STATISTICS

Descriptive Statistics

Chapter Outline

- 2.1 Frequency Distributions and Their Graphs
- 2.2 More Graphs and Displays
- 2.3 Measures of Central Tendency
- 2.4 Measures of Variation
- 2.5 Measures of Position

Section 2.1

Frequency Distributions and Their Graphs

Frequency Distribution (1 of 3)

Frequency Distribution

- A table that shows **classes** or **intervals** of data with a count of the number of entries in each class.
- The **frequency, *f*,** of a class is the number of data entries in the class.

Class	Frequency, *f*
1 - 5	5
6 - 10	8
11 - 15	6
16 - 20	8
21 - 25	5
26 - 30	4

Frequency Distribution (2 of 3)

- Each class has a **lower class limit,** which is the least number that can belong to the class, and an
- **upper class limit,** which is the greatest number that can belong to the class.

Class	Frequency, *f*
1 - 5	5
6 - 10	8
11 - 15	6
16 - 20	8
21 - 25	5
26 - 30	4

Lower class limits

Upper class limits

Frequency Distribution (3 of 3)

- The **class width** is the distance between lower (or upper) limits of consecutive classes.
- The difference between the maximum and minimum data entries is called the **range.**

Class width

$6-1=5$

Class	Frequency, *f*
1 - 5	5
6 - 10	8
11 - 15	6
16 - 20	8
21 - 25	5
26 - 30	4

Constructing a Frequency Distribution (1 of 3)

1. Decide on the number of classes.
   - Usually between 5 and 20; otherwise, it may be difficult to detect any patterns.
2. Find the class width.
   - Determine the range of the data.
   - Divide the range by the number of classes.
   - *Round up to the next convenient number.*

Constructing a Frequency Distribution (2 of 3)

3. Find the class limits.
   - You can use the minimum data entry as the lower limit of the first class.
   - Find the remaining lower limits (add the class width to the lower limit of the preceding class).
   - Find the upper limit of the first class. Remember that classes cannot overlap.
   - Find the remaining upper class limits.

Constructing a Frequency Distribution (3 of 3)

4. Make a tally mark for each data entry in the row of the appropriate class.
5. Count the tally marks to find the total frequency *f* for each class.

Example: Constructing a Frequency Distribution

The data set lists the out-of-pocket prescription medicine expenses (in dollars) for 30 adults in a recent year. Construct a frequency distribution that has seven classes.

200 239 155 252 384 165 296 405 303 400
307 241 256 315 330 317 352 266 276 345
238 306 290 271 345 312 293 195 168 342

Solution: Constructing a Frequency Distribution (1 of 4)

200 239 155 252 384 165 296 405 303 400
307 241 256 315 330 317 352 266 276 345
238 306 290 271 345 312 293 195 168 342

1. $\text{Number of classes} = 7(\text{given})$
2. Find the class width

$$\frac{\text{Range}}{\#\,\text{classes}} = \frac{\max - \min}{\#\text{classes}} = \frac{405 - 155}{7} = \frac{250}{7} \approx 35.71$$

Round up to 36

Solution: Constructing a Frequency Distribution (2 of 4)

3. Use 155 (minimum value) as first lower limit. Add the class width of 36 to get the lower limit of the next class.

   $155 + 36 = 191$

   Find the remaining lower limits.

$\text{Class width} = 36$

Lower limit	Upper limit
155 $+36 =$	190
191	226
227	262
263	298
299	334
335	370
371	406

Solution: Constructing a Frequency Distribution (3 of 4)

The upper limit of the first class is 190 (one less than the lower limit of the second class).

Add the class width of 36 to get the upper limit of the next class.

$$190 + 36 = 226$$

Find the remaining upper limits.

Lower limit	Upper limit
155	190
191	226
227	262
263	298
299	334
335	370
371	406

$\text{Class width} = 36$

Solution: Constructing a Frequency Distribution (4 of 4)

4. Make a tally mark for each data entry in the row of the appropriate class.
5. Count the tally marks to find the total frequency *f* for each class.

Expenses

Number of adults

Class	Tally	Frequency, f
155–190	|||	3
191–226	||	2
227–262	卌	5
263–298	卌 |	6
299–334	卌 ||	7
335–370	||||	4
371–406	|||	3
		$\Sigma f = 30$

Check that the sum of the frequencies equals the number in the sample.

Determining the Midpoint

Midpoint of a class

$$\frac{(\text{Lower class limit}) + (\text{Upper class limit})}{2}$$

$$\text{Midpoint} = \frac{155 + 190}{2} = 172.5$$ Midpoint of first class.

Using the class width of 36, the remaining mid-points are

$172.5 + 36 = 208.5$ Midpoint of second class

$208.5 + 36 = 244.5$ Midpoint of third class

and so on.

Determining the Relative Frequency

Relative Frequency of a class

- Portion or percentage of the data that falls in a particular class.
- $\text{relative frequency} = \dfrac{\text{class frequency}}{\text{Sample size}} = \dfrac{f}{n}$

Note that $n = \Sigma f$.

Determining the Cumulative Frequency

Cumulative frequency of a class

- The sum of the frequency for that class and all previous classes.
- The cumulative frequency of the last class is equal to the sample size *n*.

Example: Finding Midpoints, Relative and Cumulative Frequencies

Using the frequency distribution in previous example, find the midpoint, relative frequency, and cumulative frequency of each class. Describe any patterns.

Solution: Finding Midpoints, Relative and Cumulative Frequencies (1 of 3)

- The midpoints, relative frequencies, and cumulative frequencies of the first five classes are calculated as follows:

Class	f	Midpoint	Relative frequency	Cumulative frequency
155–190	3	$\frac{155+190}{2} = 172.5$	$\frac{3}{30} = 0.1$	3
191–226	2	$\frac{191+226}{2} = 208.5$	$\frac{2}{30} \approx 0.07$	$3 + 2 = 5$
227–262	5	$\frac{227+262}{2} = 244.5$	$\frac{5}{30} \approx 0.17$	$5 + 5 = 10$
263–298	6	$\frac{263+298}{2} = 280.5$	$\frac{6}{30} = 0.2$	$10 + 6 = 16$
299–334	7	$\frac{299+334}{2} = 316.5$	$\frac{7}{30} \approx 0.23$	$16 + 7 = 23$

Solution: Finding Midpoints, Relative and Cumulative Frequencies (2 of 3)

- The remaining midpoints, relative frequencies, and cumulative frequencies are shown in the expanded frequency distribution below.

Frequency Distribution for Out-of-Pocket Prescription Medicine Expenses (in dollars)

Expenses

Number of adults

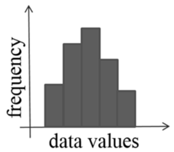
Portion of adults

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
155–190	3	172.5	0.1	3
191–226	2	208.5	0.07	5
227–262	5	244.5	0.17	10
263–298	6	280.5	0.2	16
299–334	7	316.5	0.23	23
335–370	4	352.5	0.13	27
371–406	3	388.5	0.1	30
	$\Sigma f = 30$		$\Sigma \frac{f}{n} = 1$	

Graphs of Frequency Distributions (1 of 4)

Frequency Histogram

- A bar graph that represents the frequency distribution.
- The horizontal scale is quantitative and measures the data values.
- The vertical scale measures the frequencies of the classes.
- Consecutive bars must touch.



Class Boundaries

Class boundaries

- Because consecutive bars of a histogram must touch, bars must begin and end at class boundaries instead of class limits.
- The numbers that separate classes *without* forming gaps between them.

Example: Constructing a Frequency Histogram

Draw a frequency histogram for the frequency distribution in the previous example. Describe any patterns.

Frequency Distribution for Out-of-Pocket Prescription Medicine Expenses (in dollars)

Expenses

Number of adults

Portion of adults

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
155–190	3	172.5	0.1	3
191–226	2	208.5	0.07	5
227–262	5	244.5	0.17	10
263–298	6	280.5	0.2	16
299–334	7	316.5	0.23	23
335–370	4	352.5	0.13	27
371–406	3	388.5	0.1	30
	$\Sigma f = 30$		$\Sigma \frac{f}{n} = 1$	

Solution: Constructing a Frequency Histogram (1 of 3)

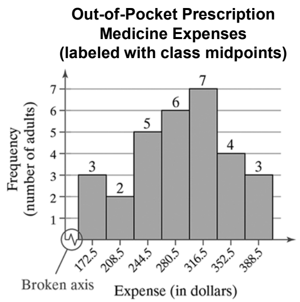
- First, find the class boundaries
- The distance from the upper limit of the first class to the lower limit of the second class is $191 - 190 = 1$.
- Half this distance is 0.5.

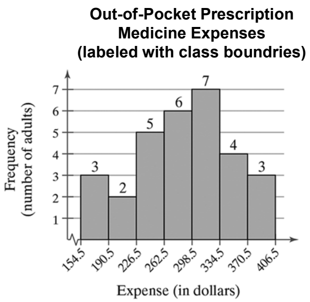
Class	Class boundaries	Frequency, f
155-190	154.5-190.5	3
191-226	190.5-226.5	2
227-262	226.5-262.5	5
263-298	262.5-298.5	6
299-334	298.5-334.5	7
335-370	334.5-370.5	4
371-406	370.5-406.5	3

- First class lower boundary $= 155 - 0.5 = 154.5$
- First class upper boundary $= 190 + 0.5 = 190.5$

Solution: Constructing a Frequency Histogram (2 of 3)

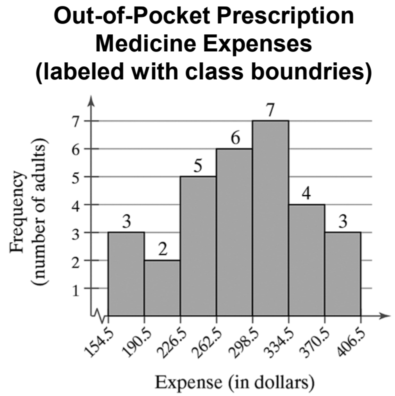
You can mark the horizontal scale either at the midpoints or at the class boundaries. Both histograms are shown below.





Solution: Constructing a Frequency Histogram (3 of 3)

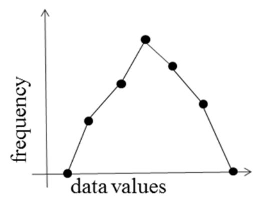
You can see that two-thirds of the adults are paying more than $262.50 for out-of-pocket prescription medicine expenses.



Graphs of Frequency Distributions (2 of 4)

Frequency Polygon

- A line graph that emphasizes the continuous change in frequencies.



Example: Constructing a Frequency Polygon

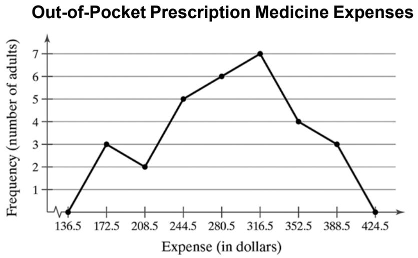
Draw a frequency polygon for the frequency distribution in previous example. Describe any patterns.

Class	Class boundaries	Frequency, *f*
155-190	154.5-190.5	3
191-226	190.5-226.5	2
227-262	226.5-262.5	5
263-298	262.5-298.5	6
299-334	298.5-334.5	7
335-370	334.5-370.5	4
371-406	370.5-406.5	3

Solution: Constructing a Frequency Polygon (1 of 2)

To construct the frequency polygon, use the same horizontal and vertical scales that were used in the histogram labeled with class midpoints.

The graph should begin and end on the horizontal axis, so extend the left side to one class width before the first class midpoint and extend the right side to one class width after the last class midpoint.



Slide 31

Graphs of Frequency Distributions (3 of 4)

Relative Frequency Histogram

- Has the same shape and the same horizontal scale as the corresponding frequency histogram.
- The vertical scale measures the **relative frequencies**, not frequencies.

Slide 33

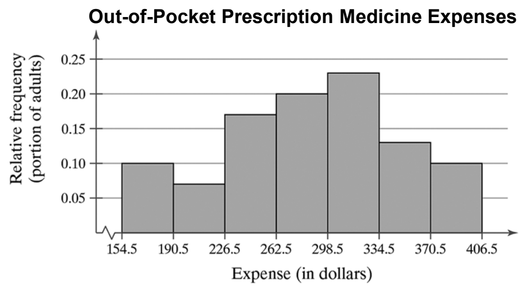
Example: Constructing a Relative Frequency Histogram

Construct a relative frequency histogram for the second example.

Class	Class boundaries	Frequency, *f*
155-190	154.5-190.5	3
191-226	190.5-226.5	2
227-262	226.5-262.5	5
263-298	262.5-298.5	6
299-334	298.5-334.5	7
335-370	334.5-370.5	4
371-406	370.5-406.5	3

Slide 34

Solution: Constructing a Relative Frequency Histogram



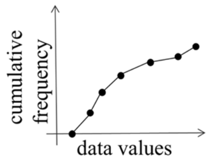
From this graph, you can quickly see that 0.2, or 20%, of the adults have expenses between \$262.50 and \$298.50.

Slide 35

Graphs of Frequency Distributions (4 of 4)

Cumulative Frequency Graph or Ogive

- A line graph that displays the cumulative frequency of each class at its upper class boundary.
- The upper boundaries are marked on the horizontal axis.
- The cumulative frequencies are marked on the vertical axis.



Constructing an Ogive (1 of 2)

1. Construct a frequency distribution that includes cumulative frequencies as one of the columns.
2. Specify the horizontal and vertical scales.
   - The horizontal scale consists of the upper class boundaries.
   - The vertical scale measures cumulative frequencies.
3. Plot points that represent the upper class boundaries and their corresponding cumulative frequencies.

Constructing an Ogive (2 of 2)

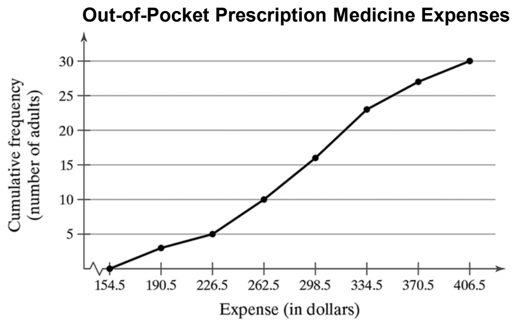
4. Connect the points in order from left to right.
5. The graph should start at the lower boundary of the first class (cumulative frequency is zero) and should end at the upper boundary of the last class (cumulative frequency is equal to the sample size).

Example: Constructing an Ogive

Construct an ogive for the second example frequency distribution.

Upper class boundary	*f*	Cumulative frequency
190.5	3	3
226.5	2	5
262.5	5	10
298.5	6	16
334.5	7	23
370.5	4	27
406.5	3	30

Solution: Constructing an Ogive



From the ogive, you can see that 10 adults had expenses of $262.50 or less. Also, the greatest increase in cumulative frequency occurs between $298.50 and $334.50.

Slide 40

Section 2.2

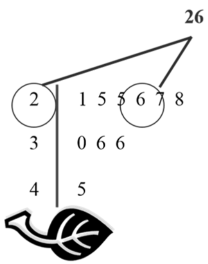
More Graphs and Displays

Slide 48

Graphing Quantitative Data Sets (1 of 2)

Stem-and-leaf plot

- Each number is separated into a **stem** and a **leaf**.
- Similar to a histogram.
- Still contains original data values.
- Provides an easy way to sort data.



Data: 21, 25, 25, **26**, 27, 28, 30, 36, 36, 45

Slide 50

Example: Constructing a Stem-and-Leaf Plot

The data set lists the numbers of text messages sent in one day by 50 cell phone users. Display the data in a stem-and-leaf plot. Describe any patterns.

Number of Text Messages Sent				
76	49	102	58	88
122	76	89	67	80
66	80	78	69	56
76	115	99	72	19
41	86	48	52	28
26	29	33	26	20
33	24	43	16	39
29	32	29	29	40
23	33	30	41	33
38	34	53	30	149

Slide 51

Solution: Constructing a Stem-and-Leaf Plot (1 of 2)

- The data entries go from a low of 16 to a high of 149.
- Use the rightmost digit as the leaf.
  - For instance,

    $76 = 7|6$ and

    $149 = 14|9$
- List the stems, 7 to 14, to the left of a vertical line.
- For each data entry, list a leaf to the right of its stem.

Number of Text Messages Sent				
76	49	102	58	88
122	76	89	67	80
66	80	78	69	56
76	115	99	72	19
41	86	48	52	28
26	29	33	26	20
33	24	43	16	39
29	32	29	29	40
23	33	30	41	33
38	34	53	30	149

Slide 52

Solution: Constructing a Stem-and-Leaf Plot (2 of 2)

Number of Text Messages Sent

Key: $10|2 = 102$

1	6 9
2	0 3 4 6 6 8 9 9 9 9
3	0 0 2 3 3 3 3 4 8 9
4	0 1 1 3 8 9
5	2 3 6 8
6	6 7 9
7	2 6 6 6 8
8	0 0 6 8 9
9	9
10	2
11	5
12	2
13	
14	9

From the display, you can see that more than 50% of the cell phone users sent between 20 and 50 text messages.

Slide 53

Example: Constructing Variations of Stem-and-Leaf Plots

Organize the data set in previous example using a stem-and-leaf plot that has two rows for each stem. Describe any patterns.

Number of Text Messages Sent				
76	49	102	58	88
122	76	89	67	80
66	80	78	69	56
76	115	99	72	19
41	86	48	52	28
26	29	33	26	20
33	24	43	16	39
29	32	29	29	40
23	33	30	41	33
38	34	53	30	149

Slide 54

Solution: Constructing Variations of Stem-and-Leaf Plots (1 of 2)

Number of Text Messages Sent

Key: $10|2 = 102$

1	
1	6 9
2	0 3 4
2	6 6 8 9 9 9 9
3	0 0 2 3 3 3 3 4
3	8 9
4	0 1 1 3
4	8 9
5	2 3
5	6 8
6	
6	6 7 9
7	2
7	6 6 6 8
8	0 0
8	6 8 9
9	
9	9
10	2
10	
11	
11	5
12	2
12	
13	
13	
14	
14	9

- List each stem twice.
- Use the leaves 0, 1, 2, 3, and 4 in the first stem row and the leaves 5, 6, 7, 8, and 9 in the second stem row.
- Notice that by using two rows per stem, you obtain a more detailed picture of the data.

Slide 55

Solution: Constructing Variations of Stem-and-Leaf Plots (2 of 2)

Number of Text Messages Sent

1		Key: 10|2 = 102
1	6 9	
2	0 3 4	
2	6 6 8 9 9 9 9	
3	0 0 2 3 3 3 3 4	
3	8 9	
4	0 1 1 3	
4	8 9	
5	2 3	
5	6 8	
6		
6	6 7 9	
7	2	
7	6 6 6 8	
8	0 0	
8	6 8 9	
9		
9	9	
10	2	
10		
11		
11	5	
12	2	
12		
13		
13		
14		
14	9	

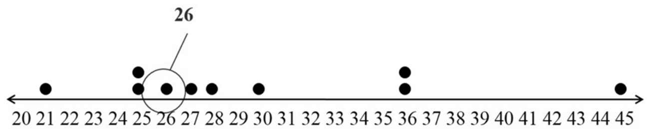
From the display, you can see that most of the cell phone users sent between 20 and 80 text messages.

Graphing Quantitative Data Sets (2 of 2)

Dot plot

- Each data entry is plotted, using a point, above a horizontal axis

Data: 21, 25, 25, **26**, 27, 28, 30, 36, 36, 45



Example: Constructing a Dot Plot

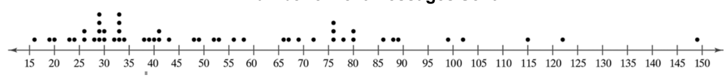
Use a dot plot to organize the data set in Example 1. Describe any patterns.

Number of Text Messages Sent									
76	49	102	58	88	122	76	89	67	80
66	80	78	69	56	76	115	99	72	19
41	86	48	52	28	26	29	33	26	20
33	24	43	16	39	29	32	29	29	40
23	33	30	41	33	38	34	53	30	149

Solution: Constructing a Dot Plot (1 of 2)

Number of Text Messages Sent									
76	49	102	58	88	122	76	89	67	80
66	80	78	69	56	76	115	99	72	19
41	86	48	52	28	26	29	33	26	20
33	24	43	16	39	29	32	29	29	40
23	33	30	41	33	38	34	53	30	149

Number of Text Messages Sent

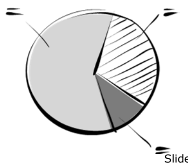


From the dot plot, you can see that most entries occur between 20 and 80 and only 4 people sent more than 100 text messages. You can also see that 149 is an unusual data entry.

Graphing Qualitative Data Sets (1 of 2)

Pie Chart

- Pie charts provide a convenient way to present qualitative data graphically as percents of a whole.
- A circle is divided into sectors that represent categories.
- The area of each sector is proportional to the frequency of each category.



Example: Constructing a Pie Chart

The numbers of earned degrees conferred (in thousands) in 2014 are shown in the table. Use a pie chart to organize the data. ***(Source: U.S. National Center for Educational Statistics)***

Earned Degrees Conferred in 2014

Type of degree	Number (in thousands)
Associate's	1003
Bachelor's	1870
Master's	754
Docoral	178

Solution: Constructing a Pie Chart (1 of 3)

- Construct the pie chart using the central angle that corresponds to each category.
  - To find the central angle, multiply $360°$ by the category's relative frequency.
  - For example, the central angle for associate's degree is

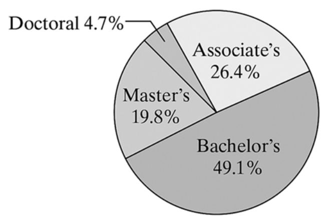
$$360°(0.264) \approx 95°$$

Solution: Constructing a Pie Chart (2 of 3)

- Find the relative frequency (percent) of each category.

Type of degree	f	Relative frequency	Angle
Associate's	1003	0.264	95°
Bachelor's	1870	0.491	177°
Master's	754	0.198	71°
Doctoral	178	0.047	17°

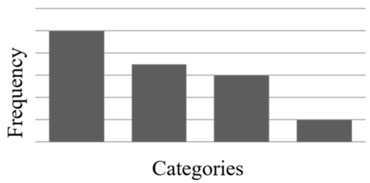
Earned Degrees Conferred in 2014



Graphing Qualitative Data Sets (2 of 2)

Pareto Chart

- A vertical bar graph in which the height of each bar represents frequency or relative frequency.
- The bars are positioned in order of decreasing height, with the tallest bar positioned at the left.



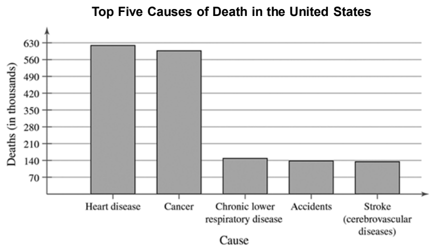
Example: Constructing a Pareto Chart

In 2014, these were the leading causes of death in the United States.

- Accidents: 136,053
- Cancer: 591,699
- Chronic lower respiratory disease: 147,101
- Heart disease: 614,348
- Stroke (cerebrovascular diseases): 133,103

Use a Pareto chart to organize the data. What was the leading cause of death in the United States in 2014? *(Source: Health, United States, 2015, Table 19)*

Solution: Constructing a Pareto Chart

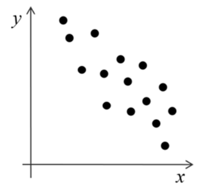


From the Pareto chart, you can see that the leading cause of death in the United States in 2014 was from heart disease. Also, heart disease and cancer caused more deaths than the other three causes combined.

Graphing Paired Data Sets (1 of 2)

Paired Data Sets

- Each entry in one data set corresponds to one entry in a second data set.
- Graph using a **scatter plot.**
  - The ordered pairs are graphed as points in a coordinate plane.
  - Used to show the relationship between two quantitative variables.

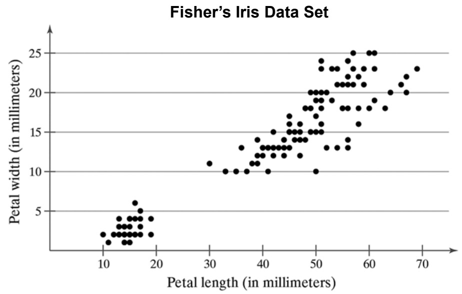


Example: Interpreting a Scatter Plot (1 of 2)

The British statistician Ronald Fisher introduced a famous data set called Fisher's Iris data set. This data set describes various physical characteristics, such as petal length and petal width (in millimeters), for three species of iris. The petal lengths form the first data set and the petal widths form the second data set. *(Source: Fisher, R. A., 1936)*

Example: Interpreting a Scatter Plot (2 of 2)

As the petal length increases, what tends to happen to the petal width?

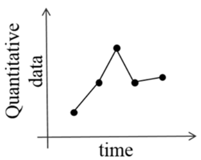


Each point in the scatter plot represents the petal length and petal width of one flower.

Graphing Paired Data Sets (2 of 2)

Time Series

- Data set is composed of quantitative entries taken at regular intervals over a period of time.
  - e.g., The amount of precipitation measured each day for one month.
- Use a **time series chart** to graph.



Example: Constructing a Time Series Chart (1 of 2)

The table lists the number of motor vehicle thefts (in millions) and burglaries (in millions) in the United States for the years 2005 through 2015. Construct a time series chart for the number of motor vehicle thefts. Describe any trends. *(Source: Federal Bureau of Investigation, Crime in the United States)*

Year	Motor vehicle thefts (in millions)	Burglaries (in millions)
2005	1.24	2.16
2006	1.20	2.19
2007	1.10	2.19
2008	0.96	2.23
2009	0.80	2.20
2010	0.74	2.17
2011	0.72	2.19
2012	0.72	2.11
2013	0.70	1.93
2014	0.69	1.71
2015	0.71	1.58

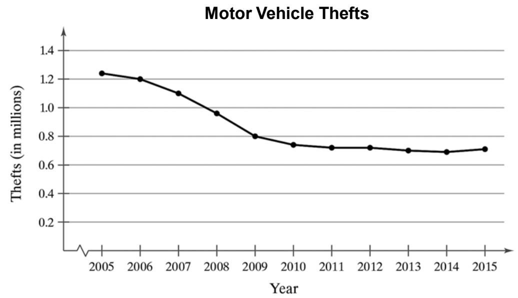
Solution: Constructing a Time Series Chart (1 of 2)

- Let the horizontal axis represent the years and let the vertical axis represent the number of motor vehicle thefts (in millions).
- Then plot the paired data and connect them with line segments

Year	Motor vehicle thefts (in millions)	Burglaries (in millions)
2005	1.24	2.16
2006	1.20	2.19
2007	1.10	2.19
2008	0.96	2.23
2009	0.80	2.20
2010	0.74	2.17
2011	0.72	2.19
2012	0.72	2.11
2013	0.70	1.93
2014	0.69	1.71
2015	0.71	1.58

Slide 75

Solution: Constructing a Time Series Chart (2 of 2)



The time series chart shows that the number of motor vehicle thefts decreased until 2011 and then remained about the same through 2015.

Slide 76

Section 2.3

Measures of Central Tendency

Slide 78

Measures of Central Tendency

Measure of central tendency

- A value that represents a typical, or central, entry of a data set.
- Most common measures of central tendency:
  - Mean
  - Median
  - Mode

Slide 80

Measure of Central Tendency: Mean

Mean (average)

- The sum of all the data entries divided by the number of entries.
- **Sigma notation**: $\sum x = \text{add all of the data entries } (x)$ in the data set.
- **Population mean**: $\mu = \dfrac{\sum x}{N}$
- **Sample mean:** $\bar{x} = \dfrac{\sum x}{n}$

Example: Finding a Sample Mean

The weights (in pounds) for a sample of adults before starting a weight-loss study are listed. What is the mean weight of the adults?

274 235 223 268 290 285 235

Solution: Finding a Sample Mean

274 235 223 268 290 285 235

- The sum of the weights is

$$\sum x = 274 + 235 + 223 + 268 + 290 + 285 + 235 = 1810$$

- To find the mean weight, divide the sum of the weights by the number of adults in the sample.

$$\bar{x} = \frac{\sum x}{n} = \frac{1810}{7} \approx 258.6$$

The mean weight of the adults is about 258.6 pounds.

Measure of Central Tendency: Median

Median

- The value that lies in the middle of the data when the data set is **ordered**.
- Measures the center of an ordered data set by dividing it into two equal parts.
- If the data set has an
  - **odd number of entries**: median is the middle data entry.
  - **even number of entries**: median is the mean of the two middle data entries.

Example: Finding the Median (1 of 2)

Find the median of the weight listed in the first example.

274 235 223 268 290 285 235

Slide 85

Solution: Finding the Median (1 of 2)

- First, order the data.

223 235 235 268 274 285 290

- There are seven entries (an odd number), the median is the middle, or fourth, data entry.

The median weight of the adults is 268 pounds.

Slide 86

Example: Finding the Median (2 of 2)

In the previous example, the adult weighing 285 pounds decides to not participate in the study. What is the median weight of the remaining adults?

223 235 235 268 274 290

Slide 87

Solution: Finding the Median (2 of 2)

- First, order the data.

223 235 235 268 274 290

- There are six entries (an even number), the median is the mean of the two middle entries.

$$\text{Median} = \frac{235 + 268}{2} = 251.5$$

The median weight of the remaining adults is 251.5 pounds.

Slide 88

Measure of Central Tendency: Mode

Mode

- The data entry that occurs with the greatest frequency.
- If no entry is repeated the data set has no mode.
- If two entries occur with the same greatest frequency, each entry is a mode (**bimodal**).

Slide 89

Example: Finding the Mode (1 of 2)

Find the mode of the weights listed in Example 1.

223 235 235 268 274 285 290

Slide 90

Solution: Finding the Mode (1 of 2)

- Ordering the data helps to find the mode.

223 235 235 268 274 285 290

- The entry of 235 occurs twice, whereas the other data entries occur only once.

The mode of the weights is 235 pounds.

Slide 91

Example: Finding the Mode (2 of 2)

At a political debate a sample of audience members was asked to name the political party to which they belong. Their responses are shown in the table. What is the mode of the responses?

Political Party	Frequency, *f*
Democrat	46
Republican	34
Independent	39
Other/don't know	5

Slide 92

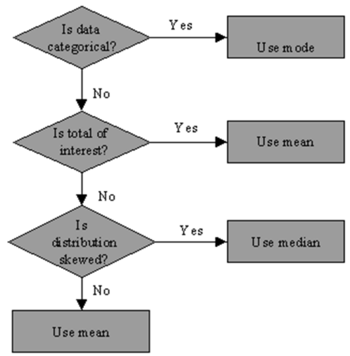
Solution: Finding the Mode (2 of 2)

Political Party	Frequency, *f*
Democrat	46
Republican	34
Independent	39
Other/don't know	5

The response occurring with the greatest frequency is Democrat. So, the mode is Democrat. In this sample, there were more Democrats than people of any other single affiliation.

Comparing the Mean, Median, and Mode

Selecting Among the Mean, Median, and Model



Comparing the Mean, Median, and Mode

- All three measures describe a typical entry of a data set.
- Advantage of using the mean:
  - The mean is a reliable measure because it takes into account every entry of a data set.
- Disadvantage of using the mean:
  - Greatly affected by **outliers** (a data entry that is far removed from the other entries in the data set).

Example: Comparing the Mean, Median, and Mode

The table shows the sample ages of students in a class. Find the mean, median, and mode of the ages. Are there any outliers? Which measure of central tendency best describes a typical entry of this data set?

Ages in a class						
20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

Solution: Comparing the Mean, Median, and Mode (1 of 3)

Ages in a class						
20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

Mean: $\bar{x} = \frac{\sum x}{n} = \frac{20+20+\ldots+24+65}{20} \approx 23.8 \text{ years}$

Median: $\frac{21+22}{2} = 21.5 \text{ years}$

Mode: 20 years (the entry occurring with the greatest frequency)

Solution: Comparing the Mean, Median, and Mode (2 of 3)

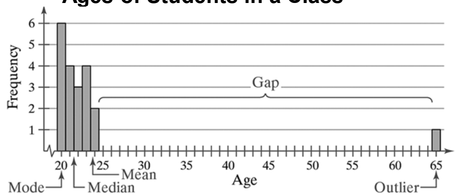
$\text{Mean} \approx 23.8 \text{ years} \quad \text{Median} = 21.5 \text{ years} \quad \text{Mode} = 20 \text{ years}$

- The mean takes every entry into account, but is influenced by the **outlier** of 65.
- The median also takes every entry into account, and it is not affected by the outlier.
- In this case the mode exists, but it doesn't appear to represent a typical entry.

Solution: Comparing the Mean, Median, and Mode (3 of 3)

Sometimes a graphical comparison can help you decide which measure of central tendency best represents a data set.

Ages of Students in a Class



In this case, it appears that the **median** best describes the data set.

Weighted Mean

Weighted Mean

- The mean of a data set whose entries have varying weights.
- The weighted mean is given by

$$\bar{x} = \frac{\sum xw}{\sum w}$$ where w is the weight of each entry x.

Example: Finding a Weighted Mean

Your grades from last semester are in the table. The grading system assigns points as follows: $A = 4$, $B = 3$, $C = 2$, $D = 1$, $F = 0$. Determine your grade point average (weighted mean).

Final Grade	Credit Hours
C	3
C	4
D	1
A	3
C	2
B	3

Slide 101

Solution: Finding a Weighted Mean

Points, x	Credit hours, w	xw
2	3	6
2	4	8
1	1	1
4	3	12
2	2	4
3	3	9
	$\Sigma w = 16$	$\Sigma(x \cdot w) = 40$

$$\bar{x} = \frac{\sum xw}{\sum w} = \frac{40}{16} = 2.5$$

Last semester, your grade point average was 2.5.

Slide 102

Mean of Grouped Data

Mean of a Frequency Distribution

- Approximated by

$$\bar{x} = \frac{\sum xf}{n} \qquad n = \sum f$$

where x and f are the midpoints and frequencies of a class, respectively.

Slide 103

Finding the Mean of a Frequency Distribution

In Words	*In Symbols*
1. Find the midpoint of each class.	$x = \frac{(\text{Lower limit}) + (\text{Upper limit})}{2}$
2. Find the sum of the products of the midpoints and the frequencies.	$\sum xf$
3. Find the sum of the frequencies.	$n = \sum f$
4. Find the mean of the frequency distribution.	$\bar{x} = \frac{\sum xf}{n}$

Slide 104

Example: Find the Mean of a Frequency Distribution

The frequency distribution shows the out-of-pocket prescription medicine expenses (in dollars) for 30 U.S. adults in a recent year. Use the frequency distribution to estimate the mean expense. Using the sample mean formula, the mean expense is $285.50. Compare this with the estimated mean.

Class midpoint, x	Frequency, f	xf
172.5	3	517.5
208.5	2	417.0
244.5	5	1222.5
280.5	6	1683.0
316.5	7	2215.5
352.5	4	1410.0
388.5	3	1165.5
	$n = 30$	$\Sigma = 8631$

Solution: Find the Mean of a Frequency Distribution

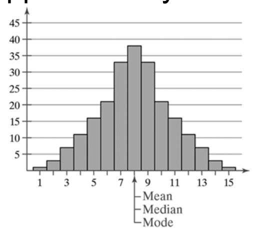
$$\bar{x} = \frac{\sum xf}{\sum n} = \frac{8631}{30} = 287.7$$

The mean expense is $287.70. This value is an estimate because it is based on class midpoints instead of the original data set.

The Shape of Distributions (1 of 4)

Symmetric Distribution

- A vertical line can be drawn through the middle of a graph of the distribution and the resulting halves are approximately mirror images.

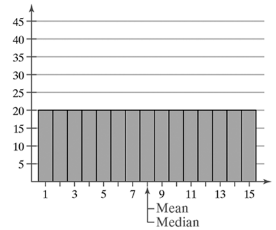


Symmetric Distribution

The Shape of Distributions (2 of 4)

Uniform Distribution (rectangular)

- All entries or classes in the distribution have equal or approximately equal frequencies.
- Symmetric.

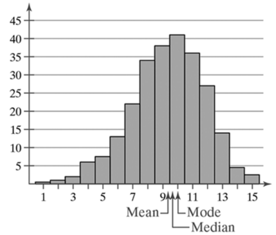


Uniform Distribution

The Shape of Distributions (3 of 4)

Skewed Left Distribution (negatively skewed)

- The “tail” of the graph elongates more to the left.
- The mean is to the left of the median.

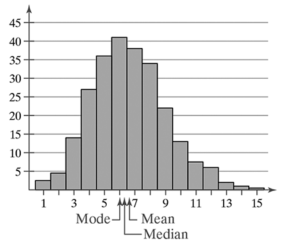


Skewed Left Distribution

The Shape of Distributions (4 of 4)

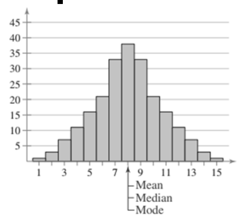
Skewed Right Distribution (positively skewed)

- The “tail” of the graph elongates more to the right.
- The mean is to the right of the median.

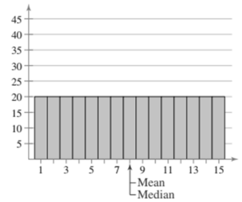


Skewed Right Distribution

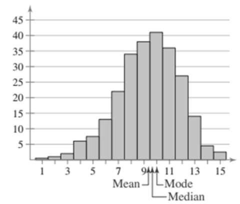
The Shape of Distributions



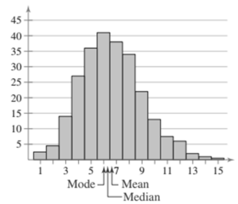
Symmetric Distribution



Uniform Distribution



Skewed Left Distribution



Skewed Right Distribution

Section 2.4

Measures of Variation

Range ● Variance and Standard Deviation ● Interpreting Standard Deviation ● Standard Deviation for Grouped Data ● Coefficient of Variation

Range

Range

- The difference between the maximum and minimum data entries in the set.
- The data must be quantitative.
- $\text{Range} = (\text{Max. data entry}) - (\text{Min. data entry})$

Example: Finding the Range

Two corporations each hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries for Corporation A.

Starting Salaries for Corporation A (in thousands of dollars)

Salary	41	38	39	45	47	41	44	41	37	42

Starting Salaries for Corporation B (in thousands of dollars)

Salary	40	23	41	50	49	32	41	29	52	58

Solution: Finding the Range

- Ordering the data helps to find the least and greatest salaries.

37 38 39 41 41 41 42 44 45 47

minimum maximum

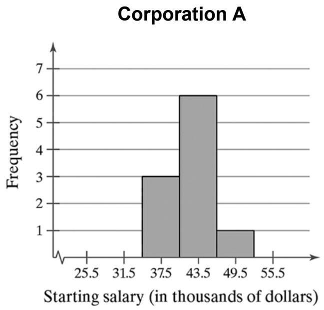
$$\begin{aligned}\text{Range} &= (\text{Max. salary}) - (\text{Min. salary}) \\ &= 47 - 37 = 10\end{aligned}$$

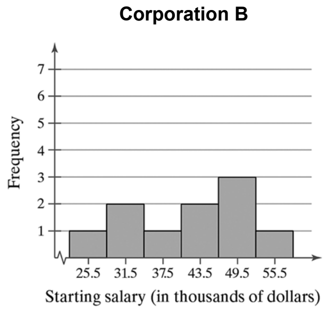
The range of starting salaries for Corporation A is 10, or $\$10,000$.

Variation (1 of 2)

- Both data sets in the last example have a mean of 41.5, or $\$41,500$, a median of 41, or $\$41,000$, and a mode of 41, or $\$41,000$. And yet the two sets differ significantly.

Variation (2 of 2)





Slide 119

Deviation, Variance, and Standard Deviation (1 of 4)

Deviation

- The difference between the data entry, *x*, and the mean of the data set.
- Population data set:
  - Deviation of $x = x - \mu$
- Sample data set:
  - Deviation of $x = x - \bar{x}$

Slide 120

Deviation, Variance, and Standard Deviation (2 of 4)

Population Variance

$$\sigma^2 = \frac{\Sigma\left(x-\mu\right)^2}{N}$$

Population Standard Deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma\left(x-\mu\right)^2}{N}}$$

Slide 121

Deviation, Variance, and Standard Deviation (3 of 4)

Observations About Standard Deviation

- The standard deviation measures the variation of the data set about the mean and has the same units of measure as the data set.
- The standard deviation is always greater than or equal to 0. When $\sigma = 0$, the data set has no variation and all entries have the same value.
- As the entries get farther from the mean (that is, more spread out), the value of σ increases.

Slide 122

Finding Population Variance & Standard Deviation

In Words	*In Symbols*
1. Find the mean of the population data set.	$\mu = \frac{\sum x}{N}$
2. Find deviation of each entry.	$x - \mu$
3. Square each deviation.	$(x-\mu)^2$
4. Add to get the sum of squares.	$SS_x = \sum(x-\mu)^2$

Slide 123

Finding the Population Variance & Standard Deviation

In Words	*In Symbols*
5. Divide by *N* to get the **population variance.**	$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$
6. Find the square root to get the **population standard deviation.**	$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$

Slide 124

Example: Finding Population Variance and Standard Deviation

Find the population variance and standard deviation of the starting salaries for Corporation A listed in the first Example.

For this data set, $N = 10$, $\sum x = 415$.

The mean is $\mu = \frac{415}{10} = 41.5$.

Slide 125

Solution: Finding Population Standard Deviation (1 of 3)

- Determine the deviation for each data entry.

Salary ($1000s), x	Deviation: $x - \mu$
41	$41 - 41.5 = -0.5$
38	$38 - 41.5 = -3.5$
39	$39 - 41.5 = -2.5$
45	$45 - 41.5 = 3.5$
47	$47 - 41.5 = 5.5$
41	$41 - 41.5 = -0.5$
44	$44 - 41.5 = 2.5$
41	$41 - 41.5 = -0.5$
37	$37 - 41.5 = -4.5$
42	$42 - 41.5 = 0.5$
$\sum x = 415$	$\sum(x - \mu) = 0$

Slide 126

Solution: Finding Population Standard Deviation (2 of 3)

- Determine SS_x

Salary, x	Deviation: $x - \mu$	Squares: $(x - \mu)^2$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
38	$38 - 41.5 = -3.5$	$(-3.5)^2 = 12.25$
39	$39 - 41.5 = -2.5$	$(-2.5)^2 = 6.25$
45	$45 - 41.5 = 3.5$	$(3.5)^2 = 12.25$
47	$47 - 41.5 = 5.5$	$(5.5)^2 = 30.25$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
44	$44 - 41.5 = 2.5$	$(2.5)^2 = 6.25$
41	$41 - 41.5 = -0.5$	$(-0.5)^2 = 0.25$
37	$37 - 41.5 = -4.5$	$(-4.5)^2 = 20.25$
42	$42 - 41.5 = 0.5$	$(0.5)^2 = 0.25$
	$\Sigma(x-\mu) = 0$	$SS_x = 88.5$

Solution: Finding Population Standard Deviation (3 of 3)

Population Variance

- $\sigma^2 = \dfrac{\Sigma(x-\mu)^2}{N} = \dfrac{88.5}{10} \approx 8.9$

Population Standard Deviation

- $\sigma = \sqrt{\sigma^2} = \sqrt{\dfrac{88.5}{10}} \approx 3.0$

The population variance is about 8.9, and the population standard deviation is about 3.0, or $\$3,000$.

Deviation, Variance, and Standard Deviation (4 of 4)

Sample Variance

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

Simple Standard Deviation

$$s = \sqrt{s^2} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

Symbols in Variance and Standard Deviation Formulas

	Population	Sample
Variance	σ^2	s^2
Standard deviation	σ	s
Mean	μ	$\bar{x}$
Number of entries	N	n
Deviation	$x - \mu$	$x - \bar{x}$
Sum of squares	$\Sigma(x-\mu)^2$	$\Sigma(x-\bar{x})^2$

Finding the Sample Variance & Standard Deviation (1 of 2)

In Words	*In Symbols*
1. Find the mean of the sample data set.	$\bar{x} = \dfrac{\Sigma x}{n}$
2. Find deviation of each entry.	$x - \bar{x}$
3. Square each deviation.	$(x-\bar{x})^2$
4. Add to get the sum of squares.	$SS_x = \Sigma(x-\bar{x})^2$

Finding the Sample Variance & Standard Deviation (2 of 2)

In Words	*In Symbols*
5. Divide by $n-1$ to get the **sample variance**.	$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$
6. Find the square root to get the **sample standard deviation.**	$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

Example: Finding Sample Variance & Standard Deviation

In a study of high school football players that suffered concussions, researchers placed the players in two groups. Players that recovered from their concussions in 14 days or less were placed in Group 1. Those that took more than 14 days were placed in Group 2. The recovery times (in days) for Group 1 are listed below. Find the sample variance and standard deviation of the recovery times.

4 7 6 7 9 5 8 10 9 8 7 10

Solution: Finding Sample Variance & Standard Deviation (1 of 2)

- Find $\sum x$.
- Find the standard deviation for each data entry, *s*.
- Find the sum of the squares, SS_x.

Time x	Deviation $x - \bar{x}$	Squares $(x - \bar{x})^2$
4	−3.5	12.25
7	−0.5	0.25
6	−1.5	2.25
7	−0.5	0.25
9	1.5	2.25
5	−2.5	6.25
8	0.5	0.25
10	2.5	6.25
9	1.5	2.25
8	0.5	0.25
7	−0.5	0.25
10	2.5	6.25
$\Sigma x = 90$		$SS_x = 39$

Solution: Finding Sample Variance & Standard Deviation (2 of 2)

For this data set, $n=12$ and $\sum x = 90$. The mean is $\bar{x} = 90/12 = 7.5$. To calculate s^2 and *s*, note that $n-1=12-1=11$. $SS_x = 39$.

- **Sample Variance**

$$s^2 = \frac{\Sigma(x-\bar{x})}{n-1} = \frac{39}{11} \approx 3.5$$

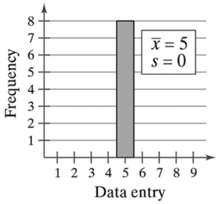
- **Sample Standard Deviation**

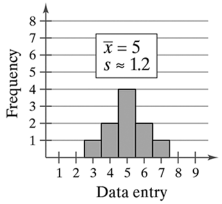
$$s = \sqrt{\frac{39}{11}} \approx 1.9$$

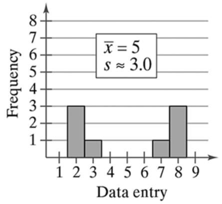
The sample variance is about 3.5, and the sample standard deviation is about 1.9 days.

Interpreting Standard Deviation

- Standard deviation is a measure of the typical amount an entry deviates from the mean.
- The more the entries are spread out, the greater the standard deviation.





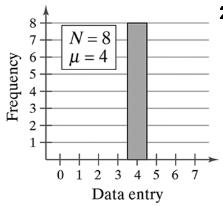


Slide 139

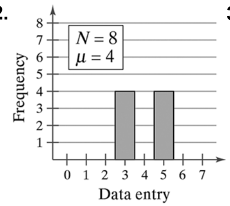
Example: Estimating Standard Deviation

Without calculating, estimate the population standard deviation of each data set.

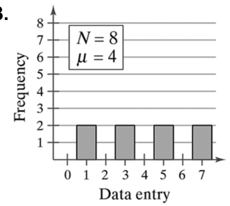
1.



2.



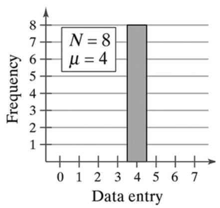
3.



Slide 140

Solution: Estimating Standard Deviation (1 of 3)

1.

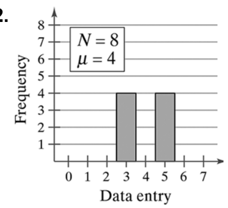


1. Each of the eight entries is 4. The deviation of each entry is 0, so $\sigma = 0$.

Slide 141

Solution: Estimating Standard Deviation (2 of 3)

2.

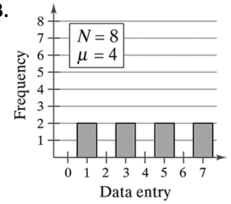


2. Each of the eight entries has a deviation of ± 1. So, the population standard deviation should be 1. By calculating, you can see that $\sigma = 1$.

Slide 142

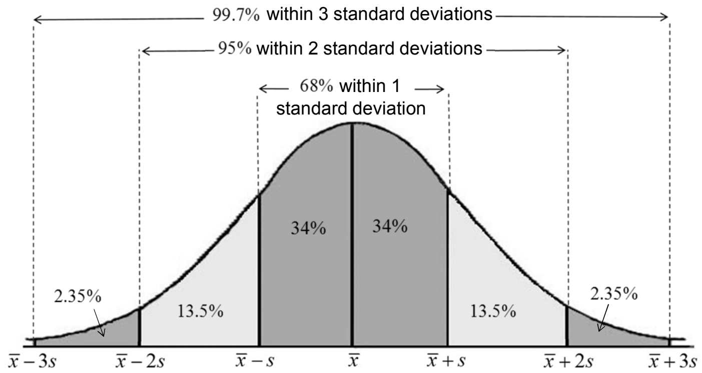
Solution: Estimating Standard Deviation (3 of 3)

3.



3. Each of the eight entries has a deviation of ± 1 or ± 3. So, the population standard deviation would be about 2. By calculating, you can see that is greater than 2, with $\sigma \approx 2.2$.

Interpreting Standard Deviation: Empirical Rule (68 – 95 – 99.7 Rule) (2 of 2)

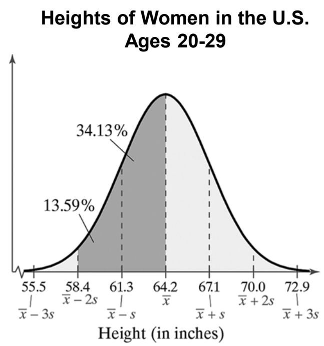


Example: Using the Empirical Rule

In a survey conducted by the National Center for Health Statistics, the sample mean height of women in the United States (ages 20-29) was 64.2 inches, with a sample standard deviation of 2.9 inches. Estimate the percent of the women whose heights are between 58.4 inches and 64.2 inches. *(Adapted from National Center for Health Statistics)*

Solution: Using the Empirical Rule (1 of 2)

- The distribution of women's heights is shown. Because the distribution is bell-shaped, you can use the Empirical Rule.
- The mean height is 64.2 inches.



Solution: Using the Empirical Rule (2 of 2)

- When you subtract two standard deviations from the mean height, you get

$$\bar{x} - 2s = 64.2 - 2(2.9) = 58.4.$$

- Because 58.4 is two standard deviations below the mean height, the percent of the heights between 58.4 and 64.2 inches is about $13.59\% + 34.13\% = 47.72\%$.

So, about 47.72% of women are between 58.4 and 64.2 inches tall.

Slide 148

Section 2.5

Measures of Position

Slide 167

Quartiles

- **Fractiles** are numbers that partition (divide) an ordered data set into equal parts.
- **Quartiles** approximately divide an ordered data set into four equal parts.
  - **First quartile, Q_1**: About one quarter of the data fall on or below Q_1.
  - **Second quartile, Q_2**: About one half of the data fall on or below Q_2 (median).
  - **Third quartile, Q_3**: About three quarters of the data fall on or below Q_3.

Slide 169

Example: Finding Quartiles

Each year in the U.S., automobile commuters waste fuel due to traffic congestion. The amounts (in gallons per year) of fuel wasted by commuters in the 15 largest U.S. urban areas are listed. Find the first, second, and third quartiles of the data set. What do you observe? *(Source: Based on 2015 Urban Mobility Scorecard)*

20 30 29 22 25 29 25 24 35 23 25 11 33 28 35

Solution:

- Q_2 divides the data set into two halves.

$$\underbrace{11\ 20\ 22\ \underset{\uparrow\, Q_1}{\mathbf{23}}\ 24\ 25\ 25}_{\text{Data entries to the left of } Q_2}\ \underset{\uparrow\, Q_2}{\mathbf{25}}\ \underbrace{28\ 29\ 29\ \underset{\uparrow\, Q_3}{\mathbf{30}}\ 33\ 35\ 35}_{\text{Data entries to the right of } Q_2}$$

Slide 170

Interquartile Range (1 of 2)

Interquartile Range (IQR)

- A measure of variation that gives the range of the middle portion (about half) of the data.
- The difference between the third and first quartiles.
- $\text{IQR} = Q_3 - Q_1$

Slide 176

Interquartile Range (2 of 2)

Using the Interquartile Range to Identify Outliers

1. Find the first (Q_1) and third (Q_3) quartiles of the data set.
2. Find the interquartile range: $\text{IQR} = Q_3 - Q_1$.
3. Multiply IQR by 1.5: 1.5(IQR).
4. Subtract 1.5(IQR) from Q_1. Any data entry less than $Q_1 - 1.5(\text{IQR})$ is an outlier.
5. Add 1.5(IQR) to Q_3. Any data entry greater than $Q_3 + 1.5(\text{IQR})$ is an outlier.

Slide 177

Example: Finding the Interquartile Range

Find the interquartile range of the data set from the first example. Are their any outliers?

Solution:

Recall $Q_1 = 23$ and $Q_3 = 30$. So, the interquartile range is $\text{IQR} = Q_3 - Q_1 = 30 - 23 = 7$.

To identify any outliers, first note that $1.5(\text{IQR}) = 1.5(7)$ $= 10.5$.

Slide 178

Solution: Finding the Interquartile Range

- There is a data entry, 11, that is less than
$$Q_1 - 1.5(\text{IQR}) = 23 - 10.5 = 12.5$$
- A data entry less than 12.5 is an outlier.
- There are no data entries greater than
$$Q_3 + 1.5(\text{IQR}) = 30 + 10.5 = 40.5$$
- A data entry greater than 40.5 is an outlier.
- So, 11 is an outlier.

In large urban areas, the amount of fuel wasted by auto commuters in the middle of the data set varies by at most 10.5 gallons. Notice that the outlier, 11, does not affect the IQR.

Slide 179

Box-and-Whisker Plot

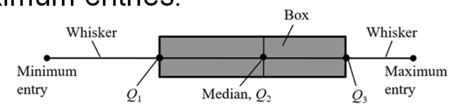
Box-and-whisker plot

- Exploratory data analysis tool.
- Highlights important features of a data set.
- Requires (**five-number summary**):
  1. Minimum entry
  2. First quartile Q_1
  3. Median Q_2
  4. Third quartile Q_3
  5. Maximum entry

Slide 180

Drawing a Box-and-Whisker Plot

1. Find the five-number summary of the data set.
2. Construct a horizontal scale that spans the range of the data.
3. Plot the five numbers above the horizontal scale.
4. Draw a box above the horizontal scale from Q_1 to Q_3 and draw a vertical line in the box at Q_2.
5. Draw whiskers from the box to the minimum and maximum entries.



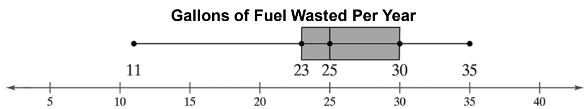
Slide 181

Example: Drawing a Box-and-Whisker Plot (1 of 2)

Draw a box-and-whisker plot that represents the data set in the first example.

$\text{Min} = 11,\ Q_1 = 23,\ Q_2 = 25,\ Q_3 = 30,\ \text{Max} = 35,$

Solution:

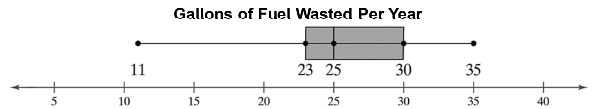


The box represents about half of the data, which are between 23 and 30.

Slide 182

Example: Drawing a Box-and-Whisker Plot (2 of 2)

Solution:



The left whisker represents about one-quarter of the data, so about 25% of the data entries are less than 23. The right whisker represents about one-quarter of the data, so about 25% of the data entries are greater than 30. Also, the length of the left whisker is much longer than the right one. This indicates that the data set has a possible outlier to the left.

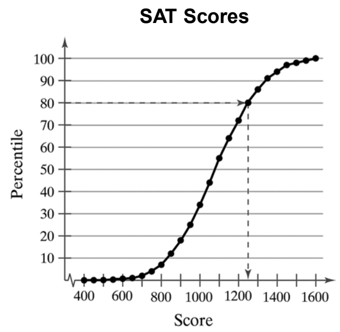
Slide 183

Percentiles and Other Fractiles

Fractiles	Summary	Symbols
Quartiles	Divides data into 4 equal parts	Q_1, Q_2, Q_3
Deciles	Divides data into 10 equal parts	$D_1, D_2, D_3, \ldots, D_9$
Percentiles	Divides data into 100 equal parts	$P_1, P_2, P_3, \ldots, P_{99}$

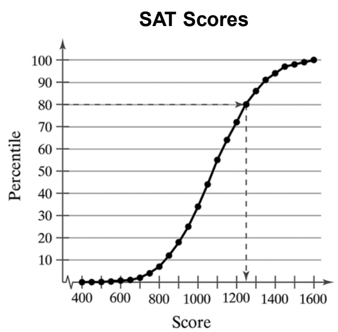
Example: Interpreting Percentiles

The ogive represents the cumulative frequency distribution for SAT test scores of college-bound students in a recent year. What test score represents the 80th percentile? *(Source: College Board)*



Solution: Interpreting Percentiles

- From the ogive, you can see that the 80th percentile corresponds to a score of 1250.
- This means that approximately 80% of the students had an SAT score of 1250 or less.



Percentile that Corresponds to a Specific Data Entry

To find the **percentile that corresponds to a specific data entry *x*,** use the formula

$$\text{Percentile of } x = \frac{\text{number of data entries less than } x}{\text{total number of data entries}} \cdot 100$$

and then round to the nearest whole number.

Example: Finding Percentiles

For the data set in the second example, find the percentile that corresponds to $\$34,000$.

Solution

- Recall that the tuition costs are in thousands of dollars, so $\$34,000$ is the data entry 34. Begin by ordering the data.

16 18 18 23 25 27 30 33 34 34 35 35 36

40 40 41 44 45 47 49 50 51 51 52 52

Slide 188

Solution: Finding Percentiles

- There are 8 data entries less than 34 and the total number of data entries is 25.

$$\text{Percentile of } 34 = \frac{\text{number of data entries less than } x}{\text{total number of data entries}} \cdot 100$$

$$= \frac{8}{25} \cdot 100 = 32$$

- The tuition cost of $\$34,000$ corresponds to the 32nd percentile.

The tuition cost of $\$34,000$ is greater than 32% of the other tuition costs.

Slide 189

The Standard Score

Standard Score (*z*-score)

- Represents the number of standard deviations a given value *x* falls from the mean μ.
- $z = \dfrac{\text{value} - \text{mean}}{\text{standard deviation}} = \dfrac{x - \mu}{\sigma}$

Slide 190

Example: Finding *z*-Scores

The mean speed of vehicles along a stretch of highway is 56 miles per hour with a standard deviation of 4 miles per hour. You measure the speeds of three cars traveling along this stretch of highway as 62 miles per hour, 47 miles per hour, and 56 miles per hour. Find the *z*-score that corresponds to each speed. Assume the distribution of the speeds is approximately bell-shaped.

Slide 191

Solution: Finding *z*-Scores

Solution

The *z*-score that corresponds to each speed is calculated below.

$x = 62 \text{ mph}$ $\quad$ $x = 47 \text{ mph}$ $\quad$ $x = 56 \text{ mph}$

$$z = \frac{62-56}{4} = 1.5 \qquad z = \frac{47-56}{4} = -2.25 \qquad z = \frac{56-56}{4} = 0$$

62 miles per hour is 1.5 standard deviations above the mean; 47 miles per hour is 2.25 standard deviations below the mean; and 56 miles per hour is equal to the mean. 47 miles per hour is unusually slow, because its speed corresponds to a *z*-score of -2.25.

Slide 192

Example: Comparing *z*-Scores from Different Data Sets

The table shows the mean heights and standard deviations for a population of men and a population of women. Compare the *z*-scores for a 6-foot-tall man and a 6-foot-tall woman. Assume the distributions of the heights are approximately bell-shaped.

Men's heights	Women's heights
$\mu = 69.9$ in.	$\mu = 64.3$ in.
$\sigma = 3.0$ in.	$\sigma = 2.6$ in.

Slide 193

Solution: Comparing *z*-Scores from Different Data Sets (1 of 2)

Solution

Note that $6 \text{ feet} = 72 \text{ inches}$. Find the *z*-score for each height.

- ***z*-score for 6-foot-tall man**

$$z = \frac{x-\mu}{\sigma} = \frac{72-69.9}{3.0} = 0.7$$

- ***z*-score for 6-foot-tall woman**

$$z = \frac{x-\mu}{\sigma} = \frac{72-64.3}{2.6} \approx 3.0$$

Slide 194

Solution: Comparing *z*-Scores from Different Data Sets (2 of 2)

Solution

The *z*-score for the 6-foot-tall man is within 1 standard deviation of the mean (69.9 inches). This is among the typical heights for a man.

The *z*-score for the 6-foot-tall woman is about 3 standard deviations from the mean (64.3 inches). This is an unusual height for a woman.

Slide 195