**Activity 7.4.1 – Exploring Distributions of Sample Means**

*Statistical inference* is the process of using sample statistics to draw conclusions about population parameters. *Sample statistics* are numerical descriptions of sample characteristics (e.g. sample mean, sample median). *Population parameters* are numerical descriptions of population characteristics (e.g. population mean, population median). Population parameters are typically unknown, so we use sample statistics to make inferences about population parameters.

To understand statistical inference, we must first understand the behavior of sample statistics from random samples. This activity will focus on sample means, how they vary, and how they relate to the underlying population mean.

**The Average Age of a Penny**

This activity explores the age of pennies currently in circulation. Each student received a bag of 20 pennies, 2 nickels, 2 dimes, and 1 quarter. Before proceeding with the activity, make sure that these are the correct counts in your bag.

1. Begin by recording the age of pennies in your bag. The penny’s age is determined by $current year-year on penny$. Create a dotplot of the ages of the pennies in your bag.



1. Now create a dotplot of the entire distribution. For this activity, we will consider this distribution to be the “population”. Place your pennies on the ages to form a class dotplot. It is not necessary that you place each penny on the actual spot that corresponds to its age.

Estimate the mean and standard deviation of the population.

1. What is the shape of this population distribution?
2. Let’s now explore sample means from random samples of size *n* = 5. Each student will calculate the mean age of a random sample of 5 pennies. Suppose we create a distribution of the sample means. What do you think the distribution will look like?
3. Select a random sample of *n* = 5 pennies from the population. Calculate the mean of the penny ages. Use the symbol $\overbar{x}(5)$ to denote this sample mean. Round your answer to one decimal place.

Sample mean

$\overbar{x}=\frac{\sum\_{}^{}x}{n}$

1. How much does $\overbar{x}(5)$ deviate from the population mean?
2. Return the pennies to your teacher. Now select a random sample of *n* = 10 pennies from the population. Calculate the mean of the penny ages. Use the symbol $\overbar{x}(10)$ to denote this sample mean. Round your answer to one decimal place.
3. How much does $\overbar{x}(10)$ deviate from the population mean?
4. Return the pennies to your teacher. Now select a random sample of *n* = 25 pennies from the population. Calculate the mean of the penny ages. Use the symbol $\overbar{x}(25)$ to denote this sample mean. Round your answer to one decimal place.
5. How much does $\overbar{x}(25)$ deviate from the population mean?

We will now construct *distributions of sample means* for each sample size. Your teacher has set up three dot plots, one for each sample size.

1. Sample means, *n* = 5. On the first dot plot, place your two nickels on the closest mean for $\overbar{x}(5)$. You are using two nickels to accentuate the height of the graph.
2. Sample means, *n* = 10. On the second dot plot, place your two dimes on the closest mean for $\overbar{x}(10)$. You are using two dimes to accentuate the height of the graph.
3. Sample means, *n* = 25. On the third dot plot, place your quarter on the closest mean for $\overbar{x}(25)$.

The dot plots above are examples of *empirical sampling distributions* – distributions of sample statistics obtained by simulating random sampling from a population. Similar to other distributions, an empirical sampling distribution has a center, shape and spread that summarize values in its distribution.

1. Estimate the means of the three empirical sampling distributions. What happens to the mean of the empirical sampling distribution as the sample size increases?
2. Estimate the standard deviation of the three empirical sampling distributions. What happens to the standard deviation of the empirical sampling distribution as the sample size increases?
3. Compare the shapes the three empirical sampling distributions. What happens to the shape of the empirical sampling distribution as the sample size increases?
4. Summarize what you have learned about distributions of sample means.

**Key Understanding:** Sampling Distributions of Sample Means

A *sampling distribution of sample means* is the distribution of *all* possible sample means from random samples of the same size.

* The mean of a sampling distribution of sample means is approximately equal to the population mean.
* The standard deviation of a sampling distribution of sample means is called the *standard error.*
* For *any* population, as the sample size increases, the sample means become less variable and the sampling distribution of sample means approaches a bell-shaped distribution.