Domain: **Statistics and Probability**

Cluster: Use random sampling to draw inferences about a population.

Standard: **7.SP.3.** Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *For example, the mean height of players on the basketball team is 10 cm greater than the mean height of players on the soccer team, about twice the variability (mean absolute deviation) on either team; on a dot plot, the separation between the two distributions of heights is noticeable.*

Standards for Mathematical Practice (MP):

MP.1. Make sense of problems and persevere in solving them.
MP.2. Reason abstractly and quantitatively.
MP.3. Construct viable arguments and critique the reasoning of others.
MP.4. Model with mathematics.
MP.5. Use appropriate tools strategically.
MP.6. Attend to precision.
MP.7. Look for and make use of structure.

Connections:

This Cluster is connected to the Grade 7 Critical Area of Focus #4, **Drawing inferences about populations based on samples**. Measures of center and variability are developed in Statistics and Probability Grade 6. This cluster expands standards 1 and 2 in Grade 7 to make inferences between populations.

Instructional Strategies

In Grade 6, students used measures of center and variability to describe sets of data. In the cluster "Use random sampling to draw inferences about a population" of Statistics and Probability in Grade 7, students learn to draw inferences about one population from a random sampling of that population. Students continue using these skills to draw informal comparative inferences about two populations. Provide opportunities for students to deal with small populations, determining measures of center and variability for each population. Then have students compare those measures and make inferences. The use of graphical representations of the same data (Grade 6) provides another method for making comparisons. Students begin to develop understanding of the benefits of each method by analyzing data with both methods.

When students study large populations, random sampling is used as a basis for the population inference. This build on the skill developed in the Grade 7 cluster "Use random sampling to draw inferences about a population" of Statistics and Probability. Measures of center and variability are used to make inferences on each of the general populations. Then the students have make comparisons for the two populations based on those inferences.

This is a great opportunity to have students examine how different inferences can be made based on the same two sets of data. Have students investigate how advertising agencies uses data to persuade customers to use their products. Additionally, provide students with two populations and have them use the data to persuade both sides of an argument.

Explanations and Examples:

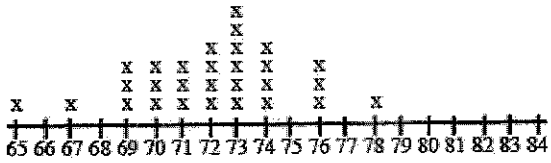
Students can readily find data as described in the example on sports team or college websites. Other sources for data include American Fact Finder (Census Bureau), Fed Stats, Ecology Explorers, USGS, or CIA World Factbook. Researching data sets provides opportunities to connect mathematics to their interests and other academic subjects. Students can utilize statistic functions in graphing calculators or spreadsheets for calculations with larger data sets or to check their computations. Students calculate mean absolute deviations in preparation for later work with standard deviations.

Example:
Jason wanted to compare the mean height of the players on his favorite basketball and soccer teams. He thinks the mean height of the players on the basketball team will be greater but doesn't know how much greater. He also wonders if the variability of heights of the athletes is related to the sport they play. He thinks that there will be a greater variability in the heights of soccer players as compared to basketball players. He used the rosters and player statistics from the team websites to generate the following lists.

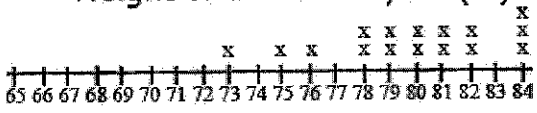
Basketball Team – Height of Players in inches for 2010-2011 Season
75, 73, 76, 78, 79, 78, 79, 81, 80, 82, 81, 84, 82, 84, 80, 84

Soccer Team – Height of Players in inches for 2010
73, 73, 73, 72, 69, 76, 72, 73, 74, 70, 65, 71, 74, 76, 70, 72, 71, 74, 71, 74, 73, 67, 70, 72, 69, 78, 73, 76, 69

To compare the data sets, Jason creates a two dot plots on the same scale. The shortest player is 65 inches and the tallest players are 84 inches.



Height of Soccer Players (in)



Height of Basketball Players (in)

In looking at the distribution of the data, Jason observes that there is some overlap between the two data sets. Some players on both teams have players between 73 and 78 inches tall. Jason decides to use the mean and mean absolute deviation to compare the data sets. Jason sets up a table for each data set to help him with the calculations.

The mean height of the basketball players is 79.75 inches as compared to the mean height of the soccer players at 72.07 inches, a difference of 7.68 inches.

The mean absolute deviation (MAD) is calculated by taking the mean of the absolute deviations for each data point. The difference between each data point and the mean is recorded in the second column of the table. Jason used rounded values (80 inches for the mean height of basketball players and 72 inches for the mean height of soccer players) to find the differences. The absolute deviation, absolute value of the deviation, is recorded in the third column. The absolute deviations are summed and divided by the number of data points in the set.

The mean absolute deviation is 2.14 inches for the basketball players and 2.53 for the soccer players. These values indicate moderate variation in both data sets. There is slightly more variability in the height of the soccer players. The difference between the heights of the teams is approximately 3 times the variability of the data sets ($7.68 \div 2.53 = 3.04$).

Soccer Players (n = 29)		
Height (in)	Deviation from Mean (in)	Absolute Deviation (in)
65	-7	7
67	-5	5
69	-3	3
69	-3	3
69	-3	3
70	-2	2
70	-2	2
70	-2	2
71	-1	1
71	-1	1
71	-1	1
72	0	0
72	0	0
72	0	0
72	0	0
73	+1	1
73	+1	1
73	+1	1
73	+1	1
73	+1	1
73	+1	1
74	+2	2
74	+2	2
74	+2	2
74	+2	2
76	+4	4
76	+4	4
76	+4	4
78	+6	6
$\Sigma = 2090$		$\Sigma = 62$

Basketball Players (n = 16)		
Height (in)	Deviation from Mean (in)	Absolute Deviation (in)
73	-7	7
75	-5	5
76	-4	4
78	-2	2
78	-2	2
79	-1	1
79	-1	1
80	0	0
80	0	0
81	1	1
81	1	1
82	2	2
82	2	2
84	4	4
84	4	4
84	4	4
$\Sigma = 1276$		$\Sigma = 40$

Mean = 2090 ÷ 29 =72 inches
MAD = 62 ÷ 29 = 2.13 inches

Mean = 1276 ÷ 16 =80 inches
MAD = 40 ÷ 16 = 2.5 inches

Common Misconceptions:
Arizona, Ohio & NC DOE