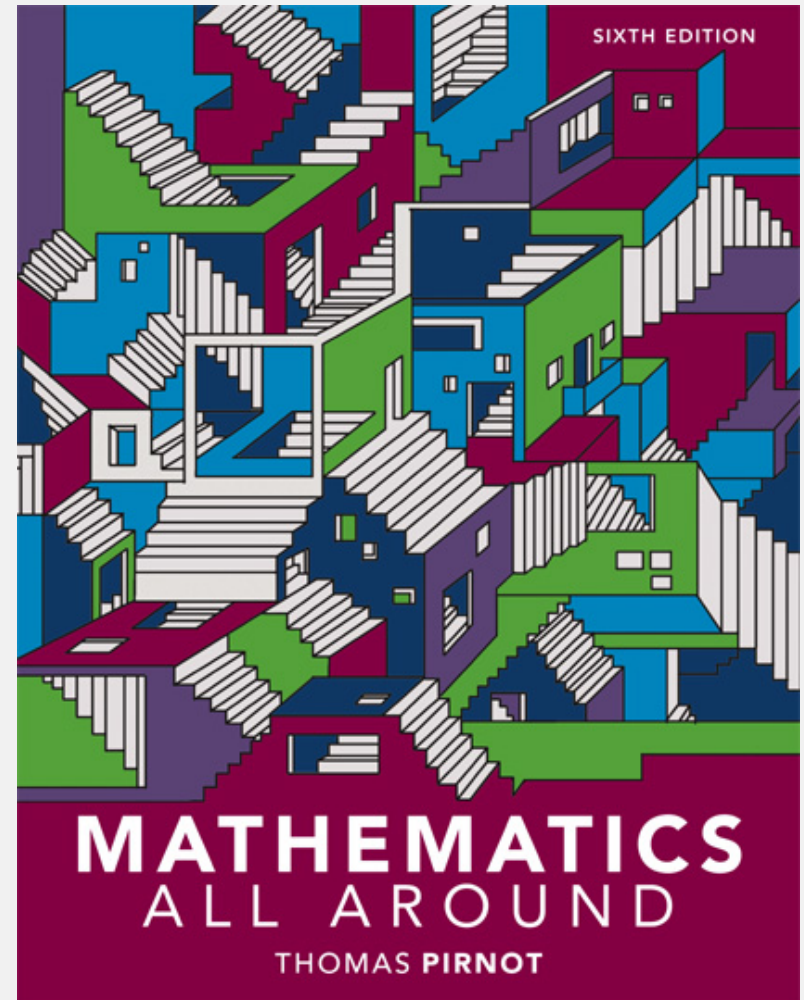


14.2

Descriptive Statistics Measures of Central Tendency



14.2 Measures of Central Tendency

- Compute the mean, median, and mode of distributions.
- Find the five-number summary of a distribution.
- Apply measures of central tendency to compare data.

The Mean and the Median

COMPUTING THE MEAN If a data set contains n data values, the **mean** \bar{x} of the data set is

$$\bar{x} = \frac{\sum x}{n}.$$

We use the Greek letter Σ (capital sigma) to indicate a sum. For example, we will write the sum of the data values 7, 2, 9, 4, and 10 by $\Sigma x = 7 + 2 + 9 + 4 + 10$.

We represent the mean of a *sample* of a population by \bar{x} (read as “x bar”), and we will use the Greek letter μ (lowercase mu) to represent the mean of the *whole population*.

Example: Finding the Mean Number of Pounds of Compost

The Bonnaroo Music and Arts Festival is an annual four-day music festival held in Manchester, Tennessee. At festivals such as Bonnaroo, organizers are committed to being green and they document their sustainability efforts. Imagine that the amounts of compost collected (in pounds) at five different recycling stations during the first day of a festival are found to be 40, 42, 65, 51, and 55. Find the mean pounds of compost generated at these five stations on the first day of the festival.

Example: Finding the Mean Number of Pounds of Compost (cont)

Solution

To calculate the mean, we add the weights from each of the stations and divide by 5, as follows:

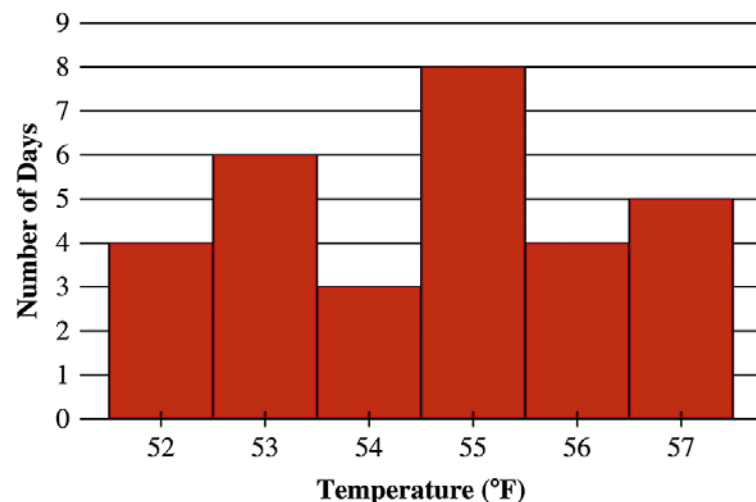
$$\bar{x} = \frac{\sum x}{n} = \frac{40 + 42 + 65 + 51 + 55}{5} = \frac{253}{5} = 50.6$$

This tells us that on the first day of the festival, there was an average of 50.6 pounds of compost at these five stations.

Example: Computing the Mean of a Frequency Distribution of Water Temperatures

The Environmental Protection Agency (EPA) suspects that hot water discharged from a nuclear power plant is responsible for a recent fish kill. To investigate this problem, the agency has recorded the water temperature

at a point downstream from the plant for the last 30 days. The graph summarizes the information obtained. What is the mean temperature for this distribution?



Example: Computing the Mean of a Frequency Distribution of Water Temperatures (cont)

Solution We see that 52 occurs four times in the data, so instead of summing $52 + 52 + 52 + 52$, we can write $52 \cdot 4$ instead. The table lists the products of the temperatures and their frequencies.

Temperature ($^{\circ}\text{F}$), x	Frequency, f	Product, $x \cdot f$
52	4	$52 \cdot 4 = 208$
53	6	$53 \cdot 6 = 318$
54	3	$54 \cdot 3 = 162$
55	8	$55 \cdot 8 = 440$
56	4	$56 \cdot 4 = 224$
57	5	$57 \cdot 5 = 285$
Totals	$\Sigma f = 30$	$\Sigma (x \cdot f) = 1,637$

sum of frequencies

sum of products

Example: Computing the Mean of a Frequency Distribution of Water Temperatures (cont)

The sum of the frequencies, $\sum f$, is the same as n , the total number of values in the distribution, and $\sum(x \cdot f)$ is the sum of the data in the distribution. Therefore, the mean of the distribution is

$$\frac{\sum(x \cdot f)}{\sum f} = \frac{\text{sum of the data}}{\text{number of values}} = \frac{1,637}{30} \approx 54.6^{\circ}\text{F}.$$

With this information, the EPA may conclude that the water is not too hot and look for another reason for the fish kill.

The Mean and the Median

COMPUTING THE MEAN OF A FREQUENCY DISTRIBUTION We use a frequency table to compute the mean of a data set as follows:

1. Write all products $x \cdot f$ of the values and their frequencies in a new column of the table.
2. Represent the sum of the products you calculated in step 1 by $\Sigma (x \cdot f)$.
3. Denote the sum of the frequencies by Σf .
4. The mean is then $\frac{\Sigma (x \cdot f)}{\Sigma f}$.



Example: The Effect of Extreme Scores on the Mean

The table lists the estimated numbers of unique visitors to the top social media sites during a recent month.

- a) What is the mean number of unique visitors?
- b) Is this mean an accurate measure of the “average” number of unique visitors to these social media sites?

Rank	Visitors (in millions)
1	1,100
2	1,000
3	310
4	225
5	250
6	120
7	110
8	100
9	85
10	80

Example: The Effect of Extreme Scores on the Mean (cont)

Solution

- a) Summing the numbers of visitors to the sites and dividing by 10 give us the mean:

$$\begin{aligned}\bar{x} &= \frac{1,100 + 1,000 + 310 + 255 + 250 + 120 + 110 + 100 + 85 + 80}{10} \\ &= 341 \text{ million}\end{aligned}$$

Example: The Effect of Extreme Scores on the Mean (cont)

- b) To answer the second question, notice that eight of the sites have a number of monthly visitors that is below the mean, and many are quite a bit lower. Therefore, the mean in this example does not give an accurate sense of what is “average” in this set of data because it is highly influenced by the top two social media sites, which have unusually high values.

The Mean and the Median

COMPUTING THE MEDIAN If we arrange a set of data values in increasing (or decreasing) order, **the median is the middle value in the list of values**. However, there are two cases to consider.

1. If there is an odd number of values, then the median is the value in the middle position.
2. If there is an even number of values, then the median is the mean of the two middle values.

Example: Finding the Median of Presidents' Ages

The table lists the ages at inauguration of the presidents who assumed office between 1901 and 2008. Find the median age for this distribution to see how Barack Obama's age at inauguration compares with that of other presidents.

T. Roosevelt	42
Taft	51
Wilson	56
Harding	55
Coolidge	51
Hoover	54
F. D. Roosevelt	51
Truman	60
Eisenhower	61
Kennedy	43
L. Johnson	55
Nixon	56
Ford	61
Carter	52
Reagan	69
G. H. W. Bush	64
Clinton	46
G. W. Bush	54
Obama	47

Example: Finding the Median of Presidents' Ages (cont)

Solution

The distribution of ages, when arranged in increasing order, is

42, 43, 46, 47, 51, 51, 51, 52, 54, 54, 55, 55, 56, 56, 60, 61, 61, 64, 69.

There are 19 scores in this distribution, the middle score is the tenth score, which is 54. Since Obama's age when he took office is well below the median, we agree with the opinion that he was young when he took office.

T. Roosevelt	42
Taft	51
Wilson	56
Harding	55
Coolidge	51
Hoover	54
F. D. Roosevelt	51
Truman	60
Eisenhower	61
Kennedy	43
L. Johnson	55
Nixon	56
Ford	61
Carter	52
Reagan	69
G. H. W. Bush	64
Clinton	46
G. W. Bush	54
Obama	47

Example: Using a Frequency Table to Find the Median

Acting on a tip from a dissatisfied customer, an agent of the Consumer Protection Agency purchased 50 quarts of a particular brand of milk at various supermarkets to see whether they contained 32 ounces of milk. What is the median for this distribution?

Number of Ounces, x	Frequency, f
27	8
28	5
29	12
30	16
31	9
	$\Sigma f = 50$

Example: Using a Frequency Table to Find the Median (cont)

Solution

Because the 50 are in increasing order, the two middle scores are in positions 25 and 26. Counting the frequencies, we see that 29 ounces is in position 25 and 30 ounces is in position 26. Therefore, the median for this distribution is

$$\frac{29 + 30}{2} = 29.5.$$

Number of Ounces, <i>x</i>	Frequency, <i>f</i>
27	8
28	5
29	12
30	16
31	9
	$\Sigma f = 50$

Example: Using a Frequency Table to Find the Median (cont)

Because the median is far below the advertised 32 ounces, the variation is probably not due to randomness in filling the containers and perhaps the company packaging this milk should be fined.



The Five Number Summary

DEFINITION The Five-Number Summary

The median divides a data set into two halves. The set of numbers below the median is called the **lower half** and the set of numbers above the median is called the **upper half**. The median of the lower half is called the **first quartile** and is indicated by Q_1 ; the **third quartile**, denoted by Q_3 , is the median of the upper half. The **five-number summary** of a set of data values consists of the following:

minimum value, Q_1 , median, Q_3 , maximum value

Example: Finding the Five-Number Summary of Presidents' Ages

Consider the list of ages of the presidents from a previous example.

42, 43, 46, 47, 51, 51, 51, 52, 54, 54, 55, 55, 56, 56, 60, 61, 61, 64, 69.

Find the following for this data set:

- a) the lower and upper halves
- b) the first and third quartiles
- c) the five-number summary

Example: Finding the Five-Number Summary of Presidents' Ages (cont)

It helps to consider the following diagram:

42,43,46,47,51,51,51,52,54, 54, 55,55,56,56,60,61,61,64,69

Lower half Upper Half

- a) Because the median is in the tenth position in the list, the lower half is 42, 43, 46, 47, 51, 51, 51, 52, 54 and the upper half is 55, 55, 56, 56, 60, 61, 61, 64, 69.

Example: Finding the Five-Number Summary of Presidents' Ages (cont)

42, 43, 46, 47, 51, 51, 51, 52, 54, 54, 55, 55, 56, 56, 60, 61, 61, 64, 69

Lower half Upper Half

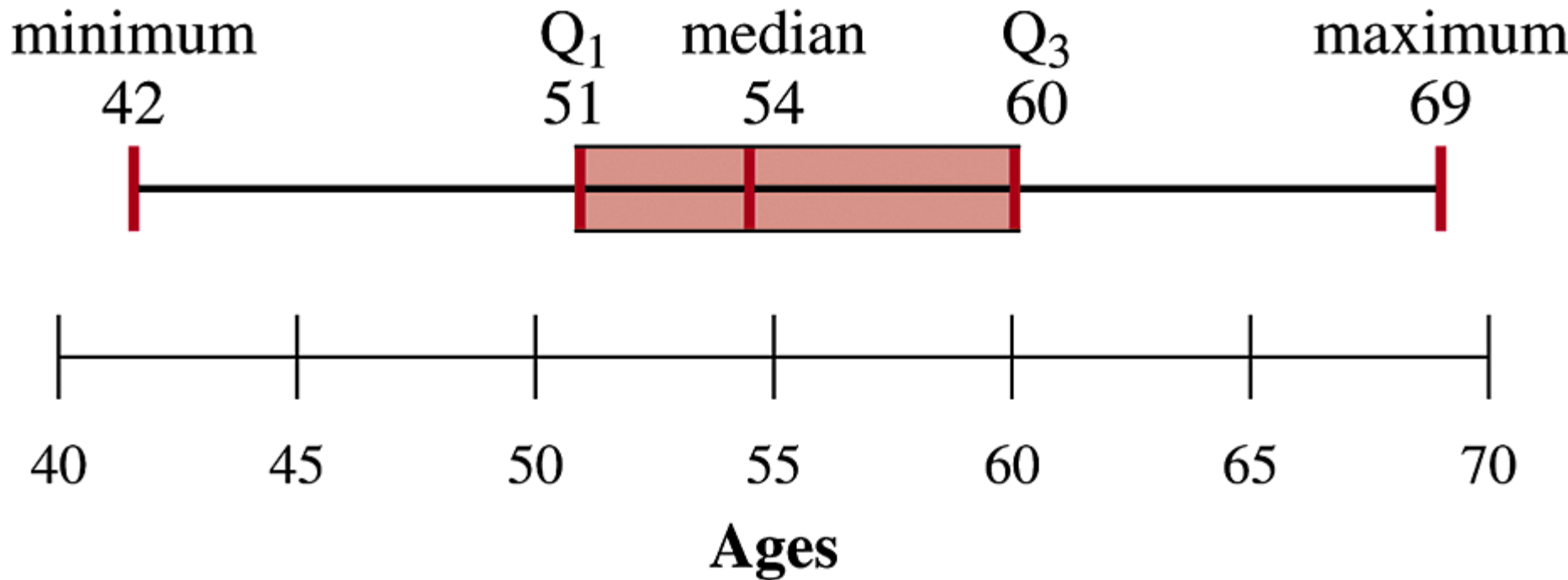
- b) The first quartile, the median of the lower half, is 51 (circled). So, $Q_1 = 51$. The third quartile is the median of the upper half, which is 60 (circled). Thus $Q_3 = 60$.
- c) The five-number summary for this data set is 42, 51, 54, 60, 69.

Constructing a Box-and-Whisker Plot

1. Determine the values for the five-number summary.
2. Draw a horizontal axis from the minimum value to the maximum value.
3. Draw a rectangular box over the axis from Q_1 to Q_3 . The height of the box is not important.
4. Draw a vertical line in the box where the median is located (note that it might not be in the middle of the box).
5. Draw horizontal lines extending from the edges of the boxes (the quartiles) to the minimum and maximum values. These are the whiskers.

The Five Number Summary

We represent the five-number summary by a graph called a *box-and-whisker plot*.



Computing the Mode

The **mode** of a set of data is the data value that occurs most frequently. It is possible for a data set to have more than one mode. If no one value occurs more often than others, the data set is said to have no mode.

Example: Finding the mode of a data set

Find the mode for each data set.

a) {2, 1, 6, 9, 6, 11}

The mode is 6 because it occurs twice.

b) {4, 6, 2, 8, 6, 9, 4, 3}

The modes are 4 and 6.

c) {2, 1, 5, 6, 8}

There is no mode.

d) {A, B, B, C, D, F}

The mode is B.

Example: Which Measure of Central Tendency Is the Best?

Assume that you are negotiating the contract for your union at Magnum Industrial Corporation. To prepare for the next negotiation session, you have gathered annual wage data and found that three workers earn \$30,000, five workers earn \$32,000, three workers earn \$44,000, and one worker earns \$50,000. In your negotiations, which measure of central tendency should you emphasize?

Example: Which Measure of Central Tendency Is the Best? (cont)

Solution

You want to make the wages seem as low as possible; therefore, you should choose the smallest measure of central tendency. We list the frequency distribution of wages.

Salary (thousands \$), x	30	32	44	50
Frequency, f	3	5	3	1

Example: Which Measure of Central Tendency Is the Best? (cont)

Mode: \$32,000

Mean:

$$\bar{x} = \frac{\sum(x \cdot f)}{\sum f} = \frac{30 \cdot 3 + 32 \cdot 5 + 44 \cdot 3 + 50 \cdot 1}{12} = \frac{432}{12} = 36$$

The mean is therefore 36.

It is to the union's benefit to present the median (or the mode) of \$32,000. No doubt management will claim that the mean of \$36,000 is the average salary.