Project-SET Sampling Variability Final Learning Trajectory [1,2]

GAISE Framework[3]

	Formulate Question	Collect Data	Analyze Data	Interpret Results	Key Developmental Understandings
Loop 1[4] **Concept of a Sampling Distribution**	a. How can we discover how a summary statistic varies from random sample to random sample taken from a population of interest? b. My friend and I just collected data from the same population. Can I expect my friend's summary statistic to be close to mine? c. Are some values of the statistic more common than others? d. How can we describe the variation of a statistic from one sample-to-another sample?	a. Describe methods that generate repeated samples of the same size from a population. b. Take repeated samples of the same size "by hand" and compute the summary statistic for each sample (e.g., collect packs of Skittles and find the proportion in each pack that are orange) *It should be noted that "by hand" simulations of this type are solely for pedagogical purposes in order to help understand computer simulations. c. Construct an approximate sampling distribution using computer-aided simulation	a. Notice that different samples give different summary statistic values for the population characteristic of interest. b. Record the distribution of the summary statistic from the different random samples by making a dot plot c. Summarize the simulated sampling distribution using shape, center, and variability also looking for potential outliers	a. Informally relate the summary statistic to the population parameter; that is, a single statistic is an estimate of the population parameter b. Link the variability in the summary statistic from sample to sample to the variability in the constructed sampling distribution and make conjectures about what might affect this variability. c. Distinguish between the population distribution, distribution of a sample, and the sampling distribution	The sampling distribution is a distribution that describes how a statistic varies for repeated samples from the same population. An approximate sampling distribution can be constructed, using simulation. This approximate sampling distribution can be described by its shape (typically symmetric or skewed), center (mean or median), and variability (typically, standard deviation).

Loop 2[5] **Using the Sampling Distribution to Examine Whether a Claimed Parameter is Plausible**		a. For a given population parameter, how can we tell which summary statistics from that population are common and which are unusual? b. Given a claim about a (unknown) population parameter, how can we decide whether it is plausible? (e.g., Claim: Twenty percent of Skittles are orange.) *It is important to note that these procedures work with certain types of claims only. When thinking of claims to implement, one must keep in mind the horizon knowledge of how to set up claims for hypothesis testing.	a. Describe how a sampling distribution can answer the posed question b. Use simulation to construct a sampling distribution sampling from a population with a parameter equal to the claimed parameter c. Take one random sample from the population with unknown parameter and compute the summary statistic of interest	a. Understand that the summary statistic varies in a predictable way, where predictable does not imply that we can predict the next summary statistic, but that we can predict the distribution of the summary statistic from repeated samples. b. Locate the summary statistic from the one sample taken on the sampling distribution.	a. Decide whether the summary statistic would be common or rare in the sampling distribution for the claimed parameter. b. If common, call that parameter a plausible one for the population from which the sample was taken. If rare, the evidence suggests the claimed parameter is not plausible. c. Write an informal conclusion concerning the claim.	Each summary statistic is a single estimate of the population parameter. The sampling distribution shows which values of the statistic are common and which are rare. If the summary statistic from a sample would be rare for the claimed parameter, that parameter isn't considered plausible as the parameter for the population from which the sample came.

Loop 3 **The Effect of Shape of Population and of the Sample Size on the Sampling Distribution**		a. Can the shape of the sampling distribution be predicted? b. How will the sampling distributions compare for samples taken from an approximately normal population? From a skewed population? From a bimodal population? c. What is the effect of the sample size on the shape, center, and variability of the sampling distribution of the mean?	a. Predict the shape, center, and variability of the sampling distributions for parts b, c, and d below. b. From an approximately normal population, construct three sampling distributions using a: (1) small sample size (less than 30), (2) medium sample size (30 to 100), and (3) large sample size (more than 100) c. Repeat part b with a skewed population. d. Repeat part b with a bimodal population.	a. For each sampling distribution, describe the shape, center and variability.	a. Describe what happens to the shape of the sampling distribution as the size of the samples increases. b. State the Central Limit Theorem c. Describe what happens to the mean of the sampling distribution as the size of the samples increases d. Describe what happens to the standard deviation of the sampling distribution as the size of the samples increases	The mean of the sampling distribution of the mean does not depend on the sample size. It is in theory always equal to the mean of the population. The variability of the sampling distribution decreases as the sample size increases, according to the formula $\sigma / \sqrt{n}$. The shape of the sampling distribution looks more approximately normal as the sample size increases.

Loop 4[6] **Sampling distributions for other statistics**		a. Do all sampling distributions behave like that of the mean? b. What is the expected behavior of sampling distribution of the median or the max or the range or the min?	a. Simulate an approximate sampling distribution for the median or any other parameter other than the mean or proportion.	b. Describe the shape, center and variability of the sampling distribution of the median.	a. Compare the approximate sampling distributions constructed to that of the mean or proportion found by the CLT. b. Review the difference between shape and variability (spread) of the sampling distribution.	Not all sampling distributions have a nice theoretical basis like the CLT.

Loop 5[7] **Using Sampling Distributions to Estimate a Parameter**		a. How can we estimate a population parameter and how can we decide if our estimate is accurate? b. How can we construct an interval of plausible parameter values using one sample? c. How well do intervals of plausible values capture the parameter? d. Given a random sample from a population with unknown parameter, what are the plausible values for that parameter?	a. Take one random sample from the population with unknown parameter and compute the summary statistic b. Simulate 1000 confidence intervals for a parameter	a. Describe the sampling distribution for the summary statistic found using the CLT b. Construct formal confidence interval for a population parameter using data from a single sample	a. Describe how knowing the sampling distribution for a sample statistic allows you to determine what specific plausible values of the parameter might be b. Describe what the simulated confidence intervals illustrate about the capture rate of the population parameter b. Understand why the unknown population parameter isn't necessarily in the confidence interval. c. Interpret the confidence interval in the context of the unknown population parameter.	A single sample can be used to produce an interval estimate of the population parameter. Confidence intervals include those parameters that have sampling distributions in which the summary statistic is common.

[1]The prerequisite knowledge needed to work through the LP consists of three items. A learners must be able to: (1) Define and understand the sample/population relationship (e.g., a sample is a portion of the population and a sample can be a representative portion of the population), (2) Define population parameters and summary statistics, (3) Define sampling methods (e.g., define how to obtain a representative sample)

[2] The Project-SET LP is organized around four dimensions: (1) the progression of the topic within a loop, (2) the sophistication on each loop (3) the alignment with GAISE, and (4) the alignment with the CCSS. The loops are meant to illustrate the different depths of the concept of sampling variability.

[3] The Project-SET LP directly aligns with the GAISE Framework. To illustrate the alignment, the LP is organized around the four GAISE components depicted as the columns of the LP.

[5] Loop 2 of the Project-SET LP aligns with CCSS S-IC.1 " Understand statistics as a process for making inferences about population parameters based on a random sample from that population."

[6] Loop 5 of the Project-SET LP aligns with CCSS S-IC.2 "Decide if a specified model is consistent with results from a given data-generating process using simulation."

[7]The entry for Interpret Results of Loop 5 directly aligns with CCSS S-CI.1 " Understand statistics as a process for making inferences about population parameters based on a random sample from that population."

[7]The entry for Interpret Results of Loop 5 directly aligns with CCSS S-CI.1 " Understand statistics as a process for making inferences about population parameters based on a random sample from that population."