Learning Goal: I will be able to find the standardized value (z-score) of an observation and use percentiles to locate individual values within distributions of data.

Where is a certain data point within the distribution?

Vocabulary:
Percentile: _____ percentile for x, the __________ of values that are _____ than x.
- Find the number of observations ______ x, divide by ___ (total number of observations)
Cumulative Relative Frequency: gives __________________ relative to ______ in sample.
Cumulative Relative Frequency plots are also called __________.
__________: A standardized value. Tells us how many standard ______________ from the mean an ____________ falls and in _______ direction.

Example 1:
Wins in Major League Baseball
The stemplot below shows the number of wins for each of the 30 Major League Baseball teams in 2009.

| Key: 5|9 represents a team with 59 wins. |
|---|---|---|---|---|
| 5 | 9 | 2455 | 00455589 | 0345667778 |
| 7 | 123557 |
| 10 | 3 |

Find the percentiles for the following teams:

(a) The Colorado Rockies, who won 92 games.

(b) The New York Yankees, who won 103 games.

(c) The Kansas City Royals and Cleveland Indians, who both won 65 games.
2.1 Describing Location in a Distribution Notes

Example 2: Relative frequency, cumulative frequency, percentiles

<table>
<thead>
<tr>
<th>Heights (in)</th>
<th>Absolute Frequency</th>
<th>Relative Frequency</th>
<th>Absolute Cum Freq</th>
<th>Relative Cum Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>62-63</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64-65</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66-67</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68-69</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70-71</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72-73</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74-75</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76-77</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What is your height if you are in the 70th percentile?

If your height is 63 in, what is your relative standing in the group?

What is the center?
Example 3: State Median Household Incomes

Here is a table showing the distribution of median household incomes for the 50 states and the District of Columbia.

<table>
<thead>
<tr>
<th>Median Income ($1000s)</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Cumulative Frequency</th>
<th>Cumulative Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 to &lt; 40</td>
<td>1</td>
<td>1/51 = 0.020</td>
<td>1</td>
<td>1/51 = 0.020</td>
</tr>
<tr>
<td>40 to &lt; 45</td>
<td>10</td>
<td>10/51 = 0.196</td>
<td>11</td>
<td>11/51 = 0.216</td>
</tr>
<tr>
<td>45 to &lt; 50</td>
<td>14</td>
<td>14/51 = 0.275</td>
<td>25</td>
<td>25/51 = 0.490</td>
</tr>
<tr>
<td>50 to &lt; 55</td>
<td>12</td>
<td>12/51 = 0.236</td>
<td>37</td>
<td>37/51 = 0.725</td>
</tr>
<tr>
<td>55 to &lt; 60</td>
<td>5</td>
<td>5/51 = 0.098</td>
<td>42</td>
<td>42/51 = 0.824</td>
</tr>
<tr>
<td>60 to &lt; 65</td>
<td>6</td>
<td>6/51 = 0.118</td>
<td>48</td>
<td>48/51 = 0.941</td>
</tr>
<tr>
<td>65 to &lt; 70</td>
<td>3</td>
<td>3/51 = 0.059</td>
<td>51</td>
<td>51/51 = 1.000</td>
</tr>
</tbody>
</table>

Here is the cumulative relative frequency graph for the income data. The point at (50, 0.49) means 49% of the states had median household incomes less than $50,000.

The point at (55, 0.725) means that 72.5% of the states had median household incomes less than $55,000. Thus, 72.5% - 49% = 23.5% of the states had median household incomes between $50,000 and $55,000 since the cumulative relative frequency increased by 0.235.

Due to rounding error, this value is slightly different than the relative frequency for the 50 to <55 category.

Use the cumulative relative frequency graph for the state income data to answer each question.

(a) At what percentile is California, with a median income of $57,445?

(b) Estimate and interpret the first quartile of this solution.
2.1 Describing Location in a Distribution Notes

**Example 4: Home run kings**

The single-season home run record for major league baseball has been set just three times since Babe Ruth hit 60 home runs in 1927. Roger Maris hit 61 in 1961, Mark McGwire hit 70 in 1998 and Barry Bonds hit 73 in 2001. In an absolute sense, Barry Bonds had the best performance of these four players, since he hit the most home runs in a single season. However, in a relative sense this may not be true. Baseball historians suggest that hitting a home run has been easier in some eras than others. This is due to many factors, including quality of batters, quality of pitchers, hardness of the baseball, dimensions of ballparks, and possible use of performance-enhancing drugs. To make a fair comparison, we should see how these performances rate relative to others hitters during the same year.

Below is the mean and standard deviation of homerun totals that year. Which player had the most outstanding performance relative to his peers? To answer this, we need to figure how we could use that information to equalize their performances.

<table>
<thead>
<tr>
<th>Year</th>
<th>Player</th>
<th>HR</th>
<th>Mean</th>
<th>SD</th>
<th>z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1927</td>
<td>Babe Ruth</td>
<td>60</td>
<td>7.2</td>
<td>9.7</td>
<td></td>
</tr>
<tr>
<td>1961</td>
<td>Roger Maris</td>
<td>61</td>
<td>18.8</td>
<td>13.4</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>Mark McGwire</td>
<td>70</td>
<td>20.7</td>
<td>12.7</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>Barry Bonds</td>
<td>73</td>
<td>21.4</td>
<td>13.2</td>
<td></td>
</tr>
</tbody>
</table>

Which player had the most outstanding performance relative to his peers?

**Z-score** –

**Example 5: Wins in Major League Baseball**

In 2009, the mean number of wins was 81 with a standard deviation of 11.4 wins. Find and interpret the z-scores for the following teams.

(a) The New York Yankees, with 103 wins.

(b) The New York Mets, with 70 wins.
2.1 Describing Location in a Distribution Notes

**Example 6: Transforming Data**
Maria measures the lengths of 5 cockroaches that she finds at school. Here are her results (inches):

1.4   2.2   1.1   1.6   1.2

a) Find the mean and standard deviation of Maria’s measurements.

b) Maria’s science teacher is furious to discover that she has measured the cockroach lengths in inches rather than centimeters. (There are 2.54 cm in 1 in) How are the mean and standard deviation affected by converting the measurements to centimeters?

c) The cockroaches crawled into some radioactive substance in the science lab and each grew 5 cm tails. How are the mean and standard deviation affected by this additional length?

**Effects of Adding (or Subtracting) a Constant:**

**Effects of Multiplying (or Dividing) by a Constant:**

**Example 7: Taxi Cabs**
In 2010, Taxi Cabs in New York City charged an initial fee of $2.50 plus $2 per mile. In equation form, \( fare = 2.50 + 2(miles) \). At the end of the month a businessman collects all of his taxi cab receipts and calculates some numerical summaries. The mean fare he paid was $15.45 with a standard deviation of $10.20. What are the mean and standard deviation of the lengths of his cab rides in miles?

**Example 8: Z-score connection to transforming data:**
What is the mean and standard deviation of a set of data that has been transformed into z-scores? Use the cockroach data, 1.4 2.2 1.1 1.6 1.2, find z-scores, and find the new mean and standard deviation.
2.1 Describing Location in a Distribution Notes

Density Curves

We’ve learned to graph data, describe the distribution visually and with number summaries. Now we need to go a step further ... to a model of the distribution called a density curve.

Actual data sample pictures

The big idea is...convert a somewhat uneven distribution into a smooth curve and the area under the curve at any given range describes its proportion in relation to the whole distribution. The total area has to represent 100% of the data so the total area must = 1.

Notation alert! Real data: ______ and ______, Mathematical model/density curve: ______ and ______

Density curve
✓
✓
✓
✓
✓
✓

Mean and Median of a Density Curve

- The median is where area is 50% on both sides
- The mean is the balancing point

Sample illustrations:

Batting Averages

The histogram to the right shows the distribution of batting average (proportion of hits) for the 432 Major League Baseball players with at least 100 plate appearances in the 2009 season. The smooth curve shows the overall shape of the distribution.

In the first graph to the right, the 4 rightmost bars represent the proportion of players who had batting averages of at least 0.270. There are 177 such players out of a total of 432, for a proportion of 0.410. In the second graph, the area under the curve to the right of 0.270 is shaded. This area is 0.391, only 0.019 away from the actual proportion of 0.410.

Summary:
What is the formula for z-score?

What are the new symbols for mean and standard deviation?