



Because learning changes everything.®

Chapter 4

Describing Data: Displaying and Exploring Data

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 Compute and interpret the correlation coefficient.

LO4-7 Develop and explain a contingency table.

Dot Plots ¹

Dot Plot Summarizes the distribution of one variable by stacking dots as points on a number line that shows all the values.

Display a dot for each observation.

If there are identical values, the dots are “piled” on top of each other.

- See the shape of distribution.
- Values where the data cluster.
- Good for smaller data sets.

Dot Plots ²

- Example: The number of vehicles served at two dealers.

Tionesta Ford Lincoln

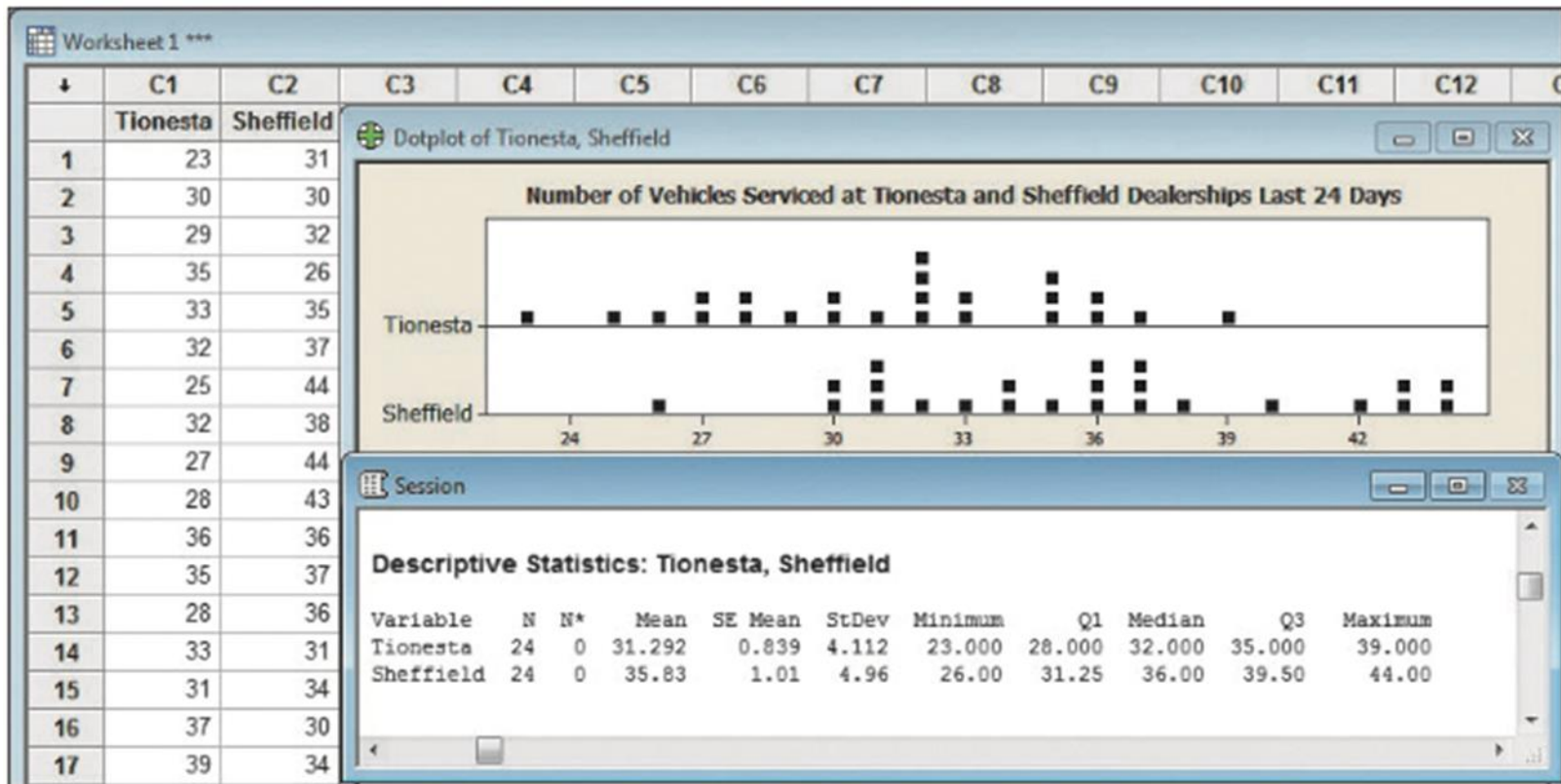
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
23	33	27	28	39	26
30	32	28	33	35	32
29	25	36	31	32	27
35	32	35	37	36	30

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 ³

- Example continued

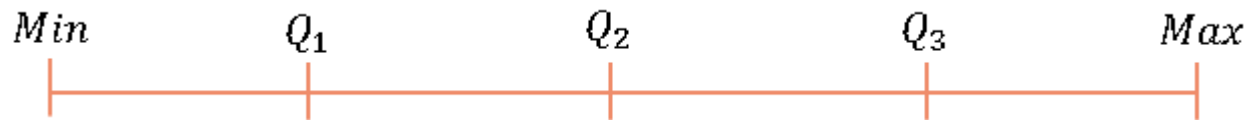


[Access the text alternative for slide images.](#)

Measures of Position ¹

The standard deviation is the most widely used measure of dispersion.

We can also determine the location of values that divide a set of observations into parts.



Quartiles: Divide a set of observations into four equal parts.

- Q_1 : Value below which 25% of the observations occur.
- Q_2 : Value below which 50% of the observations occur (median).
- Q_3 : Value below which 75% of the observations occur.

Measures of Position ₂

- Deciles: Divide a set of observations into 10 equal parts.
- Percentiles: Divide a set of observations into 100 equal parts.
- Let L_p refer to the location of a desired interval.
- For n observations, the p^{th} percentile is given by:

$$L_p = (n + 1) \frac{p}{100}$$

- Example: The 92nd percentile would be L_{92}

Measures of Position ³

- Example: The commissions earned last month by a sample of 15 brokers.

\$2,038	\$1,758	\$1,721	\$1,637	\$2,097	42,047	\$2,205	\$1,787	\$2,287
1,940	2,311	2,054	2,406	1,471	1,460			

- Locate the median, the first quartile and the third quartile of the commissions.
- 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 ⁴

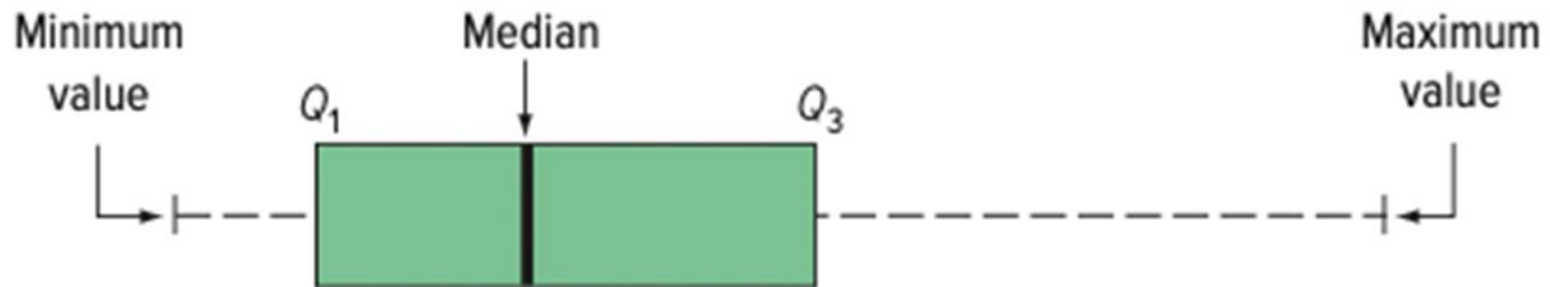
- Example continued
- $Q_1 : L_{25} = (16 + 1) \frac{25}{100} = 4$
- Median: $L_{50} = (16 + 1) \frac{50}{100} = 8$
- $Q_3 : L_{75} = (n + 1) \frac{75}{100} = 12$

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

- $Q_1 = \$1,721$, Median = \$2,038 and $Q_3 = \$2,205$

Box Plots ¹

Box Plot A graphic display that shows the general shape of a variable's distribution. It is based on five descriptive statistics: the maximum and minimum values, the first and third quartiles, and the median.



[Access the text alternative for slide images.](#)

Box Plots ₂

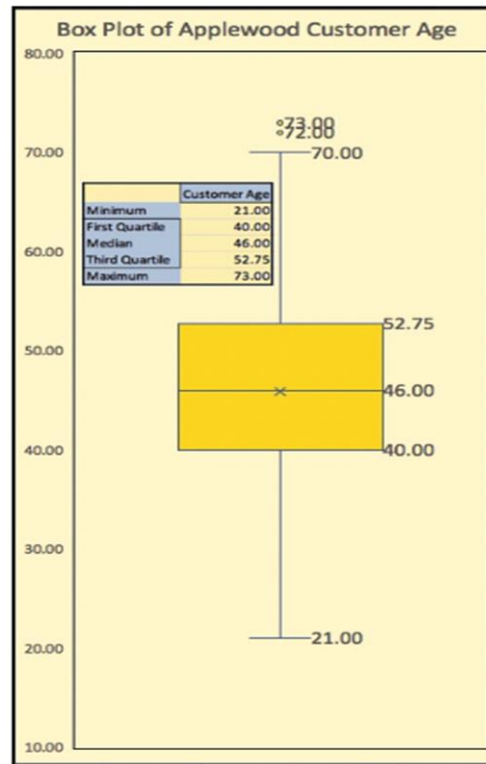
- The interquartile range is $Q_3 - Q_1$
- Range of values between the first and third quartiles.
- 50% of a distribution's values.

Outlier A data point that is unusually far from the other.

- Greater than upper boundary: $Q_3 + 1.5(Q_3 - Q_1)$
- Lower than lower boundary: $Q_1 - 1.5(Q_3 - Q_1)$
- Marked as asterisks beyond last value less than the boundary.
- No outliers: minimum or maximum should be the end lines.

Box Plots ³

- Example: The Applewood Auto Group data and the buyer age variable.



[Access the text alternative for slide images.](#)

Skewness ¹

- There are four common shapes.



[Access the text alternative for slide images.](#)

Skewness ²

Pearson: $s_k = \frac{3(\bar{x} - \text{Median})}{s}$

- Between -3 and 3.
- Near -3: considerable negative skewness.
- Near 0: symmetric.
- Near 3: considerable positive skewness.

Software skewness: $s_k = \frac{n}{(n-1)(n-2)} \left[\sum \left(\frac{x - \bar{x}}{s} \right)^3 \right]$

- Focus on the brackets with the mean and standard deviation.
- Called standardizing (discussed later).
- Values are larger than the mean: positive and skewed right.
- Values are less than the mean: negative and skewed left.

Skewness ³

- Example: The earnings per share for software companies.

\$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	

- Compute the mean, median, and standard deviation.
- Find the coefficient of skewness using Pearson's estimate and the software methods.
- What is your conclusion regarding the shape of the distribution?

Skewness ⁴

- Example continued
- Mean: $\bar{x} = \$4.95$;
Median: the middle value of \$3.18
- Standard deviation: $s = \$5.22$.

$$s_k = \frac{n}{(n-1)(n-2)} \left[\sum \left(\frac{x - \bar{x}}{s} \right)^3 \right]$$

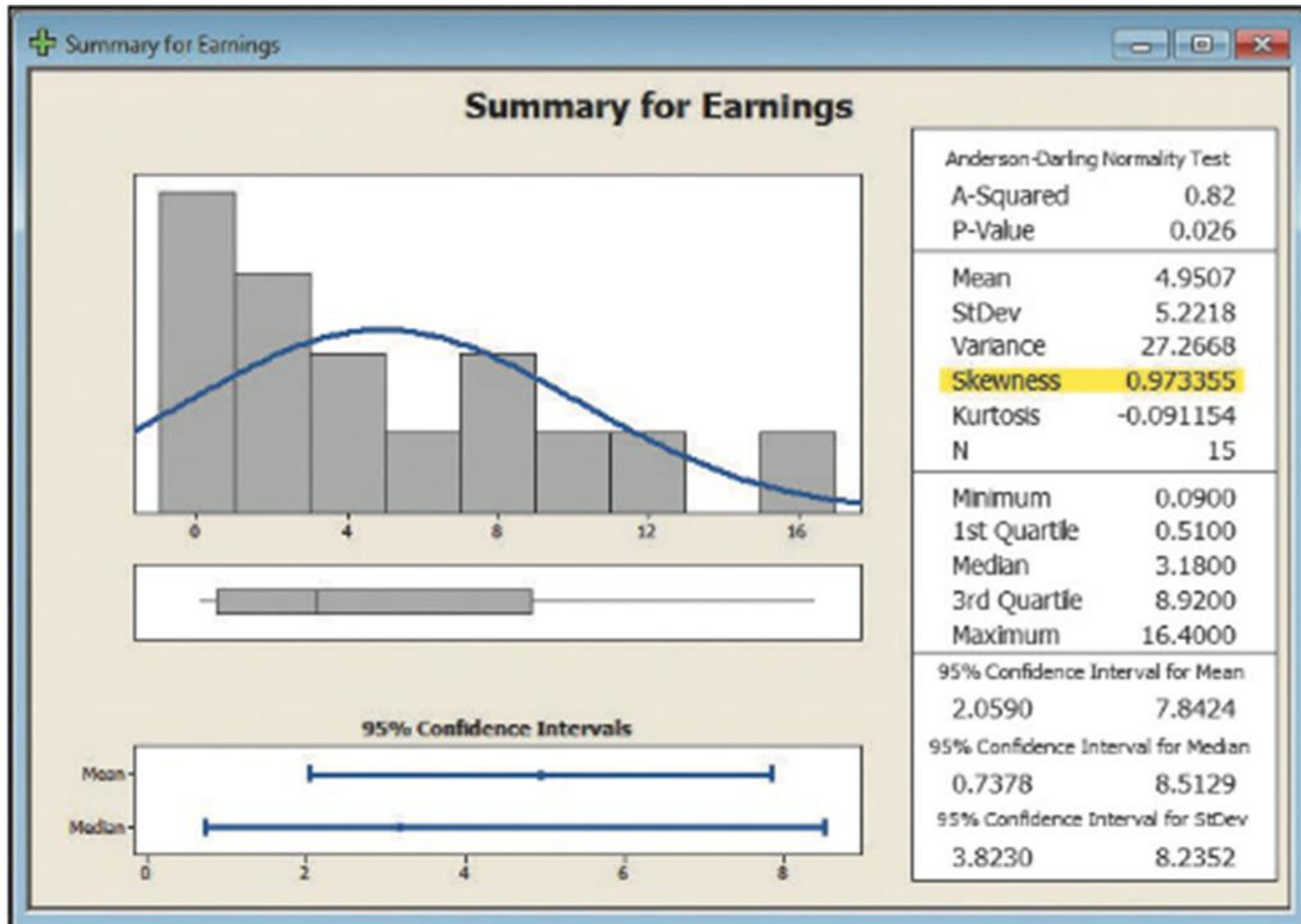
$$= \frac{15}{(15-1)(15-2)} (11.8274)$$

$$= 0.975$$

- Skewed right.
- Note the mean > median.

Earnings per Share	$(x - \bar{x}) \div s$	$((x - \bar{x}) \div s)^3$
0.09	-0.9310	-0.8070
0.13	-0.9234	-0.7873
0.41	-0.8697	-0.6579
0.51	-0.8506	-0.6154
1.12	-0.7337	-0.3950
1.20	-0.7184	-0.3708
1.49	-0.6628	-0.2912
3.18	-0.3391	-0.0390
3.50	-0.2778	-0.0214
6.36	0.2701	0.0197
7.83	0.5517	0.1679
8.92	0.7605	0.4399
10.13	0.9923	0.9772
12.99	1.5402	3.6539
16.40	2.1935	10.5537
		11.8274

Skewness 5

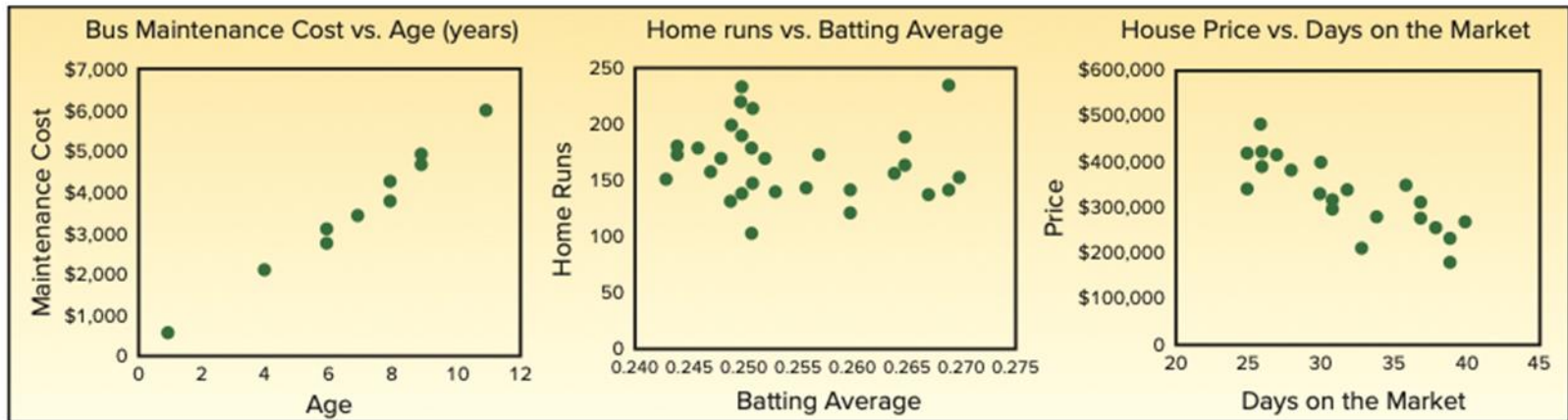


[Access the text alternative for slide images.](#)

Describing the Relationship Between Two Variables ₁

Scatter Diagram Graphical technique used to show the relationship between two variables measured with interval or ratio scales.

- Study the relationship between two variables (bivariate).
- Usually one variable on the other.



[Access the text alternative for slide images.](#)

Describing the Relationship Between Two Variables ²

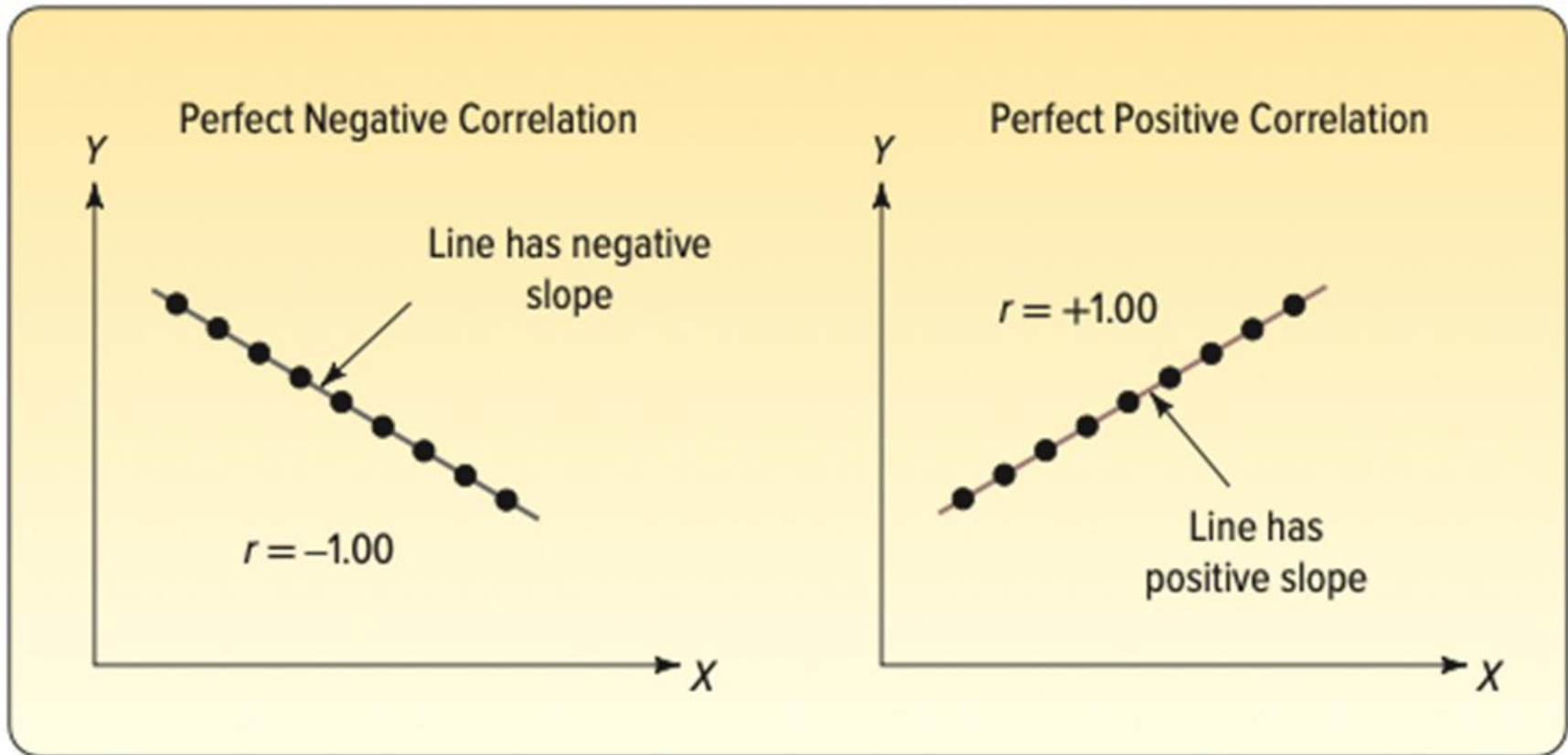
Compute a statistic called the correlation coefficient.

Measures the direction and strength of the relationship.

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n - 1)s_y s_x}$$

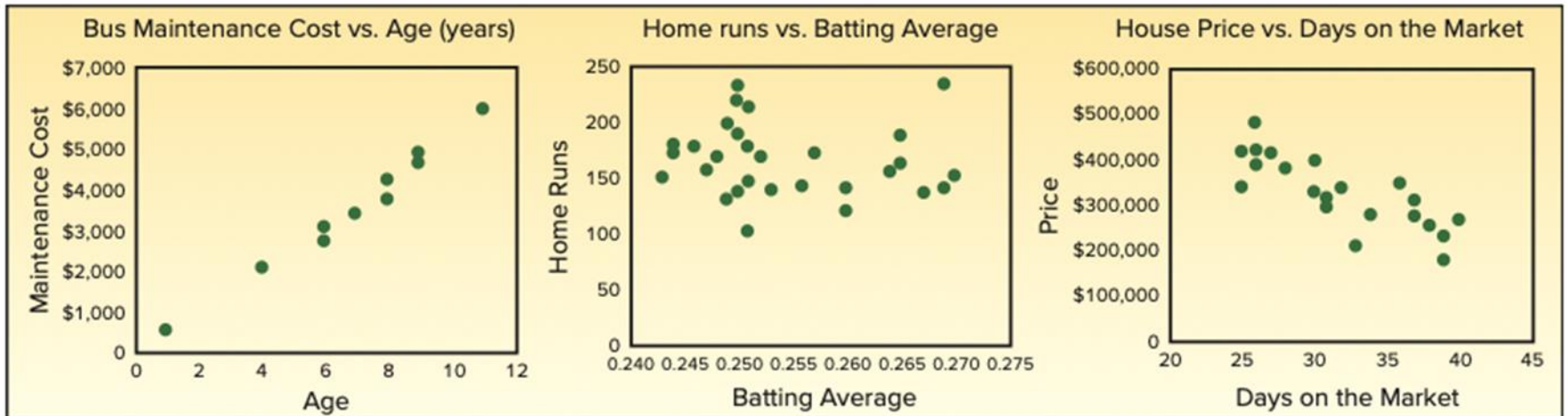
- Ranges from -1.0 to $+1.0$.
- The closer the coefficient is to -1.0 or $+1.0$, the stronger the relationship.
- If r is close to 0.0 , we can say that there is no relationship between the variables.
- Positive indicates a positive relationship.
- Negative indicates a negative relationship.

Describing the Relationship Between Two Variables ³



[Access the text alternative for slide images.](#)

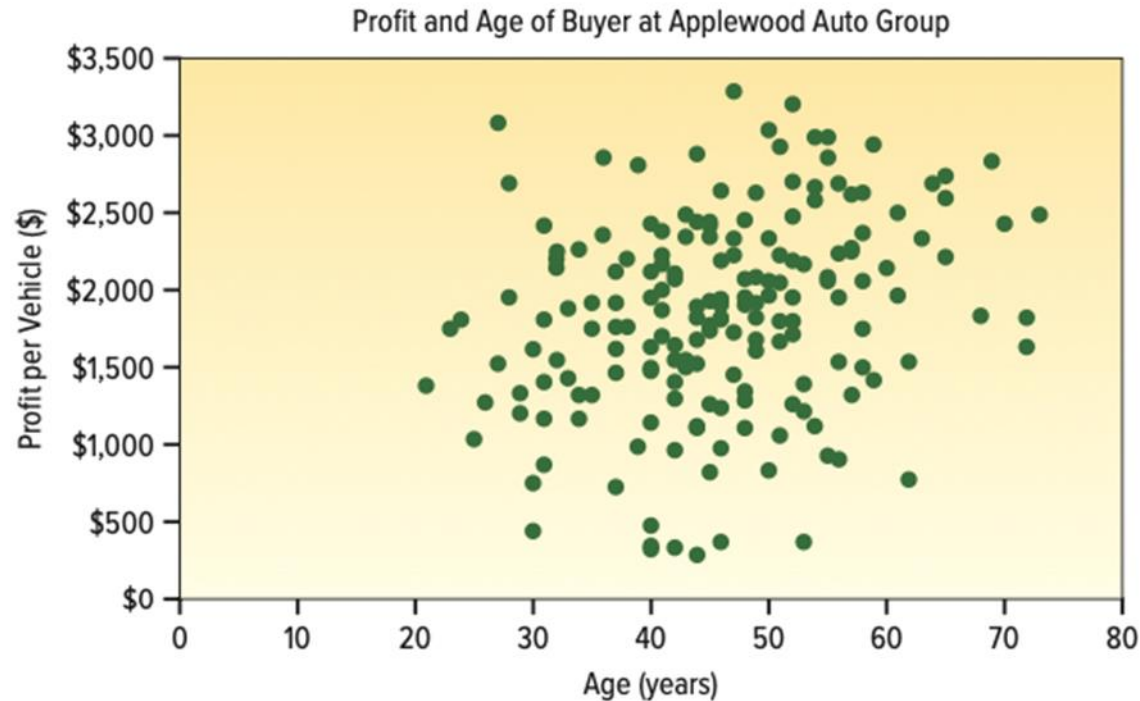
Describing the Relationship Between Two Variables ⁴



[Access the text alternative for slide images.](#)

Describing the Relationship Between Two Variables ⁵

- Example: The relationship between vehicle profit and age.
- The correlation is 0.26.



[Access the text alternative for slide images.](#)

Contingency Tables ¹

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 Tables ²

- Example: The profit for four dealerships.

Contingency Table Showing the Relationship between Profit and Dealership

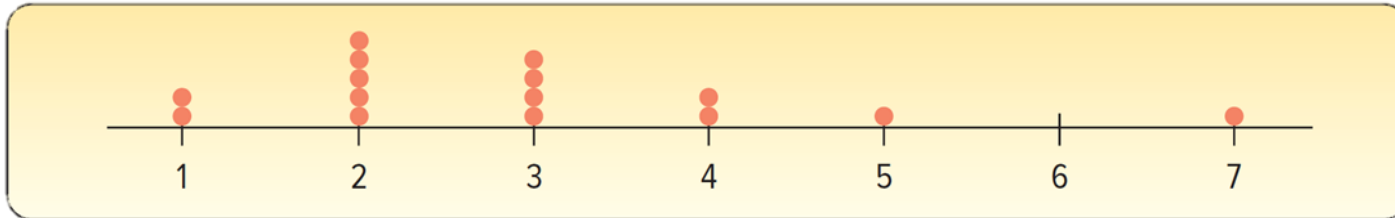
Above/Below Median Profit	Kane	Olean	Sheffield	Tionesta	Total
Above	25	20	19	26	90
Below	27	20	26	17	30
Total	52	40	45	43	180

Chapter 4 Practice Problems

Question 3

LO4-1

Consider the following chart.



- What is this chart called?
- How many observations are in the study?
- What are the maximum and the minimum values?
- Around what values do the observations tend to cluster?

[Access the text alternative for slide images.](#)

Question 7

LO4-2

The Thomas Supply Company Inc. is a distributor of gas-powered generators. As with any business, the length of time customers take to pay their invoices is important. Listed below, arranged from smallest to largest, is the time, in days, for a sample of the Thomas Supply Company Inc. invoices.

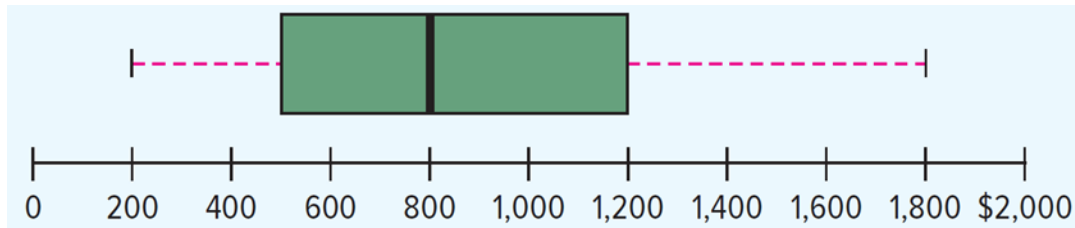
13	13	13	20	26	27	31	34	34	34	35	35	36	37	38
41	41	41	45	47	47	47	50	51	53	54	56	62	67	82

- Determine the first and third quartiles.
- Determine the second decile and the eighth decile.
- Determine the 67th percentile.

Question 9

LO4-3

The box plot below shows the amount spent for books and supplies per year by students at four-year public colleges.



- Estimate the median amount spent.
- Estimate the first and third quartiles for the amount spent.
- Estimate the interquartile range for the amount spent.
- Beyond what point is a value considered an outlier?
- Identify any outliers and estimate their values.
- Is the distribution symmetrical or positively or negatively skewed?

[Access the text alternative for slide images.](#)

Question 15

LO4-4

Listed below are the commissions earned (\$000) last year by the 15 sales representatives at Furniture Patch Inc.

\$3.9	\$5.7	\$7.3	\$10.6	\$13.0	\$13.6	\$15.1	\$15.8	\$17.1
17.4	17.6	22.3	38.6	43.2	87.7			

- Determine the mean, median, and the standard deviation.
- Determine the coefficient of skewness using Pearson's method.
- Determine the coefficient of skewness using the software method.

Question 17

LO4-5,6

Create a scatter diagram and compute a correlation coefficient. How would you describe the relationship between the values?

x-Value	y-Value	x-Value	y-Value
10	6	11	6
8	2	10	5
9	6	7	2
11	5	7	3
13	7	11	7

Question 19

LO4-7

The Director of Planning for Devine Dining Inc. wishes to study the relationship between the time of day a customer dined and whether the guest orders dessert. To investigate the relationship, the manager collected the following information on 200 recent customers.

	Time of Day		
Dessert Ordered	Lunch	Dinner	Total
Yes	32	85	117
No	<u>68</u>	<u>15</u>	<u>83</u>
Total	100	100	200

- What is the level of measurement of the two variables?
- What is the above table called?
- Does the data suggest that customer are more likely to order dessert? Explain why.
- Does the data suggest that customers at lunch time are more likely to order dessert? Explain why.
- Does the data suggest that customers at dinner time are more likely to order dessert? Explain why.



Because learning changes everything.®

www.mheducation.com