**Activity 7.3.1 – Exploring Distributions of Sample Proportions**

*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 proportion, sample mean). *Population parameters* are numerical descriptions of population characteristics (e.g. population proportion, population mean). When population parameters are unknown we use sample statistics to make inferences about population parameters.

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

**Population of High School Students**

The table on the following page displays a hypothetical population of *N* = 50 high school students. Each student has identified his or her favorite season. Of the 50 students, 24 prefer summer, so the population proportion of all students who prefer summer is 0.48. We use *p* to denote population proportion, so ***p* = 0.48**.

1. Select a random sample of *n* = 10 students from the population. Count the number of students who prefer summer and calculate the sample proportion. Use the symbol , pronounced p-hat, to denote the sample proportion. Express it as a fraction and decimal.

In this situation, a “success” is a student who prefers summer.

1. Find a second random sample of *n* = 10 students from the population. Count the number of students who prefer summer and calculate the sample proportion. Use the symbol to denote the sample proportion. Express it as a fraction and decimal.

**Table:** Population of *N* = 50 high school students

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student** | **Favorite Season** |  | **Student** | **Favorite Season** |
| 1 | Summer |  | 26 | Fall |
| 2 | Spring |  | 27 | Summer |
| 3 | Fall |  | 28 | Summer |
| 4 | Summer |  | 29 | Fall |
| 5 | Summer |  | 30 | Spring |
| 6 | Fall |  | 31 | Summer |
| 7 | Summer |  | 32 | Summer |
| 8 | Spring |  | 33 | Summer |
| 9 | Summer |  | 34 | Winter |
| 10 | Fall |  | 35 | Summer |
| 11 | Fall |  | 36 | Summer |
| 12 | Summer |  | 37 | Winter |
| 13 | Winter |  | 38 | Fall |
| 14 | Summer |  | 39 | Fall |
| 15 | Fall |  | 40 | Spring |
| 16 | Summer |  | 41 | Winter |
| 17 | Winter |  | 42 | Summer |
| 18 | Spring |  | 43 | Winter |
| 19 | Spring |  | 44 | Summer |
| 20 | Summer |  | 45 | Spring |
| 21 | Summer |  | 46 | Summer |
| 22 | Fall |  | 47 | Summer |
| 23 | Summer |  | 48 | Winter |
| 24 | Spring |  | 49 | Summer |
| 25 | Fall |  | 50 | Summer |

*Note*: This is a hypothetical population of high school students. The data are from actual high school students who participated in the *Census at Schools* project.

1. You and your classmates have generated a distribution of sample proportions. Plot your sample proportions and your classmates’ sample proportions on the dot plot below.

**Distribution of Sample Proportions, *n* = 10**



The dot plot above is an example of an *empirical sampling distribution* – a distribution 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 or calculate the mean and standard deviation of the sample proportions in the empirical sampling distribution.
2. What is the shape of the empirical sampling distribution?
3. Which sample proportions appear to be unusual? Explain.
4. How does the mean of the empirical sampling distribution compare the population proportion *p* = 0.48?
5. *Sampling variability* refers to the fact that the distribution of sample proportions from random samples varies in a predictable way. How has this been exemplified in this activity?

**The Role of Sample Size**

What happens if we increase the size of the random samples? For example, what would happen to the distribution of sample proportions if the sample size increased from *n* = 10 to *n* = 20?

1. Make a conjecture to answer the previous question.
2. Select two random samples of *n* = 20 students from the population on page 2. For each sample, count the number of students who prefer summer and calculate the sample proportion. Use the symbol , pronounced p-hat, to denote the sample proportion. Express each as fraction and decimal.

Sample proportion 1:

Sample proportion 2:

1. You and your classmates have generated a distribution of sample proportions. Plot your sample proportions and your classmates’ sample proportions on the dot plot below.

**Distribution of Sample Proportions, *n* = 20**



1. Estimate or calculate the mean and standard deviation of the sample proportions in the empirical sampling distribution.
2. What is the shape of the empirical sampling distribution?
3. Which sample proportions appear to be unusual? Explain.
4. How did the center, shape, and variability of the distribution of sample proportions change when the sample size increased from *n* = 10 to *n* = 20?

The summary below highlights the key concepts that we explored in this activity.

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

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

* The mean of a sampling distribution of sample proportions is the population proportion.
* The standard deviation of a sampling distribution of sample proportions is called the *standard error*.
* As the sample size increases, the standard error of the sampling distribution decreases.
* As the sample size increases, the sampling distribution of sample proportions approaches a bell-shaped and symmetric distribution.

1. Sample proportions are *estimates* of the population proportion. If we need to estimate a population proportion with a sample proportion, why would it be beneficial to select a large-sized random sample?
2. If the proportion of the student population who prefer summer to the other seasons is 0.48, would it be unlikely to find a random sample of *n* = 10 students in which the sample proportion prefer summer is 0.70 or higher? Explain.
3. If the proportion of the student population who prefer summer to the other seasons is 0.48, would it be unlikely to find a random sample of *n* = 20 students in which the sample proportion prefer summer is 0.70 or higher? Explain.