Time Frame: Approximately 2-4 Weeks

Connections to Previous Learning:

Students in Grade 6 learn the concepts of ratio and unit rate as well as the precise mathematical language used to describe these relationships. They learn to solve problems using ratio and rate reasoning using a variety of tools such as tables, tape diagrams, double number lines, and equations.

Focus of this Unit:

Students build upon their understanding of statistics by examining how selected data can be used to draw conclusions, make predictions, and compare populations.

Connections to Subsequent Learning:

Students will apply their experiences with data, coordinate planes, and linear functions to the study of variables related to a question of interest. They will analyze bivariate data through linear models and frequency tables.

From the 6-8 Statistics and Probability Progression document, pp. 8-10:

Random sampling: In earlier grades students have been using data, both categorical and measurement, to answer simple statistical questions; however, they have paid little attention to how the data were selected. A primary focus for Grade 7 is the process of selecting a random sample, and the value of doing so. If three students are to be selected from the class for a special project, students recognize that a fair way to make the selection is to put all the student names in a box, mix them up, and draw out three names "at random." Individual students realize that they may not get selected, but that each student has the same chance of being selected. In other words, random sampling is a fair way to select a subset (a sample) of the set of interest (the population). A statistic computed from a random sample, such as the mean of the sample, can be used as an estimate of that same characteristic of the population from which the sample was selected. This estimate must be viewed with some degree of caution because of the variability in both the population and sample data. A basic tenet of statistical reasoning, then, is that random sampling allows results from a sample to be generalized to a much larger body of data, namely, the population from which the sample was selected.

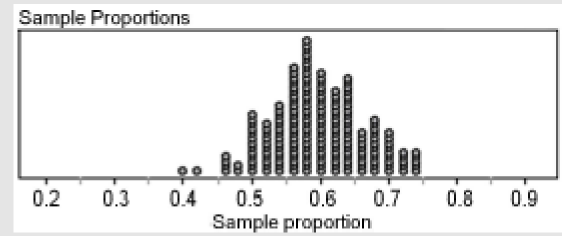
Female life expectancies in African and European countries

"What proportion of students in the seventh grade of your school chooses football as their favorite sport?" Students realize that they do not have the time and energy to interview all seventh graders, so the next best way to get an answer is to select a random sample of seventh graders and interview them on this issue. The sample proportion is the best estimate of the population proportion, but students realize that the two are not the same and a different sample will give a slightly different estimate. In short, students realize that conclusions drawn from random samples generalize beyond the sample to the population from which the sample was selected, but a sample *statistic* is only an estimate of a corresponding population *parameter* and there will be some discrepancy between the two. Understanding variability in sampling allows the investigator to gauge the expected size of that discrepancy.

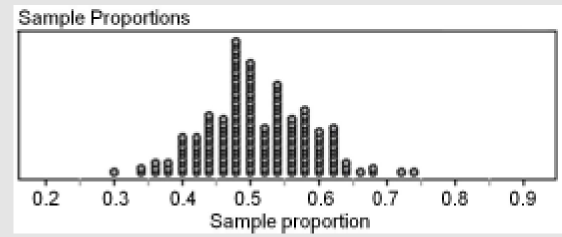
The variability in samples can be studied by means of simulation. Students are to take a random sample of 50 seventh graders from a large population of seventh graders to estimate the proportion having football as their favorite sport. Suppose, for the moment, that the true proportion is 60%, or 0.60. How much variation can be expected among the sample proportions? The scenario of selecting samples from this population can be simulated by constructing a “population” that has 60% red chips and 40% blue chips, taking a sample of 50 chips from that population, recording the number of red chips, replacing the sample in the population, and repeating the sampling process. (This can be done by hand, with the aid of technology, or by a combination of the two.) Record the proportion of red chips in each sample and plot the results.

The dot plots show results for 200 such random samples of size 50 each. Note that the sample proportions pile up around 0.60, but it is not too rare to see a sample proportion down around 0.45 or up around .0.75. Thus, we might expect a variation of close to 15 percentage points in either direction. Interestingly, about that same amount of variation persists for true proportions of 50% and 40%, as shown in the dot plots. Students can now reason that random samples of size 50 are likely to produce sample proportions that are within about 15 percentage points of the true population value. They should now conjecture as to what will happen if the sample size is doubled or halved, and then check out the conjectures with further simulations. Why are sample sizes in public opinion polls generally around 1000 or more, rather than as small as 50?

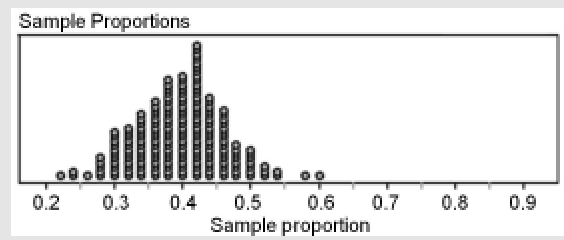
Results of simulations



Proportions of red chips in 200 random samples of size 50 from a population in which 60% of the chips are red.



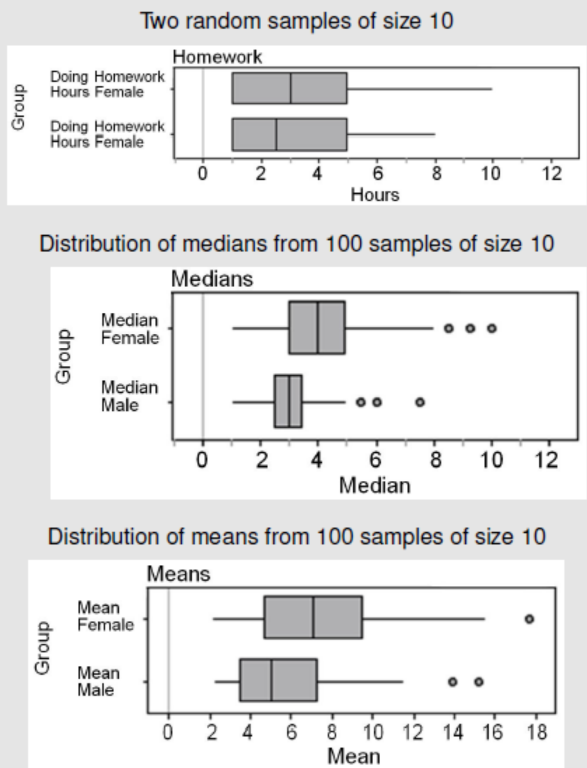
Proportions of red chips in 200 random samples of size 50 from a population in which 50% of the chips are red.



Proportions of red chips in 200 random samples of size 50 from a population in which 40% of the chips are red.

Informal comparative inference: To estimate a population mean or median, the best practice is to select a random sample from that population and use the sample mean or median as the estimate, just as with proportions. However, many of the practical problems dealing with measures of center are comparative in nature, such as comparing average scores on the first and second exam or comparing average salaries between female and male employees of a firm. Such comparisons may involve making conjectures about population parameters and constructing arguments based on data to support the conjectures (MP3).

Hours spent on homework per week



Source: Census at Schools Project, `amstat.org/censusatschool/`

If all measurements in a population are known, no sampling is necessary and data comparisons involve the calculated measures of center. Even then, students should consider variability. The figures in the margin show the female life expectancies for countries of Africa and Europe. It is clear that Europe tends to have the higher life expectancies and a much higher median, but some African countries are comparable to some of those in Europe. The mean and MAD for Africa are 53.6 and 9.5 years, respectively, whereas those for Europe are 79.3 and 2.8 years. In Africa, it would not be rare to see a country in which female life expectancy is about ten years away from the mean for the continent; however in Europe the life expectancy in most countries is within three years of the mean.

For random samples, students should understand that medians and means computed from samples will vary from sample to sample and that making informed decisions based on such sample statistics requires some knowledge of the amount of variation to expect. Just as for proportions, a good way to gain this knowledge is through simulation, beginning with a population of known structure.

The following examples are based on data compiled from nearly 200 middle school students in the Washington, DC area participating in the Census at Schools Project. Responses to the question, "How many hours per week do you usually spend on homework?," from a random sample of 10 female students and another of 10 male students from this population gave the results plotted to the left.

Females have a slightly higher median, but students should realize that there is too much variation in the sample data to conclude that, in this population, females have a higher median homework time. An idea of how much variation to expect in samples of size 10 is needed.

Simulation to the rescue! Students can take multiple samples of size 10 from the Census of Schools data to see how much the sample medians themselves tend to vary. The sample medians for 100 random samples of size 10 each, with 100 samples of males and 100 samples of females, is shown to left. This plot shows that the sample medians vary much less than the homework hours themselves and provides more convincing evidence that the female median homework hours is larger than that for males. Half of the female sample medians are within one hour of 4 while half of the male sample medians are within half hour of 3, although there is still overlap between the two groups.

A similar analysis based on sample means gave the results seen in the margin. Here, the overlap of the two distributions is more severe and the evidence weaker for declaring that the females have higher mean study hours than males.

Desired Outcomes

Standard(s):

Use random sampling to draw inferences about a population.

- 7.SP.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.
- 7.SP.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.*

Draw informal comparative inferences about two populations

- 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.*
- 7.SP.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.*

Supporting Standards:

- 7.NS.1 Apply and extend previous understandings of addition and subtraction to add and subtract rational numbers; represent addition and subtraction on a horizontal or vertical number line diagram.
  a) Describe situations in which opposite quantities combine to make 0. *For example, a hydrogen atom has 0 charge because its two constituents are oppositely charged.*
  b) Understand $p + q$ as the number located a distance $|q|$ from p, in the positive or negative direction depending on whether q is positive or negative. Show that a number and its opposite have a sum of 0 (are additive inverses). Interpret sums of rational numbers by describing real-world contexts.
  c) Understand subtraction of rational numbers as adding additive inverse, $p - q = p + (-q)$. Show that the distance between two rational numbers on the number line is the absolute value of their difference, and apply this principle in real-world contexts.
  d) Apply properties of operations as strategies to add and subtract rational numbers.
- 7. NS.2 Apply and extend previous understandings of multiplication and division and of fractions to multiply and divide rational numbers.
  a) Understand that multiplication is extended from fractions to rational numbers by requiring that operations continue to satisfy the properties of operations, particularly the distributive property, leading to products such as (-1)(-1) = 1 and the rules for multiplying signed numbers, interpret products of rational numbers by describing real-world contexts.
  b) Understand that integers can be divided, provided that the divisor is not zero, and every quotient of integers (with non-zero divisor) is a rational number. If p and q are integers, then $-(p/1) = (-p)/q = p/(-q)$. Interpret quotients of rational numbers by describing real-world contexts.
  c) Apply properties of operations as strategies to multiply and divide rational numbers.
  d) Convert a rational number to a decimal using long division; know that the decimal form of a rational number terminates in 0s or eventually repeats.
- 7. NS.3 Solve real-world and mathematical problems involving the four operations with rational number.
- 7.EE.2 Understand that rewriting an expression in different forms in a problem context can shed light on the problem and how the quantities in it are related. *For example, $a + 0.05a = 1.05a$ means that "increase by 5% is the same as multiply by 1.05."*
- 7.EE.3 Solve multi-step real-life and mathematical problems posed with positive and negative rational numbers in any form (whole numbers, fractions, and decimals), using tools strategically. Apply properties of operations to calculate with numbers in any form; convert between forms as appropriate; and assess the reasonableness of answers using mental computation and estimation strategies. *For example: If a woman making $25 an hour gets a 10% raise, she will make an additional 1/10 of her salary an hour, or $2.50 for a new salary of $27.50. If you want to place a towel bar 9 ¾ inches long in the center of a door that is 27 ½ inches wide, you will need to place the bar about 9 inches from each edge; this estimate can be used as a check on the exact computation.*

Transfer:

Students will apply concepts and procedures to draw inferences about a population or two populations using random sampling or informal comparisons.

Ex: In a survey of 30 students, 21 like pizza. If there are 600 students in the school, predict how many students like pizza?

Understandings: *Students will 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.
- Random sampling tends to produce representative samples and support valid inferences.
- Two data distributions can be compared using visual and numerical representations based upon measures of center and measures of variability to draw conclusions.

Essential Questions:

- How can two data distributions be compared?
- How can statistics be used to gain information about a sample population?
- How can a random sample of a larger population be used to draw inferences?

Mathematical Practices: (Practices to be explicitly emphasized are indicated with an *.)

1. **Make sense of problems and persevere in solving them.** Students make sense of information by connecting visual, tabular, and symbolic representations of sample populations in real-life contexts.

* 2. **Reason abstractly and quantitatively.** Students' reason about the values in data representations based upon their relationship to the real number line.

* 3. **Construct viable arguments and critique the reasoning of others.** Students use data to make inferences from sample sets. They construct viable arguments by referring to representations as evidence of their inferences and question each other regarding these inferences.

* 4. **Model with mathematics.** Students generate representative samples in real-world contexts and represent these visually, in tables, and symbolically to gain information from sample sets.

5. **Use appropriate tools strategically.** Students choose appropriate mathematical and visual representations, including technology-based tools, to represent the data distributions.
6. **Attend to precision.** Students use precision to collect accurate measurement information from sample populations and precise language when generating and interpreting data.
7. **Look for and make use of structure.** Students interpret data representations in tables, histograms, box plots, and scatter plots by examining the features of those representations.
8. **Look for and express regularity in repeated reasoning.** Students use their prior experiences with visual representations of data to interpret new data sample sets.

Prerequisite Skills/Concepts:	Advanced Skills/Concepts:
Students should already be able to: • Develop understanding of statistical variability. (6.SP.1-3) • Summarize and describe distributions. (6.SP.4-5)	*Some students may be ready to:* • Investigate patterns of association in bivariate data (8.SP.1-4)
Knowledge: *Students will know…* All standards in this unit go beyond the knowledge level.	**Skills:** *Students will be able to …* • Recognize and identify that different sampling techniques must be used in real life situations, because it is very difficult to survey an entire population. (7.SP.1) • Select appropriate sample sizes based on a population in real-life situations and explain why generalizations about a population from a sample are valid only if the sample is random and representative of that population. (7.SP.1) Collect data from a sample population in order to predict information about a population. (7.SP.1) • Interpret data from a random sample to draw inferences about a population with an unknown characteristic of interest. (7.SP.2) Generate multiple samples (or simulated samples) of the same size to determine the variation in estimates or predictions by comparing the samples. (7.SP.2) • Identify the degree of overlap between two numerical sets of data. (7.SP.3) • Visually compare two numerical data distributions with like ranges. (7.SP.3) Measure the difference between the centers of two different data distributions and express this difference as a multiple of a measure of variability. (7.SP.3) • Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. (7.SP.4)

WIDA Standard: (English Language Learners)

English language learners communicate information, ideas, and concepts necessary for academic success in the content area of Mathematics.
English language learners benefit from:

- Explicit instruction regarding statistical language with regard to visual data representations.
- Clear examples of bias expressed using simple language and supported with visual representations.

Academic Vocabulary:

Critical Terms:	Supplemental Terms:
Random sample Biased sample Unbiased sample Histogram Box plot Dot plot Double box plot Double dot plot	Statistics Mean Median Mode

Assessment

Pre-Assessments	Formative Assessments	Summative Assessments	Self-Assessments
• Displaying and Analyzing Data Stations • Is it Statistical?	• Survey Analysis • Random Sampling • Critiquing the Analysis of Random Sampling • Rock and Roll Illinois- Visual Overlap • Mean Absolute Deviation • Height and Feet	• Peer Survey	• Peer Survey Rubric

Sample Lesson Sequence

1. **What is sampling/random sampling – one population 7.SP.1 and 7.SP.2**
2. Analyzing Data from Random Sampling – one and two populations 7.SP.3 and 7.SP.4

Major Standards Supporting Standards Additional Standards