**Activity 7.3.4 – Hypothesis Tests on Population Proportions**

*Statistical inference* is the process of using sample statistics to make conclusions about population parameters. This activity explores how a sample proportion can be used to assess whether a claim about the value of a population proportion is reasonable.

**Using Sample Evidence to Test a Claim about an Unknown Population Proportion**

A teacher announces that she will randomly select students to come to the board to present their work. She doesn’t describe how, but simply states that she will use a random process to choose students so that each student is equally likely to be chosen on each selection. There are 5 students in the class.

One day, Kevin – one of the students in the class – is selected to come to the board two times in a row. Kevin finds it hard to believe that this could have occurred solely due to chance. Does this particular sequence provide evidence that the teacher is not using a random process to select students?

1. Assuming the teacher randomly selects students, what is the probability that Kevin is selected to present?
2. Assuming the teacher randomly selects students, what is the probability that Kevin is selected to present two times in a row?
3. Does Kevin’s sequence provide evidence that the teacher is not using a random process to select students?
4. Suppose Kevin was selected to present three times in a row. Would this sequence provide evidence that the teacher is not using a random process to select students? Explain.
5. If Kevin was selected to present three times in a row, can we conclude with absolute certainty that the teacher is not using a random process to select students? Explain.

The situation above can be addressed by a *hypothesis test*. In a hypothesis test we use sample statistics to test a claim made about the value of a population parameter. In doing so, we assume that the value of the parameter equals a specific value and then assess the likelihood of observing the sample results under this assumption. Hypothesis tests involve uncertainty since we never actually know that value of the population parameter.

**Randomization Hypothesis Test**

A *randomization hypothesis test* assesses the likelihood of a sample statistic occurring using a randomization distribution. A *randomization distribution* is a distribution of sample statistics randomly obtained via simulations from a population with a fixed parameter.

**Who Is Paying for College?**

A 2015 Junior Achievement USA® and The Allstate Foundation survey reported that 48% of teenagers in the United States think their parents will help them pay for college. A school counselor obtains a random sample of 20 teenagers at her school and finds that 70% of them think their parents will help them pay for college. Does the sample provide evidence that the proportion of all teenagers at this school who hold this view is actually higher than 48%?

**Claim**

**Sample 1**

To answer this question we will make an unusual step: we will assume the population proportion *p* is 0.48. This is the *hypothesis* we test. Then, we will find the probability of obtaining a sample proportion of 0.70 or higher, assuming the population proportion *p* is 0.48. We do this via a randomization distribution of sample proportions.

**Constructing a Randomization Distribution of Sample Proportions**

We can construct a randomization distribution of sample proportions by:

* Defining a hypothetical population with proportion equal to the hypothesized proportion,
* Randomly sample (with replacement) from the hypothetical population to generate a randomized distribution of sample proportions

We can model a population with proportion *p* = 0.48 using random numbers as follows:

* The population consists of all two-digit numbers: 00 to 99
* Numbers from 00 to 47 correspond to “successes” – teenagers who think their parents will help pay for college (48% of the population)
* Numbers from 48 – 99 correspond to “failures” – teenagers who do not think their parents will help pay for college (52% of the population)

1. *Random Sample 1* – Randomly select 20 two-digit numbers. Count the number of two-digit numbers that correspond to “successes” – teenagers who think their parents will help pay for college. Find the sample proportion of “successes”.
2. *Random Sample 2* – Randomly select 20 two-digit numbers. Count the number of two-digit numbers that correspond to “successes” – teenagers who think their parents will help pay for college. Find the sample proportion of “successes”.
3. Create a dotplot of sample proportions using all the sample proportions from your class.



The dot plot above is a *randomization distribution.* It was formed under the assumption that the population proportion is *p* = 0.48. The variability in the sample proportions is due to sampling variability (random chance).

1. Use technology to determine the mean and standard deviation of sample proportions in the randomization distribution.
2. The *observed* sample proportion is . According to the randomization distribution, what is the probability of obtaining a random sample with a sample proportion of 0.70 or higher?

A *P*-value is the probability of obtaining a sample statistic as or more extreme as the one observed assuming the population parameter is equal to a specific value.

* When a *P*-value is less than 5%, we say the sample statistic is *statistically significant*. This means that that it is likely that the statistic did not emerge due to chance alone. We reject the assumption about the population parameter.
* When a *P*-value is greater than or equal to 5%, we say the sample statistic is *not statistically significant*. This means that the sample statistic could have occurred solely due to chance. We do not reject the assumption about the population parameter.

1. Is the observed sample proportion statistically significant? Explain.
2. We tested the hypothesis that the proportion of all teenagers at the school who think their parents will help them pay for college is 0.48. What can we conclude about this hypothesis? What can we conclude about the population proportion *p*?

**Hopeful About the Present and Future**

A 2013 Gallup Poll of approximately 590,000 students in grades 5 – 12 in the United States found that 54% of students are hopeful about the present and future. Suppose we believe the proportion of students at our school that feel this way is greater than 54%. We randomly survey 80 students and find that 52 are hopeful about the present and future. Does this sample provide evidence that the population proportion at our school who are hopeful about the present and future is greater than 0.54?

We can conduct a randomization test as follows:

* Assume the proportion of students at our school who are hopeful about the present and future is *p* = 0.54. This is the hypothesis we test.
* Construct a randomization distribution of sample proportions from random samples of size

*n* = 80.

* Find the probability of observing a sample proportion as or more extreme than the one found.
* State a conclusion about the population proportion.

The following graph shows a randomization distribution of 200 sample proportions from random samples of size *n* = 80. The population proportion is assumed to be *p* = 0.54.

Randomization Distribution of Sample Proportions



1. The observed sample proportion is . According to the randomization distribution, what is the probability of obtaining a random sample with a sample proportion of 0.65 or more?
2. Is the observed sample proportion statistically significant? Explain.
3. We tested the hypothesis that the proportion of all students at the school who are hopeful about the present and future is 0.54. What can we conclude about this hypothesis? What can we conclude about the population proportion *p*?