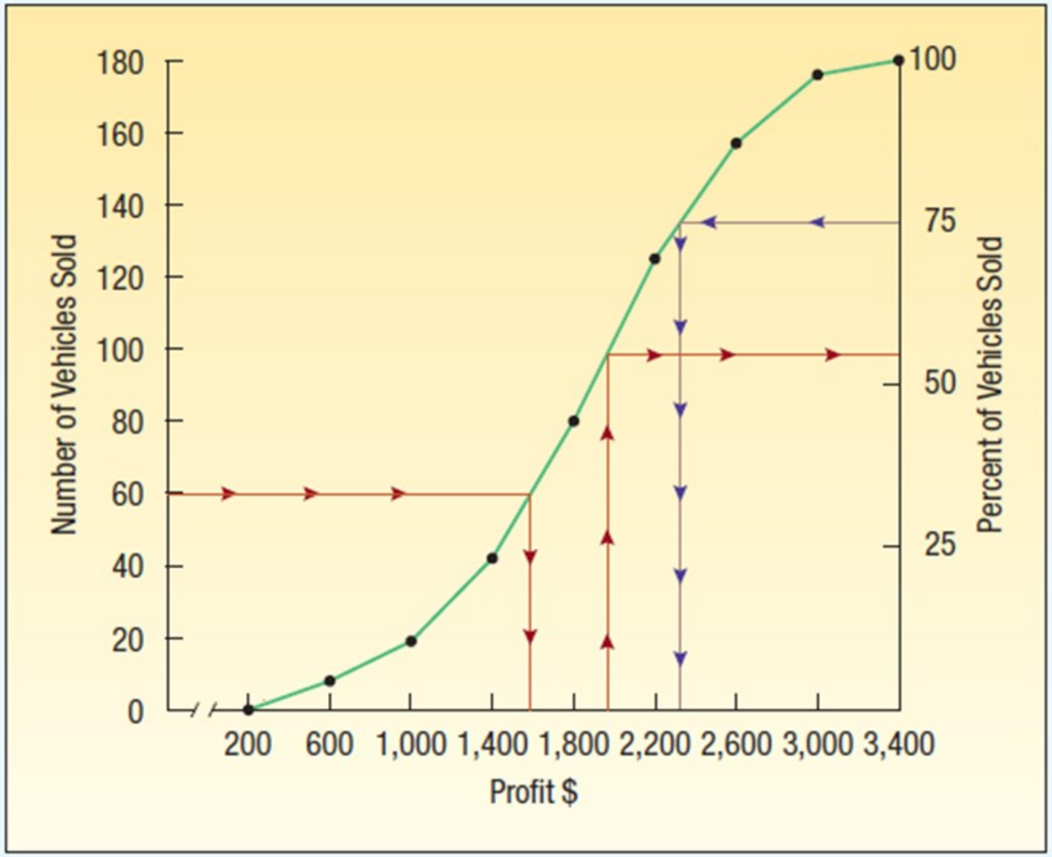


CHART 2–7 Cumulative Frequency Polygon for Profit on Vehicles Sold Last Month at Applewood Auto Group

Describing Data: Numerical Measures

Chapter 3

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 range, variance, and standard deviation

LO3-4 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

Population Mean

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.

- A measurable characteristic of a population is a parameter

PARAMETER A characteristic of a population.

Example: Population Mean

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

1. Why is this information a population?

2. What is the mean number of miles between exits?

Example: Population Mean Continued

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

1. Why is this information a population?

This is a population because we are considering **all** of the exits in Kentucky.

2. What is the mean number of miles between exits?

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

Sample Mean

SAMPLE MEAN $$\bar{x} = \frac{\Sigma x}{n} \qquad [3\text{–}2]$$

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.

- A measurable characteristic of a sample is a statistic

STATISTIC A characteristic of a sample.

Example: Sample Mean

Verizon 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?

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

$$\bar{x} = \frac{\Sigma x}{n} = \frac{90 + 77 + \cdots + 83}{12} = \frac{1{,}170}{12} = 97.5$$

Properties of the Arithmetic Mean

- Interval or ratio scale of measurement is required
- All the data values are used in the calculation
- It 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

The Median

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

Prices Ordered from Minimum to Maximum		Prices 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

Characteristics of the Median

- 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
- It is unique to a set of data

Finding the Median

- To find the median for an even numbered data set
- Sort the observations and calculate the average of the two middle values

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

3 5 7 5 9 1 3 9 17 10

Arranging the data in ascending order gives:

1 3 3 5 5 7 9 9 10 17

Thus, the median is 6.

The Mode

MODE The value of the observation that occurs most frequently.

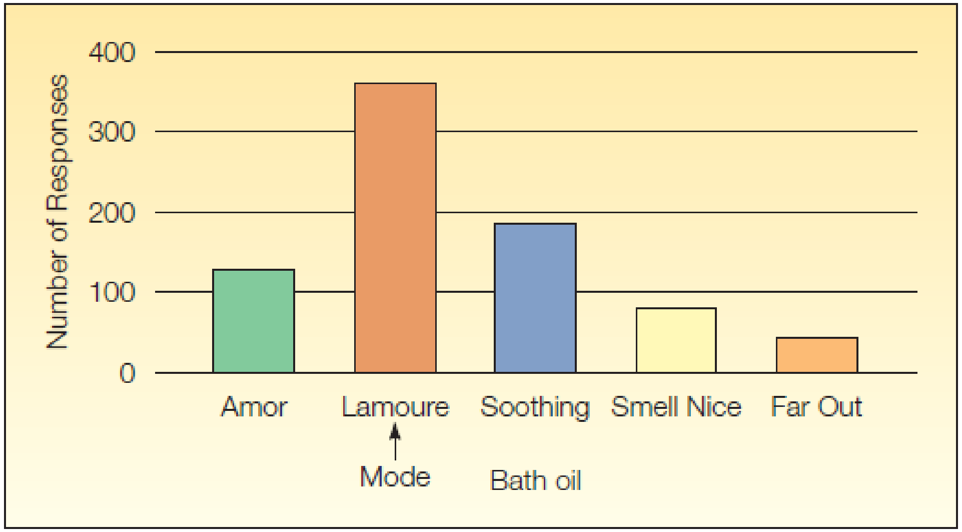


CHART 3–1 Number of Respondents Favoring Various Bath Oils

- Major Characteristics of the mode:
  - The mode can be found for nominal level data
  - A set of data can have more than one mode

Relative Positions of Mean, Median, and Mode

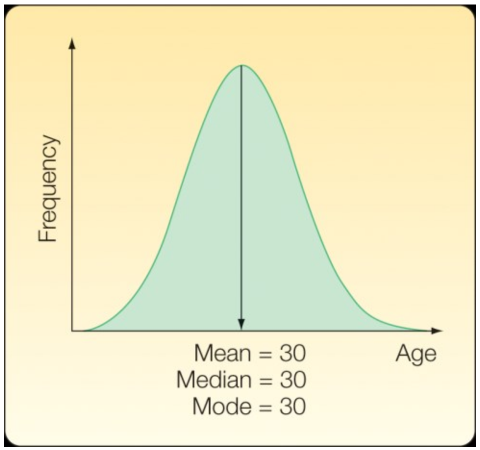


Chart 3-2 A Symmetric Distribution

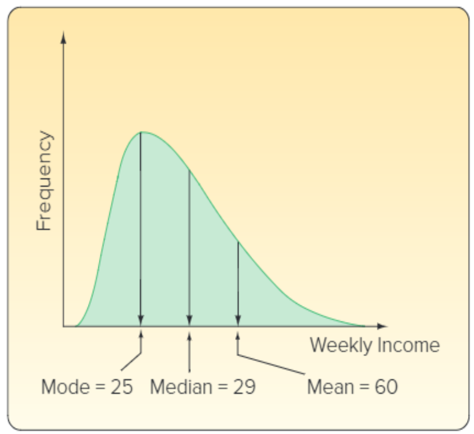


Chart 3-3 A Positively Skewed Distribution

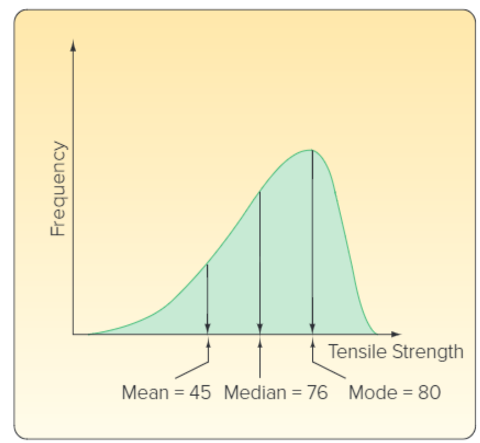


Chart 3-4 A Negatively Skewed Distribution

The Weighted Mean

- The weighted mean is found by multiplying each observation, x, by its corresponding weight, w

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

- 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?

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

Why Study Dispersion?

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

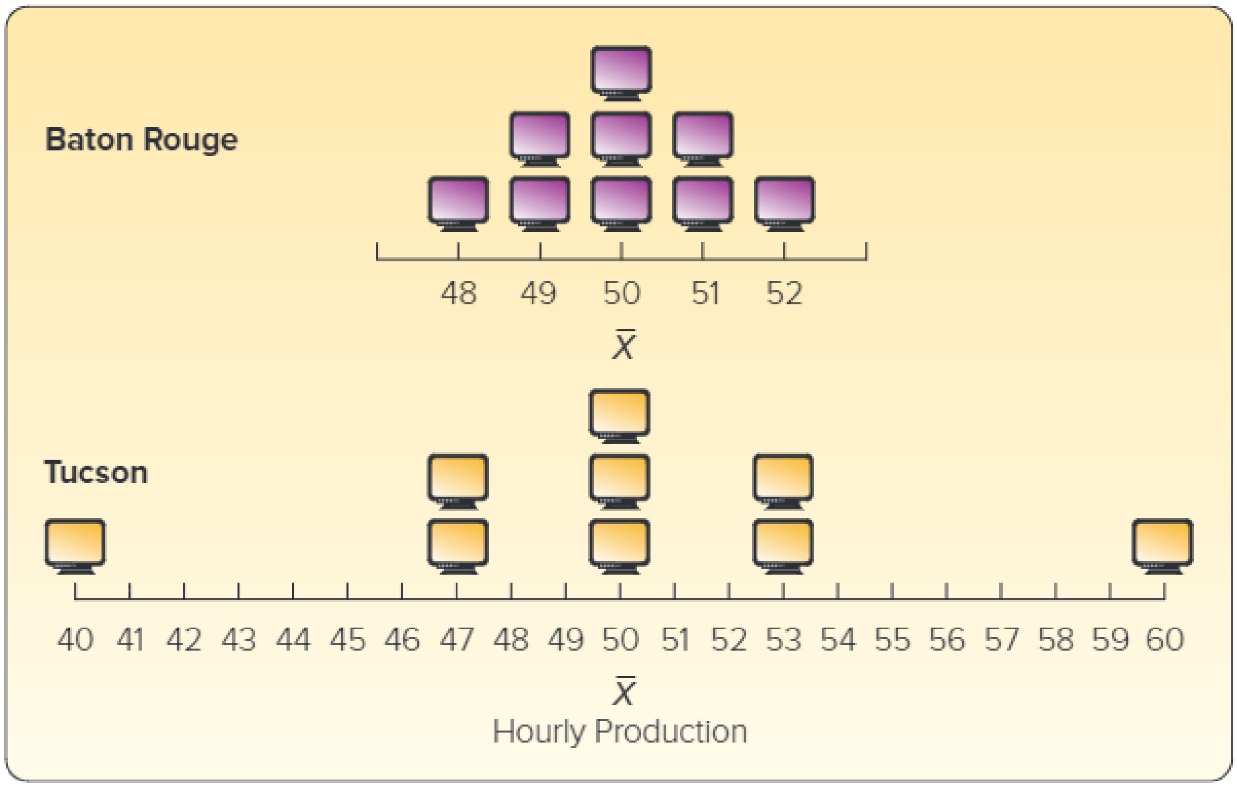


CHART 3–5 Hourly Production of Computer Monitors at the Baton Rouge and Tucson Plants

Why Study Dispersion?

- The range is the difference between the maximum and minimum values in a set of data
- The formula for range is

RANGE	Range = Maximum value − Minimum value	**(3–4)**

- The major characteristics of the range are
  - Only two values are used in its calculation
  - It is influenced by extreme values
  - It is easy to compute and to understand

Population Variance

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

POPULATION VARIANCE $$\sigma^2 = \frac{\Sigma(x - \mu)^2}{N} \qquad \textbf{(3–5)}$$

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 a particular observation in the population.
- μ is the arithmetic mean of the population.
- N is the number of observations in the population.

- Major characteristics of the variance are:
  - All observations are used in the calculation
  - The units are somewhat difficult to work with, they are the original units squared

Population Standard Deviation

POPULATION STANDARD DEVIATION

$$\sigma = \sqrt{\frac{\Sigma(x - \mu)^2}{N}} \qquad \textbf{(3–6)}$$

where:
σ is the population standard deviation
x is the value of each observation in the sample
μ is the mean of the population
N is the number of observations in the population

- Taking the square root of the population variance transforms it to the same unit of measurement used for the original data

Sample Variance

SAMPLE VARIANCE $$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} \quad \textbf{(3–7)}$$

where:

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

- Note the n-1 in the denominator. Using n tends to underestimate the population variance. The use of n-1 in the denominator provides the appropriate correction for this tendency.

Example: Sample Variance

The hourly wages for a sample of part-time employees at Pickett's Hardware store are: \$12, \$20, \$16, \$18, and \$19.

The sample mean is \$17

What is the sample variance?

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

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

$= 10$ in dollars squared

Sample Standard Deviation

SAMPLE STANDARD DEVIATION $$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} \qquad (3\text{–}8)$$

where:
s is the sample standard deviation.
x is the value of each observation in the sample.
$\bar{x}$ is the mean of the sample.
n is the number of observations in the sample.

- The sample standard deviation is used as an estimator of the population standard deviation.

Standard Deviation

- The major characteristics of the standard deviation are:
  - It is in the same units as the original data
  - It is the square root of the average squared distance from the mean
  - It cannot be negative
  - It is the most widely used measure of dispersion

Interpretations and Uses of the Standard Deviation

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 value greater than 1.

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?

About 92%, found by

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

Interpretations and Uses of the Standard Deviation

THE EMPIRICAL RULE For a symmetrical, bell-shaped frequency distribution, approximately 68% of the observations will lie within plus and minus one standard deviation of the mean, about 95% of the observations will lie within plus or minus 2 standard deviations of the mean, and practically all (99.7%) will lie within 3 standard deviations of the mean.

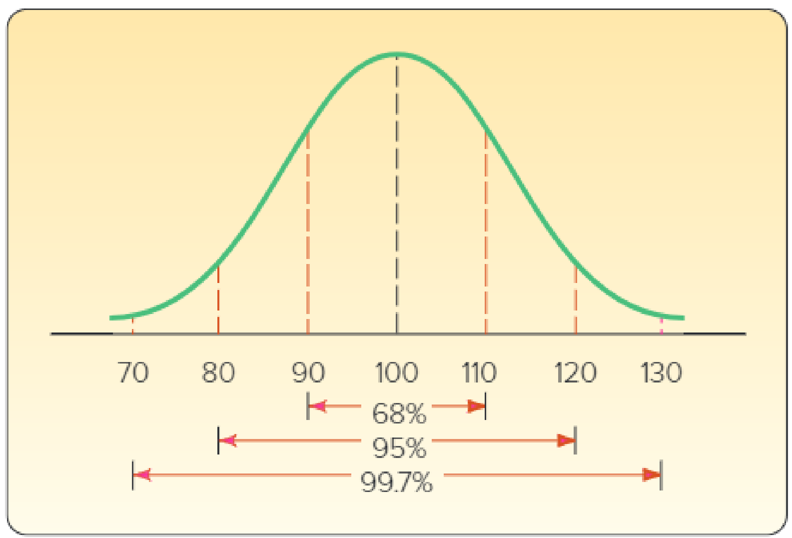


CHART 3–6 A Symmetrical, Bell-Shaped Curve Showing the Relationships between the Standard Deviation and the Percentage of Observations

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
- Important to maintain an independent and principled point of view
- Statistical reporting requires objective and honest communication of any results

Describing Data: Displaying and Exploring Data

Chapter 4

Learning Objectives

LO4-1 Construct and interpret a dot plot

LO4-2 Identify and compute measures of position

LO4-3 Construct and analyze a box plot

LO4-4 Compute and interpret the coefficient of skewness

LO4-5 Create and interpret a scatter diagram

LO4-6 Develop and explain a contingency table

Dot Plots Example

- Use dot plots to compare the two data sets like these of the number of vehicles serviced last month for two different dealerships

Tionesta Ford Lincoln Mercury					
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
23	33	27	28	39	26
30	32	28	33	35	32
29	25	36	31	32	27
35	32	35	37	36	30

Sheffield Motors Inc.					
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
31	35	44	36	34	37
30	37	43	31	40	31
32	44	36	34	43	36
26	38	37	30	42	33

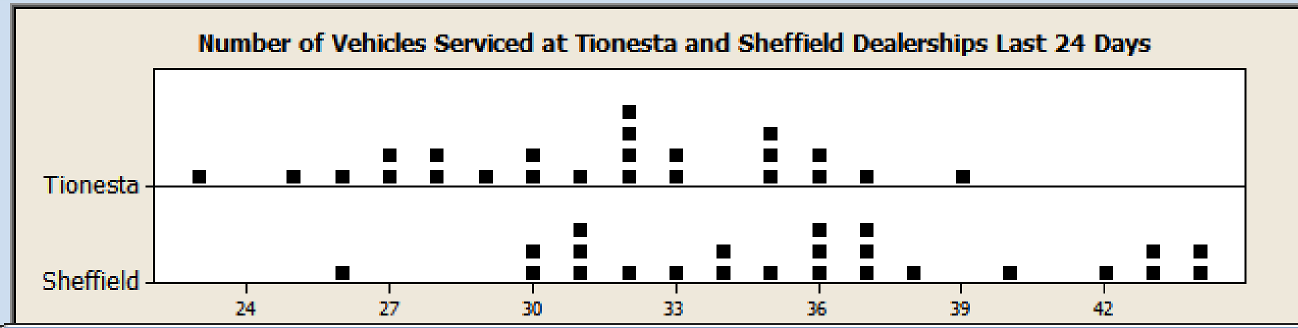
Dot Plots Example

- Minitab provides dot plots and summary statistics

Worksheet 1 ***

↓	C1	C2
	Tionesta	Sheffield
1	23	31
2	30	30
3	29	32
4	35	26
5	33	35
6	32	37
7	25	44
8	32	38
9	27	44
10	28	43
11	36	36
12	35	37
13	28	36
14	33	31
15	31	34
16	37	30
17	39	34

Dotplot of Tionesta, Sheffield



Session

Descriptive Statistics: Tionesta, Sheffield

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Tionesta	24	0	31.292	0.839	4.112	23.000	28.000	32.000	35.000	39.000
Sheffield	24	0	35.83	1.01	4.96	26.00	31.25	36.00	39.50	44.00

Measures of Position

- Measures of location also describe the shape of the distribution and can be expressed as percentiles

LOCATION OF A PERCENTILE $$L_p = (n + 1)\frac{P}{100} \quad [4\text{–}1]$$

- Quartiles divide a set of observations into four equal parts
  - The interquartile range is the difference between the third quartile and the first quartile
- Deciles divide a set of observations into 10 equal parts
- Percentiles divide a set of observations into 100 equal parts

Measures of Position Example

- Morgan Stanley is an investment company with offices located throughout the United States. Listed below are the commissions earned last month by a sample of 15 brokers.

\$2,038	\$1,758	\$1,721	\$1,637	\$2,097	\$2,047	\$2,205	\$1,787	\$2,287
1,940	2,311	2,054	2,406	1,471	1,460			

- First, sort the data from smallest to largest

\$1,460	\$1,471	\$1,637	\$1,721	\$1,758	\$1,787	\$1,940	\$2,038
2,047	2,054	2,097	2,205	2,287	2,311	2,406	

Measures of Position Example

- Next, find the median
- L_{50} = (15+1)*50/100 = 8
- So the median is $2,038, the value at position 8

$$L_{25} = (15+1)\frac{25}{100} = 4 \qquad L_{75} = (15+1)\frac{75}{100} = 12$$

Therefore, the first and third quartiles are located at the 4th and 12th positions, respectively: $L_{25} = \$1,721; L_{75} = \$2,205$

$1,460	$1,471	$1,637	$1,721	$1,758	$1,787	$1,940	$2,038
2,047	2,054	2,097	2,205	2,287	2,311	2,406	

Box Plots

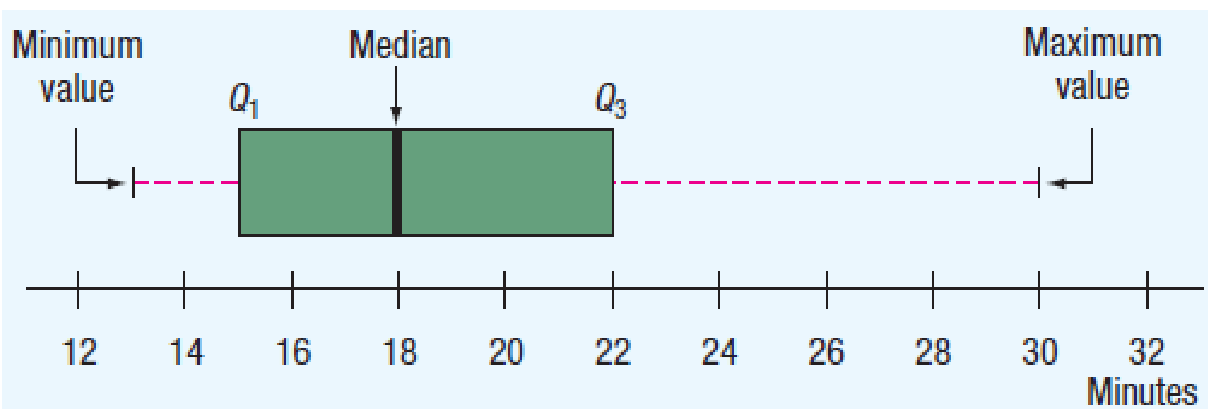
- A box plot is a graphical display using quartiles
- A box plot is based on five statistics:
  - Minimum value
  - 1st quartile
  - Median
  - 3rd quartile
  - Maximum value
- The interquartile range is Q3 – Q1
- Outliers are values that are inconsistent with the rest of the data and are identified with asterisks in box plots

Box Plot Example

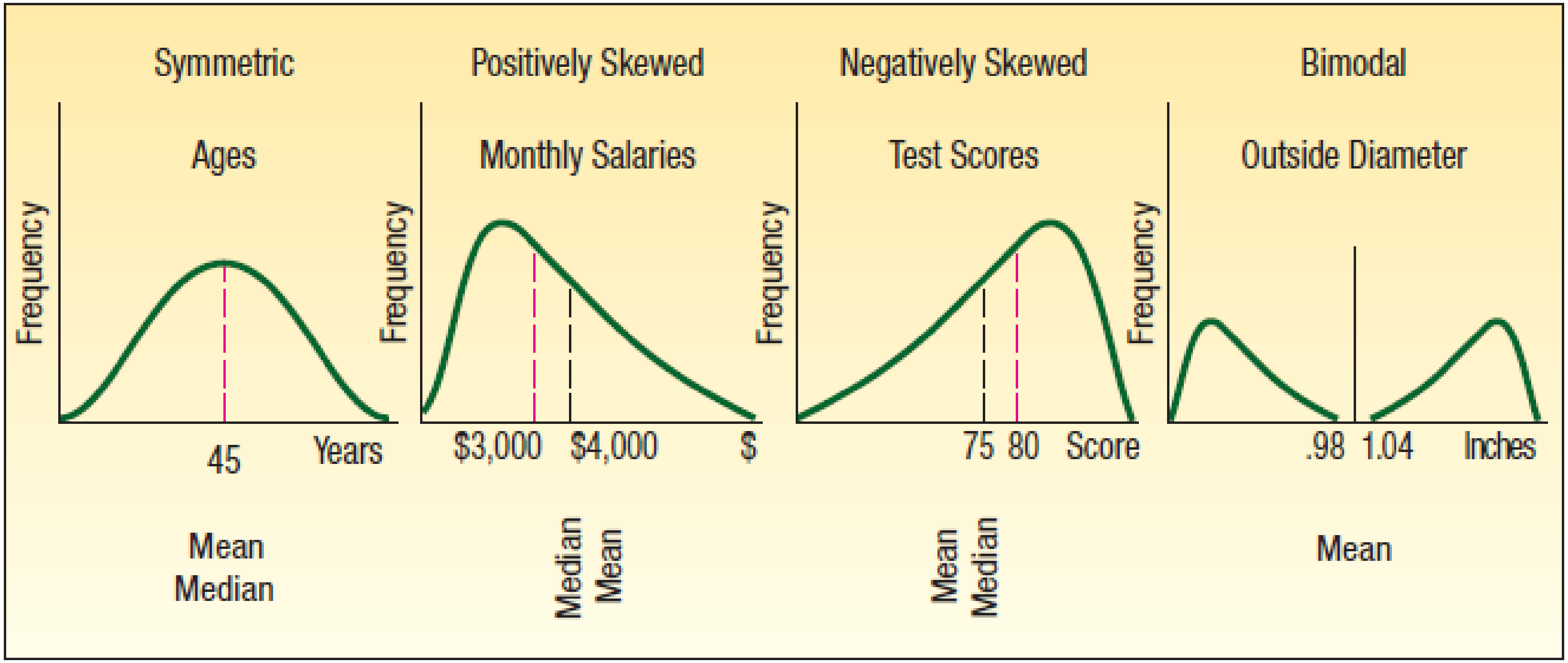
- Alexander's Pizza offers free delivery of its pizza within 15 miles. How long does a typical delivery take? Within what range will most deliveries be completed?
- Using a sample of 20 deliveries, Alexander determined the following:
  - Minimum value = 13 minutes
  - Q1 = 15 minutes
  - Median = 18 minutes
  - Q3 = 22 minutes
  - Maximum value = 30 minutes
- Develop a box plot for delivery times

Box Plot Example Continued

- Begin by drawing a number line using an appropriate scale
- Next, draw a box that begins at Q1 (15 minutes) and ends at Q3 (22 minutes)
- Draw a vertical line at the median (18 minutes)
- Extend a horizontal line out from Q3 to the maximum value (30 minutes) and out from Q1 to the minimum value (13 minutes)



Common Shapes of Data



Skewness

- The coefficient of skewness is a measure of the symmetry of a distribution
- Two formulas for coefficient of skewness

PEARSON'S COEFFICIENT OF SKEWNESS $$sk = \frac{3(\bar{x} - \text{Median})}{s} \quad [4\text{–}2]$$

- The coefficient of skewness can range from -3 to +3
- A value near -3 indicates considerable negative skewness
- A value of 1.63 indicates moderate positive skewness
- A value of 0 means the distribution is symmetrical

Skewness Example

- Following are the earnings per share for a sample of 15 software companies for the year 2017. The earnings per share are arranged from smallest to largest.

$0.09	$0.13	$0.41	$0.51	$ 1.12	$ 1.20	$ 1.49	$3.18
3.50	6.36	7.83	8.92	10.13	12.99	16.40	

- Begin by finding the mean, median, and standard deviation. Find the coefficient of skewness.

- What do you conclude about the shape of the distribution?

Skewness Example

Step 1: Compute the Mean

$$\overline{X} = \frac{\sum X}{n} = \frac{\$74.26}{15} = \$4.95$$

Step 2: Compute the Standard Deviation

$$s = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{n-1}} = \sqrt{\frac{(\$0.09 - \$4.95)^2 + \ldots + (\$16.40 - \$4.95)^2)}{15-1}} = \$5.22$$

Step 3: Find the Median

The middle value in the set of data, arranged from smallest to largest is 3.18

Step 4: Compute the Skewness

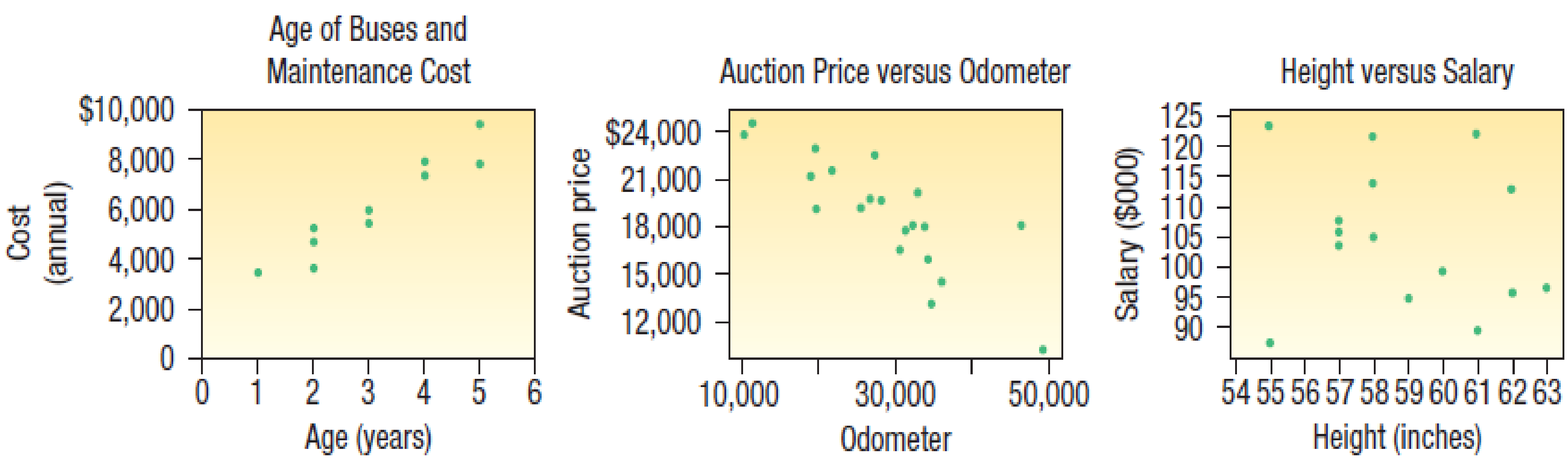
$$sk = \frac{3(\overline{X} - Median)}{s} = \frac{3(\$4.95 - \$3.18)}{\$5.22} = 1.017$$

- What do you conclude about the shape of the distribution?

Describing the Relationship Between Two Variables

- A scatter diagram is a graphical tool to portray the relationship between two variables or bivariate data
- Both variables are measured with interval or ratio level scale
- If the scatter of points moves from the lower left to the upper right, the variables under consideration are directly or positively related
- If the scatter of points moves from the upper left to the lower right, the variables are inversely or negatively related

Scatter Diagrams



Contingency Tables

- A contingency table is used to classify nominal scale observations according to two characteristics

CONTINGENCY TABLE A table used to classify observations according to two identifiable characteristics.

- It is a cross-tabulation that simultaneously summarizes two variables of interest
- Both variables need only be nominal or ordinal

Contingency Table Example

- Applewood Auto Group's profit comparison

Contingency Table Showing the Relationship between Profit and Dealership					
Above/Below Median Profit	Kane	Olean	Sheffield	Tionesta	Total
Above	25	20	19	26	90
Below	27	20	26	17	90
Total	52	40	45	43	180

- 90 of the 180 cars sold had a profit above the median and half below. This meets the definition of median.
- The percentage of profits above the median are Kane 48%, Olean 50%, Sheffield 42% , and Tionesta 60%.