Grade 6 Math Unit 5- Statistics and Probability

UNIT OVERVIEW

In Grade 6, instructional time should focus on four critical areas. This unit addresses **Critical Focus Area #4, Developing understanding of statistical thinking. (See Connections for explanation)**

This unit addresses the following clusters:
- Develop understanding of statistical variability
- Summarize and describe distributions

STANDARDS

CC_Common Core State Standards - Mathematics (2010) - Grade 6
Domain 6.SP Statistics and Probability
Cluster Statement: Develop understanding of statistical variability.
Standard 6.SP.1 Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.
Standard 6.SP.2 Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
Standard 6.SP.3 Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

Cluster Statement: Summarize and describe distributions.
Standard 6.SP.4 Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
Standard 6.SP.5 Summarize numerical data sets in relation to their context, such as by:
6.SP.5.a Reporting the number of observations.
6.SP.5.b Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
6.SP.5.c Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
6.SP.5.d Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

CONTENT ELABORATIONS

6.SP.1	Statistics are numerical data relating to a group of individuals; statistics is also the name for the science of collecting, analyzing and interpreting such data. A statistical question anticipates an answer that varies from one individual to the next and is written to account for the variability in the data. Data are the numbers produced in response to a statistical question. Data are frequently collected from surveys or other sources (i.e., documents) Students differentiate between statistical questions and those that are not. A statistical question is one that collects information that addresses differences in a population. The question is framed so that the responses will allow for the differences. For example, the question "How tall am I?" is not a statistical question because there is only one response; however, the question, "How tall are the students in my class?" is a statistical question since the responses anticipates variability by providing a variety of possible anticipated responses that have numerical answers. Questions can result in a narrow or wide range of numerical values. Students might want to know about the fitness of the students at their school. Specifically, they want to know about the exercise habits of the students. So rather than asking, "Do you exercise?" they should ask about the amount of exercise the students at their school get per week. A statistical question for this study could be: "How many hours per week on average do students at Jefferson Middle School exercise?" To collect this information, students might design a survey question that anticipates variability by providing a variety of possible responses that have numerical answers, such as: 3 hours per week, 4 hours per week, and so on. Be sure that students ask questions that have specific numerical answers. **MP.1 MP.3, MP.6** should be emphasized.
6.SP.2	The distribution is the arrangement of the values of a data set. Distribution can be described using center (median or mean), and spread. Data collected can be represented on graphs, which will show the shape of the distribution of the data. Students examine the distribution of a data set and discuss the center, spread and overall shape with dot plots, histograms and box plots. Mode as a measure of center and range as a measure of variability are not addressed in the CCSS and as such are not a focus of instruction. These concepts can be introduced during instruction as needed. **MP.2, MP.4, MP.5, MP.6, MP.7** should be emphasized.
6.SP.3	Data sets contain many numerical values that can be summarized by one number such as a measure of center. The measure of center gives a numerical value to represent the center of the data (i.e. midpoint of an ordered list or the balancing point). Another characteristic of a data set is the variability (or spread) of the values. Example questions and possible solutions: - How many students are represented in the data set? (19 students are represented in the data set.) - What are the mean and median of the data set? What do these values mean? How do they compare? (The mean of the data set is 3.5. The median is 3. The mean indicates that if the values were equally distributed, all students would score a 3.5. The median indicate that 50% of the students scored a 3 or higher; 50% of the students scored a 3 or lower.) - What is the range of the data? What does this value mean? (The range of the data is 6, indicating that the values vary 6 points between the lowest and highest scores.) **MP.2, MP.4, MP.5, MP.6, MP.7** should be emphasized.

6.SP.4	Students display data sets using number lines. Dot plots, histograms and box plots are three graphs to be used. A dot plot is a graph that uses a point (dot) for each piece of data. The plot can be used with data sets that include fractions and decimals. A histogram shows the distribution of continuous data using intervals on the number line. The range of each bar represents the number of data values in that interval. Box plots are another useful way to display data and are plotted horizontally or vertically on a number line. These values give a summary of the shape of a distribution of values in a data set by dividing the set into quartiles. The box plot is constructed from the five-number summary (minimum, lower quartile, median, upper quartile, and maximum). Students understand that the size of the box or whiskers represent the middle 50% of the data. Students can readily compare two sets of data if they are displayed with side by side box plots on the same scale. Box plots display the degree of spread of the data and the skewness of the data. Dot plots are simple plots on a number line where each dot represents a piece of data in the data set. Dot plots are suitable for small to moderate size data sets and are useful for highlighting the distribution of the data including clusters, gaps and outliers. In most real data sets, there is a large amount of data and many numbers will be unique. A graph (such as a dot plot) showing how many ones, how many twos, etc. would not be meaningful; however, a histogram can be used. Students organize the data into convenient ranges and use these intervals to generate a frequency table and histogram. Note that changing the size of the range changes the appearance of the graph and the conclusions you may draw from it. In order to display numerical data in dot plots, histograms or box plots, students need to make decisions and perform calculations. Students are expected to display data graphically on a format appropriate for the data set as well reading data from graphs generated by other students or contained in reference materials. Students can use applets to create data displays. **MP.2, MP.4, MP.5, MP.6, MP.7** should be emphasized.
6.SP.5	Students record the number of observations. Using histograms, students determine the number of values between specified intervals. Given a box plot and the total number of data values, students identify the number of data points that are represented by the box. Reporting of the number of observations must consider the attribute of the data sets, including units (when applicable). Consideration may need to be given to how the data was collected (i.e. random sampling) Given a set of data values, students summarize the measure of center with the median or mean. The median is the value in the middle of a ordered list of data. This value means that 50% of the data is greater than or equal to it and that 50% of the data is less than or equal to it. The mean is the arithmetic average or balance point of a distribution. The mean is the sum of the values in a data set divided by how many values there are in the data set. The mean represents the value if all pieces of the data set had the same value. As a balancing point, the mean is the value where the data values above and the data values below have the same value. Measures of variation can be described using the inter-quartile range or the Mean Absolute Deviation. The inter-quartile range describes the variability between the middle 50% of a data set. It is found by subtracting the lower quartile from the upper quartile. it represents the length of the box in a box plot and is not affected by the outliers. The Mean Absolute Deviation describes the variablility of the data set by determining the absolute value deviation (the distance) of each data piece from the mean and then finding the average of these deviations. Both the inter-quartile range and the Mean Absolute Deviation are represented by a single numerical value. Higher values represent a greater variability in the data. Students understand how the measures of center and measures of variability are represented by the graphical display. Students describe the context of the data, using the shape of the data and are able to use this information to determine an appropriate measure of center and measure of variability. Students summarize numerical data by providing background information about the attribute being measured, methods and unit of measurement, the context of data collection activities, the number of observations, and summary statistics. Summary statistics include quantitative measures of center spread, and variability including extreme values (minimum and maximum, mean, median, mode, range, quartiles, inter-quartile ranges, and mean absolute deviation. The measure of center that a student chooses to describe a data set will depend upon the shape of the data distribution and context of data collection. The mode is the value in the data set that occurs most frequently. The mode is the least frequently used as a measure of center because data sets may not have a mode, may have more than one mode, or the mode may not be descriptive of the data set. The mean be be affected greatly by a few data points that are very low or very high. In this case, the median or middle value of the data set might be more descriptive. In data sets that are symmetrically distributed, the mean and median will be very close to the same. In data sets that are skewed, the mean and median will be different, with the median frequently providing a better overall description of the data set. **MP.2, MP.3, MP.4, MP.5, MP.6, MP.7** should be emphasized.

UNIT VOCABULARY

average	outliers	dot plot
first quartile	quartiles	frequency distribution
interquartile range	range	gap
mean	statistical question	histogram
mean absolute deviation	third quartile	line graph
measure of center	box plot	line plot
measures of variation	cluster	peak
median	distribution	symmetric
mode		

BIG IDEAS

ENDURING UNDERSTANDINGS

•Recognize that statistical questions and the answers account for variability in the data.
•Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
•Recognize that a measure of center for a numerical data set summarizes all of its values
with a single number, while a measure of
variation describes how its values vary with a
single number.
• Understand that numerical data can be displayed in plots on a number line, including dot plots, histograms, and box plots.
•Summarize numerical data sets in relation to their context, such as by:
Reporting the number of observations.
Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
 Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference
to the context in which the data were gathered.
 Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

ESSENTIALS QUESTIONS

Choose a few questions based on the needs of your students

•What is the best way to organize a set of data?
•What kinds of graphs will best represent a given set of data?
•How can I describe the center of a set of
data?
•How can I describe the spread of a set of data?
•How can I use data to compare different groups?
•How do I choose and create appropriate graphs to represent data?
What conclusions can be drawn from data?
•How can I recognize when a question is
statistical and when it is not?
•What is the difference in a measure of center and a measure of variation?

CONNECTIONS

In **Critical Focus Area # 4**, building on and reinforcing their understanding of number, students begin to develop their ability to think statistically. Students recognize that a data distribution may not have a definite center and that different ways to measure center yield different values. The mean measures center in the sense that it is the value that each data point would take on if the total of the data values were redistributed equally, and also in the sense that it is a balance point. Students recognize that a measure of variability (inter-quartile range or mean absolute deviation) can also be useful for summarizing data because two very different sets of data can have the same mean and median yet be distinguished by their variability. Students learn to describe and summarize numerical data sets, identifying clusters, peaks, gaps and symmetry, considering the context in which the data were collected.

Measures of center and measures of variability are used to draw informal comparative inferences about two populations in **7.SP.4**

Standards for Mathematical Practice (SMP)
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 (Deductive reasoning)
MP.8 Look for and express regularity in repeated reasoning (Inductive Reasoning)

DEVELOP UNDERSTANDING OF STATISTICAL VARIABILITY

	CONTENT	SKILLS
6.RP.1	Recognize a statistical question.	Recognize a statistical question. 1. Recognize that data can have variability. 2. Recognize a statistical question (examples versus non-examples) 3. Write a statistical question.
6.RP.2	Collect data sets and examine distribution.	Collect data sets and examine distribution. 1. Know that a set of data has a distribution. 2. Describe a set of data by its center, e.g., mean and median. 3. Describe a set of data by its spread and overall shape, e.g. by identifying data clusters, peaks, gaps and symmetry.
6.RP.3	Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.	Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number. 1. Recognize there are measures of central tendency for a data set, e.g., mean, median, mode. 2. Recognize there are measures of variances for a data set, e.g., range, interquartile range, mean absolute deviation. 3. Recognize measures of central tendency for a data set summarizes the data with a single number. 4. Recognize measures of variation for a data set describes how its values vary with a single number.

SUMMARIZE AND DESCRIBE DISTRIBUTIONS

	CONTENT	SKILLS

6.SP.4	Display numerical data in plots on a number line.	Display numerical data in plots on a number line. 1. Identify the components of dot plots, histograms, and box plots. 2. Find the median, quartile and interquartile range of a set of data. 3. Analyze a set of data to determine its variance. 4. Create a dot plot to display a set of numerical data. 5. Create a histogram to display a set of numerical data. 6. Create a box plot to display a set of numerical data.
6.SP.5	Summarize numerical data sets in relation to their context.	Summarize numerical data sets in relation to their context. 1. Organize and display data in tables and graphs. 2. Report the number of observations in a data set or display. 3. Describe the data being collected, including how it was measured and its units of measurement. 4. Calculate quantitative measures of center, e.g., mean, median, mode. 5. Calculate quantitative measures of variance, e.g., range, interquartile range, mean absolute deviation. 6. Identify outliers. 7. Determine the effect of outliers on quantitative measures of a set of data, e.g., mean, median, mode, range, interquartile range, mean absolute deviation. 8. Choose the appropriate measure of central tendency to represent the data. 9. Analyze the shape of the data distribution and the context in which the data were gathered to choose the appropriate measures of central tendency and variability and justify why this measure is appropriate in terms of the context.
Common Core Model Curriculum McGraw-Hill, **Glencoe Math** Chapters 11-12 Georgia Math frameworks, Grade 6 Unit 6 Manipulatives Smart Board Resources Hands-On Standards Box Plot Tool on NCTM's Illuminations Histogram Tool on NCTM's Illuminations		