

## 7<sup>th</sup> CCSS D5 Statistics and Probability Conceptual Foundation (4-5 weeks)

### Domain 5: Statistics and Probability 7.SP

#### D5 Cluster 1: Use random sampling to draw inferences about a population.

1. Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.
2. Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. *For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or prediction might be.*

5.1.1 I can make generalizations from statistical data about a population sample.

5.1.2 I can answer questions related to sample size and validity: for example, “How large is a large enough sample size?”, “What makes a sample valid?”

5.1.3 I can make reasonable arguments about whether or not conclusions drawn from a sample are valid.

#### D5 Cluster 2: Draw informal comparative inferences about two populations.

3. Informally assess the degree of visual overlap of two numerical data distributions with similar variability, 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.*

4. Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *For example, decide whether the words in a chapter of a seventh-grade science book are generally longer than the words in a chapter of a fourth-grade science book.*

5.2.1 I can compare and draw informal inferences about two populations using measures of center (median, mean) and measures of variation (range, quartiles, interquartile range), visual overlap, and mean absolute deviation.

5.2.2 I can compare the degree of visual overlap of two data plots. I can describe the visual difference and explain what that difference means.

#### D5 Cluster 3: Investigate chance processes and develop, use, and evaluate probability models.

5. Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely, and a probability near 1 indicates a likely event.

6. Approximate the probability of a chance event by collecting data on the chance process that produces it and observing its long-run relative frequency, and predict the approximate relative frequency given the probability.

*For example, when rolling a number cube 600 times, predict that a 3 or 6 would be rolled roughly 200 times, but probably not exactly 200 times.*

7. Develop a probability model and use it to find probabilities of events. Compare probabilities from a model to observed frequencies; if the agreement is not good, explain possible sources of the discrepancy.

a. Develop a uniform probability model by assigning equal probability to all outcomes, and use the model to determine probabilities of events.

b. Develop a probability model (which may not be uniform) by observing frequencies in data generated from a chance process.

*For example, if a student is selected at random from a class, find the probability that Jane will be selected and the probability that a girl will be selected.*

*For example, find the approximate probability that a spinning penny will land heads up or that a tossed paper cup will land open-end down. Do the outcomes for the spinning penny appear to be equally likely based on the observed frequencies?*

8. Find probabilities of compound events using organized lists, tables, tree diagrams, and simulation.

a. Understand that, just as with simple events, the probability of a compound event is the fraction of outcomes in the sample space for which the compound event occurs.

b. Represent sample spaces for compound events using methods such as organized lists, tables and tree diagrams. For an event described in everyday language (e.g., “rolling double sixes”), identify the outcomes in the sample space which compose the event.

c. Design and use a simulation to generate frequencies for compound events.

*For example, use random digits as a simulation tool to approximate the answer to the question: If 40% of donors have type A blood, what is the probability that it will take at least 4 donors to find one with type A blood?*

5.3.1 I can explain why the numeric probability of an event must be between 0 and 1. I can explain the likelihood of an event occurring based on probability near 0, 1/2, and 1.

5.3.2 I can predict probability from collecting data (experimental). I can also establish the theoretical probability by determining the sample space.

5.3.3 I can find the probability of compound events by constructing models, i.e., lists, tables, tree diagrams, and simulation.

5.3.4 I can design and use a simulation to generate frequencies for compound events.

\*Please note: The following conceptual foundation highlights important concepts and instructional support, but is not all inclusive. Please refer to “I Can” statements above for a more complete picture of content.

## D5 Cluster 1: Use random sampling to draw inferences about a population.

**NOTE: Students should collect and interpret real data. They should experience the following steps multiple times and observe and question their peers' data and conclusions using the vocabulary below.**

1. Collect real data about a population. CONTEXT is important--interpreting data and graphs relies on real-life/world situations.
2. Start with a question which involves polling people or pulling random data from the larger population
3. Plan the random sampling and ask the following questions:
  - "How large is a large enough sample size?"
  - "What makes a sample valid?"
4. Organize the data using box plots and/or dot plots.
5. Interpret the data and the plots using the tools and vocabulary below.
6. Draw conclusions about the data and the population as related to the sample.
7. Observe and question their peers' data, plots, and conclusions

Lessons and data collection ideas can be found at: <http://middlemathccss.wordpress.com/7th-grade-math/7th-d5-statistics-and-probability/random-sampling-about-a-population-sample-size-and-validity-generalize-from-data-support-conclusions/>

### Important vocabulary:

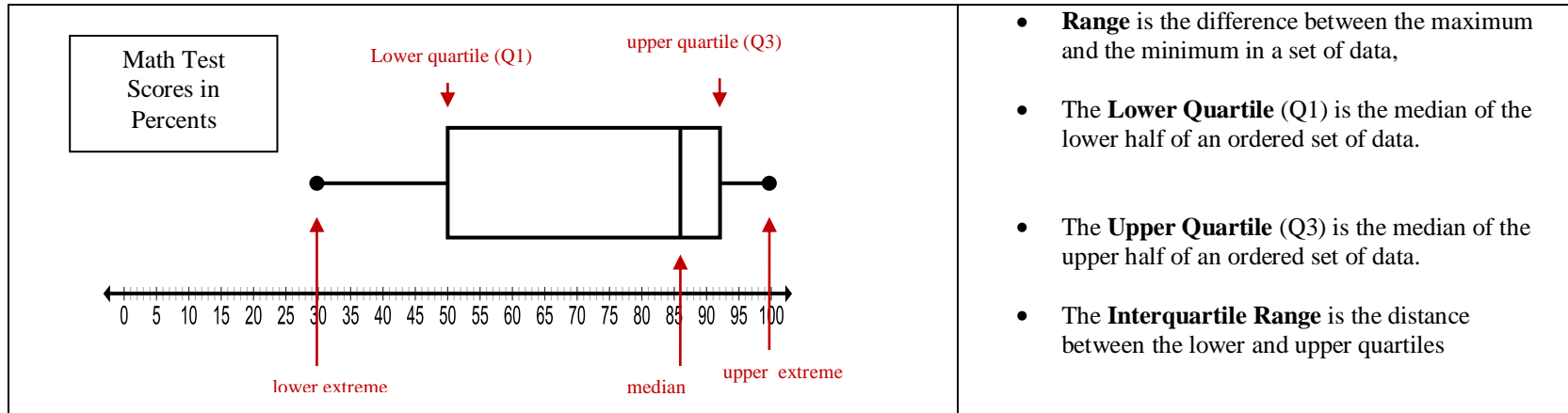
\* Access Vocabulary cards at <http://www.graniteschools.org/departments/teachinglearning/Curriculum/math/Pages/MathematicsVocabulary.aspx>

- **Inference:** The act of drawing a conclusion by deductive reasoning. The conclusion drawn is also called an inference. When students observe data and graphs and formulate observations and answer questions from the data, they are making inferences. (Inference riddles at: <http://www.philtulga.com/Riddles.html>)
- **Population:** A population can be defined as the set of all people or objects with the desired characteristic. Because there is very rarely enough time or money to gather information from everyone or everything in a population (census), the goal becomes finding a representative sample (or subset) of that population. Thus population relates to the sample size and validity (next).
- **Sample and sample validity:** A sample collected in such a way as to eliminate bias in the results. (*Sample validity example: If a magazine wants feedback on changes they made to their layout and they're collecting data based on a voluntary response card included in the magazine, it is not representative of the whole magazine readership population, because it is a voluntary response, typically only those who feel strongly will provide feedback.*)
- **Generalize:** To use statistics from samples to make broader conclusions (inferences) about the populations from which they came. Students might observe the data and graphs and generalize their inferences more broadly. The larger the sample size, typically the more valid the generalizability.
- **Justify:** To prove or show to be just, right, or reasonable. Students will need to justify the conclusions they make from graphs and data.
- **Measures of center:** Median and mean (if extreme values or outliers are present, the median may be a better measure of center)

Continued Below

## Important Vocabulary continued:

- **Measure of variability or spread** (use range, quartiles, interquartile range, or mean absolute deviation)



### To Create a Box Plot

- 1) Write the data in order from least to greatest.
  - 2) Draw and label a horizontal number line that can show the data in equal intervals.
  - 3) Find the median of the data set and mark it on the number line.
  - 4) Find the median of the upper half of the data. This is called the upper quartile (Q3). Mark it on the number line.
  - 5) Find the median of the lower half of the data. This is called the lower quartile (Q1). Mark it on the number line.
  - 6) Mark the lower extreme (minimum) on the number line.
  - 7) Mark the upper extreme (maximum) on the number line.
  - 8) Draw a box between the lower quartiles and the upper quartile. Draw a vertical line through the median to split the box.
  - 9) Draw a “whisker” from the lower quartile to the lower extreme.
  - 10) Draw a “whisker” from the upper quartile to the upper extreme.
- **Mean absolute deviation:** is an average of how far each data point in a set is from the mean of the set of data. Example: The weights of the three people are 56 kgs, 78 kgs, and 88 kgs.
    - Step 1: Find the mean.  $(56+78+88)/3 = 74$
    - Step 2: Determine the deviation of each variable from the mean.  $56 - 74 = -18$ ,  $78 - 74 = 4$ ,  $90 - 74 = 16$
    - Step 3: Make the deviation ‘absolute’ by taking the absolute value of each deviation.(just considering the distance from 0)
    - Step 4:  $(18 + 4 + 16)/3 = 12.67$  is the mean absolute deviation.

**D5 Cluster 2: Draw informal comparative inferences about two populations.**

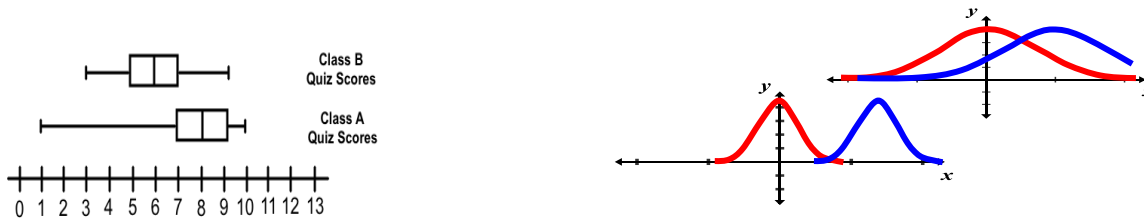
**NOTE: Students should experience the following multiple times.**

1. Collect real data about a population. CONTEXT is important--interpreting data and graphs relies on real-life/-world situations.
2. Start with a question which involves polling people or pulling random data from the larger population
3. Plan the random sampling and ask the following questions:
  - “How large is a large enough sample size?”
  - “What makes a sample valid?”
4. Organize the data using box plots and/or dot plots.
5. Interpret the data and the plots using the tools and vocabulary listed.
6. Draw conclusions about the data and the populations.
7. Observe and question their peers’ data, plots, and conclusions

Lessons and data collection ideas can be found at: <http://middlemathccss.wordpress.com/7th-grade-math/7th-d5-statistics-and-probability/random-sampling-about-a-population-sample-size-and-validity-generalize-from-data-support-conclusions/>

**Additional Vocabulary for two populations:**

- **Degree of visual overlap:** From a dot plot or histogram, pencil in a curve over the data, then observe and describe how the curve of the data overlaps or doesn’t overlap. From two box plots, you can visualize the overlap. What do the overlaps tell you about the data sets?



## Statistical Questions Students Should Be Able to Answer

### 6<sup>th</sup> Grade: (6<sup>th</sup>—Histograms, Dot Plots, Box Plots, Frequency Tables (one-variable numerical))

- What was the statistical question asked?
- How many observations were made or people were surveyed to gather the data?
- What was measured or counted by the statistical question?
- How was the attribute measured or counted?
- What unit of measurement is being used to describe the data?
- What is the shape of the data?
- Where are there clusters of data or gaps in the data?
- Are there any outliers in the data?
- Is mean or median a better descriptor for the measure of center for a particular set of data?
- What is the mean absolute deviation of the data points in the set?
- What is the range of the data?
- What is the lower quartile (Q1) or median of the lower half of the data?
- What is the upper quartile (Q3) or median of the upper half of the data?
- What can be observed about the statistical question asked?

### 7<sup>th</sup> Grade—same questions as 6<sup>th</sup> grade, but comparing two populations (7<sup>th</sup>—Dot/Box Plots)

- What are the measures of variation found in the data? (range, quartiles, interquartile range)
- Describe the visual overlap of the two populations? What can we learn about the populations from the visual overlap? How are they related or not related?
- What can be observed about the statistical question asked?
- What are the patterns of association
- What are our conclusions?

### 8<sup>th</sup> Grade—most of the questions from 6<sup>th</sup>-7<sup>th</sup> are appropriate for 8<sup>th</sup> concepts also (8<sup>th</sup>—Scatter Plots/Line of Fit, Frequency Tables)

#### Scatter Plots/Line of Fit

- What is the relationship between the two variables?
- What are the patterns of association shown on the graph of the data? (clustering, outliers, positive/negative association, linear/nonlinear association)
- What is the rate of change in the line of fit?
- What is the y intercept and what does it mean?
- What predictions might be made from the data and graph?

#### Frequency Tables (the frequency tables in 6<sup>th</sup> were one variable—numerical—for dot, histograms, box plots)

- What are the two variables
- What are the patterns of association between the two variables?
- What inferences might be made from observing the patterns of association between the two variables?
- How can you and justify the inferences made from the data and graphs?

## Investigate chance processes and develop, use, and evaluate probability models.

**NOTE:** Once, again, it's important to collect and use real data.

### Important vocabulary:

- **Likelihood of an event** Talk about the likelihood of an event as close to 0—not likely, close to 1—likely, close to  $\frac{1}{2}$ --slightly above or below (somewhat likely or somewhat unlikely), etc.
- **Sample Space** The total possible outcomes of an event.
- **Experimental and theoretical probability** Experimental—produces data, real or simulated. Theoretical—considers the outcome as related to the sample space.
- **Probability of Compound Events**—two or more independent events considered together, for example: what is the probability of rolling a head and rolling a 3?



$$\frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$$

- **Models for compound events** (lists, tables, tree diagrams)
- **Designing a simulation** to generate frequencies for compound events. For example, using a random number generator, pulling cubes from a bag to represent real objects, etc.