

1. Which statistic is more resistant to outliers (or extreme data values)?

A) Mean B) Median

Ans: B Difficulty: Medium L.O.: 2.2.4; 2.3.6

2. Which statistic is more resistant to outliers (or extreme data values)?

A) Interquartile Range B) Standard Deviation

Ans: A Difficulty: Medium L.O.: 2.3.6

Use the following to answer questions 3-8:

February 12, 2009 marked the 200th anniversary of Charles Darwin's birth. To celebrate, Gallup, a national polling organization, surveyed 1,018 Americans about their education level and their beliefs about the theory of evolution. The survey results are displayed in the provided two-way table.

	High School or Less	Some College	College Graduate	Postgraduate	Total
Believe	80	133	121	63	397
Do Not Believe	103	94	48	9	254
No Opinion	197	98	59	13	367
Total	380	325	228	85	1,018

3. What proportion of respondents have a college degree? Round your answer to three decimal places.

Ans: 0.224

Difficulty: Easy L.O.: 2.1.4

4. What proportion of respondents have no opinion on the theory of evolution? Round your answer to two decimal places.

Ans: 0.36

Difficulty: Easy L.O.: 2.1.4

5. What proportion of non-believers have a high school education or less? Use four decimal places in your answer.

Ans: 0.4055

Difficulty: Medium L.O.: 2.1.4

6. What proportion of college graduates believe in the theory of evolution? Use four decimal places in your answer.

Ans: 0.5307

Difficulty: Medium L.O.: 2.1.4

7. Find the proportion of respondents who believe in evolution for each education level (round each to three decimal places). Does there seem to be an association between education level and belief in evolution? If so, in what direction?

Ans: Proportion of respondents who believe in evolution for each education level:

HS or less: 0.211

Some College: 0.409

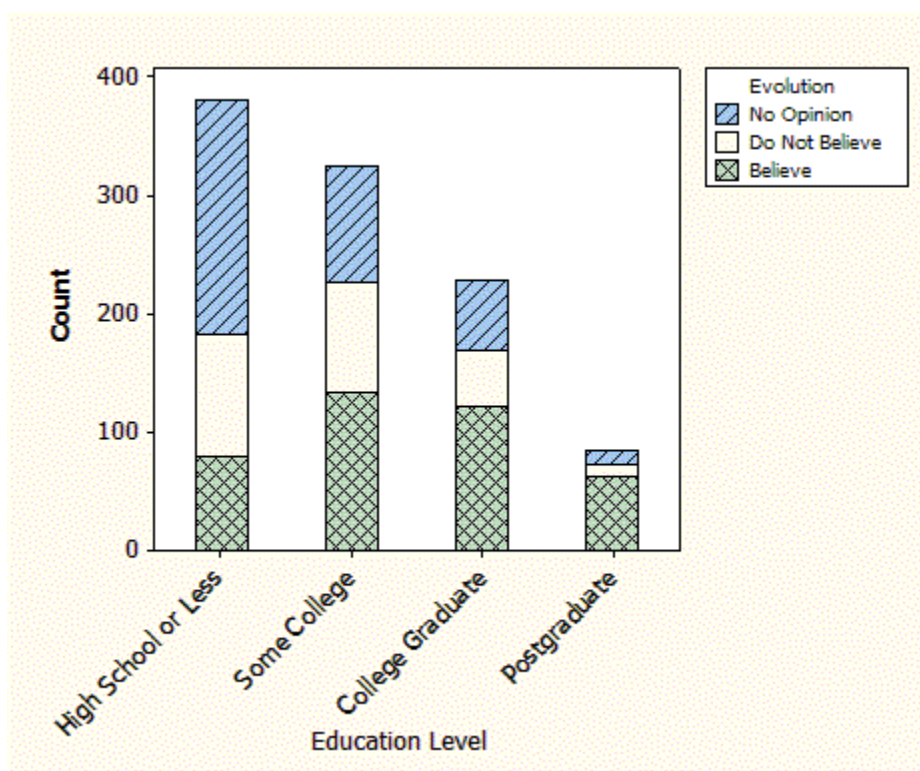
College Graduate: 0.531

Postgraduate: 0.741

There does seem to be an association between education level and belief in evolution. Individuals with more education are more likely to believe in evolution.

Difficulty: Challenging L.O.: 2.1.4

8. The survey results are displayed in the segmented bar chart. Does there appear to be an association between education level and belief in the theory of evolution? If so, what does it mean about these two variables?



Ans: Individuals with a high school education or less are least likely to believe in the theory of evolution and are also most likely to have no opinion. The percentage of people who believe in the theory of evolution increases as education level increases, and the percentage with no opinion decreases as education level increases.

Students might also notice that there are more respondents with a high school education or less, and that the number of respondents in each education category decreases as education level increases, however, these observations, though useful and informative, are not about the relationship between the two variables.

Difficulty: Challenging L.O.: 2.1.5

Use the following to answer questions 9-12:

A Fun-Size bag of M&M's contains 5 green, 3 red, 3 yellow, 4 orange, 4 blue, and 2 brown candies.

9. What proportion of the M&M's are green? Use four decimal places in your answer.

Ans: 0.2381

Difficulty: Easy L.O.: 2.1.2

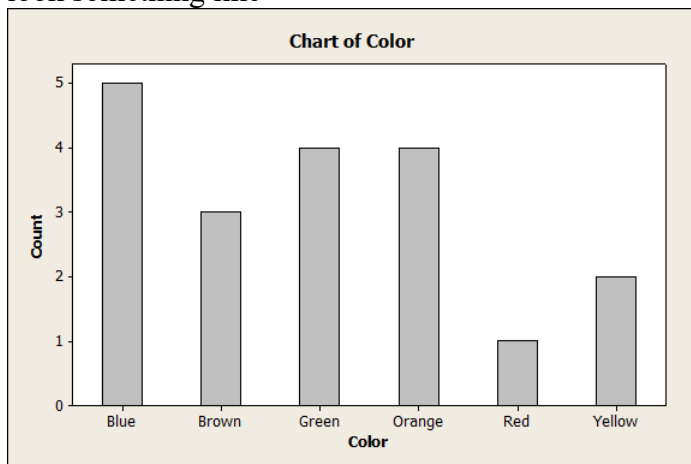
10. What proportion of the candies are yellow or orange? Use four decimal places in your answer.

Ans: 0.3333

Difficulty: Easy L.O.: 2.1.2

11. Sketch a bar chart of the data.

Ans: Will depend on values of the values randomly chosen for each color, but should look something like



Difficulty: Easy L.O.: 2.1.1

12. Construct a relative frequency table of the results. Use two decimal places in your relative frequencies.

Ans:

Color	Proportion
Green	0.24
Red	0.14
Yellow	0.14
Orange	0.19
Blue	0.19
Brown	0.10
Total	1.00

Difficulty: Easy L.O.: 2.1.1

Use the following to answer questions 13-19:

In a survey conducted by the Gallup organization September 6-9, 2012, 1,017 adults were asked "In general, how much trust and confidence do you have in the mass media - such as newspapers, TV, and radio - when it comes to reporting the news fully, accurately, and fairly?" 81 said that they had a "great deal" of confidence, 325 said they had a "fair amount" of confidence, 397 said they had "not very much" confidence, and 214 said they had "no confidence at all".

13. Display the results in a frequency table.

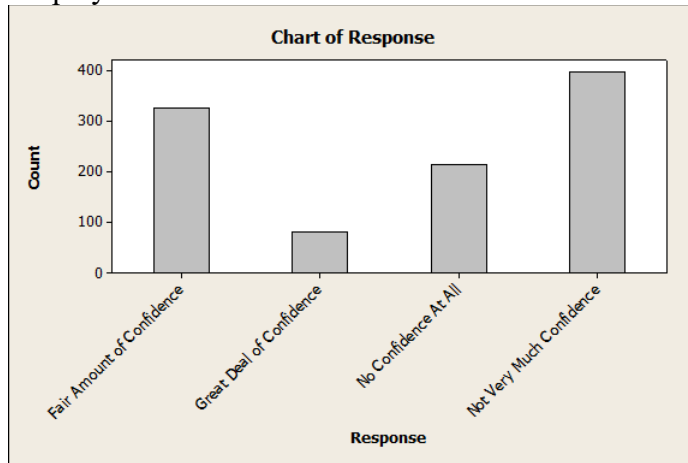
Ans:

Response	Count
Great Deal of Confidence	81
Fair Amount of Confidence	325
Not Very Much Confidence	397
No Confidence At All	214
Total	1,017

Difficulty: Easy L.O.: 2.1.1

14. Sketch a bar chart of the data.

Ans: Displayed below



Difficulty: Easy L.O.: 2.1.1

15. Give a relative frequency table of the data. Use two decimal places in your relative frequencies.

Ans:

Response	Proportion
Great Deal of Confidence	0.08
Fair Amount of Confidence	0.32
Not Very Much Confidence	0.39
No Confidence At All	0.21
Total	1.00

Difficulty: Easy L.O.: 2.1.1

16. What proportion of respondents have a fair amount of confidence in the media? Use two decimal places in your answer.

Ans: 0.32

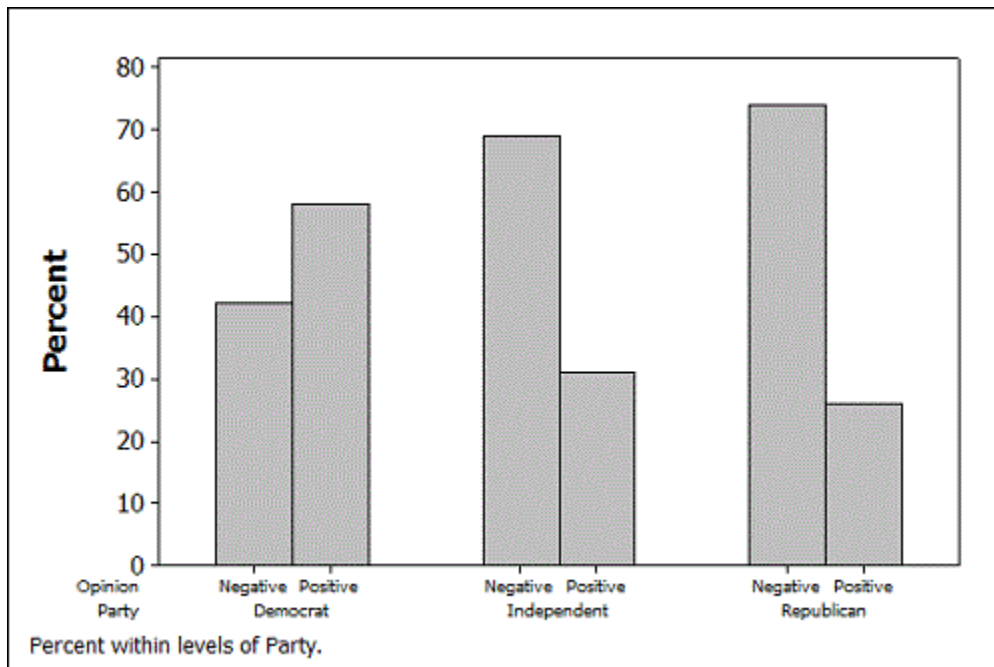
Difficulty: Easy L.O.: 2.1.2

17. What proportion of respondents have a negative opinion (not very much confidence/none at all) about the mass media? Use two decimal places in your answer.

Ans: 0.60

Difficulty: Medium L.O.: 2.1.2

18. In the article, they discuss the association between political party and opinion about the media ("positive" = great deal/fair amount of confidence and "negative" = not very much confidence/none at all). The results, as percentages, are displayed in the side-by-side bar charts. Describe the association between political party and opinion about the media. Use the bar chart to estimate the proportion of individuals with positive opinions about the media for each political party.



Ans: Democrats are more likely to have a positive opinion about the media, while the majority of Republicans and Independents have negative opinions about the media. The Republicans are the group that are least likely to have a positive opinion about the media. About 58% of Democrats have a positive opinion about the media, while roughly 31% of Independents and 26% of Republicans have positive opinions about the media (*answers may vary slightly*).

Difficulty: Medium L.O.: 2.1.5

19. In the same article, they compare opinions about the media in 2012 to those in the previous election year (2008). In 2008, 60% of Democrats, 27% of Republicans, and 41% of Independents had positive opinions (great deal/fair amount of confidence) about the media. For each political party, find the difference in the proportion of positive opinions in 2008 and the proportion of positive opinions in 2012. Comment on if/how opinions have changed for the political parties over the past four years.

Ans: *Differences may vary slightly as they are estimating the 2012 proportions from the bar chart.*

Democrats: $\hat{p}_{2008} - \hat{p}_{2012} = 0.60 - 0.58 = 0.02$

Republicans: $\hat{p}_{2008} - \hat{p}_{2012} = 0.27 - 0.26 = 0.01$

Independents: $\hat{p}_{2008} - \hat{p}_{2012} = 0.41 - 0.31 = 0.10$

The proportion of people with positive views on the media have decreased over the last four years for all parties. The Democrats and Republicans saw only small changes in this proportion, 0.02 and 0.01 respectively. The Independents had the largest change in the proportion of people with positive views on the media, with a decrease of 10% (0.10).

Difficulty: Medium L.O.: 2.1.1

Use the following to answer questions 20-22:

In a 2012 survey, Gallup asked a sample of U.S. adults if they would prefer to have a job outside the home, or if they would prefer to stay home to care for the family and home. Partial results for the individuals who expressed a preference, broken down by gender, are displayed in the two-way table.

	Job Outside of the Home	Stay at Home	Total
Male	391	???	504
Female	254	219	473
Total	645	332	977

20. Find the number of males who would prefer to stay at home.

Ans: $504 - 391 = 113$ or $332 - 219 = 113$

Difficulty: Medium L.O.: 2.1.3

21. What proportion of respondents would prefer a job outside of the home? Round your answer to two decimal places.

Ans: 0.66

Difficulty: Easy L.O.: 2.1.4

22. Compute the difference in the proportion of men who would prefer a job outside of the home and the proportion of females who would prefer a job outside of the home. Use two decimal places in your answer.

Ans: $\hat{p}_M - \hat{p}_F = 0.78 - 0.54 = 0.24$

Difficulty: Easy L.O.: 2.1.4

23. Students in a small statistics class were asked which was their dominant hand and if they were in a STEM (science, technology, engineering, and math) major. Their results are listed below. Use the results to construct a two-way table.

Student	Hand	STEM?	Student	Hand	STEM?
1	Right	No	9	Right	Yes
2	Right	Yes	10	Left	No
3	Left	Yes	11	Right	Yes
4	Right	Yes	12	Right	Yes
5	Right	No	13	Right	No
6	No	No	14	Right	No
7	Right	No	15	Right	Yes
8	Left	No	16	Right	No

Ans:

	Right	Left	Total
STEM, Yes	6	1	7
STEM, No	7	2	9
Total	13	3	16

Difficulty: Medium L.O.: 2.1.3

Use the following to answer questions 24-29:

Use this dataset to compute each of the following summary statistics.

2 26 16 10 7 20 17

24. Median

Ans: 16

Difficulty: Easy L.O.: 2.2.2

25. Mean (rounded to two decimal places)

Ans: 14.00

Difficulty: Easy L.O.: 2.2.2

26. Q1

Ans: 7

Difficulty: Easy L.O.: 2.3.1

27. Q3

Ans: 20

Difficulty: Easy L.O.: 2.3.1

28. IQR

Ans: 13

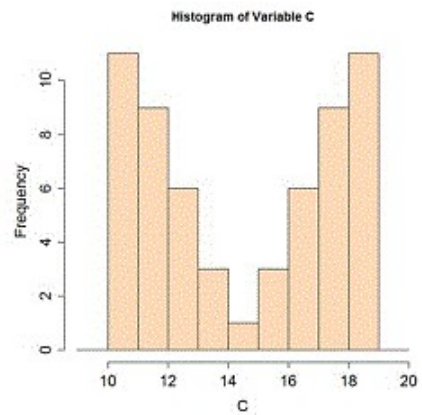
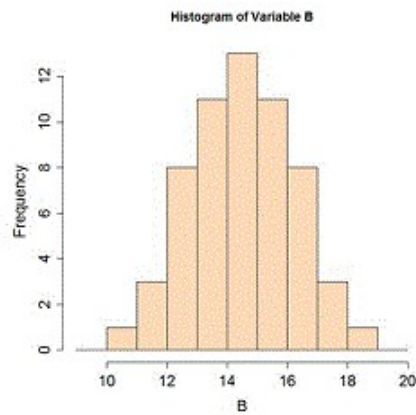
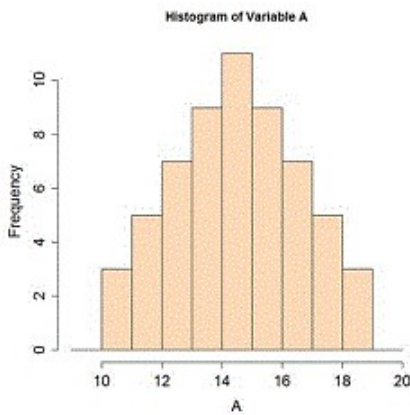
Difficulty: Easy L.O.: 2.3.5

29. Range

Ans: 24

Difficulty: Easy L.O.: 2.3.5

30. Each of the variables displayed in the histograms below has a mean of 14.5, a range of 8, and 59 observations. Rank the three variables according to their standard deviations, from the smallest to the largest.



A) A, B, C B) B, A, C C) C, B, A D) A, C, B

Ans: B Difficulty: Challenging L.O.: 2.3.2

Use the following to answer questions 31-34:

One of the symptoms of the flu is an elevated pulse rate. Pulse rates (in beats per minute) for $n = 23$ patients with the flu are provided.

75	80	81	82	82	83	84	85	86	88	88	90
90	90	91	92	93	93	95	97	99	101	110	

31. Give the sample mean pulse rate. Use two decimal places in your answer.

Ans: 89.35 beats per minute

Difficulty: Easy L.O.: 2.2.2

32. Find the standard deviation of the pulse rates. Use two decimal places in your answer.

Ans: 7.85 beats per minute

Difficulty: Medium L.O.: 2.3.1

33. Give the five-number summary of these pulse rates.

Ans: (Answers may vary slightly due to software used; R was used to compute these summaries)

Min.	1st Qu.	Median	3rd Qu.	Max.
75.00 bpm	83.50 bpm	90.00 bpm	93.00 bpm	110.00 bpm

bpm = beats per minute

Summaries from Minitab:

Min.	1st Qu.	Median	3rd Qu.	Max.
75.00 bpm	83.00 bpm	90.00 bpm	93.00 bpm	110.00 bpm

bpm = beats per minute

Difficulty: Easy L.O.: 2.3.1; 2.3.4

34. Are there any outliers? If so, which data points? Clearly show your work to justify your answer.

Ans: *Answers for IQR could vary slightly, depending on method/software used. Regardless, 110 should be the only outlier.*

Using Q1 and Q3 found in R:

Use the 1.5IQR rule to detect outliers.

$$\text{IQR} = 93 - 83.50 = 9.5$$

$$1.5\text{IQR} = 14.25$$

An observation is an outlier if it is smaller than

$$83.5 - 14.25 = 69.25 \text{ bpm}$$

or larger than

$$93 + 14.25 = 107.25 \text{ bpm.}$$

There is only one outlier, 110 bpm.

Difficulty: Medium L.O.: 2.4.1

Use the following to answer questions 35-36:

The distribution of waiting times at the student health center is bell-shaped with a mean of 10 minutes and a standard deviation of 3.

35. Give an interval that is likely to contain about 95% of wait times.

Ans: 4 to 16 minutes

Difficulty: Medium L.O.: 2.3.2

36. Find and interpret the z-score of someone who waits 5 minutes. Round your z-score to two decimal places. Be sure to specifically indicate if a wait time of 5 minutes is unusual.
Ans: -1.67

Interpretations will vary:

An individual who waits 5 minutes has a wait time that is 1.666667 standard deviations below the mean wait time.

If the z-score is less than two standard deviations below the mean, the answer should indicate that this wait time is not unusual.

If the z-score is more than two standard deviations below the mean, the answer should indicate that this wait time is unusual.

Difficulty: Medium L.O.: 2.3.3

Use the following to answer questions 37-46:

Identify which graphical display might be appropriate in each case.

The graphs you may choose from are:

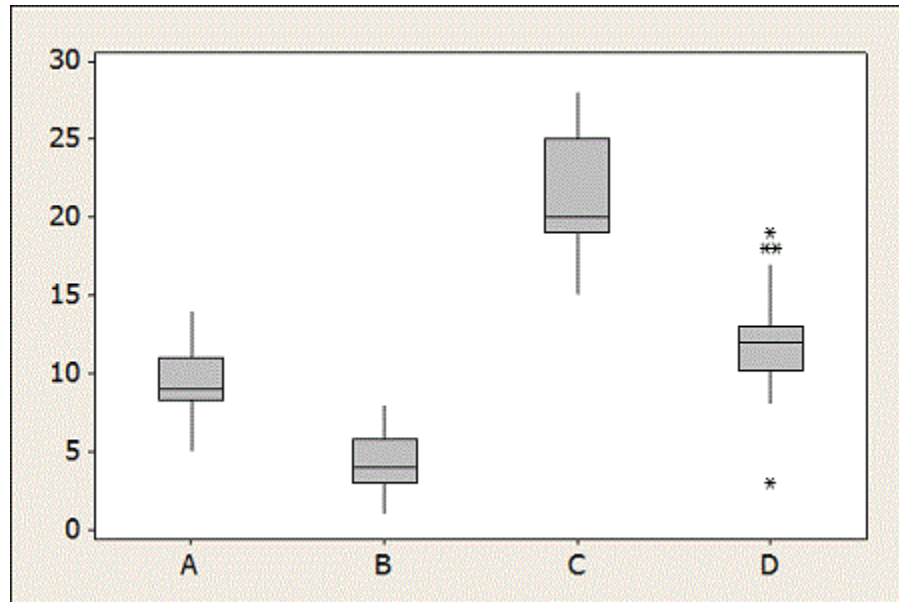
- Bar chart
- Pie chart
- Side-by-side bar chart
- Segmented bar chart
- Histogram
- Dotplot
- Side-by-side boxplots
- Scatterplot

37. Investigate the number of Facebook friends students in your class have.
Ans: Histogram or dotplot
Difficulty: Medium L.O.: 2.2.0
38. Compare the number of points scored for all games played in a season for all football teams in the Big 10 conference.
Ans: Side-by-side boxplots
Difficulty: Medium L.O.: 2.4.3
39. Investigate the relationship between pulse rate (in beats per minute) and systolic blood pressure (the top number in a blood pressure reading, measured in mm Hg = millimeters of mercury) for patients at the student health center.
Ans: Scatterplot
Difficulty: Medium L.O.: 2.5.0

40. Investigate the favorite type of music (country, rock, classical, etc) for the students in your class.
Ans: Bar chart
Difficulty: Medium L.O.: 2.1.0
41. Compare the percentage of people in favor of Barack Obama and Mitt Romney in the 2012 Presidential election for the different regions in the U.S. (Northeast, Southeast, Midwest, West).
Ans: Side-by-side bar chart or segmented bar chart
Difficulty: Medium L.O.: 2.1.0
42. Investigate the relationship between the length of the right foot (in cm) and the length of the right forearm (in cm) for students in your class.
Ans: Scatterplot
Difficulty: Medium L.O.: 2.5.0
43. Investigate the length of songs on your iPod.
Ans: Histogram or Dotplot
Difficulty: Medium L.O.: 2.2.0
44. Investigate the number of text messages sent yesterday by students in your class.
Ans: Histogram or dotplot
Difficulty: Medium L.O.: 2.2.0
45. Investigate the relationship between gender and left/right handedness for students in the class.
Ans: Side-by-side bar chart or segmented bar chart
Difficulty: Medium L.O.: 2.1.0
46. Investigate the relationship between the number of hours of exercise per week and athlete or not for a sample of students at a small university.
Ans: Side-by-side boxplots
Difficulty: Medium L.O.: 2.4.3

47. Use the following to answer questions a-d:

Match the five-number summary to the appropriate boxplot.



a. _____ 15, 19, 20, 25, 28

Ans: C

Difficulty: Easy L.O.: 2.3.4; 2.4.2

b. _____ 1, 3, 4, 6, 8

Ans: B

Difficulty: Easy L.O.: 2.3.4; 2.4.2

c. _____ 3, 10, 12, 13, 19

Ans: D

Difficulty: Easy L.O.: 2.3.4; 2.4.2

d. _____ 5, 8, 9, 11, 14

Ans: A

Difficulty: Easy L.O.: 2.3.4; 2.4.2

48. If a distribution is heavily skewed to the left, which relationship between the mean and median is most likely?

A) Mean < Median B) Mean $\approx$ Median C) Mean > Median

Ans: A Difficulty: Medium L.O.: 2.2.4

49. If a distribution is roughly symmetric, which relationship between the mean and median is likely true?

A) Mean < Median B) Mean $\approx$ Median C) Mean > Median

Ans: B Difficulty: Medium L.O.: 2.2.4

50. If a distribution is heavily skewed to the right, which relationship between the mean and median is likely true?

A) Mean < Median B) Mean $\approx$ Median C) Mean > Median

Ans: C Difficulty: Medium L.O.: 2.2.4

Use the following to answer questions 51-52:

Use this dataset to calculate the following summary statistics. Report each with one decimal place.

1	17	13	19	27	21	10	5
---	----	----	----	----	----	----	---

51. Median

Ans: 15.0

Difficulty: Easy L.O.: 2.2.2

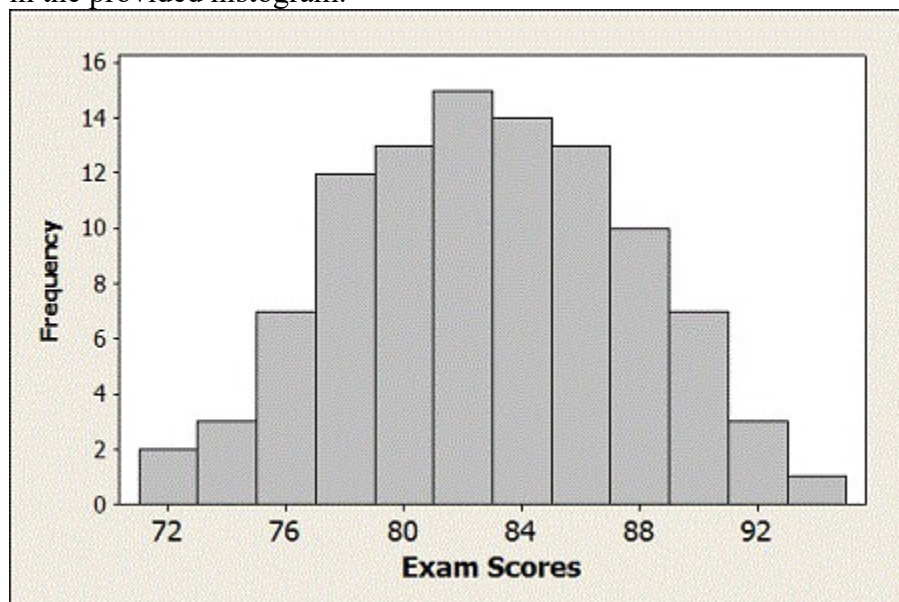
52. Mean

Ans: 14.1

Difficulty: Easy L.O.: 2.2.2

Use the following to answer questions 53-56:

Scores on an exam (out of 100 points) given in a large introductory statistics course are displayed in the provided histogram.



53. Which best describes the shape of the distribution of exam scores?
 A) Approximately symmetric B) Skewed to the left C) Skewed to the right
 Ans: A Difficulty: Easy L.O.: 2.2.1
54. Based on the histogram, which value is likely the mean exam score?
 A) 82 B) 88 C) 76 D) 92
 Ans: A Difficulty: Easy L.O.: 2.2.3
55. Based on the histogram of exam scores, which value is likely the median exam score?
 A) 92 B) 88 C) 82 D) 72
 Ans: C Difficulty: Easy L.O.: 2.2.3
56. Based on the histogram, the standard deviation of the exam scores is likely closest to which of these values?
 A) 0.5 B) 10 C) 5 D) 1
 Ans: C Difficulty: Challenging L.O.: 2.3.2; 2.3.5

Use the following to answer questions 57-64:

The provided Minitab output displays descriptive statistics for the amount of financial aid, in thousands of dollars, awarded to a sample of students at a large university.

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Aid (\$1000)	120	0	15.142	0.672	7.362	3.400	9.925	13.500	19.275	35.000

57. How many students are included in the sample?

Ans: 120

Difficulty: Easy L.O.: 2.3.1

58. Based on the mean and median financial aid amounts displayed in the summary, which of the following most likely describes the shape of the distribution of financial aid amounts?

A) Slightly skewed to the left C) Slightly skewed to the right

B) Roughly symmetric

Ans: C Difficulty: Medium L.O.: 2.2.4

59. Based on the output, give an interval that is certain to contain the 15th percentile of the distribution of financial aid amounts.

Ans: Answers may vary. The lower bound must be the minimum and the upper bound can be Q1 or anything larger. Thus one possible answer is \$3,400 to \$9,925.

Difficulty: Challenging L.O.: 2.3.4

60. Based on the output, give an interval that is certain to contain the 60th percentile of the distribution of financial aid amounts.

Ans: Answers may vary. Based on the available information, the smallest interval that contains the 60th percentile is the median to Q3: \$13,500 to \$19,275.

Difficulty: Challenging L.O.: 2.3.4

61. What is the range of financial aid amounts?

Ans: Range = Max - Min = \$35,000 - \$3,400 = \$31,600

Difficulty: Easy L.O.: 2.3.5

62. What is the IQR of financial aid amounts?

Ans: IQR = Q3 - Q1 = \$19,275 - \$9,925 = \$9,350

Difficulty: Medium L.O.: 2.3.5

63. Is the largest financial aid amount an outlier? Show your work to justify your answer.

Ans: $IQR = \$9350$ or 9.35

$$1.5IQR = 1.5 * 9.35 = 14.025$$

$$Q3 + 1.5IQR = 19.275 + 14.025 = 33.3$$

Since the maximum value of 35 is above the cut-off for outliers on the high end of the distribution, 35 or \$35,000 is an outlier.

Difficulty: Medium L.O.: 2.4.1

64. Find and interpret the z-score for the smallest financial aid amount.

Ans: $z = (3.4 - 15.142)/7.362 = -1.59$

Someone receiving \$3,400 in financial aid is 1.59 standard deviations below the sample mean financial aid amount.

Difficulty: Medium L.O.: 2.3.3

Use the following to answer questions 65-71:

Trying to determine the number of students to accept is a tricky task for universities. The Admissions staff at a small private college wants to use data from the past few years to predict the number of students enrolling in the university from those who are accepted by the university. The data are provided in the following table.

Number Accepted	Number Enrolled
2,440	611
2,800	708
2,720	637
2,360	584
2,660	614
2,620	625

65. What is the explanatory (X) variable?

Ans: Number of students accepted

Difficulty: Medium L.O.: 1.1.3; 2.6.0

66. What is the response (Y) variable?

Ans: Number of students enrolling

Difficulty: Medium L.O.: 1.1.3; 2.6.0

67. Find the correlation between the number of students accepted and enrolled. Use two decimal places in your answer.

Ans: $r = 0.83$

Difficulty: Medium L.O.: 2.5.4

68. Find the least squares regression line for predicting the number enrolled from the number accepted.

Ans: $\hat{E} = 89 + 0.208\text{accepted}$, where E = enrolled

Difficulty: Medium L.O.: 2.6.1

69. Interpret the slope in context.

Ans: The enrollment is predicted to increase by 0.208 students for each additional student accepted. (The enrollment is predicted to increase by 2.08 students for each additional 10 students accepted.)

Difficulty: Medium L.O.: 2.6.3

70. Interpret the intercept of the line in context. Does the interpretation make sense?

Ans: When 0 students are accepted, 89 are enrolled. (Note that this does not make sense because we are extrapolating.)

Difficulty: Medium L.O.: 2.6.3; 2.6.5

71. Suppose Admissions has announced that 2,575 students have been accepted this year. Use your regression equation to predict the number of students that will enroll.

Ans: Let y = # of students enrolled

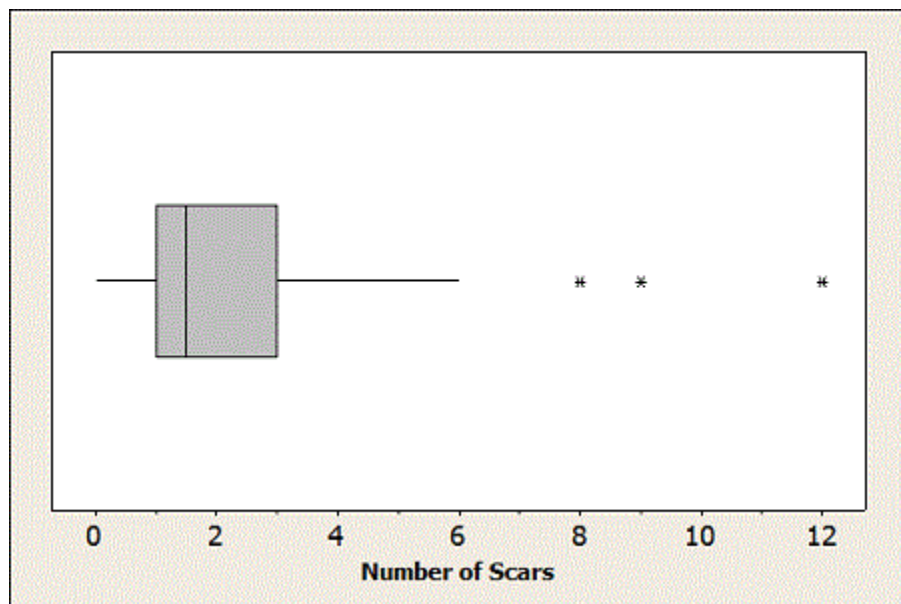
$$\hat{y} = 89 + 0.208(2,575) = 624.6$$

The model predicts that if 2,575 are accepted, then 624.6 students will enroll at the college.

Difficulty: Medium L.O.: 2.6.2

Use the following to answer questions 72-78:

Students in an introductory statistics course were asked to count the number of scars on their dominant hand (the one they write with the most). The results are displayed in the provided boxplot.



72. From the boxplot you can identify how many students are in the class.

Ans: False Difficulty: Medium L.O.: 2.4.2

73. Use the boxplot to estimate the median number of scars that students in the class have on their dominant hand.

Ans: 1.5

Difficulty: Medium L.O.: 2.4.2

74. The distribution of the number of scars would be classified as

A) skewed to the left. B) roughly symmetric. C) skewed to the right.

Ans: C Difficulty: Medium L.O.: 2.4.2

75. Calculate the IQR for the distribution of the number of scars students have.

Ans: $Q3 = 3$ and $Q1 = 1$, so $IQR = 2$

Difficulty: Medium L.O.: 2.3.5; 2.4.2

76. Which answer best describes the following conclusion?

"There are no students with 7 scars on their dominant hand."

A) True B) False C) Cannot be determined

Ans: A Difficulty: Challenging L.O.: 2.4.2

77. Which answer best describes the following conclusion?

"There are no students with 2 scars on their dominant hand."

- A) True B) False C) Cannot be determined

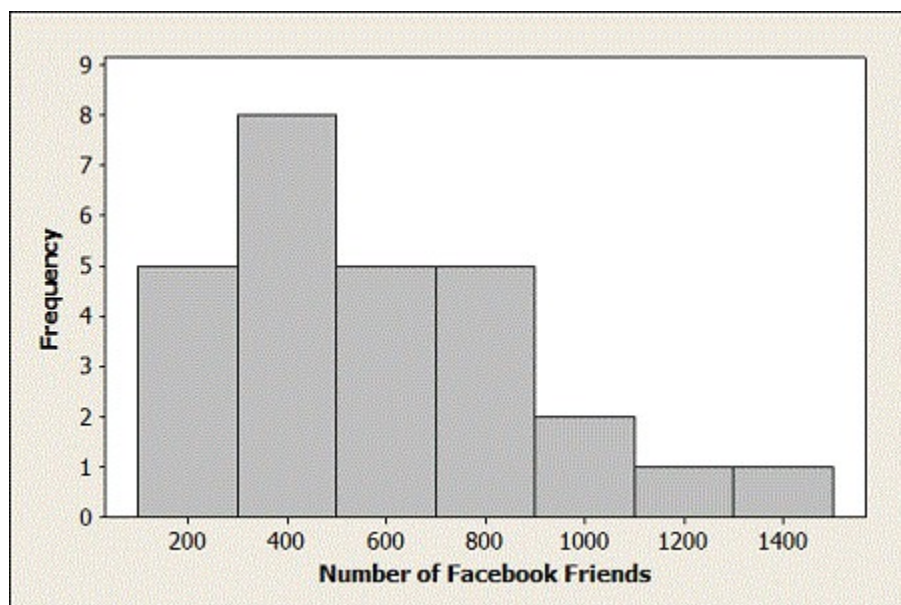
Ans: C Difficulty: Challenging L.O.: 2.4.2

78. The mean can be determined exactly from the boxplot.

Ans: False Difficulty: Medium L.O.: 2.4.2

Use the following to answer questions 79-81:

The provided histogram displays the number of Facebook friends for students in a small statistics class.



79. Which of the following best describes the shape of the distribution of the number of Facebook friends?

- A) Skewed to the left B) Roughly symmetric C) Skewed to the right

Ans: C Difficulty: Easy L.O.: 2.2.1

80. The mean number of Facebook friends is closest to which value?

- A) 220 friends B) 560 friends C) 810 friends D) 1,000 friends

Ans: B Difficulty: Medium L.O.: 2.2.3

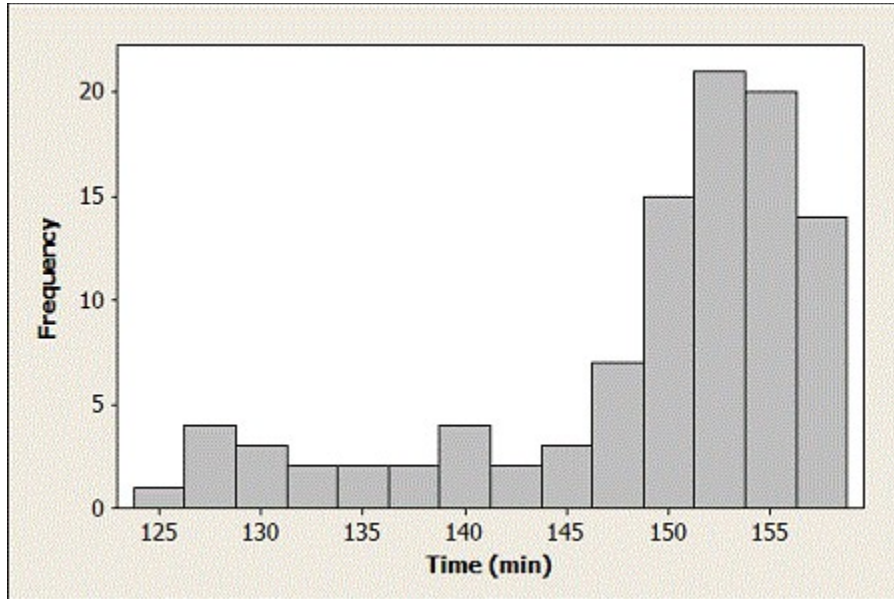
81. The median number of Facebook friends is likely closest to which value?

- A) 300 B) 500 C) 700 D) 800

Ans: B Difficulty: Medium L.O.: 2.2.3

Use the following to answer questions 82-84:

The finishing time for the top 100 men in the 2011 New York Marathon are displayed in the provided figure.



82. Which of the following best describes the distribution of times for the top 100 male finishers in the 2011 New York Marathon?

- A) Skewed to the left B) Roughly symmetric C) Skewed to the right

Ans: A Difficulty: Easy L.O.: 2.2.1

83. The mean time for the top 100 males is closest to which value?

- A) 148 minutes B) 151 minutes C) 140 minutes D) 135 minutes

Ans: A Difficulty: Medium L.O.: 2.2.3

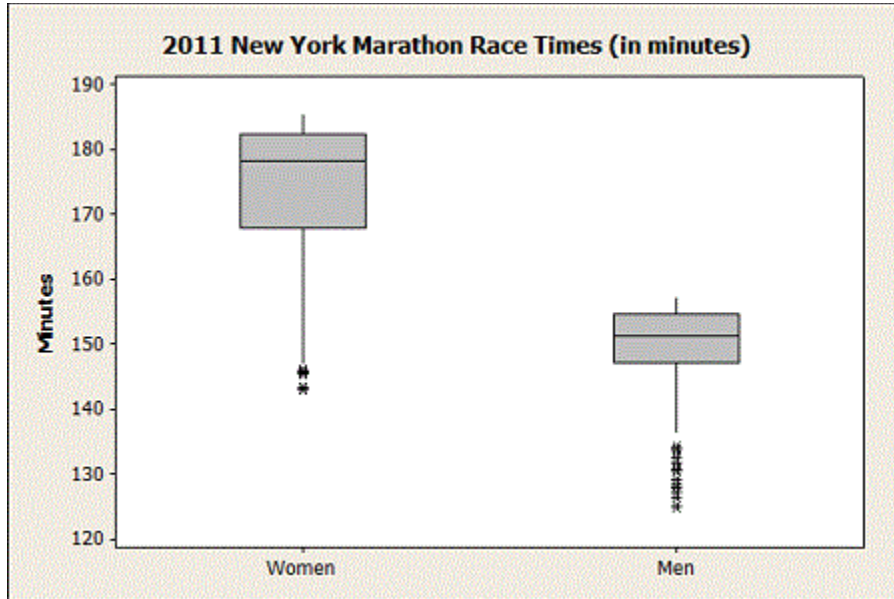
84. The median time for the top 100 males is closest to which value?

- A) 156 minutes B) 140 minutes C) 145 minutes D) 151 minutes

Ans: D Difficulty: Medium L.O.: 2.2.3

Use the following to answer questions 85-87:

The side-by-side boxplots compare the top 100 men's and women's finishing times in the 2011 New York Marathon.



85. Which group tends to finish the race faster?

- A) Men B) Women

Ans: A Difficulty: Medium L.O.: 2.4.4

86. Which group has the larger spread in its race times?

- A) Men B) Women

Ans: B Difficulty: Medium L.O.: 2.4.4

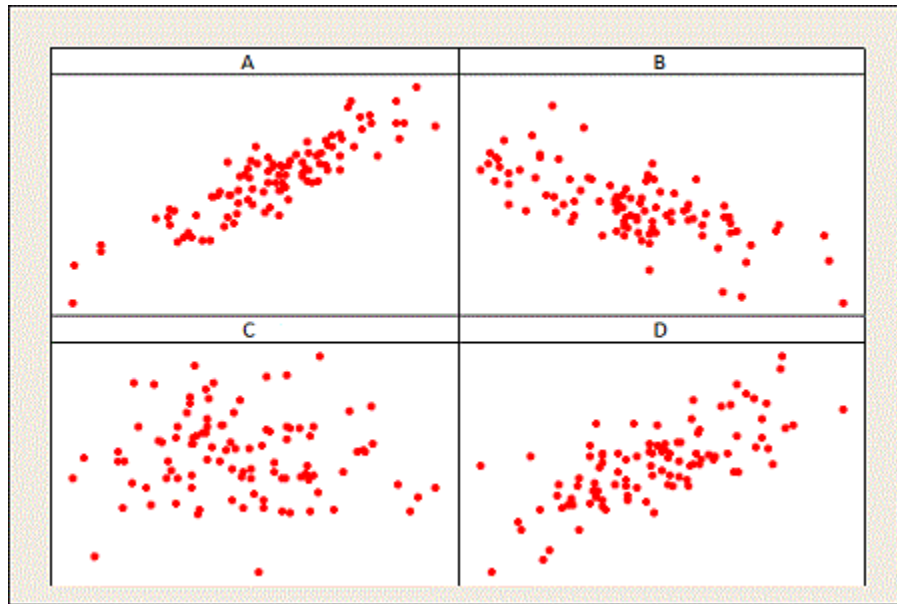
87. Does there appear to be an association between gender and race time? Explain.

Ans: Yes, it appears that men tend to be faster than women. The median race time for men is well below that for women, and there is very little overlap between the distributions for the two groups.

Difficulty: Medium L.O.: 2.4.4

88. Use the following to answer questions a-d:

Match the correlation to the corresponding scatterplot.



a. _____ 0.719

Ans: D

Difficulty: Easy L.O.: 2.5.3

b. _____ -0.064

Ans: C

Difficulty: Easy L.O.: 2.5.3

c. _____ 0.889

Ans: A

Difficulty: Easy L.O.: 2.5.3

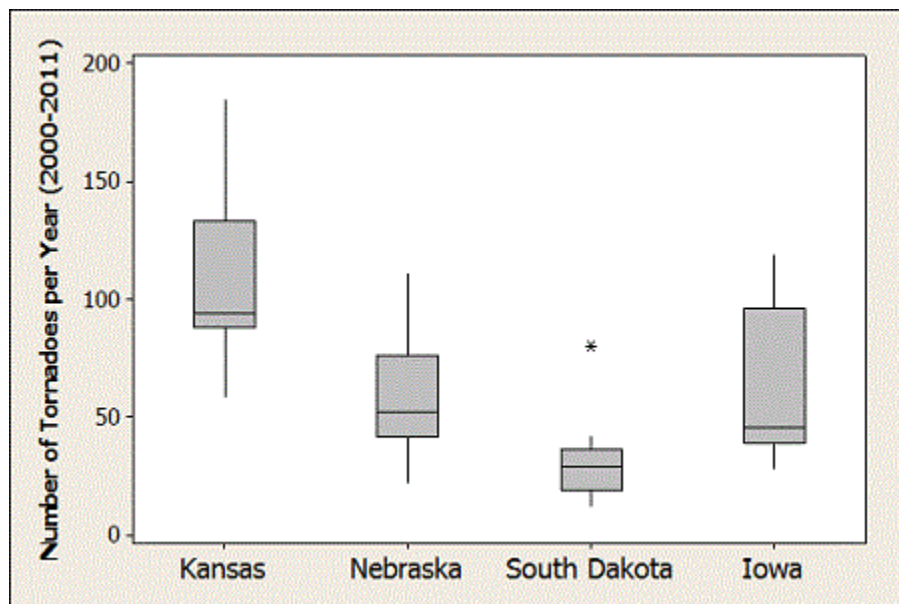
d. _____ -0.701

Ans: B

Difficulty: Easy L.O.: 2.5.3

Use the following to answer questions 89-94:

The states located in the Midwestern region of the country typically experience a large number of tornadoes every year. The number of tornadoes from 2000-2011 for four Midwestern states (Kansas, Nebraska, South Dakota, and Iowa) are displayed in the side-by-side boxplots.



89. Which state tends to see the most tornadoes per year?
 A) Kansas B) Nebraska C) South Dakota D) Iowa
 Ans: A Difficulty: Easy L.O.: 2.4.3
90. Which state has the largest range?
 A) Kansas B) Nebraska C) South Dakota D) Iowa
 Ans: A Difficulty: Medium L.O.: 2.43
91. Which state tends to see the fewest tornadoes per year?
 A) Kansas B) Nebraska C) South Dakota D) Iowa
 Ans: C Difficulty: Easy L.O.: 2.4.3
92. Which state has the largest IQR?
 A) Kansas B) Nebraska C) South Dakota D) Iowa
 Ans: D Difficulty: Medium L.O.: 2.4.3
93. Which state has an outlier?
 A) Kansas B) Nebraska C) South Dakota D) Iowa
 Ans: C Difficulty: Easy L.O.: 2.4.3

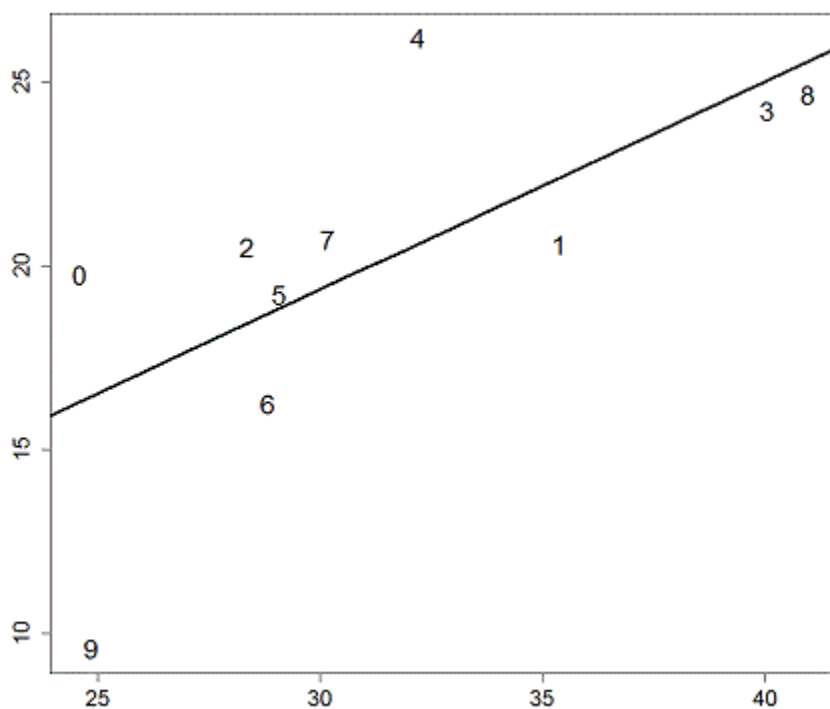
94. (58, 88, 94, 133, 185) is the five number summary for which state?

A) Kansas B) Nebraska C) South Dakota D) Iowa

Ans: A Difficulty: Medium L.O.: 2.3.4; 2.4.3

Use the following to answer questions 95-97:

The least squares regression line is displayed on the provided scatterplot. Note that the points are displayed with numbers (each point having its own number), rather than points.



95. Which point has the most extreme negative residual?

A) 0 B) 1 C) 4 D) 9

Ans: D Difficulty: Challenging L.O.: 2.6.4

96. Which point has the most extreme positive residual?

A) 0 B) 4 C) 8 D) 9

Ans: B Difficulty: Challenging L.O.: 2.6.4

97. Which point has the residual that is closest to 0?

A) 1 B) 4 C) 5 D) 9

Ans: C Difficulty: Challenging L.O.: 2.6.4

Use the following to answer questions 98-106:

Students in a small statistics course collected data to determine if the length of the forearm could be used to predict the length of the foot (both measured in centimeters). Their data are displayed in the provided table.

Forearm (cm)	29	28	27	23	26	29.5	36	29	30	24	27	29.5	32
Foot (cm)	26	23	24	23	25	27	29	28	23	23	24	26	31

98. Based on their goal (to predict foot length from forearm length), which variable is the explanatory variable?

A) Forearm length B) Foot length

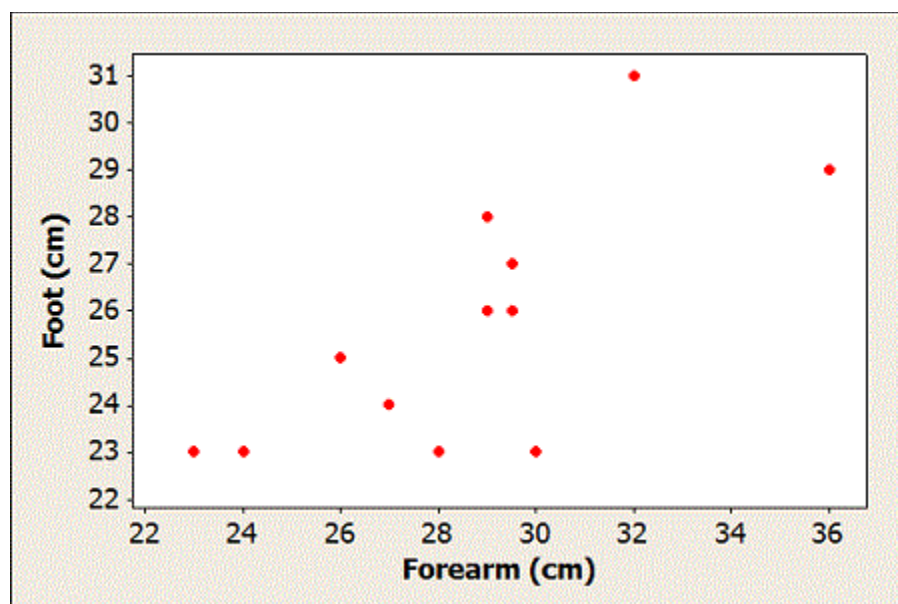
Ans: A Difficulty: Medium L.O.: 1.1.3; 2.5.0; 2.6.0

99. Which of the following would you expect to be true about the association between the length of the forearm and the length of the foot?

A) Positive association B) Negative association C) No association

Ans: A Difficulty: Easy L.O.: 2.5.2

100. A scatterplot of the data collected by the students is provided. Does there appear to be a positive or negative association between these two variables? What does this mean for these two variables?



Ans: There is a positive association between forearm and foot length. This means that people with longer forearms tend to have larger feet (and similarly people with shorter forearms tend to have smaller feet).

Difficulty: Easy L.O.: 2.5.1; 2.5.2

101. Find the correlation between forearm and foot length. Use three decimal places in your answer.

Ans: $r = 0.739$

Difficulty: Medium L.O.: 2.5.4

102. If the forearm and foot lengths had been measured in inches instead of centimeters the correlation would be different.

Ans: False Difficulty: Challenging L.O.: 2.5.3

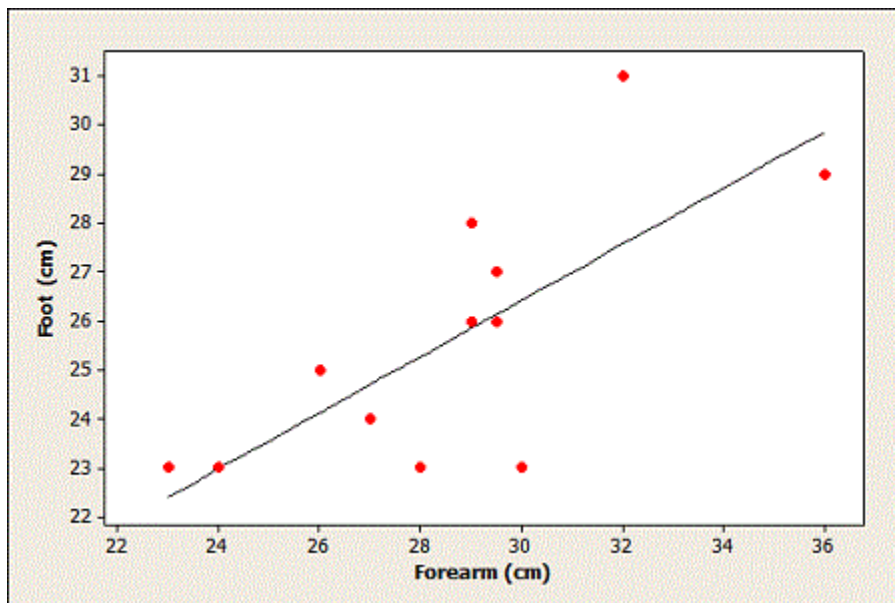
103. Find the least squares regression equation for predicting foot length from forearm length.

Ans: $y = \text{foot}$

$$\hat{y} = 9.216 + 0.5735 \text{ forearm}$$

Difficulty: Medium L.O.: 2.6.1

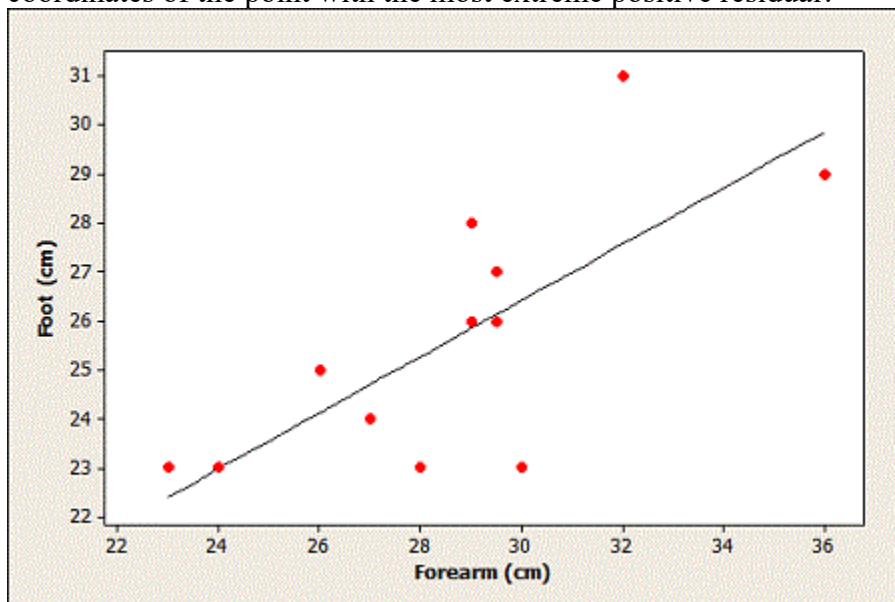
104. A scatterplot of the data with the least squares regression line is shown. What are the coordinates of the point with the most extreme negative residual?



- A) (28, 23) B) (30, 23) C) (32, 31) D) (36, 29)

Ans: B Difficulty: Medium L.O.: 2.6.4

105. A scatterplot of the data with the least squares regression line is shown. What are the coordinates of the point with the most extreme positive residual?



- A) (32, 31) B) (30, 23) C) (28, 23) D) (23, 23)

Ans: A Difficulty: Medium L.O.: 2.6.4

106. Using your model to predict the foot length for an individual with a forearm length of 45 cm would be _____.

Ans: extrapolation

Difficulty: Medium L.O.: 2.6.5

Use the following to answer questions 107-112:

A biologist collected data on a sample of 20 porcupines. She wants to be able to predict the body mass of a porcupine (in grams) based on the length of the porcupine (in cm). Her least squares regression equation is $\hat{M} = -3089 + 175.6 \text{Length}$, where M is the mass.

107. Interpret the slope of the least squares regression line, in the context of the situation.

Ans: The body mass of porcupines is predicted to increase by 175.6 g for every centimeter of length.

Difficulty: Medium L.O.: 2.6.3

108. If it would make sense, provide a clear interpretation of the intercept of the regression line, in context. Otherwise, explain why the interpretation does not make sense.

Ans: The interpretation of the intercept does not make sense in this situation. A real porcupine would not have a length of 0 cm. Making a prediction for a porcupine with a body length of 0 cm would be extrapolation because 0 cm is far outside the range of the original data.

Difficulty: Medium L.O.: 2.6.3; 2.6.5

109. Predict the body mass of a porcupine that is 51 cm long. Report your answer using one decimal place.

Ans: $\hat{M} = -3089 + 175.6(51) = 5,866.6$

A porcupine that is 51 cm long is predicted to have a body mass of 5,866.6 g.

Difficulty: Medium L.O.: 2.6.2

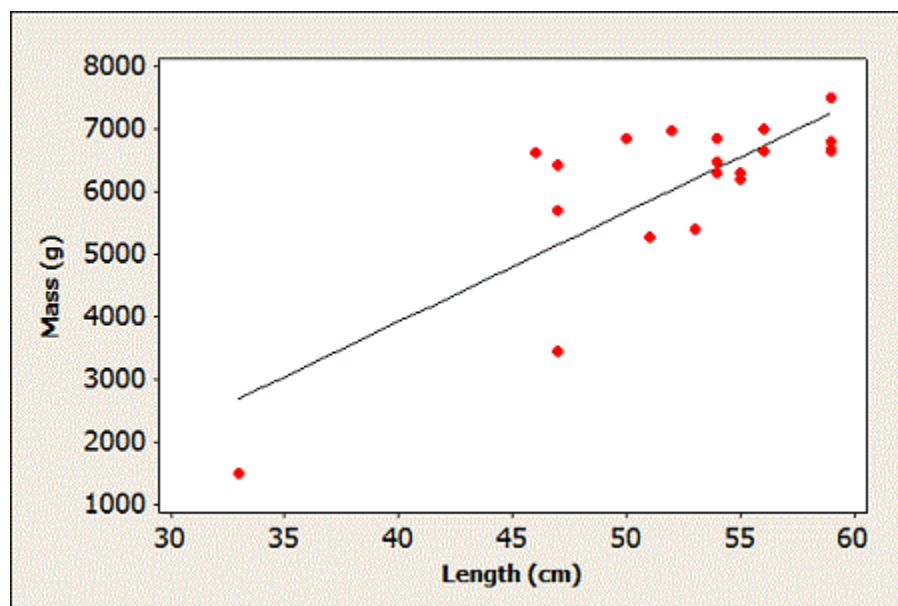
110. One of the porcupines in the dataset had a body length of 51 cm and a body mass of 5,281 g. Calculate the residual for this porcupine. Use one decimal place in your calculation.

Ans: $\hat{M} = -3089 + 175.6(51) = 5,866.6$ g is the predicted body mass for a porcupine that is 51 cm long.

The residual for this porcupine is $e = y - \hat{y} = 5,281 - 5,866.6 = -585.6$ g.

Difficulty: Medium L.O.: 2.6.4

111. A scatterplot of the biologist's data, with the least squares regression line, is provided. There is a clear outlier in the lower left corner of the plot. How would removing this point from the dataset most likely affect the correlation between body length (cm) and body mass (g)?



- A) It would make the correlation stronger.
 B) It would make the correlation weaker.
 C) It would have no impact on the correlation between body length and body mass.

Ans: B Difficulty: Challenging L.O.: 2.5.6; 2.6.6

112. Another variable that the biologist recorded was the chest circumference (in cm) of the porcupines. Explain what both a negative and a positive association between body mass and chest circumference would mean. Which is more plausible in this situation?

Ans: A positive correlation would mean that porcupines that have a larger chest circumference tend to have a higher body mass while those with a smaller chest circumference would tend to have a lower body mass.

A negative correlation would mean that porcupines that have a larger chest circumference tend to have a lower body mass while those with the smaller chest circumference would tend to have a higher body mass.

The positive correlation is most plausible in this situation - the porcupines that are "bigger" in the chest area are likely bigger overall and thus should have the higher body mass.

Difficulty: Medium L.O.: 2.5.2; 2.5.3

Use the following to answer questions 113-114:

The midrange is another way to measure of the center of a distribution. The midrange of a dataset is defined to be the average of the minimum and maximum values in the dataset.

113. Calculate the midrange of this dataset.

4 10 12 2 7 5 9 8

Ans: midrange = $(12+2)/2 = 7$

Difficulty: Medium L.O.: 2.2.0

114. In general, would you think that the midrange should be a resistant statistic? Briefly explain.

Ans: No, it would not be resistant in general because it is calculated with the minimum and maximum in the dataset. If there is an outlier in the dataset, it will be one of these values. This would affect the average of the minimum and maximum (since averages are not resistant).

Difficulty: Challenging L.O.: 2.2.4

Use the following to answer questions 115-117:

A student working an independent research project wants to investigate if there is an association between the amount of sleep someone gets and their body mass index (BMI) - an indicator of body fatness. For a sample of 45 students, she records their BMI and the average amount of sleep they get on weeknights over a two-week period.

115. What would it mean for average amount of sleep and BMI to be positively correlated?

Ans: It would mean that the more sleep individuals get, the higher their BMI tends to be.

Difficulty: Medium L.O.: 2.5.2

116. What would it mean for average amount of sleep to be negatively correlated with BMI?

Ans: It would mean that individuals who get more sleep tend to have lower BMIs.

Difficulty: Medium L.O.: 2.5.2

117. Suppose the student found a correlation of -0.413 between amount of sleep and BMI.

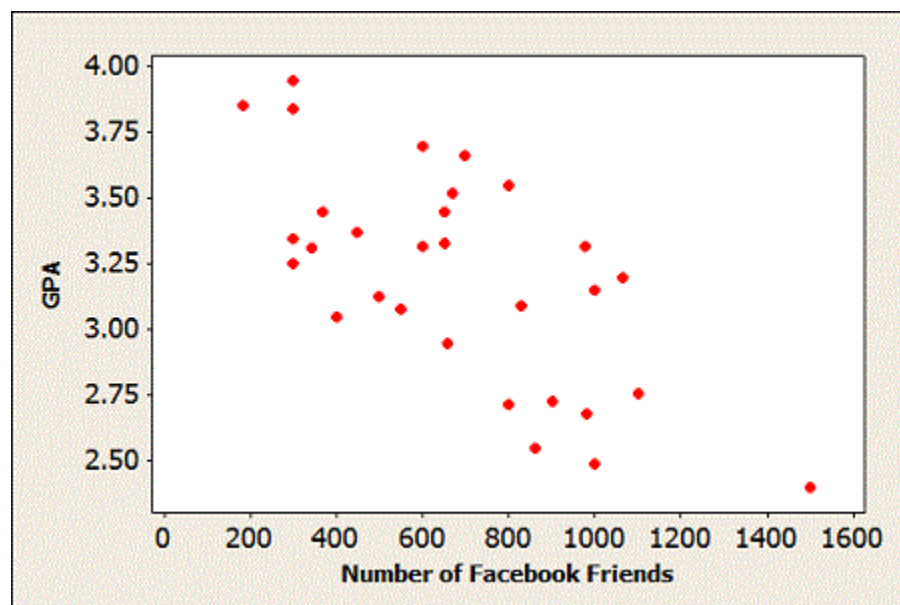
Would it be appropriate for her to conclude that getting more sleep causes individuals to have a lower BMI? Briefly justify your answer.

Ans: No, because correlation does not imply causation.

Difficulty: Medium L.O.: 2.5.5

Use the following to answer questions 118-120:

The scatterplot shows the relationship between GPA and the number of Facebook friends for 30 students in a class.



118. Discuss the information contained in the scatterplot. What does it mean about GPA and number of Facebook friends?

Ans: There seems to be a moderately strong negative linear association between GPA and the number of Facebook friends. The negative association indicates that individuals who have more Facebook friends tend to have a lower GPA, while individuals with fewer Facebook friends tend to have a higher GPA. There is a potential outlier where one student had roughly 1,500 Facebook friends and a GPA a little below 2.50.

Difficulty: Medium L.O.: 2.5.1

119. For each corner of the scatterplot (top left, top right, bottom left, bottom right), describe a student whose responses place him or her in that corner.

Ans: Top Left: A student in this corner of the scatterplot would have a small number of Facebook friends (less than 200) and a GPA fairly close to 4.0, i.e., a very strong student with relatively few Facebook friends.

Top Right: A student in this corner of the scatterplot would have a large number of Facebook friends (more than 1,000) and a high GPA (close to 4.0), i.e., a very strong student with lots of Facebook friends.

Bottom Left: A student in this corner of the scatterplot would have a small number of Facebook friends (less 200) and a low GPA (maybe 2.75 or less), i.e., a somewhat weak student with relatively few Facebook friends.

Bottom Right: A student in this corner of the scatterplot would have a large number of Facebook friends (more than 1,000) and a low GPA (maybe 2.75 or less), i.e., a somewhat weak student with lots of Facebook friends.

Difficulty: Medium L.O.: 2.5.0

120. The correlation between GPA and the number of Facebook friends is -0.686. Should you go "unfriend" some of your Facebook friends if you want to improve your GPA (i.e., can you conclude that having more Facebook friends lowers GPA)? Explain briefly.

Ans: No! Correlation does not imply causation! A possible lurking variable could be amount of time spent studying - maybe individuals with more Facebook friends spend more time online and less time studying (hence the lower GPA).

Difficulty: Medium L.O.: 2.5.5