

Chapter 3

Descriptive Statistics: Numerical Measures

Learning Objectives

1. Understand the purpose of measures of location.
2. Be able to compute the mean, weighted mean, geometric mean, median, mode, quartiles, and various percentiles.
3. Understand the purpose of measures of variability.
4. Be able to compute the range, interquartile range, variance, standard deviation, and coefficient of variation.
5. Understand skewness as a measure of the shape of a data distribution. Learn how to recognize when a data distribution is negatively skewed, roughly symmetric, and positively skewed.
6. Understand how z scores are computed and how they are used as a measure of relative location of a data value.
7. Know how Chebyshev's theorem and the empirical rule can be used to determine the percentage of the data within a specified number of standard deviations from the mean.
8. Learn how to construct a 5-number summary and a box plot.
9. Be able to compute and interpret covariance and correlation as measures of association between two variables.
10. Understand the role of summary measures in data dashboards.

Solutions:

$$1. \quad \bar{x} = \frac{\sum x_i}{n} = \frac{75}{5} = 15$$

10, 12, 16, 17, 20

Median = 16 (middle value)

$$2. \quad \bar{x} = \frac{\sum x_i}{n} = \frac{96}{6} = 16$$

10, 12, 16, 17, 20, 21

$$\text{Median} = \frac{16+17}{2} = 16.5$$

$$3. \quad \text{a.} \quad \bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{6(3.2) + 3(2) + 2(2.5) + 8(5)}{6 + 3 + 2 + 8} = \frac{70.2}{19} = 3.69$$

$$\text{b.} \quad \frac{3.2 + 2 + 2.5 + 5}{4} = \frac{12.7}{4} = 3.175$$

4.

Period	Rate of Return (%)
1	-6.0
2	-8.0
3	-4.0
4	2.0
5	5.4

The mean growth factor over the five periods is:

$$\bar{x}_g = \sqrt[n]{(x_1)(x_2) \cdots (x_5)} = \sqrt[5]{(0.940)(0.920)(0.960)(1.020)(1.054)} = \sqrt[5]{0.8925} = 0.9775$$

So the mean growth rate $(0.9775 - 1)100\% = -2.25\%$.

5. 15, 20, 25, 25, 27, 28, 30, 34

$$L_{20} = \frac{P}{100}(n+1) = \frac{20}{100}(8+1) = 1.8$$

$$20\text{th percentile} = 15 + .8(20 - 15) = 19$$

$$L_{25} = \frac{P}{100}(n+1) = \frac{25}{100}(8+1) = 2.25$$

$$25\text{th percentile} = 20 + .25(25 - 20) = 21.25$$

$$L_{65} = \frac{P}{100}(n+1) = \frac{65}{100}(8+1) = 5.85$$

$$65\text{th percentile} = 27 + .85(28 - 27) = 27.85$$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(8+1) = 6.75$$

$$75\text{th percentile} = 28 + .75(30 - 28) = 29.5$$

6. $\text{Mean} = \frac{\sum x_i}{n} = \frac{657}{11} = 59.73$

Median = 57 6th item

Mode = 53 It appears 3 times

7. a. The mean commute time is $1291.5/48 = 26.91$ minutes.

b. The median commute time is 25.95 minutes.

c. The data are bimodal. The modes are 23.4 and 24.8.

d. $L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(48+1) = 36.75$

$$75\text{th percentile} = 28.5 + .75(28.5 - 28.5) = 28.5$$

8. a. Median = 80 or \$80,000. The median salary for the sample of 15 middle-level managers working at firms in Atlanta is slightly lower than the median salary reported by the Wall Street Journal.

b. $\bar{x} = \frac{\sum x_i}{n} = \frac{1260}{15} = 84$

Mean salary is \$84,000. The sample mean salary for the sample of 15 middle-level managers is greater than the median salary. This indicates that the distribution of salaries for middle-level managers working at firms in Atlanta is positively skewed.

c. The sorted data are as follows:

53 55 63 67 73 75 77 80 83 85 93 106 108 118 124

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(16) = 4$$

First quartile or 25th percentile is the value in position 4 or 67.

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(16) = 12$$

Third quartile or 75th percentile is the value in position 12 or 106.

9. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{30.57}{12} = 2.55$

b. Order the data from low 1.88 to high 5.00

Use 6th and 7th positions.

$$\text{Median} = \frac{2.34 + 2.19}{2} = 2.265$$

$$\text{c. } L_{.25} = \left(\frac{25}{100}\right)(12+1) = 3.25 \quad Q_1 = 1.97 + .25(2.09-1.97) = 2.0$$

$$L_{.75} = \left(\frac{75}{100}\right)(12+1) = 9.75 \quad Q_3 = 2.56 + .75(2.91-2.56) = 2.8225$$

$$10. \text{ a. } \bar{x} = \frac{\sum x_i}{n} = \frac{1318}{20} = 65.9$$

Order the data from the lowest rating (42) to the highest rating (83)

Position	Rating	Position	Rating
1	42	11	67
2	53	12	67
3	54	13	68
4	61	14	69
5	61	15	71
6	61	16	71
7	62	17	76
8	63	18	78
9	64	19	81
10	66	20	83

$$L_{.50} = \frac{P}{100}(n+1) = \frac{50}{100}(20+1) = 10.5$$

Median or 50th percentile = $66 + .5(67 - 66) = 66.5$

Mode is 61.

$$\text{b. } L_{.25} = \frac{P}{100}(n+1) = \frac{25}{100}(20+1) = 5.25$$

First quartile or 25th percentile = 61

$$L_{.75} = \frac{P}{100}(n+1) = \frac{75}{100}(20+1) = 15.75$$

Third quartile or 75th percentile = 71

$$\text{c. } L_{.90} = \frac{P}{100}(n+1) = \frac{90}{100}(20+1) = 18.9$$

90th percentile = $78 + .9(81 - 78) = 80.7$

90% of the ratings are 80.7 or less; 10% of the ratings are 80.7 or greater.

11. a. The median number of hours worked per week for high school science teachers is 54.
- b. The median number of hours worked per week for high school English teachers is 47.
- c. The median number of hours worked per week for high school science teachers is greater than the median number of hours worked per week for high school English teachers; the difference is $54 - 47 = 7$ hours.
12. a. The minimum number of viewers that watched a new episode is 13.3 million, and the maximum number is 16.5 million.
- b. The mean number of viewers that watched a new episode is 15.04 million or approximately 15.0 million; the median also 15.0 million. The data is multimodal (13.6, 14.0, 16.1, and 16.2 million); in such cases the mode is usually not reported.
- c. The data are first arranged in ascending order.

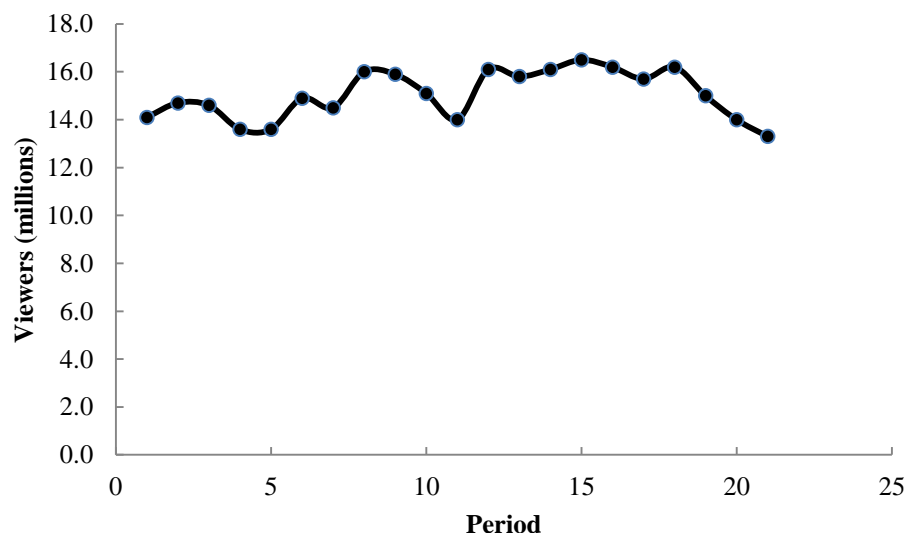
$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(21+1) = 5.50$$

First quartile or 25th percentile = $14 + .50(14.1 - 14) = 14.05$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(21+1) = 16.5$$

Third quartile or 75th percentile = $16 + .5(16.1 - 16) = 16.05$

- d. A graph showing the viewership data over the air dates follows. Period 1 corresponds to the first episode of the season, period 2 corresponds to the second episode, and so on.



This graph shows that viewership of *The Big Bang Theory* has been relatively stable over the 2011–2012 television season.

13. Using the mean we get $\bar{x}_{\text{city}} = 15.58$, $\bar{x}_{\text{highway}} = 18.92$

For the samples we see that the mean mileage is better on the highway than in the city.

City

13.2 14.4 15.2 15.3 15.3 15.3 15.9 16 16.1 16.2 16.2 16.7 16.8
 \uparrow
 Median

Mode: 15.3

Highway

17.2 17.4 18.3 18.5 18.6 18.6 18.7 19.0 19.2 19.4 19.4 20.6 21.1
 \uparrow
 Median

Mode: 18.6, 19.4

The median and modal mileages are also better on the highway than in the city.

14. For March 2011:

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(50+1) = 12.75$$

$$\text{First quartile or 25th percentile} = 6.8 + .75(6.8 - 6.8) = 6.8$$

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(50+1) = 25.5$$

$$\text{Second quartile or median} = 8 + .5(8 - 8) = 8$$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(50+1) = 38.25$$

$$\text{Third quartile or 75th percentile} = 9.4 + .25(9.6 - 9.4) = 9.45$$

For March 2012:

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(50+1) = 12.75$$

$$\text{First quartile or 25th percentile} = 6.2 + .75(6.2 - 6.2) = 6.2$$

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(50+1) = 25.5$$

$$\text{Second quartile or median} = 7.3 + .5(7.4 - 7.3) = 7.35$$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(50+1) = 38.25$$

$$\text{Third quartile or 75th percentile} = 8.6 + .25(8.6 - 8.6) = 8.6$$

It may be easier to compare these results if we place them in a table.

	March 2011	March 2012
First Quartile	6.80	6.20
Median	8.00	7.35
Third Quartile	9.45	8.60

The results show that in March 2012 approximately 25% of the states had an unemployment rate of 6.2% or less, lower than in March 2011. And, the median of 7.35% and the third quartile of 8.6% in March 2012 are both less than the corresponding values in March 2011, indicating that unemployment rates across the states are decreasing.

15. To calculate the average sales price we must compute a weighted mean. The weighted mean is

$$\frac{501(34.99) + 1425(38.99) + 294(36.00) + 882(33.59) + 715(40.99) + 1088(38.59) + 1644(39.59) + 819(37.99)}{501 + 1425 + 294 + 882 + 715 + 1088 + 1644 + 819}$$

$$= 38.11$$

Thus, the average sales price per case is \$38.11.

16. a.

Grade x_i	Weight w_i
4 (A)	9
3 (B)	15
2 (C)	33
1 (D)	3
0 (F)	0
	60 Credit Hours

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{9(4) + 15(3) + 33(2) + 3(1)}{9 + 15 + 33 + 3} = \frac{150}{60} = 2.50$$

- b. Yes; satisfies the 2.5 grade point average requirement

17. a.
$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{9191(4.65) + 2621(18.15) + 1419(11.36) + 2900(6.75)}{9191 + 2621 + 1419 + 2900}$$

$$= \frac{126,004.14}{16,131} = 7.81$$

The weighted average total return for the Morningstar funds is 7.81%.

- b. If the amount invested in each fund was available, it would be better to use those amounts as weights. The weighted return computed in part (a) will be a good approximation, if the amount invested in the various funds is approximately equal.

c. Portfolio Return =
$$\frac{2000(4.65) + 4000(18.15) + 3000(11.36) + 1000(6.75)}{2000 + 4000 + 3000 + 1000}$$

$$= \frac{122,730}{10,000} = 12.27$$

The portfolio return would be 12.27%.

18.

Assessment	Deans	$w_i x_i$	Recruiters	$w_i x_i$
5	44	220	31	155
4	66	264	34	136
3	60	180	43	129
2	10	20	12	24
1	0	0	0	0
Total	180	684	120	444

$$\text{Deans: } \bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{684}{180} = 3.8$$

$$\text{Recruiters: } \bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{444}{120} = 3.7$$

Business school deans rated the overall academic quality of master's programs slightly higher than corporate recruiters did.

19. To calculate the mean growth rate we must first compute the geometric mean of the five growth factors:

Year	% Growth	Growth Factor x_i
2007	5.5	1.055
2008	1.1	1.011
2009	-3.5	0.965
2010	-1.1	0.989
2011	1.8	1.018

$$\bar{x}_g = \sqrt[n]{(x_1)(x_2) \cdots (x_5)} = \sqrt[5]{(1.055)(1.011)(0.965)(0.989)(1.018)} = \sqrt[5]{1.036275} = 1.007152$$

The mean annual growth rate is $(1.007152 - 1)100 = 0.7152\%$.

20.

Year	Stivers		Trippi	
	End of Year Value	Growth Factor	End of Year Value	Growth Factor
Year 1	\$11,000	1.100	\$5,600	1.120
Year 2	\$12,000	1.091	\$6,300	1.125
Year 3	\$13,000	1.083	\$6,900	1.095
Year 4	\$14,000	1.077	\$7,600	1.101
Year 5	\$15,000	1.071	\$8,500	1.118
Year 6	\$16,000	1.067	\$9,200	1.082
Year 7	\$17,000	1.063	\$9,900	1.076
Year 8	\$18,000	1.059	\$10,600	1.071

For the Stivers mutual fund we have:

$$18000=10000[(x_1)(x_2)\cdots(x_8)], \text{ so } [(x_1)(x_2)\cdots(x_8)]=1.8 \text{ and}$$

$$\bar{x}_g = \sqrt[8]{(x_1)(x_2)\cdots(x_8)} = \sqrt[8]{1.80} = 1.07624$$

So the mean annual return for the Stivers mutual fund is $(1.07624 - 1)100 = 7.624\%$

For the Trippi mutual fund we have:

$$10600=5000[(x_1)(x_2)\cdots(x_8)], \text{ so } [(x_1)(x_2)\cdots(x_8)]=2.12 \text{ and}$$

$$\bar{x}_g = \sqrt[8]{(x_1)(x_2)\cdots(x_8)} = \sqrt[8]{2.12} = 1.09848$$

So the mean annual return for the Trippi mutual fund is $(1.09848 - 1)100 = 9.848\%$.

While the Stivers mutual fund has generated a nice annual return of 7.6%, the annual return of 9.8% earned by the Trippi mutual fund is far superior.

$$21. \quad 3500=5000[(x_1)(x_2)\cdots(x_9)], \text{ so } [(x_1)(x_2)\cdots(x_9)] = 0.7, \text{ and so}$$

$$\bar{x}_g = \sqrt[9]{(x_1)(x_2)\cdots(x_9)} = \sqrt[9]{0.7} = .961144$$

So the mean annual growth rate is $(.961144 - 1)100 = -3.8856\%$

$$22. \quad 25,000,000=10,000,000[(x_1)(x_2)\cdots(x_6)], \text{ so } [(x_1)(x_2)\cdots(x_6)]=2.50, \text{ and so}$$

$$\bar{x}_g = \sqrt[6]{(x_1)(x_2)\cdots(x_6)} = \sqrt[6]{2.50} = 1.165$$

So the mean annual growth rate is $(1.165 - 1)100 = 16.5\%$

$$23. \quad \text{Range } 20 - 10 = 10$$

$$10, 12, 16, 17, 20$$

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(6) = 1.5$$

$$\text{First Quartile or } Q_1 = 10 + .5(12-10) = 11$$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(6) = 4.5$$

$$\text{Third Quartile or } Q_3 = 17 + .5(20-17) = 18.5$$

$$\text{IQR} = Q_3 - Q_1 = 18.5 - 11 = 7.5$$

$$24. \quad \bar{x} = \frac{\sum x_i}{n} = \frac{75}{5} = 15$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{64}{4} = 16$$

$$s = \sqrt{16} = 4$$

25. 15, 20, 25, 25, 27, 28, 30, 34 Range = 34 - 15 = 19

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(9) = 2.25$$

$$\text{First Quartile or } Q_1 = 20 + .25(25-20) = 21.25$$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(9) = 6.75$$

$$\text{Third Quartile or } Q_3 = 28 + .75(30-28) = 29.5$$

$$\text{IQR} = Q_3 - Q_1 = 29.5 - 21.25 = 8.25$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{204}{8} = 25.5$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{242}{7} = 34.57$$

$$s = \sqrt{34.57} = 5.88$$

26. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{74.4}{20} = 3.72$

b. $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1.6516}{20-1}} = \sqrt{.0869} = .2948$

- c. The average price for a gallon of unleaded gasoline in San Francisco is much higher than the national average. This indicates that the cost of living in San Francisco is higher than it would be for cities that have an average price close to the national average.

27. a. The mean price for a round-trip flight into Atlanta is \$356.73, and the mean price for a round-trip flight into Salt Lake City is \$400.95. Flights into Atlanta are less expensive than flights into Salt Lake City. This possibly could be explained by the locations of these two cities relative to the 14 departure cities; Atlanta is generally closer than Salt Lake City to the departure cities.
- b. For flights into Atlanta, the range is \$290.0, the variance is 5517.41, and the standard Deviation is \$74.28. For flights into Salt Lake City, the range is \$458.8, the variance is 18933.32, and the standard deviation is \$137.60.

The prices for round-trip flights into Atlanta are less variable than prices for round-trip flights into Salt Lake City. This could also be explained by Atlanta's relative nearness to the 14 departure cities.

28. a. The mean serve speed is 180.95, the variance is 21.42, and the standard deviation is 4.63.
- b. Although the mean serve speed for the twenty Women's Singles serve speed leaders for the 2011 Wimbledon tournament is slightly higher, the difference is very small. Furthermore, given the variation in the twenty Women's Singles serve speed leaders from the 2012 Australian Open and the twenty Women's Singles serve speed leaders from the 2011 Wimbledon tournament, the difference in the mean serve speeds is most likely due to random variation in the players' performances.

29. a. Range = $60 - 28 = 32$

28 42 45 48 49 50 55 58 60

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(10) = 2.5 \quad \text{First Quartile or } Q_1 = 42 + .5(45-42) = 43.5$$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(10) = 7.5 \quad \text{Third Quartile or } Q_3 = 55 + .5(58-55) = 56.5$$

$$\text{IQR} = Q_3 - Q_1 = 56.5 - 43.5 = 13$$

b. $\bar{x} = \frac{435}{9} = 48.33$

$$\Sigma(x_i - \bar{x})^2 = 742$$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{742}{8} = 92.75$$

$$s = \sqrt{92.75} = 9.63$$

- c. The average air quality is about the same. But, the variability is greater in Anaheim.

30. Dawson Supply: Range = $11 - 9 = 2$

$$s = \sqrt{\frac{4.1}{9}} = 0.67$$

J.C. Clark: Range = $15 - 7 = 8$

$$s = \sqrt{\frac{60.1}{9}} = 2.58$$

31. a.

	18–34	35–44	45+
mean	1368.0	1330.1	1070.4
median	1423.0	1382.5	1163.5
standard deviation	540.8	431.7	334.5

- b. The 45+ group appears to spend less on coffee than the other two groups, and the 18–34 and 35–44 groups spend similar amounts of coffee.

32. a. Automotive: $\bar{x} = \frac{\sum x_i}{n} = \frac{39201}{20} = 1960.05$

Department store: $\bar{x} = \frac{\sum x_i}{n} = \frac{13857}{20} = 692.85$

b. Automotive: $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{4,407,720.95}{19}} = 481.65$

Department store: $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{456804.55}{19}} = 155.06$

c. Automotive: $2901 - 598 = 2303$
 Department Store: $1011 - 448 = 563$

d. Order the data for each variable from the lowest to highest.

	Automotive	Department Store
1	598	448
2	1512	472
3	1573	474
4	1642	573
5	1714	589
6	1720	597
7	1781	598
8	1798	622
9	1813	629
10	2008	669
11	2014	706
12	2024	714
13	2058	746
14	2166	760
15	2202	782
16	2254	824
17	2366	840
18	2526	856
19	2531	947
20	2901	1011

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(21) = 5.25$$

Automotive: First quartile or 25th percentile = $1714 + .25(1720 - 1714) = 1715.5$

Department Store: First quartile or 25th percentile = $589 + .25(597 - 589) = 591$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(21) = 15.75$$

Automotive: Third quartile or 75th percentile = $2202 + .75(2254 - 2202) = 2241$

Department Store: First quartile or 75th percentile = $782 + .75(824 - 782) = 813.5$

Automotive IQR = $Q_3 - Q_1 = 2241 - 1715.5 = 525.5$

Department Store IQR = $Q_3 - Q_1 = 813.5 - 591 = 222.5$

- e. Automotive spends more on average, has a larger standard deviation, larger max and min, and larger range than Department Store. Autos have all new model years and may spend more heavily on advertising.

33. a. For 2014

$$\bar{x} = \frac{\sum x_i}{n} = \frac{608}{8} = 76$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{30}{7}} = 2.07$$

For 2015

$$\bar{x} = \frac{\sum x_i}{n} = \frac{608}{8} = 76$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{194}{7}} = 5.26$$

- b. The mean score is 76 for both years, but there is an increase in the standard deviation for the scores in 2015. The golfer is not as consistent in 2015 and shows a sizeable increase in the variation with golf scores ranging from 71 to 85. The increase in variation might be explained by the golfer trying to change or modify the golf swing. In general, a loss of consistency and an increase in the standard deviation could be viewed as a poorer performance in 2015. The optimism in 2015 is that three of the eight scores were better than any score reported for 2014. If the golfer can work for consistency, eliminate the high score rounds, and reduce the standard deviation, golf scores should show improvement.

34. Quarter milers

$$s = 0.0564$$

$$\text{Coefficient of Variation} = (s/\bar{x})100\% = (0.0564/0.966)100\% = 5.8\%$$

Milers

$$s = 0.1295$$

$$\text{Coefficient of Variation} = (s/\bar{x})100\% = (0.1295/4.534)100\% = 2.9\%$$

Yes; the coefficient of variation shows that as a percentage of the mean the quarter milers' times show more variability.

$$35. \quad \bar{x} = \frac{\sum x_i}{n} = \frac{75}{5} = 15$$

$$s^2 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{64}{4}} = 4$$

$$10 \quad z = \frac{10-15}{4} = -1.25$$

$$20 \quad z = \frac{20-15}{4} = +1.25$$

$$12 \quad z = \frac{12-15}{4} = -.75$$

$$17 \quad z = \frac{17-15}{4} = +.50$$

$$16 \quad z = \frac{16-15}{4} = +.25$$

$$36. \quad z = \frac{520-500}{100} = +.20$$

$$z = \frac{650-500}{100} = +1.50$$

$$z = \frac{500-500}{100} = 0.00$$

$$z = \frac{450-500}{100} = -.50$$

$$z = \frac{280-500}{100} = -2.20$$

$$37. \quad a. \quad z = \frac{20-30}{5} = -2, \quad z = \frac{40-30}{5} = 2 \quad 1 - \frac{1}{2^2} = .75 \text{ At least 75\%}$$

$$b. \quad z = \frac{15-30}{5} = -3, \quad z = \frac{45-30}{5} = 3 \quad 1 - \frac{1}{3^2} = .89 \text{ At least 89\%}$$

$$c. \quad z = \frac{22-30}{5} = -1.6, \quad z = \frac{38-30}{5} = 1.6 \quad 1 - \frac{1}{1.6^2} = .61 \text{ At least 61\%}$$

$$d. \quad z = \frac{18-30}{5} = -2.4, \quad z = \frac{42-30}{5} = 2.4 \quad 1 - \frac{1}{2.4^2} = .83 \text{ At least 83\%}$$

$$e. \quad z = \frac{12-30}{5} = -3.6, \quad z = \frac{48-30}{5} = 3.6 \quad 1 - \frac{1}{3.6^2} = .92 \text{ At least } 92\%$$

38. a. Approximately 95%
 b. Almost all
 c. Approximately 68%
39. a. This is from 2 standard deviations below the mean to 2 standard deviations above the mean.

With $z = 2$, Chebyshev's theorem gives:

$$1 - \frac{1}{z^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = .75$$

Therefore, at least 75% of adults sleep between 4.5 and 9.3 hours per day.

- b. This is from 2.5 standard deviations below the mean to 2.5 standard deviations above the mean.

With $z = 2.5$, Chebyshev's theorem gives:

$$1 - \frac{1}{z^2} = 1 - \frac{1}{2.5^2} = 1 - \frac{1}{6.25} = .84$$

Therefore, at least 84% of adults sleep between 3.9 and 9.9 hours per day.

- c. With $z = 2$, the empirical rule suggests that 95% of adults sleep between 4.5 and 9.3 hours per day. The percentage obtained using the empirical rule is greater than the percentage obtained using Chebyshev's theorem.
40. a. \$3.33 is one standard deviation below the mean and \$3.53 is one standard deviation above the mean. The empirical rule says that approximately 68% of gasoline sales are in this price range.
- b. Part (a) shows that approximately 68% of the gasoline sales are between \$3.33 and \$3.53. Since the bell-shaped distribution is symmetric, approximately half of 68%, or 34%, of the gasoline sales should be between \$3.33 and the mean price of \$3.43. \$3.63 is two standard deviations above the mean price of \$3.43. The empirical rule says that approximately 95% of the gasoline sales should be within two standard deviations of the mean. Thus, approximately half of 95%, or 47.5%, of the gasoline sales should be between the mean price of \$3.43 and \$3.63. The percentage of gasoline sales between \$3.33 and \$3.63 should be approximately $34\% + 47.5\% = 81.5\%$.
- c. \$3.63 is two standard deviations above the mean and the empirical rule says that approximately 95% of the gasoline sales should be within two standard deviations of the mean. Thus, $1 - 95\% = 5\%$ of the gasoline sales should be more than two standard deviations from the mean. Since the bell-shaped distribution is symmetric, we expected half of 5%, or 2.5%, would be more than \$3.63.
41. a. 647 is one standard deviation above the mean. Approximately 68% of the scores are between 447 and 647 with half of 68%, or 34%, of the scores between the mean of 547 and 647. Also, since the distribution is symmetric, 50% of the scores are above the mean of 547. With 50% of the scores above 547 and with 34% of the scores between 547 and 647, $50\% - 34\% = 16\%$ of the scores are 647 or higher.
- b. 747 is two standard deviations above the mean. Approximately 95% of the scores are between 347 and 747 with half of 95%, or 47.5%, of the scores between the mean of 547 and 747. Also, since the distribution is symmetric, 50% of the scores are above the mean of 547. With 50% of the scores

above 547 and with 47.5% of the scores between 547 and 747, $50\% - 47.5\% = 2.5\%$ of the scores are 747 or higher.

- c. Approximately 68% of the scores are between 447 and 647 with half of 68%, or 34%, of the scores are between 447 and the mean of 547.
- d. Approximately 95% of the scores are between 347 and 747 with half of 95%, or 47.5%, of the scores between 347 and the mean of 547. Approximately 68% of the scores are between 447 and 647 with half of 68%, or 34%, of the scores between the mean of 547 and 647. Thus, $47.5\% + 34\% = 81.5\%$ of the scores are between 347 and 647.

42. a. $z = \frac{x - \mu}{\sigma} = \frac{2300 - 3100}{1200} = -.67$

b. $z = \frac{x - \mu}{\sigma} = \frac{4900 - 3100}{1200} = 1.50$

- c. \$2300 is .67 standard deviations below the mean. \$4900 is 1.50 standard deviations above the mean. Neither is an outlier.

d. $z = \frac{x - \mu}{\sigma} = \frac{13000 - 3100}{1200} = 8.25$

\$13,000 is 8.25 standard deviations above the mean. This cost is an outlier.

43. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{106.5}{10} = 10.65$

Median = 10.7

Mode = 10.7

b. Range = $11.8 - 8.3 = 3.5$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{8.085}{9} = .8983$$

$$s = \sqrt{.8983} = .9478$$

c. For the response time of 8.3, the z -score is $z = \frac{8.3 - 10.65}{.9478} = -2.4794$

Thus, this observation does not appear to be an outlier.

- d. The national standard of six minutes is not being met for this neighborhood. The city should consider making changes in its response strategy including relocating stations to reduce the travel time.

44. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{765}{10} = 76.5$

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{442.5}{10-1}} = 7.01$$

$$\text{b. } z = \frac{x - \bar{x}}{s} = \frac{84 - 76.5}{7.01} = 1.07$$

Approximately one standard deviation above the mean. Approximately 68% of the scores are within one standard deviation. Thus, half of the remaining 32%, or 16%, of the games should have a winning score of more than one standard deviation above the mean or a score of 84 or more points.

$$z = \frac{x - \bar{x}}{s} = \frac{90 - 76.5}{7.01} = 1.93$$

Approximately two standard deviations above the mean. Approximately 95% of the scores are within two standard deviations. Thus, half of the remaining 5%, or 2.5%, of the games should have a winning score of more than two standard deviations above the mean or a score of more than 90 points.

$$\text{c. } \bar{x} = \frac{\sum x_i}{n} = \frac{122}{10} = 12.2$$

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{559.6}{10-1}} = 7.89$$

$$\text{Smallest margin 3: } z = \frac{x - \bar{x}}{s} = \frac{3 - 12.2}{7.89} = -1.17$$

$$\text{Largest margin 24: } z = \frac{x - \bar{x}}{s} = \frac{24 - 12.2}{7.89} = 1.50. \text{ No outliers.}$$

45. a. The data in ascending order follow:

36 38 43 44 44 45 50 55 56 56 57 60 62 65 69

$$\bar{x} = \frac{\sum x_i}{n} = \frac{780}{15} = 52$$

$$L_{50} = \frac{P}{100}(n+1) = \frac{50}{100}(15+1) = 8$$

Median is the value in position 8 or 55.

$$\text{b. } L_{25} = \frac{P}{100}(n+1) = \frac{25}{100}(15+1) = 4$$

First quartile or 25th percentile is the value in position 4 or 44.

$$L_{75} = \frac{P}{100}(n+1) = \frac{75}{100}(15+1) = 12$$

Third quartile or 75th percentile is the value in position 12 or 60.

- c. The range is $69 - 36 = 33$ and the interquartile range is $60 - 44 = 16$.

d.
$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{1402}{14} = 100.1429$$

$$s = \sqrt{100.1429} = 10.0071$$

- e. The z-score values do not indicate any outliers.

- f. The sample mean of 52% indicates that Wal-Mart does appear to be meeting its goal of reducing the number of hourly employees by about 50%.

46. 15, 20, 25, 25, 27, 28, 30, 34

Smallest = 15

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(8+1) = 2.25$$

First quartile or 25th percentile = $20 + .25(25 - 20) = 21.25$

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(8+1) = 4.5$$

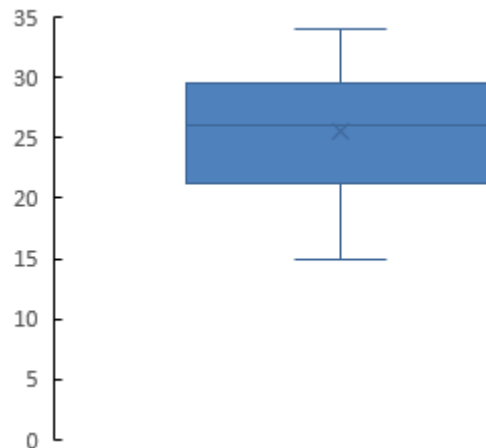
Second quartile or median = $25 + .5(27 - 25) = 26$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(8+1) = 6.75$$

Third quartile or 75th percentile = $28 + .75(30 - 28) = 29.5$

Largest = 34

47. A box plot created using Excel's Box and Whisker Statistical Chart follows.



48. 5, 6, 8, 10, 10, 12, 15, 16, 18

Smallest = 5

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(9+1) = 2.5$$

First quartile or 25th percentile = $6 + .5(8 - 6) = 7$

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(9+1) = 5.0$$

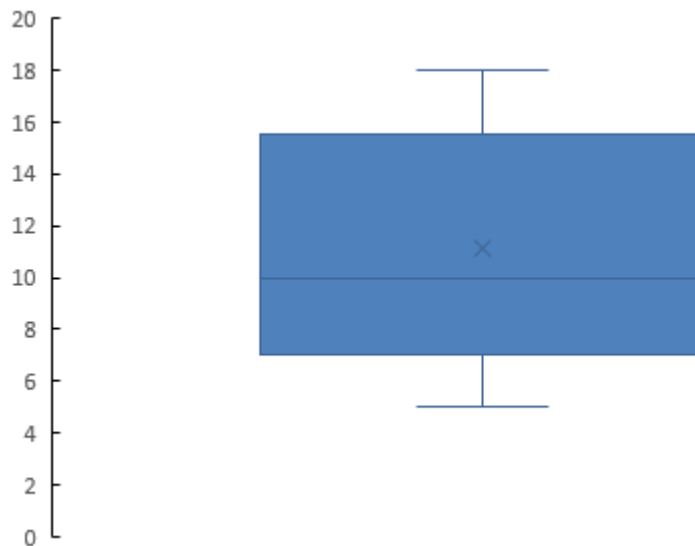
Second quartile or median = 10

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(9+1) = 7.5$$

Third quartile or 75th percentile = $15 + .5(16 - 15) = 15.5$

Largest = 18

A box plot created using Excel's Box and Whisker Statistical Chart follows.



49. $IQR = 50 - 42 = 8$

Lower Limit: $Q_1 - 1.5 IQR = 42 - 12 = 30$

Upper Limit: $Q_3 + 1.5 IQR = 50 + 12 = 62$

65 is an outlier

50. a. The first place runner in the men's group finished $109.03 - 65.30 = 43.73$ minutes ahead of the first place runner in the women's group. Lauren Wald would have finished in 11th place for the combined groups.
- b. Using Excel's MEDIAN function the results are as follows:

Men	Women
109.64	131.67

Using the median finish times, the men's group finished $131.67 - 109.64 = 22.03$ minutes ahead of the women's group.

Also note that the fastest time for a woman runner, 109.03 minutes, is approximately equal to the median time of 109.64 minutes for the men's group.

- c. Using Excel's QUARTILE.EXC function the quartiles are as follows:

Quartile	Men	Women
1	83.1025	122.080
2	109.640	131.670
3	129.025	147.180

Excel's MIN and MAX functions provided the following values.

	Men	Women
Minimum	65.30	109.03
Maximum	148.70	189.28

Five number summary for men: 65.30, 83.1025, 109.640, 129.025, 148.70

Five number summary for women: 109.03, 122.08, 131.67, 147.18, 189.28

- d. Men: $IQR = 129.025 - 83.1025 = 45.9225$

$$\text{Lower Limit} = Q_1 - 1.5(IQR) = 83.1025 - 1.5(45.9225) = 14.22$$

$$\text{Upper Limit} = Q_3 + 1.5(IQR) = 129.025 + 1.5(45.9225) = 197.91$$

There are no outliers in the men's group.

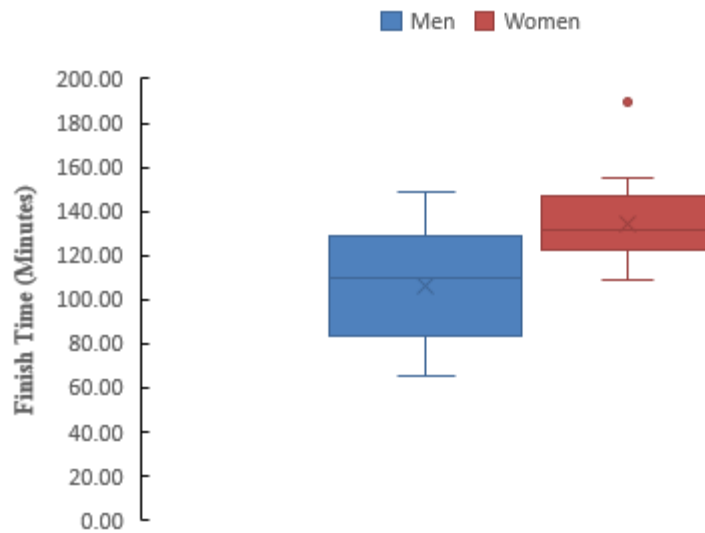
$$\text{Women: } IQR = 147.18 - 122.08 = 25.10$$

$$\text{Lower Limit} = Q_1 - 1.5(IQR) = 122.08 - 1.5(25.10) = 84.43$$

$$\text{Upper Limit} = Q_3 + 1.5(IQR) = 147.18 + 1.5(25.10) = 184.83$$

The two slowest women runners with times of 189.27 and 189.28 minutes are outliers in the women's group.

- e. A box plot created using Excel's Box and Whisker Statistical Chart follows.



The box plots show the men runners with the faster or lower finish times. However, the box plots show the women runners with the lower variation in finish times. The interquartile ranges of 45.9225 minutes for men and 25.10 minutes for women support this conclusion.

51. a. Smallest = 608

$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(21+1) = 5.5$$

First quartile or 25th percentile = $1850 + .5(1872 - 1850) = 1861$

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(21+1) = 11.0$$

Second quartile or median = 4019

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(21+1) = 16.5$$

Third quartile or 75th percentile = $8305 + .5(8408 - 8305) = 8356.5$

Largest = 14138

Five-number summary: 608, 1861, 4019, 8365.5, 14138

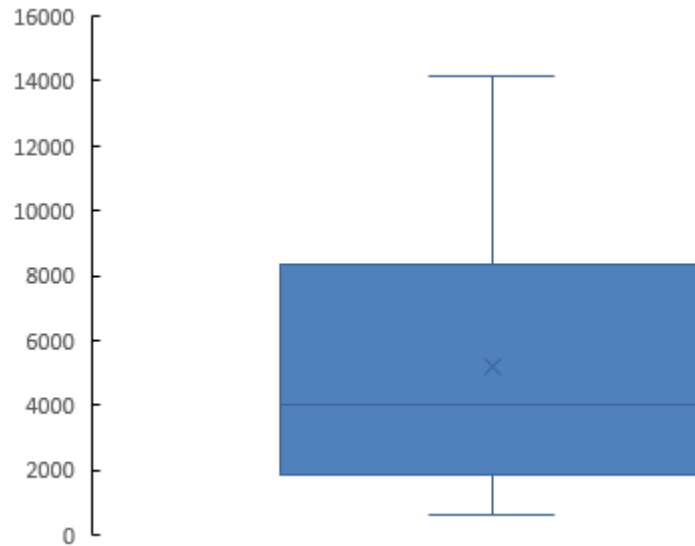
- b. Limits:

$$\text{IQR} = Q_3 - Q_1 = 8356.5 - 1861 = 6495.5$$

$$\text{Lower Limit: } Q_1 - 1.5(\text{IQR}) = 1861 - 1.5(6495.5) = -7,882.25$$

$$\text{Upper Limit: } Q_3 + 1.5(\text{IQR}) = 8356.5 + 1.5(6495.5) = 18,099.75$$

- c. There are no outliers, all data are within the limits.
- d. Yes, if the first two digits in Johnson and Johnson's sales were transposed to 41,138, sales would have shown up as an outlier. A review of the data would have enabled the correction of the data.
- e. A box plot created using Excel's Box and Whisker Statistical Chart follows.



52. Excel's MIN, QUARTILE.EXC, and MAX functions provided the following five-number summaries:

	AT&T	Sprint	T-Mobile	Verizon
Minimum	66	63	68	75
First Quartile	68	65	71.25	77
Median	71	66	73.5	78.5
Third Quartile	73	67.75	74.75	79.75
Maximum	75	69	77	81

- a. Median for T-Mobile is 73.5
- b. 5- number summary: 68, 71.25, 73.5, 74.75, 77
- c. $\text{IQR} = Q_3 - Q_1 = 74.75 - 71.25 = 3.5$

$$\text{Lower Limit} = Q_1 - 1.5(\text{IQR})$$

$$= 71.25 - 1.5(3.5) = 66$$

$$\text{Upper Limit} = Q_3 + 1.5(\text{IQR})$$

$$= 74.75 + 1.5(3.5) = 80$$

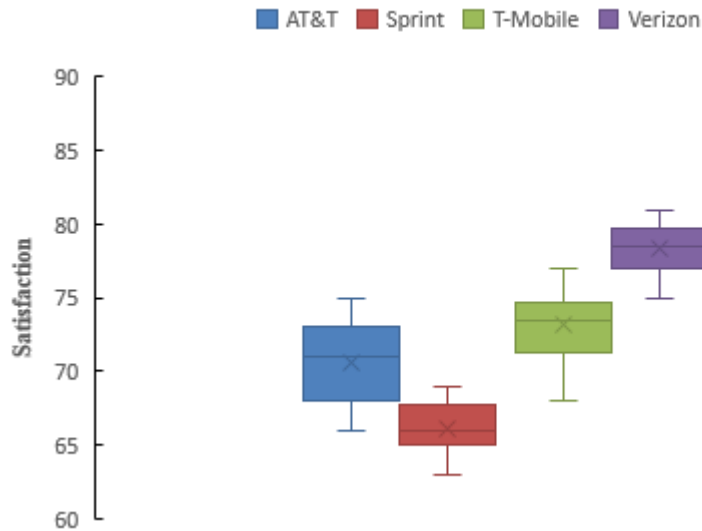
All ratings are between 66 and 80. There are no outliers for the T-Mobile service.

- d. Using the five number summaries shown initially, the limits for the four cell-phone services are as follows:

	AT&T	Sprint	T-Mobile	Verizon
Minimum	66	63	68	75
First Quartile	68	65	71.25	77
Median	71	66	73.5	78.5
Third Quartile	73	67.75	74.75	79.75
Maximum	75	69	77	81
IQR	5	2.75	3.5	2.75
1.5(IQR)	7.5	4.125	5.25	4.125
Lower Limit	60.5	60.875	66	72.875
Upper Limit	80.5	71.875	80	83.875

There are no outliers for any of the cell-phone services.

- e. A box plot created using Excel's Box and Whisker Statistical Chart follows.



The box plots show that Verizon is the best cell-phone service provider in terms of overall customer satisfaction. Verizon's lowest rating is better than the highest AT&T and Sprint ratings and is better than 75% of the T-Mobile ratings. Sprint shows the lowest customer satisfaction ratings among the four services.

53. a. The median one year – total return is 13.9% (obtained from average the 25th and 26th positions).

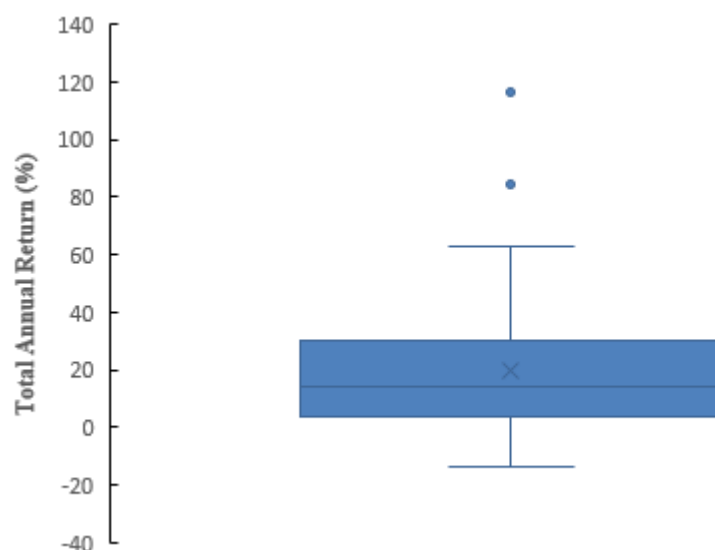
- b. 20 of the 50 companies or 40% had a one year – total return greater than the S&P average return.
- c. The five-number summary is -13.9, 3.7, 13.9, 30.0, 117.1
- d. Returns less than -35.75 or greater than 69.45 are considered outliers. Thus, the following observations are outliers:

Southwest Airlines: Return = 84.5

Delta Air Lines: Return = 116.6

Facebook: Return = 117.1

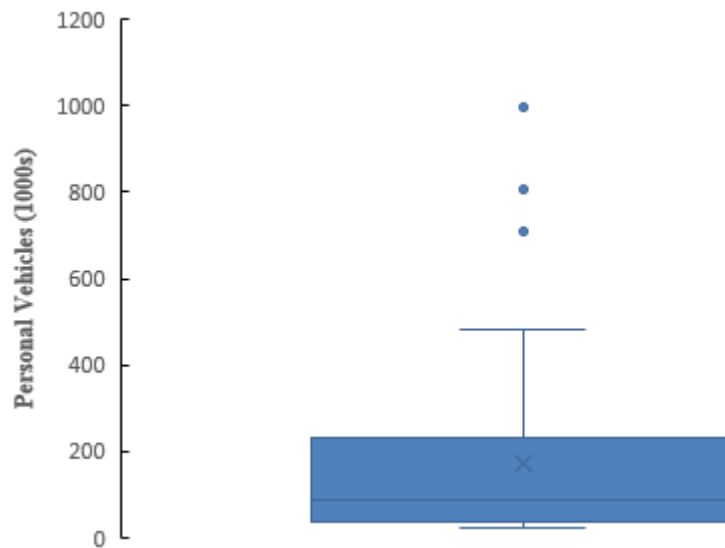
- e. A box plot created using Excel's Box and Whisker Statistical Chart follows.



54. Excel's AVERAGE, MIN, QUARTILE.EXC, and MAX functions provided the following results; values for IQR and the upper and lower limits are also shown.

Personal Vehicles (1000s)	
Mean	173.24
Minimum	21
First Quartile	38.5
Second Quartile	89.5
Third Quartile	232
Maximum	995
IQR	193.5
1.5(IQR)	290.25
Lower Limit	-251.75
Upper Limit	522.25

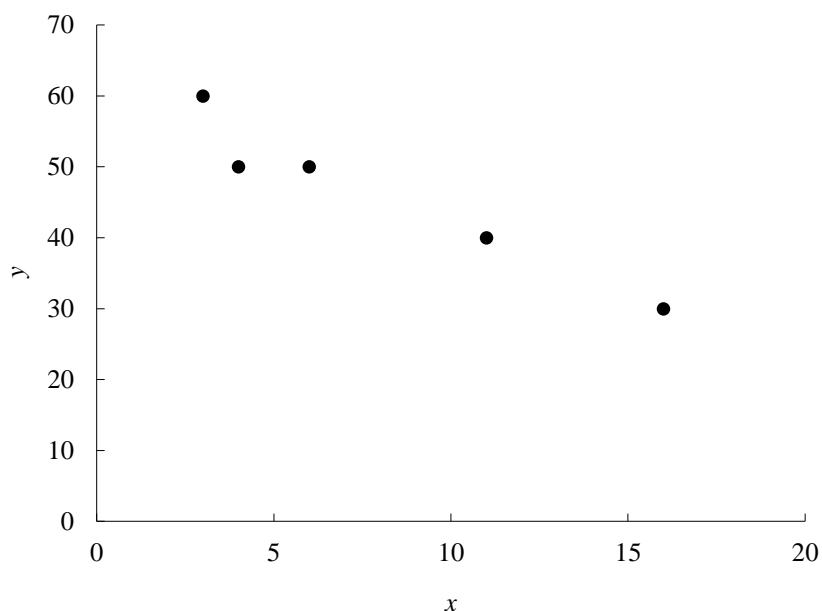
- Mean = 173.24 and median (second quartile) = 89.5
- First quartile = 38.5 and the third quartile = 232
- 21, 38.5, 89.5, 232, 995
- A box plot created using Excel's Box and Whisker Statistical Chart follows.



The box plot shows the distribution of number of personal vehicle crossings is skewed to the right (positive). Three ports of entry are considered outliers:

NY: Buffalo-Niagara Falls	707
TX: El Paso	807
CA: San Ysidro	995

55. a.



b. Negative relationship

$$\text{c/d. } \Sigma x_i = 40 \quad \bar{x} = \frac{40}{5} = 8 \quad \Sigma y_i = 230 \quad \bar{y} = \frac{230}{5} = 46$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -240 \quad \Sigma(x_i - \bar{x})^2 = 118 \quad \Sigma(y_i - \bar{y})^2 = 520$$

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-240}{5-1} = -60$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{118}{5-1}} = 5.4314$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{520}{5-1}} = 11.4018$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-60}{(5.4314)(11.4018)} = -.969$$

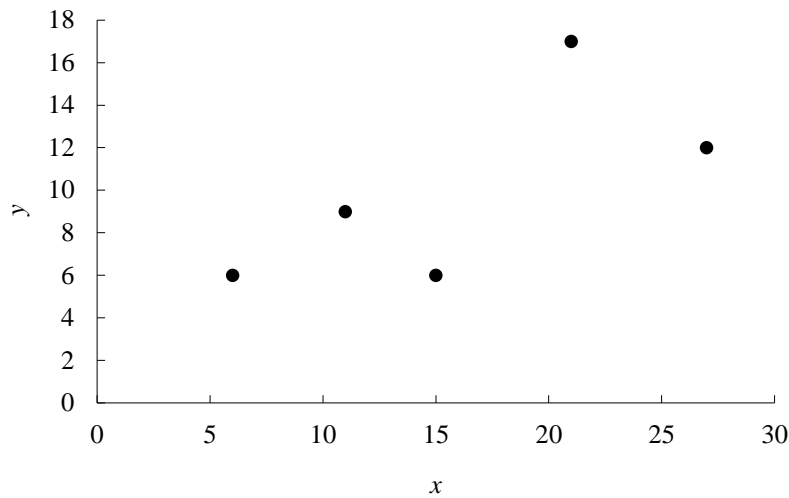
Sample covariance = -60

The negative value of the sample covariance indicates a negative linear relationship

Sample correlation coefficient = -.969

There is a strong negative linear relationship.

56. a.



b. Positive relationship

$$\text{c/d. } \Sigma x_i = 80 \quad \bar{x} = \frac{80}{5} = 16 \quad \Sigma y_i = 50 \quad \bar{y} = \frac{50}{5} = 10$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 106 \quad \Sigma(x_i - \bar{x})^2 = 272 \quad \Sigma(y_i - \bar{y})^2 = 86$$

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{106}{5-1} = 26.5$$

Sample covariance = 26.5

The positive value of the sample covariance indicates a positive linear relationship.

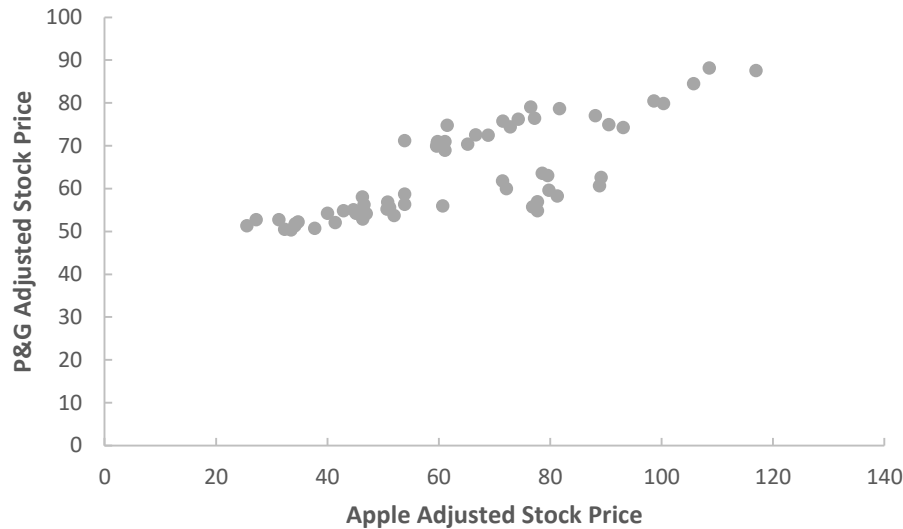
$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{272}{5-1}} = 8.2462$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{86}{5-1}} = 4.6368$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{26.5}{(8.2462)(4.6368)} = .693$$

Sample correlation coefficient = .693 which indicates a moderately strong positive linear relationship

57. a.



- b. The scatter diagram shows that there is a positive linear relationship (as one stock price increases, the other tends to increase as well).

- c. Let x = the Apple adjusted stock price and y = the P&G adjusted stock price

$$\Sigma x_i = 3810.19 \quad \bar{x} = \frac{3810.19}{60} = 63.50 \quad \Sigma y_i = 3825.59 \quad \bar{y} = \frac{3825.59}{60} = 63.76$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 11,118.99$$

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{11,118.99}{60-1} = 188.46$$

A positive covariance shows a positive linear relationship between adjusted stock price for Apple and the adjusted stock price for P&G.

- d.

$$\Sigma(x_i - \bar{x})^2 = 28,685.62 \quad \Sigma(y_i - \bar{y})^2 = 6,986.42$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{28685.62}{60-1}} = 22.05$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{6986.42}{60-1}} = 10.88$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{188.46}{(22.05)(10.88)} = .79$$

The strong positive correlation shows that the adjusted stock prices of Apple and P&G tend to move together. That is, when the adjusted stock price of Apple increases, so does the adjusted stock price of P&G. Because the adjusted stock prices for these two companies tend to increase and decrease together, holding only Apple and P&G stock in a portfolio could lead to a lot of volatility in the value of the portfolio. To minimize volatility, a financial analyst will often recommend acquiring at least some stocks that have negatively correlated adjusted prices.

58. Let x = miles per hour and y = miles per gallon

$$\Sigma x_i = 420 \quad \bar{x} = \frac{420}{10} = 42 \quad \Sigma y_i = 270 \quad \bar{y} = \frac{270}{10} = 27$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -475 \quad \Sigma(x_i - \bar{x})^2 = 1660 \quad \Sigma(y_i - \bar{y})^2 = 164$$

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-475}{10-1} = -52.7778$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1660}{10-1}} = 13.5810$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{164}{10-1}} = 4.2687$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-52.7778}{(13.5810)(4.2687)} = -.91$$

A strong negative linear relationship exists. For driving speeds between 25 and 60 miles per hour, higher speeds are associated with lower miles per gallon.

59. a. We would expect there would be a negative relationship between smoke detector use and deaths from home fires. That is, as more households have smoke detectors, warning of a fire would help the inhabitants of the home escape and lead to fewer deaths from home fires.

b.

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{14.28}{17} = .84 \quad \bar{y} = \frac{\Sigma y_i}{n} = \frac{257.4}{17} = 15.1412$$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
0.50	22.9	-0.34	7.7588	0.1156	60.1993	-2.6380
0.67	20.8	-0.17	5.6588	0.0289	32.0223	-0.9620
0.74	17.3	-0.10	2.1588	0.0100	4.6605	-0.2159

0.76	20.5	-0.08	5.3588	0.0064	28.7170	-0.4287
0.77	19.4	-0.07	4.2588	0.0049	18.1376	-0.2981
0.82	18.9	-0.02	3.7588	0.0004	14.1288	-0.0752
0.81	17.6	-0.03	2.4588	0.0009	6.0458	-0.0738
0.85	16.2	0.01	1.0588	0.0001	1.1211	0.0106
0.86	13.6	0.02	-1.5412	0.0004	2.3752	-0.0308
0.88	14.4	0.04	-0.7412	0.0016	0.5493	-0.0296
0.90	14.3	0.06	-0.8412	0.0036	0.7076	-0.0505
0.92	13.0	0.08	-2.1412	0.0064	4.5846	-0.1713
0.95	10.9	0.11	-4.2412	0.0121	17.9876	-0.4665
0.96	10.9	0.12	-4.2412	0.0144	17.9876	-0.5089
0.97	10.2	0.13	-4.9412	0.0169	24.4152	-0.6424
0.96	8.4	0.12	-6.7412	0.0144	45.4435	-0.8089
0.96	8.1	0.12	-7.0412	0.0144	49.5782	-0.8449
Total				0.2514	328.6612	-8.2350

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-8.2350}{17-1} = -.5147$$

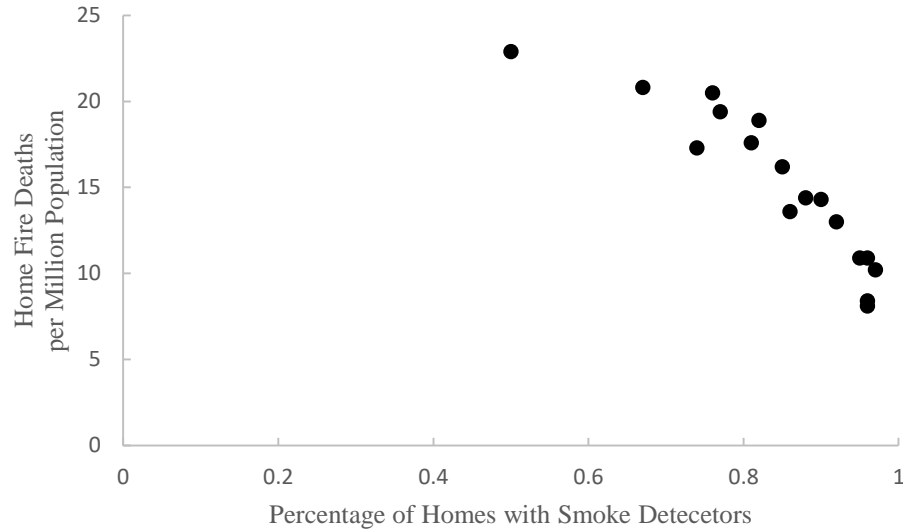
$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{.2514}{17-1}} = .1253$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{328.6612}{17-1}} = 4.5323$$

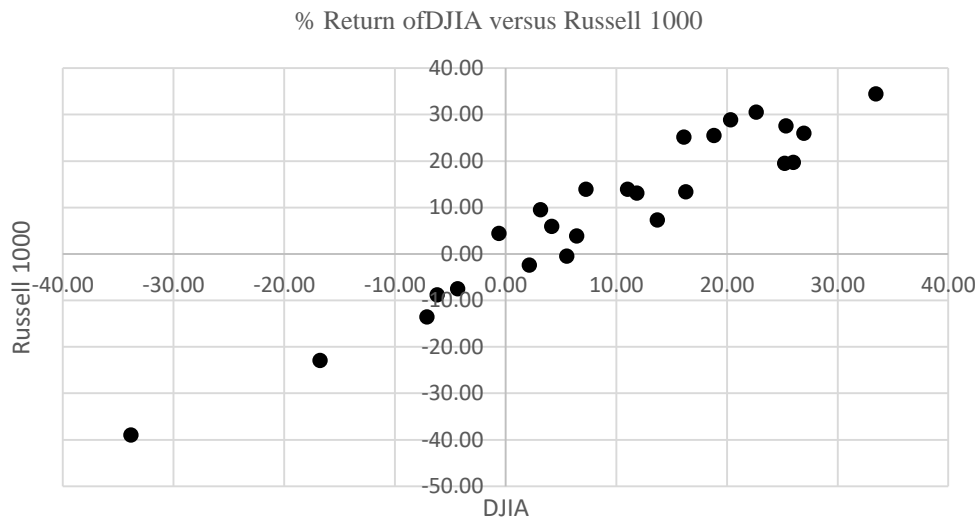
$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-.5147}{(.1253)(4.5323)} = -.9060$$

There is a strong negative correlation between smoke detector use and deaths from home fires.

c.



60. a.



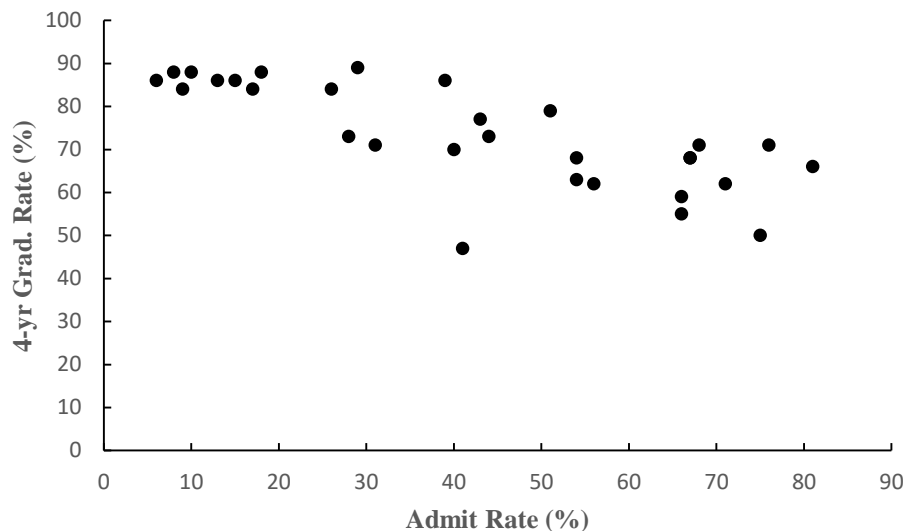
b. DJIA : $\bar{x} = \frac{\sum x_i}{n} = \frac{227.57}{25} = 9.10$ $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{5672.61}{24}} = 15.37$

Russell 1000: $\bar{x} = \frac{\sum x_i}{n} = \frac{227.29}{25} = 9.09$ $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{7679.81}{24}} = 17.89$

c. $r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{263.611}{(15.37)(17.89)} = .959$

- d. Based on this sample, the two indexes are very similar. They have a strong positive correlation. The variance of the Russell 1000 is slightly larger than that of the DJIA.

61. a.



There appears to be a negative linear relationship between the two variables. In other words, higher admission rates are associated with lower graduation rates.

- b. The sample correlation coefficient is -0.76 (which can be easily attained from the CORREL function in Excel). This indicates a negative linear relationship between the two variables.

62. The data in ascending order follow.

Position	Value	Position	Value
1	0	11	3
2	0	12	3
3	1	13	3
4	1	14	4
5	1	15	4
6	1	16	5
7	1	17	5
8	2	18	6
9	3	19	6
10	3	20	7

- a. The mean is 2.95 and the median is 3.

b.
$$L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(20+1) = 5.25$$

First quartile or 25th percentile = $1 + .25(1 - 1) = 1$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(20+1) = 15.75$$

Third quartile or 75th percentile = $4 + .75(5 - 4) = 4.75$

- c. The range is 7 and the interquartile range is $4.75 - 1 = 3.75$.
- d. The variance is 4.37 and standard deviation is 2.09.
- e. Because most people dine out a relatively few times per week and a few families dine out very frequently, we would expect the data to be positively skewed. The skewness measure of 0.34 indicates the data are somewhat skewed to the right.
- f. The lower limit is -4.625 and the upper limit is 10.375 . No values in the data are less than the lower limit or greater than the upper limit, so there are no outliers.
63. a. $L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(24) = 12$ (12th position)
- Previous Coach Median = 850,000
- New Coach Median = 1,150,000
- b. Range: Previous Coach Base: $3,500,000 - 267,800 = 3,232,200$
- New Coach Base: $3,200,000 - 280,000 = 2,920,000$
- c. Standard Deviation:
- Previous Coach Base: $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{2.2209 \times 10^{13}}{22}} = 1,004,740$
- New Coach Base: $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1.7894 \times 10^{13}}{22}} = 901,873$
- d. The new coaches have a higher median annual salary, but a smaller range and standard deviation.
64. a. The mean and median patient wait times for offices with a wait tracking system are 17.2 and 13.5, respectively. The mean and median patient wait times for offices without a wait tracking system are 29.1 and 23.5, respectively.
- b. The variance and standard deviation of patient wait times for offices with a wait tracking system are 86.2 and 9.3, respectively. The variance and standard deviation of patient wait times for offices without a wait tracking system are 275.7 and 16.6, respectively.
- c. Offices with a wait tracking system have substantially shorter patient wait times than offices without a wait tracking system.
- d. $z = \frac{37 - 29.1}{16.6} = 0.48$

e. $z = \frac{37 - 17.2}{9.3} = 2.13$

As indicated by the positive z -scores, both patients had wait times that exceeded the means of their respective samples. Even though the patients had the same wait time, the z -score for the sixth patient in the sample who visited an office with a wait tracking system is much larger because that patient is part of a sample with a smaller mean and a smaller standard deviation.

- f. The z -scores for all patients follow.

Without Wait Tracking System	With Wait Tracking System
-0.31	1.49
2.28	-0.67
-0.73	-0.34
-0.55	0.09
0.11	-0.56
0.90	2.13
-1.03	-0.88
-0.37	-0.45
-0.79	-0.56
0.48	-0.24

The z -scores do not indicate the existence of any outliers in either sample.

65. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{148}{20} = 7.4$

b. $s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)} = \frac{58.8}{19} = 3.09$ $s = \sqrt{3.09} = 1.76$

66. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{20665}{50} = 413.3$ This is slightly higher than the mean for the study.

b. $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{69424.5}{49}} = 37.64$

c. $L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(51) = 12.75$ First Quartile or $Q_1 = 374 + .75(384-374) = 381.5$

$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(51) = 38.25$ Third Quartile or $Q_3 = 445 + .25(445-445) = 445$

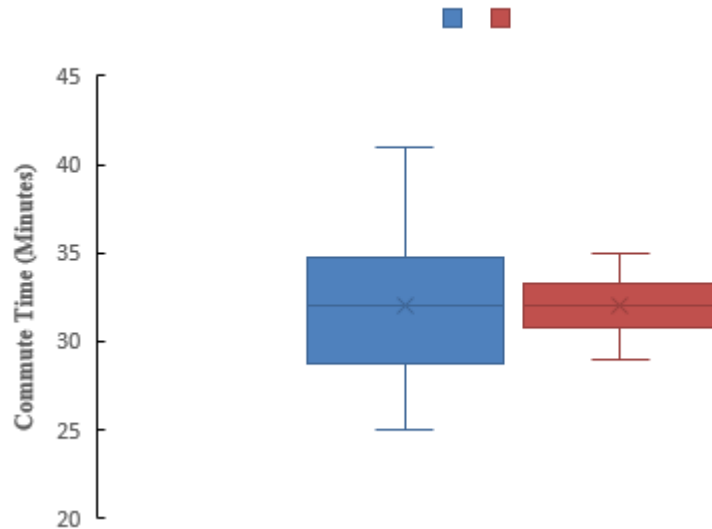
$IQR = 445 - 381.5 = 63.5$

$LL = Q_1 - 1.5 IQR = 381.5 - 1.5(63.5) = 286.25$

$$UL = Q_3 + 1.5 IQR = 445 + 1.5(63.5) = 540.25$$

There are no outliers.

67. a. Public Transportation: $\bar{x} = \frac{320}{10} = 32$
 Automobile: $\bar{x} = \frac{320}{10} = 32$
- b. Public Transportation: $s = 4.64$
 Automobile: $s = 1.83$
- c. Prefer the automobile. The mean times are the same, but the auto has less variability.
- d. A box plot created using Excel's Box and Whisker Statistical Chart follows.



The box plot shows lower variability with automobile transportation and supports the conclusion in part c.

68. Excel's MIN, QUARTILE.EXC, and MAX functions provided the following results; values for the IQR and the upper and lower limits are also shown.

Annual Household Income	
Minimum	46.5
First Quartile	50.75
Second Quartile	52.1
Third Quartile	52.6
Maximum	64.5
IQR	1.85

1.5(IQR)	2.775
Lower Limit	47.975
Upper Limit	55.375

- a. The data in ascending order follow:

46.5 48.7 49.4 51.2 51.3 51.6 52.1 52.1 52.2 52.4 52.5 52.9 53.4 64.5

$$L_{50} = \frac{p}{100}(n+1) = \frac{50}{100}(14+1) = 7.5$$

$$\text{Median or 50th percentile} = 52.1 + .5(52.1 - 52.1) = 52.1$$

b. Percentage change = $\left(\frac{52.1 - 55.5}{55.5} \right) 100 = -6.1\%$

c. $L_{25} = \frac{p}{100}(n+1) = \frac{25}{100}(14+1) = 3.75$

$$25\text{th percentile} = 49.4 + .75(51.2 - 49.4) = 50.75$$

$$L_{75} = \frac{p}{100}(n+1) = \frac{75}{100}(14+1) = 11.25$$

$$75\text{th percentile} = 52.5 + .25(52.9 - 52.5) = 52.6$$

d. 46.5 50.75 52.1 52.6 64.5

e. $\bar{x} = \frac{\sum x_i}{n} = \frac{730.8}{14} = 52.2$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{208.12}{13} = 16.0092$$

$$s = \sqrt{16.0092} = 4.0012$$

The z-scores = $\frac{x_i - \bar{x}}{s}$ are shown below:

-1.42 -0.87 -0.70 -0.25 -0.22 -0.15 -0.02 -0.02 0.00 0.05 0.07 0.17 0.30 3.07

The last household income (64.5) has a z-score > 3 and is an outlier.

$$\text{Lower Limit} = Q_1 - 1.5(\text{IQR}) = 50.75 - 1.5(52.6 - 50.75) = 47.98$$

$$\text{Upper Limit} = Q_3 + 1.5(\text{IQR}) = 52.6 + 1.5(52.6 - 50.75) = 55.38$$

Using this approach the first observation (46.5) and the last observation (64.5) would be consider outliers.

The two approaches will not always provide the same results.

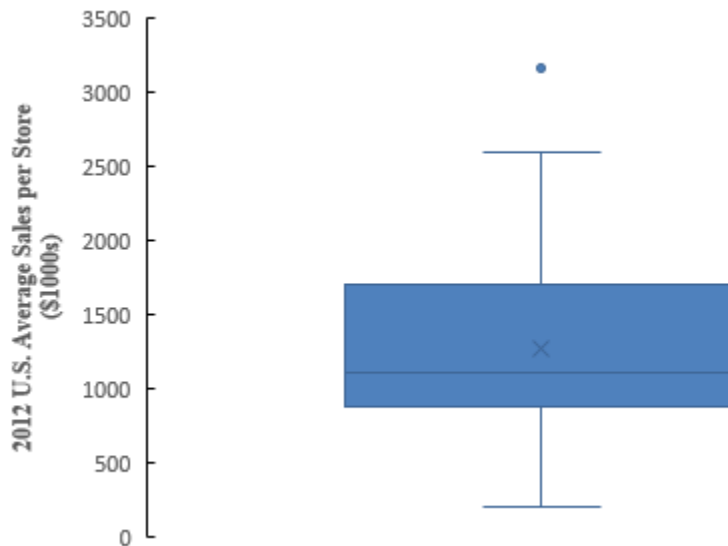
69. Excel's MIN, QUARTILE.EXC, and MAX functions provided the following results; values for the IQR and the upper and lower limits are also shown.

2012 U.S. Average Sales per Store (\$1000s)	
Minimum	207
First Quartile	879
Second Quartile	1103
Third Quartile	1700
Maximum	3158
IQR	821
1.5(IQR)	1231.5
Lower Limit	-352.5
Upper Limit	2931.5

- a. Mean = 1275.2
- b. First quartile = 879 and the third quartile = 1700

25% of the restaurants have an average sales per unit less than or equal to 879 and 25% of the restaurants have an average sales per unit greater than or equal to 1700.

- c. A box plot created using Excel's Box and Whisker Statistical Chart follows..



There is one outlier corresponding to Chick-fil-A.

d. The table follows:

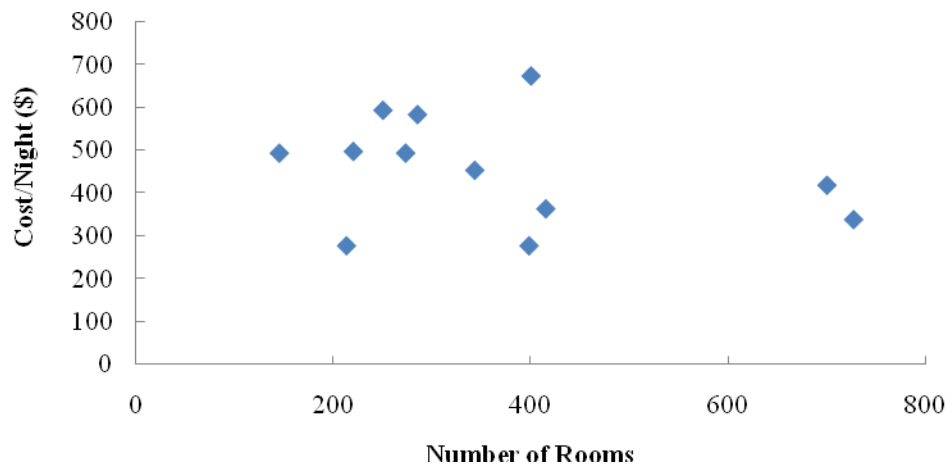
Segment	Average Sales per unit (\$1000s)
Burger	1440
Chicken	1457
Mexican	1329
Pizza/Pasta	730
Sandwich	1280
Snacks	1088

The table shows that the lowest average sales per unit is for the Pizza/Pasta segment.

70. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{4368}{12} = 364$ rooms

b. $\bar{y} = \frac{\sum y_i}{n} = \frac{5484}{12} = \457

c.



It is difficult to see much of a relationship. When the number of rooms becomes larger, there is no indication that the cost per night increases. The cost per night may even decrease slightly.

d.

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
220	499	-144	42	20,736	1,764	-6,048
727	340	363	-117	131,769	13,689	-42,471
285	585	-79	128	6,241	16,384	-10,112
273	495	-91	38	8,281	1,444	-3,458
145	495	-219	38	47,961	1,444	-8,322
213	279	-151	-178	22,801	31,684	26,878
398	279	34	-178	1,156	31,684	-6,052
343	455	-21	-2	441	4	42
250	595	-114	138	12,996	19,044	-15,732
414	367	50	-90	2,500	8,100	-4,500
400	675	36	218	1,296	47,524	7,848
700	420	336	-37	<u>112,896</u>	<u>1,369</u>	<u>-12,432</u>
Total				369,074	174,134	-74,359

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-74,359}{11} = -6759.91$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{369,074}{11}} = 183.17$$

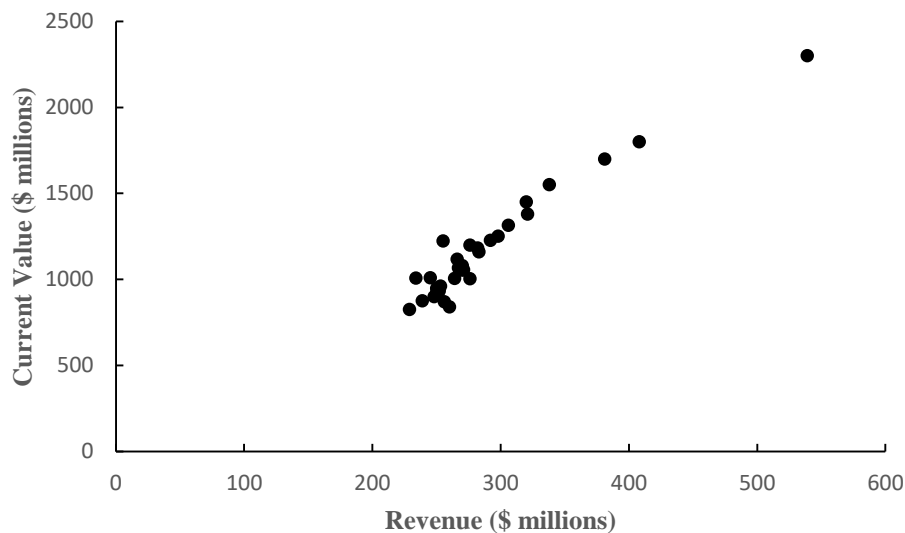
$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{174,134}{11}} = 125.82$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-6759.91}{(183.17)(125.82)} = -.293$$

There is evidence of a slightly negative linear association between the number of rooms and the cost per night for a double room. Although this is not a strong relationship, it suggests that the higher room rates tend to be associated with the smaller hotels.

This tends to make sense when you think about the economies of scale for the larger hotels. Many of the amenities in terms of pools, equipment, spas, restaurants, and so on exist for all hotels in the *Travel + Leisure* top 50 hotels in the world. The smaller hotels tend to charge more for the rooms. The larger hotels can spread their fixed costs over many room and may actually be able to charge less per night and still achieve a nice profit. The larger hotels may also charge slightly less in an effort to obtain a higher occupancy rate. In any case, it appears that there is a slightly negative linear association between the number of rooms and the cost per night for a double room at the top hotels.

71. a.



There appears to be a positive linear relationship between two variables.

- b. The sample correlation coefficient is .96 (which can be easily attained from the CORREL function in Excel). This indicates a very strong linear relationship between the two variables.

72. a.

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
.407	.422	-.1458	-.0881	.0213	.0078	.0128
.429	.586	-.1238	.0759	.0153	.0058	-.0094
.417	.546	-.1358	.0359	.0184	.0013	-.0049
.569	.500	.0162	-.0101	.0003	.0001	-.0002
.569	.457	.0162	-.0531	.0003	.0028	-.0009
.533	.463	-.0198	-.0471	.0004	.0022	.0009
.724	.617	.1712	.1069	.0293	.0114	.0183
.500	.540	-.0528	.0299	.0028	.0009	-.0016
.577	.549	.0242	.0389	.0006	.0015	.0009
.692	.466	.1392	-.0441	.0194	.0019	-.0061
.500	.377	-.0528	-.1331	.0028	.0177	.0070
.731	.599	.1782	.0889	.0318	.0079	.0158
.643	.488	.0902	-.0221	.0081	.0005	-.0020
.448	.531	-.1048	.0209	.0110	.0004	-.0022
Total				.1617	.0623	.0287

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{.0287}{14-1} = .0022$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{.1617}{14-1}} = .1115$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{.0623}{14-1}} = .0692$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{.0022}{.1115(.0692)} = +.286$$

- b. There is a low positive correlation between a major league baseball team's winning percentage during spring training and its winning percentage during the regular season. The spring training record should not be expected to be a good indicator of how a team will play during the regular season.

Spring training consists of practice games between teams with the outcome as to who wins or who loses not counting in the regular season standings or affecting the chances of making the playoffs. Teams use spring training to help players regain their timing and evaluate new players. Substitutions are frequent with the regular or better players rarely playing an entire spring training game. Winning is not the primary goal in spring training games. A low correlation between spring training winning percentage and regular season winning percentage should be anticipated.

$$73. \quad \bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{20(20) + 30(12) + 10(7) + 15(5) + 10(6)}{20 + 30 + 10 + 15 + 10} = \frac{965}{85} = 11.4 \text{ days}$$

74.

w_i	x_i	$w_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$w_i(x_i - \bar{x})^2$	$(x_i - \bar{x})^2$	$w_i(x_i - \bar{x})^2$
10	47	470	-13.68	187.1424	1871.42	187.26	1872.58
40	52	2080	-8.68	75.3424	3013.70	75.42	3016.62
150	57	8550	-3.68	13.5424	2031.36	13.57	2036.01
175	62	10850	+1.32	1.7424	304.92	1.73	302.98
75	67	5025	+6.32	39.9424	2995.68	39.89	2991.69
15	72	1080	+11.32	128.1424	1922.14	128.05	1920.71
<u>10</u>	<u>77</u>	<u>770</u>	<u>+16.32</u>	<u>266.3424</u>	<u>2663.42</u>	<u>266.20</u>	<u>2662.05</u>
475		28,825			14,802.64		14802.63

Columns 5 and 6 are calculated with rounding, while columns 7 and 8 are based on unrounded calculations.

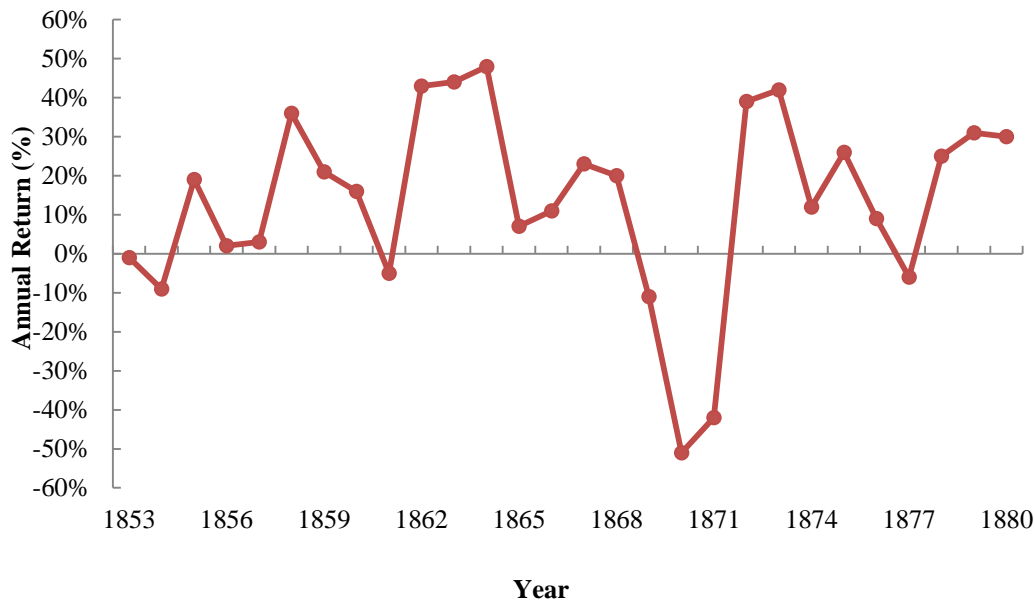
$$a. \quad \bar{x} = \frac{28,825}{475} = 60.68$$

$$b. \quad s^2 = \frac{14,802.64}{474} = 31.23$$

$$s = \sqrt{31.23} = 5.59$$

75. a.

Chapter 3



It appears the Panama Railroad Company outperformed the New York Stock Exchange annual average return of 8.4%, but the large drop in returns in 1870-71 makes it difficult to be certain.

- b. The geometric mean is

$$\bar{x}_g = \sqrt[28]{(x_1)(x_2) \cdots (x_{28})} = \sqrt[28]{16.769} = 1.106$$

where each x represents each year's growth factor and is equal to $1 +$ the return on stock for that year. Therefore:

$x_1 = 1 + (-1\%) = 99\% = .99$ for year 1853
 $x_2 = 1 + (-9\%) = 91\% = .91$ for year 1854
 $x_3 = 1 + 19\% = 119\% = 1.19$ for year 1855
 $x_4 = 1 + 2\% = 102\% = 1.02$ for year 1856
 etc

So the mean annual return on Panama Railroad Company stock is 10.6%. During the period of 1853–1880, the Panama Railroad Company stock yielded a return superior to the 8.4% earned by the New York Stock Exchange.

Note that we could also calculate the geometric mean with Excel. If the growth factors for the individual years are in cells C2:C29, then typing =GEOMEAN(C2:C29) into an empty cell will yield the geometric mean.