

MEASURES OF CENTRAL TENDENCY AND DATA ANALYSIS

MEASURES OF CENTRAL TENDENCY

The **measures of central tendency** are numbers that locate or approximate the “center” of a set of data. Mean, median, and mode are the most common measures of central tendency.

MEAN

The **mean** is the arithmetic average of a data set. Add all of the values in a set and divide this sum by the number of values in the set.

Example 1

Find the mean of this set of data:

34, 31, 37, 44, 38, 34, 42, 34, 43, and 41.

- $34 + 31 + 37 + 44 + 38 + 34 + 42 + 34 + 43 + 41 = 378$
- $378 \div 10 = 37.8$

The mean of this set of data is 37.8.

Example 2

Find the mean of this set of data:

92, 82, 80, 92, 78, 75, 95, and 77.

- $92 + 82 + 80 + 92 + 78 + 75 + 95 + 77 + 77 = 748$
- $748 \div 9 \approx 83.1$

The mean of this set of data is about 83.1.

Problems

Find the mean of each set of data.

1. 29, 28, 34, 30, 33, 26, and 34.
2. 25, 34, 35, 27, 31, and 30.
3. 80, 89, 79, 84, 95, 79, 78, 89, 76, 82, 76, 92, 89, 81, and 123.
4. 116, 104, 101, 111, 100, 107, 113, 118, 113, 101, 108, 109, 105, 103, and 91.

MODE

The **mode** is the value in a data set that occurs most often. Data sets may have more than one mode, or no mode at all.

Example 3

Find the mode of this set of data:

34, 31, 37, 44, 34, 42, 34, 43, and 41.

- The mode of this data set is 34 since there are three 34's and only one of each of the other numbers.

Example 4

Find the mode of this set of data:

92, 82, 80, 92, 78, 75, 95, 77, and 77.

- The modes of this set of data are 77 and 92 since there are two of each of these numbers and only one of each of the other numbers. This data set is said to be bimodal since it has two modes.

Problems

Find the mode of each set of data.

5. 29, 28, 34, 30, 33, 26, and 34.

6. 25, 34, 35, 27, 25, 31, and 30.

7. 80, 89, 79, 84, 95, 79, 89, 76, 82, 76, 92, 89, 81, and 123.

8. 116, 104, 101, 111, 100, 107, 113, 118, 113, 101, 108, 109, 105, 103, and 91.

MEDIAN

The **median** is the middle number in a set of data arranged in numerical order. If there are an even number of values, the median is the mean of the two middle numbers.

Example 5

Find the median of this set of data:
34, 31, 37, 44, 38, 34, 43, and 41.

- Arrange the data in order:
31, 34, 34, 34, 37, 38, 41, 43, and 44.
- Find the middle value(s): 37 and 38.
- Since there are two middle values, find their mean: $37 + 38 = 75$, $75 \div 2 = 37.5$. Therefore, the median of this data set is 37.5.

Example 6

Find the median of this set of data:
92, 82, 80, 92, 78, 75, 95, 77, and 77.

- Arrange the data in order:
75, 77, 77, 78, 80, 82, 92, 92, and 95.
- Find the middle value(s): 80. Therefore, the median of this data set is 80.

Problems

Find median of each set of data.

- | | | | |
|-----|--|-----|---|
| 9. | 29, 28, 34, 30, 33, 26, and 34. | 10. | 25, 34, 27, 25, 31, and 30. |
| 11. | 80, 89, 79, 84, 95, 79, 78, 89, 76, 82, 76, 92, 89, 81, and 123. | 12. | 116, 104, 101, 111, 100, 107, 113, 118, 113, 101, 108, 109, 105, 103, and 91. |

RANGE

The **range** of a set of data is the difference between the highest value and the lowest value.

Example 7

Find the range of this set of data: 114, 109, 131, 96, 140, and 128.

- The highest value is 140.
- The lowest value is 96.
- $140 - 96 = 44$.
- The range of this set of data is 44.

Example 8

Find the range of this set of data: 37, 44, 36, 29, 78, 15, 57, 54, 63, 27, and 48.

- The highest value is 78.
- The lowest value is 27.
- $78 - 27 = 51$.
- The range of this set of data is 51.

Problems

Find range of each set of data.

13. 29, 28, 34, 30, 33, 26, and 34.

14. 25, 34, 27, 25, 31, and 30.

15. 80, 89, 79, 84, 95, 79, 78, 89, 76, 82, 76, 92, 89, 81, and 123.

16. 116, 104, 101, 111, 100, 107, 113, 118, 113, 101, 108, 109, 105, 103, and 91.

DATA ANALYSIS

OUTLIERS

Outliers are numbers in a data set that are either much higher or much lower than the other numbers in the set.

Example 9

Find the outlier of this set of data: 88, 90, 96, 93, 87, 12, 85, and 94.

- The outlier is 12.

Example 10

Find the outlier of this set of data: 67, 54, 49, 76, 64, 59, 60, 72, 123, 44, and 66.

- The outlier is 123.

Problems

Find the outlier for each set of data.

17. 70, 77, 75, 68, 98, 70, 72, and 71.

19. 1376, 1645, 1783, 1455, 3754, 1790, 1384, 1643, 1492, and 1776.

18. 14, 22, 17, 61, 20, 16, and 15.

20. 62, 65, 93, 51, 55, 14, 79, 85, 55, 72, 78, 83, 91, and 76.

STEM-AND-LEAF PLOTS

A **stem-and-leaf plot** is a way to display data that shows the individual values from a set of data and how the values are distributed. This type of display clearly shows median, mode, range, and outliers. The “stem” part on the graph represents the leading digit(s) of the number. The “leaf” part of the graph represents the other digit(s).

Example 12

Make a stem-and-leaf plot of this set of data:
34, 31, 37, 44, 38, 29, 34, 42, 43, 34, 52, and
41.

$$\begin{array}{r|l} 2 & 9 \\ 3 & 1\ 4\ 4\ 4\ 7\ 8 \\ 4 & 1\ 2\ 3\ 4 \\ 5 & 2 \end{array}$$

Example 13

Make a stem-and-leaf plot of this set of data:
92, 82, 80, 92, 78, 75, 95, 77, and 77.

$$\begin{array}{r|l} 7 & 5\ 7\ 7\ 8 \\ 8 & 0\ 2 \\ 9 & 2\ 2\ 5 \end{array}$$

Problems

Make a stem-and-leaf plot of each set of data.

21. 29, 28, 34, 30, 33, 26, 18, and 34.

22. 25, 34, 27, 25, 19, 31, 42, and 30.

23. 80, 89, 79, 84, 95, 79, 89, 67, 82, 76,
92, 89, 81, and 123.

24. 116, 104, 101, 111, 100, 107, 113, 118,
113, 101, 108, 109, 105, 103, and 91.

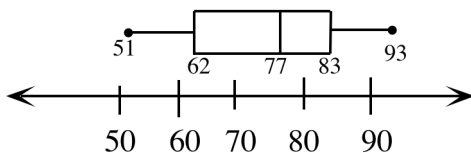
BOX-AND-WHISKER PLOTS

A way to display data that shows how the data is grouped or clustered is a **box-and-whisker plot**. The box-and-whisker plot displays the data using **quartiles**.

Example 14

Display this data in a box-and-whisker plot:
51, 55, 55, 62, 65, 72, 76, 78, 79, 82, 83, 85,
91, and 93.

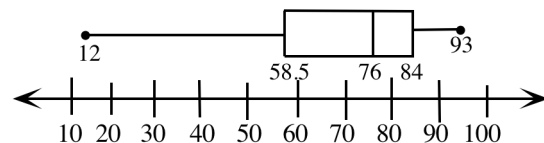
- Since this data is already in order from least to greatest, it can be seen that the range is $93 - 51 = 42$. Thus you start with a number line with equal intervals from 50 to 95.
- The median of the set of data is 77. A line is drawn at this value above the number line.
- The median of the lower half of the data (**the lower quartile**) is 62. A line is drawn at this value above the number line.
- The median of the upper half of the data (**the upper quartile**) is 83. A line is drawn at this value above the number line.
- A box is drawn between the upper and lower quartiles.
- Place a dot at the minimum value (51) and a dot at the maximum value (93). The lines which connect these dots to the box are called the whiskers.



Example 15

Display this data in a box-and-whisker plot:
62, 65, 93, 51, 12, 79, 85, 55, 72, 78, 83, 91,
and 76.

- Place the data in order from least to greatest: 12, 51, 55, 62, 65, 72, 76, 78, 79, 83, 85, 91, 93. The range is $93 - 12 = 81$. Thus you want a number line with equal intervals from 10 to 100.
- Find the median of the set of data: 76. Draw the line.
- Find the lower quartile: $55 + 62 = 117$; $117 \div 2 = 58.5$. Draw the line.
- Find the upper quartile: $83 + 85 = 168$; $168 \div 2 = 84$. Draw the line.
- Draw the box connecting the upper and lower quartiles. Place a dot at the minimum value (12) and a dot at the maximum value (93). Draw the whiskers.



Problems

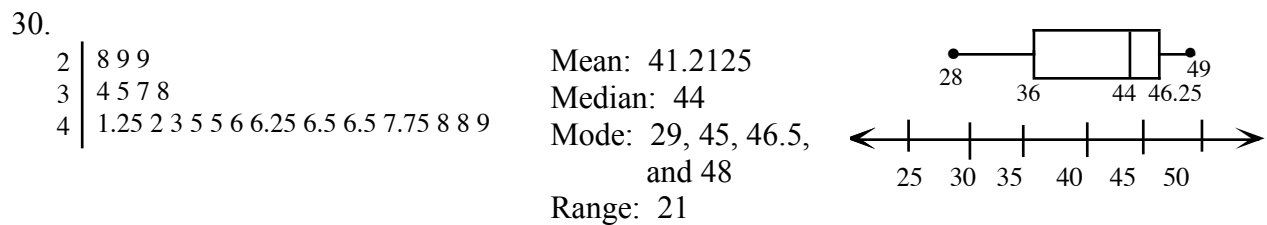
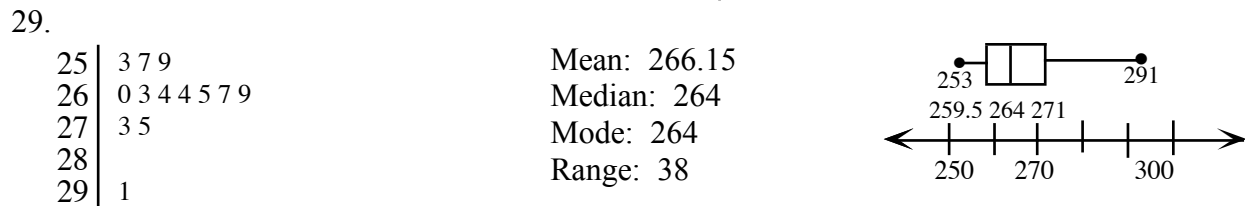
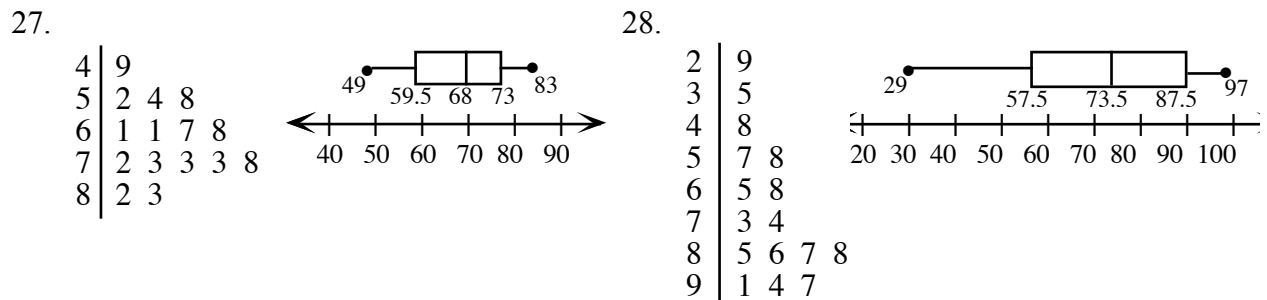
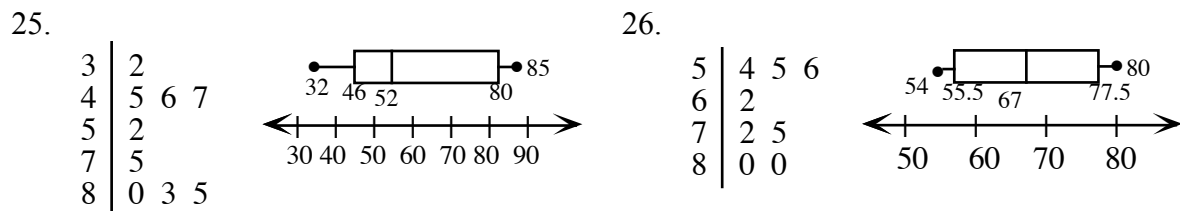
Create a stem-and-leaf plot and a box-and-whisker plot for each set of data.

25. 45, 47, 52, 85, 46, 32, 83, 80, and 75. 26. 75, 62, 56, 80, 72, 55, 54, and 80.
27. 49, 54, 52, 58, 61, 72, 73, 78, 73, 82, 28. 65, 35, 48, 29, 57, 87, 94, 68, 86, 73,
83, 73, 61, 67, and 68. 58, 74, 85, 91, 88, and 97.
29. Given a set of data: 265, 263, 269, 259, 267, 264, 253, 275, 264, 260, 273, 257, and 291.
- a) Make a stem-and-leaf plot of this data.
 - b) Find the mean, median, and mode of this data.
 - c) Find the range of this data.
 - d) Make a box-and-whisker plot for this data.
30. Given a set of data: 48, 42, 37, 29, 49, 46, 38, 28, 45, 45, 35, 46.25, 34, 46, 46.5, 43, 46.5, 48, 41.25, 29, and 47.75.
- a) Make a stem-and-leaf plot of this data.
 - b) Find the mean, median, and mode of this data.
 - c) Find the range of the data.
 - d) Make a box-and-whisker plot for this data.

Answers

1. 30.57 2. 29.57 3. ≈ 86.1 4. ≈ 106.7
 5. 34 6. 25 7. 89 8. 101 and 113
 9. 30 10. 30 11. 82 12. 7
 13. 8 14. 9 15. 47 16. 27
 17. 93 18. 61 19. 3754 20. 14

21.
$$\begin{array}{r|l} 2 & 689 \\ 3 & 0344 \end{array}$$
 22.
$$\begin{array}{r|l} 2 & 557 \\ 3 & 0145 \end{array}$$
 23.
$$\begin{array}{r|l} 7 & 66899 \\ 8 & 0124999 \\ 9 & 25 \\ 10 & \\ 11 & \\ 12 & 3 \end{array}$$
 24.
$$\begin{array}{r|l} 9 & 1 \\ 10 & 011345789 \\ 11 & 13368 \end{array}$$



SURVEYS AND SAMPLING POPULATIONS

Before drawing conclusions from data, it is important to know that the data is not biased and is drawn from a representative sample.

A **biased** question will influence the results by favoring one position over others. Consider the following pairs of questions:

- A. Do you support the President's education plan?
- B. Do you support the President's education plan since it will ensure success for more students in school?

- A. Do you exercise?
- B. Moderate exercise is necessary to stay healthy. Do you exercise?

- A. Should teachers make more money?
- B. Don't you agree that teachers should make more money?

In each case, the second question deliberately attempts to influence the results.

A **population** is a collection of objects or a group of people about whom information is gathered. A **sample** is a subgroup of the population. For example, if you want to conduct a survey at your school about what foods to serve in the cafeteria, the population would be the entire student body. If you choose to sample 10% of the students, you need to include the correct fraction of students from each grade and an equal number of male and female students. A **representative sample** is a subgroup of the population that matches the general characteristics of the entire population. Asking players on the basketball team "*Do you watch sports on television?*" would not be a representative sample.

Problems

Determine the unbiased and better survey question.

1.
 - A. Do you think the frequent occurrence of brutal violence in today's movies affects young impressionable children?
 - B. Do you think violence in today's movies affects children?

2.
 - A. Who should be school president?
 - B. Don't you agree that Max should be school president?

Determine if each question is fair and unbiased.

3. Do you think that cruel people who hurt defenseless animals should be imprisoned?
4. Which soda is your favorite?
5. Which shirt color is your favorite? 1) blue 2) beige 3) brown 4) tan

Determine if this appears to be a representative sample for the question asked. If not, why not?

6. Ask every tenth person leaving the mall *"What is your favorite store in the mall?"*
7. Call random people at home at 10 a.m. *"What is your favorite cartoon?"*
8. Ask your classmates *"What is the best bank in town?"*
9. Ask random Junior class homerooms *"When should we have the Junior Prom?"*
10. Ask the boys P.E. class *"What is the best food in the cafeteria?"*

Answers

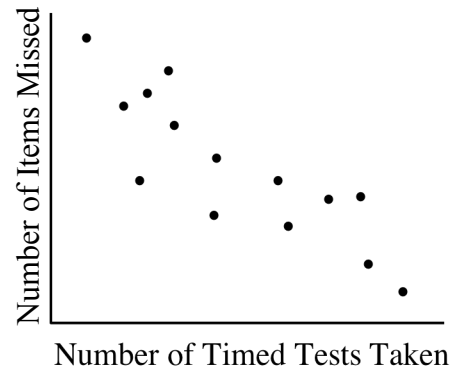
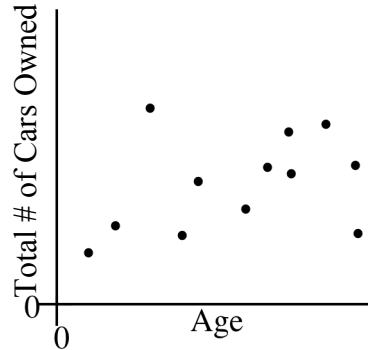
- | | | | | |
|--------|-------|-----------|---------|-----------|
| 1. B | 2. A | 3. biased | 4. fair | 5. biased |
| 6. yes | 7. no | 8. no | 9. yes | 10. no |

CORRELATION

A **scatter plot**, like the one at the top right, appears to have a **positive correlation**, since people seem to own more cars as they increase in age.

The second scatter plot shows a **negative correlation**, where the number of incorrect items on tests decreases as more tests are taken. A positive correlation exists if the items on both axes tend to increase. A negative correlation exists if the items on the y -axis decrease as the items on the x -axis increase. The closer the points come to approaching a straight line, the stronger the correlation.

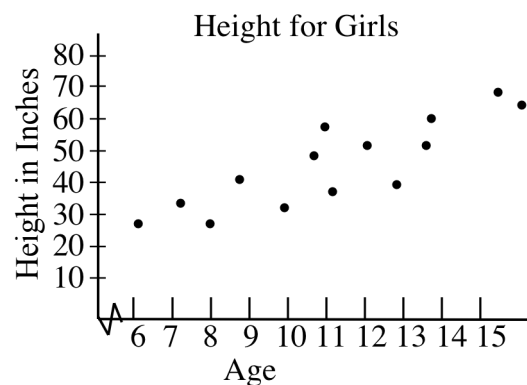
One caution: It is easy to jump to the conclusion that if there is a strong correlation, one factor **causes** the other. This is not necessarily true. For example, a scatter plot showing the relationship of height to reading ability might show a very strong correlation but increased height does not increase reading ability, nor does increased reading ability cause an increase in height. In this case both height and reading ability are age related.



Problems

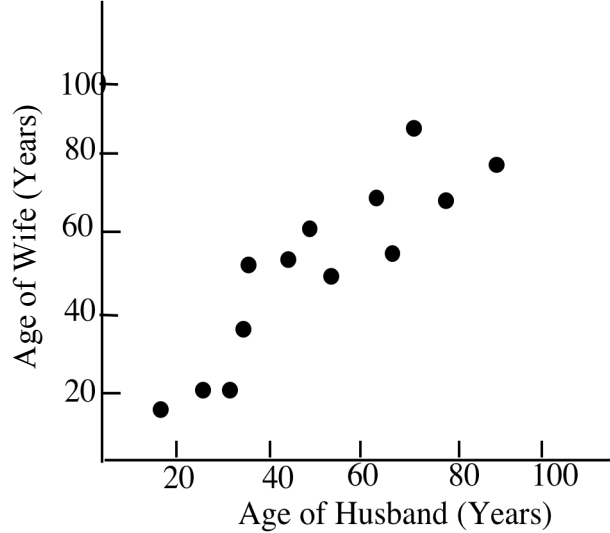
Use the scatter plot at right to answer the following questions:

1. Is there a correlation between age and height in girls?
2. If there is a correlation, is it positive or negative?
3. What does your answer to question 2 mean?
4. What will happen to the scatter plot if the x -axis is increased to age 40?



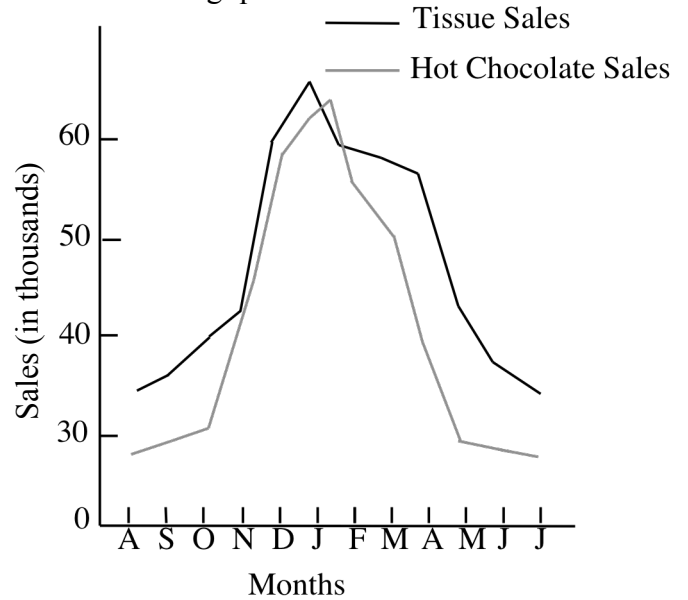
Look at the scatter plot below showing a husband's age compared to his wife's age.

5. Is this a positive or negative correlation?
A **negative correlation** means that as one variable increases in value, the other variable decreases in value.
6. What does the graph tell you about the age of a husband compared to the age of his wife?
7. Does an increase in the husband's age cause an increase in the wife's age?



In the preceding problem, the events are not caused by each other; each one is caused by another factor. Look at the double-line graph below to answer the following questions.

8. Do the sales of tissues and hot chocolate appear to be correlated or somehow connected?
9. Does the sale of tissue cause the increase in the sale of hot chocolate? Explain.
10. Why does the data seem to be related as it appears in the line graph?



Answers

1. yes
2. positive
3. Girls increase in height as they age.
4. The correlation would continue until it leveled off somewhere between ages 15 and 20. From that point, height should stay the same since growth has ended.
5. positive
6. The husband's age goes up and so does the age of the wife.
7. No; but people tend to marry someone about their age, and we all get older at the same rate.
8. yes
9. No; the data shows that the sales go up and down at the same time of year, but one event does not cause the other.
10. Because of the time of year; in the winter there is more need for tissue because of colds, and in the winter people also drink more hot chocolate.