

## 15.1 Mean, Median, and Mode

---

### Answers

1. Mean: 1.76, Median: 1, Mode: 1
2. Because there are two large numbers at the end that contribute more heavily to the mean, but doesn't impact the median.
3. The median or mode would make the most sense in this case because they are not impacted by the outliers of 10 siblings. Also, you can't have 1.76 siblings, so 1.76 siblings would not make sense as the "typical number of siblings".
4. 885
5. 16
6. 154 miles
7. 5
8. Mean: 19.2, Median: 19, Mode: 15
9. Mean: 7.083 , Median: 6, Mode: 5
10. Mean: 3.889, Median: 4, Mode: None
11. Mean: 69.75, Median: 88.5
12. Mean: 31.67, Median: 19.5
13. #12 because there was one number much larger than the rest in the data.
14. #11 because there was one number much smaller than the rest in the data.
15. In #11, 12 is the outlier. In #12, 98 is the outlier. If you were to remove the outliers, the mean would be closer to the median in each case.

## 15.2 Expected Value and Payoffs

---

### Answers

1. To calculate expected value, multiply the value of each outcome by the probability of that outcome and find the sum of all of these products.
2. False, \$0.50 is your expected average amount of winnings if you were to play the game many times.
3. True
4.  $\frac{1}{6}$  or 17 cents
5. 36 cents
6. They should charge people more than 65 cents (the expected value) to play.
7. 18.02
8. 81.3%
9. 70.5%
10. 93.0%
11. 79.5%
12. 76.3%
13. Answers vary
14. You should charge more than \$1.15 in order to theoretically make a profit by the end of the night.
15. Answers vary
16. Casinos need to design games that people have a chance of winning, but that ultimately the casino will make money off of. Expected value helps to determine what their profit will be on average for each game and can help them to determine what should be charged for each game.

## 15.3 Five Number Summary

---

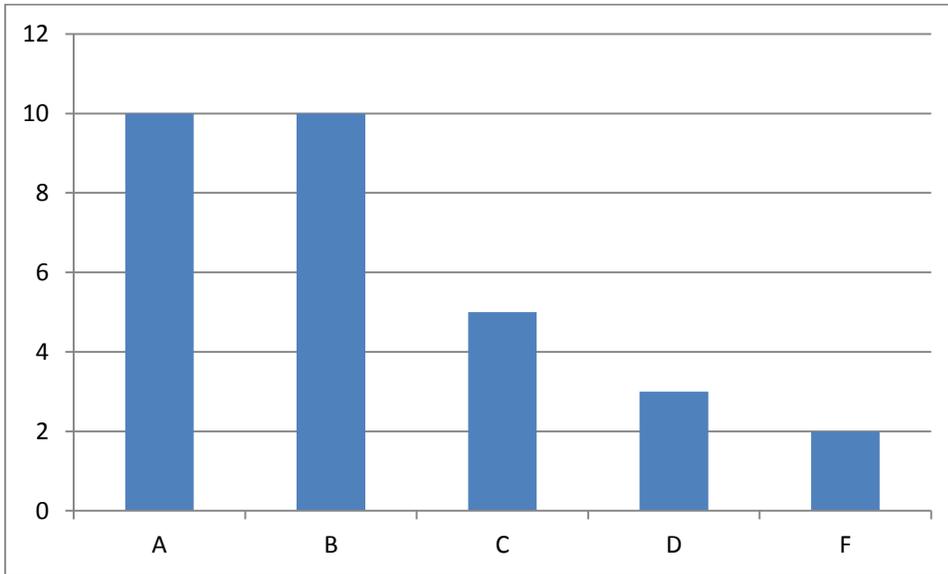
### Answers

1. Minimum: 0.08; Q1: 0.18; Q2: 0.235; Q3: 0.27; Maximum: 0.32
2. Minimum: 77; Q1: 79.5; Q2: 86.5; Q3: 90.5; Maximum: 99
3. Minimum: 53; Q1: 79.5; Q2: 84.5; Q3: 92.5; Maximum: 98
4. Minimum: 51; Q1: 72; Q2: 85; Q3: 91; Maximum: 96
5. Minimum: 185; Q1: 220.5; Q2: 281; Q3: 363.5; Maximum: 518
6. Minimum: 33; Q1: 38; Q2: 40; Q3: 48.5; Maximum: 71
7. Minimum: 3; Q1: 6; Q2: 12; Q3: 16; Maximum: 21
8. Minimum: 6; Q1: 16.5; Q2: 26; Q3: 37; Maximum: 49
9. Minimum: 5; Q1: 9; Q2: 10.5; Q3: 12; Maximum: 17
10. Minimum: 49; Q1: 50; Q2: 53; Q3: 57; Maximum: 67
11. Minimum: 5; Q1: 18.5; Q2: 20.5; Q3: 22.5; Maximum: 24
12. Minimum: 620; Q1: 800; Q2: 850; Q3: 900; Maximum: 1070
13. Minimum: 12; Q1: 13.5; Q2: 17; Q3: 22.5; Maximum: 42
14. Minimum: 15; Q1: 17.5; Q2: 21; Q3: 32.5; Maximum: 55
15. Minimum: 120; Q1: 122; Q2: 124.5; Q3: 129; Maximum: 149

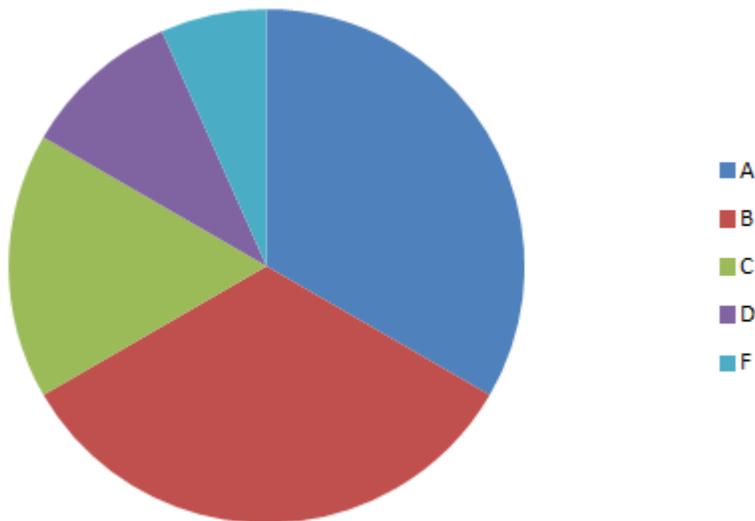
## 15.4 Graphic Displays of Data

### Answers

1. pie charts and bar graphs
2. histograms and boxplots
- 3.

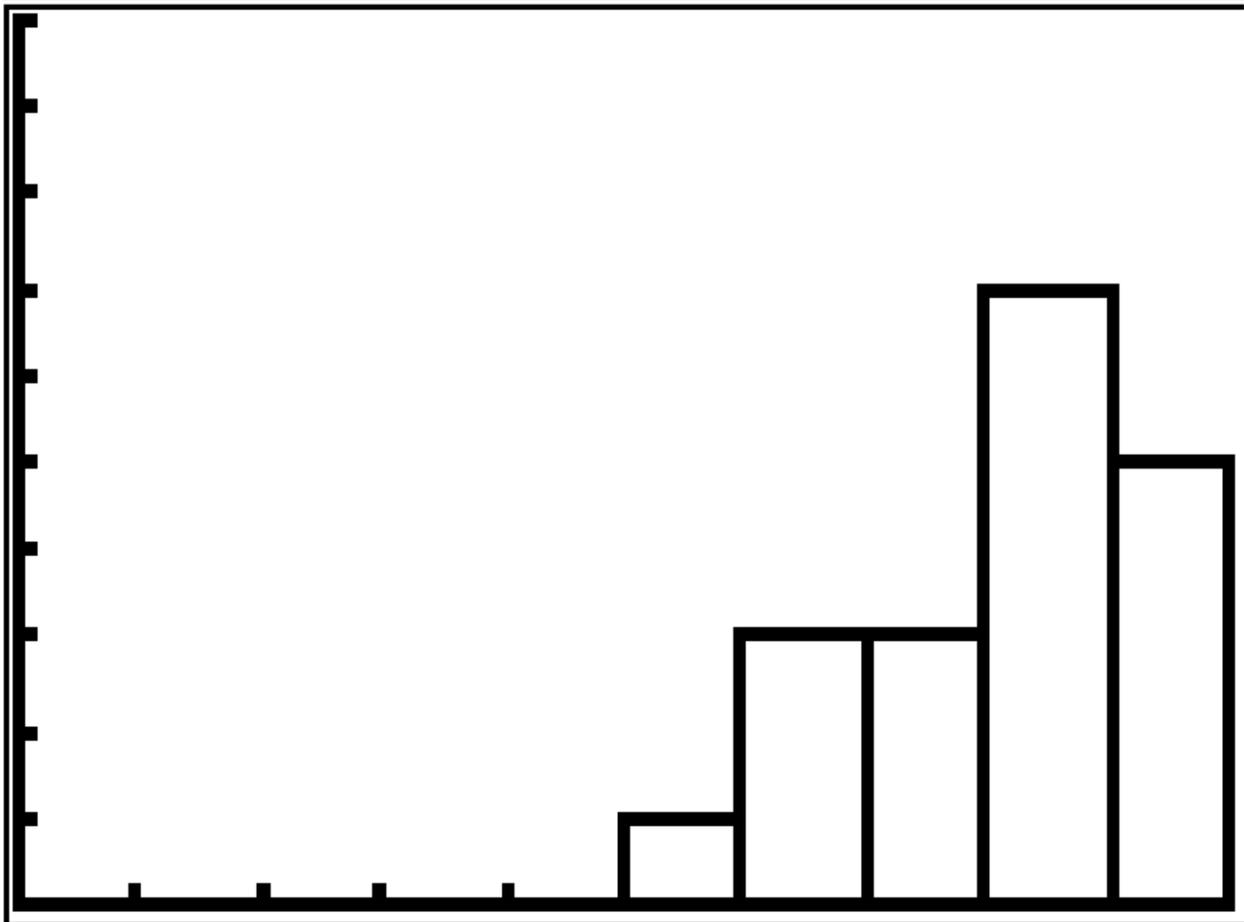


- 4.



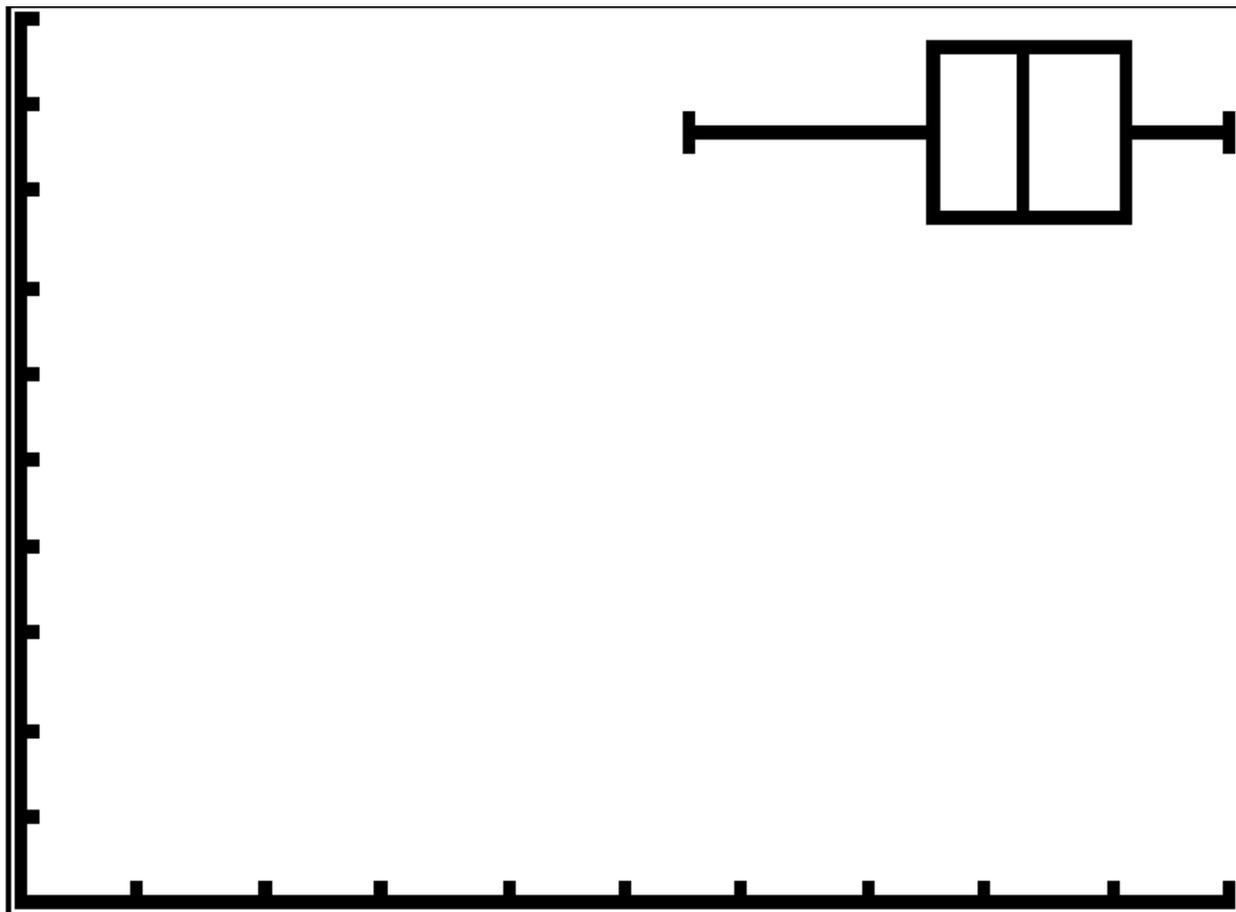
5. Answers vary

6.



7. {55, 75, 82.5, 91.5, 100}

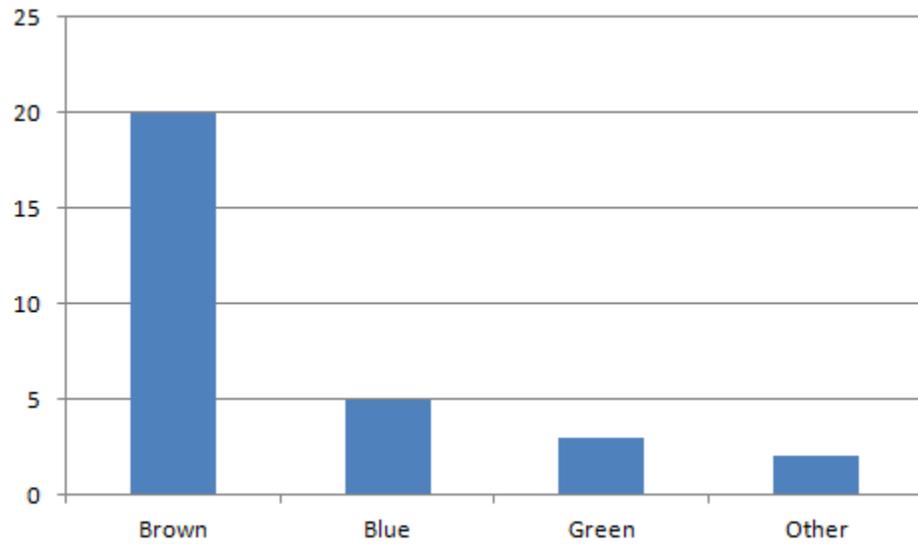
8.



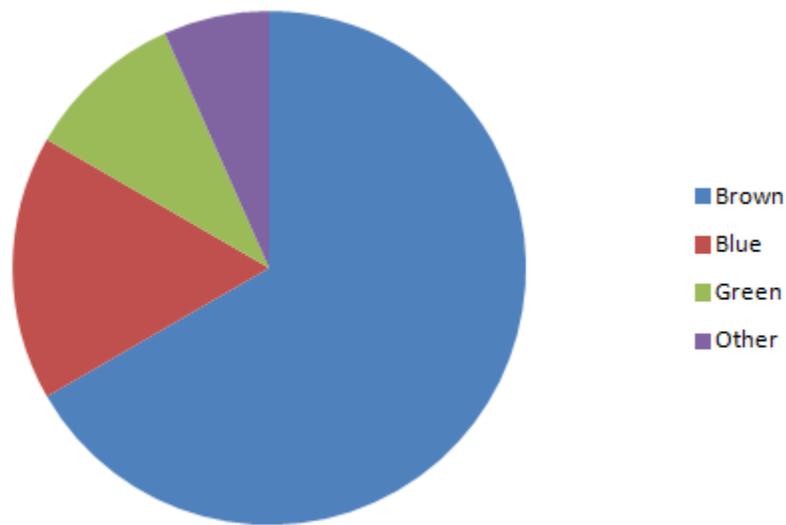
9. Answers vary. A student might notice that there is a bigger range in the first half of the data than in the second half of the data.

10. Answers vary. A student might notice that more of the data is in the left side of the graph than the right side of the graph.

11.



12.



13. Answers vary.

14. Bar graph or pie chart.

15. Histogram or boxplot.

## 15.5 Variance

---

### Answers

1. Standard deviation and variance are both measures of spread, but variance is a larger number. Standard deviation is the square root of variance.
2. Data Set A has data that is more varied and spread out than the data in Data Set B.
3. 33.2967
4. 16.5432
5. 254.568
6. 33.2967
7. 16.5432
8. 254.568
9. 4
10. c
11. If variance is large, data will be spread out on a histogram. There might be a lot of data at low values and a lot of data at high values, with not much in between. If variance is small, all data will be close together near the mean.
12. No, you can only calculate the variance for quantitative data, not categorical data. Bar graphs only show categorical data.
13. Mean: 81.5; Sample Variance: 149.105; Sample Standard Deviation: 12.2109
14. Mean: 5.15; Population Variance: 12.0275; Population Standard Deviation: 3.46807
15. It is often not realistic or possible to find data from the whole population, and then you have to be satisfied with only having a sample of the population. For example, it would be impossible to find a piece of information from every person in the world, but you might be able to get data samples from every country.

## 15.6 The Normal Curve

---

### Answers

1. 0
2. 1
3. 84.13%
4. 15.87%
5. 2.28%
6. 95.45%
7. 64.69%
8. 0
9. 90.21%
10. 1.15%
11. 1.64 standard deviations below the mean.
12. 49.38%
13. Approximately 68% of adult women are between 63 and 67 inches tall.
14. 69.15%
15. 0.644%

## 15.7 Linear Correlation

---

### Answers

1. Very strong positive correlation. The data is perfectly linear with a positive slope. As the value of one variable increases, the value of the other variable increases.
2. Mild negative correlation. The data is somewhat linear with a negative slope. As the value of one variable increases, the value of the other variable tends to decrease.
3. Very strong negative correlation. The data is perfectly linear with a negative slope. As the value of one variable increases, the value of the other variable decreases.
4. No correlation. There is no relationship between the two variables. As the value of one variable increases, there is no pattern to what happens to the value of the other variable.
5. Strong positive correlation. The data is strongly linear with a positive slope. As the value of one variable increases, the value of the other variable increases.
6.  $\hat{y} = 0.0974125588 + 0.005546124x$ ;  $r = 0.94668$
7. 2.87. This seems to fit with the data.
8. The relevant domain is between 200 and 800, which is the range of possible SAT scores. The regression equation will be most accurate between 400 and 700, where all of the data points are.
9. No, just because the two variables are correlated doesn't mean that a high SAT score would *cause* a high GPA.
10.  $\hat{y} = 7.164812942 + 0.5995955511x$ ;  $r = 0.5682806756$
11. 18.56. This is reasonable, but because the correlation coefficient is only about 0.5, there isn't a strong correlation and the regression equation isn't a great way to make predictions.
12. Because the data is only mildly positive correlated, you can't use this data to make predictions that you can be very confident about. Some students' scores went up from quiz 1 to quiz 2 and some students' scores went down from quiz 1 to quiz 2.
13. Answers vary. Possible answer: Correlation is a relationship between multiple variables while causation is when a change in one variable causes a change in another variable.
14. Answers vary. Possible answer: The correlation coefficient measures how strong the correlation is between two variables and whether the correlation is positive or negative.
15. Answers vary. Possible answer: If you have a larger sample size you will have more data points and will be closer to have data from the full population.

## 15.8 Modeling with Regression

---

### Answers

- Two good choices are natural log regression and logistic regression. With natural log regression the equation is  $\hat{y} = 11.4146888 \ln(x) + 27.61585$  and with logistic regression the equation is  $\hat{y} = \frac{65.8877}{1 + 1.422 e^{-210439}}$ .
- The logistic function is a better fit because it levels off when the height as the height of the women levels off.
- The natural log regression does not have a y-intercept and the logistic equation has a y-intercept at 26.7. This would be the height of a baby when it was born. The typical length of a baby at birth is 20 inches, so the model is a bit unreasonable.
- With the logistic equation, the predicted height is 64.88768 inches. This is reasonable because women aren't growing anymore when they are 70. To be perfectly realistic, the model should level off for awhile in the middle and then start to decrease, as women will tend to actually slowly lose height as they get older after about age 40.
- The equation is  $\hat{y} = 58.9 \cdot 0.514^x$ .
- Exponential regression makes sense because she is losing approximately half her height with each sip. This is exponential decay with a common ratio of  $\frac{1}{2}$ .
- About 5 sips.
- 1.0866 inches.
- The equation is  $\hat{y} = \frac{385.3269}{1 + 74425077 e^{-3.55923}}$ .
- The logistic model is appropriate because there is a maximum for how many people can know the rumor (400 students), so the model needs to level off around 400.
- The model says 1 person will know the rumor after about 3.4 days. This doesn't fit with the actual data which found that after 3 days, 29 people knew the rumor. After 5 days, the model tends to fit the data much better.
- Sine regression works very well. The equation is  $\hat{y} = 4.033 \cdot \sin(0.51745x + 0.0307) + 8.94582$ .
- The predicted depth is 5.392634697 feet. The actual depth is 5.4 feet. The residual is 0.073653035.

14. The cubic regression equation is  $\hat{y} = 0.052x^3 - 0.92837995x^2 + 3.737x + 8.6615$ . This model fits the data points well, but doesn't make as much sense out of the domain of 0 to 10 hours. The cubic model shows the depth continually increasing after 10 hours, which it of course wouldn't actually do. The sine model makes more sense because it is periodic, just as the depth of the water will be.

15. Modeling with regression allows you to quickly make predictions and generalizations about relationships between variables. When there is a lot of data, modeling helps to summarize the data visually and algebraically.