About the Mathematics

Do senior citizens and college students have different memories about high school? The activity *Confidence Intervals: 2-Sample Proportions* involves investigating random samples from two populations from a large Midwestern city with respect to the question: 'When you were in high school, did you have a hero?' How can responses from two random samples from each population be used to estimate the responses of all senior citizens and all college students in this city?

This activity will give students the following opportunities:

1. Determine that the assumptions concerning the samples are met (large population relative to the sample, independence, differences are normally distributed)

2. Generate simulated sampling distributions for known differences between the proportions of two populations who responded "yes" to the question.

3. Compare the simulated sampling distributions to the observed difference in proportions for the two random samples to determine a set of population differences for which the observed 2-sample difference could contain the difference of the proportions of the two populations. This set is a confidence interval for the true difference in proportions for the two populations.

TI-Nspire™ Technology Skills:
- Open a document
- Move between Pages
- Click on a minimized slider

Lesson Materials:
Confidence_Intervals_2-Sample_ Proportions.tns

Visit www.mathnspired.com for lesson updates.

The approach to confidence intervals in this activity is intended to provide the conceptual understanding of a **confidence interval** for the difference in two proportions. Interpreting this interval and what it indicates about the differences in the two populations is the focus of the investigation developed with the .tns file. The formula for the confidence interval for the difference in two proportions that appears in most texts will produce an interval based on the observed difference between two samples but can obscure the interpretation of the results. This activity provides a foundation for that interpretation.
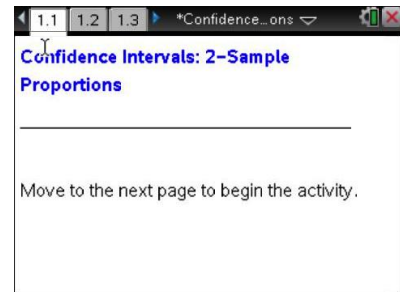
Math Objectives
- Students will find and interpret an estimated confidence interval for the difference between two proportions from samples randomly selected from two independent populations.
- Students will reason abstractly and quantitatively (CCSS Mathematical Practices).

The Problem Situation
Do senior citizens recall their high school memories in the same way as college-aged students? The following given information is summarized on Page 1.5 of the .tns file:

- The question, "When you were in high school, did you have a hero?", was asked of 22 randomly-selected senior citizens (identified as generation 1 or g1) and 62 randomly-selected college students (identified as generation 2 or g2) from a large Midwestern city.

- Twenty senior citizens responded "yes" to this question while 10 college students responded "yes".

- The **proportion** of senior citizens who answered "yes" was approximately 0.909; the **proportion** of college students who answered "yes" was approximately 0.161.

- The **difference** in the proportion of senior citizens who responded "yes" and the college students who responded "yes" is approximately 0.747.

Question: What does this observed difference in proportions from the samples indicate about the difference between the proportions of the population of senior citizens and the population of college students who would respond "yes" to the survey question?

The task in this activity is to determine a **confidence interval** to make an inference about the real difference in the proportions of all senior citizens and all college students who would answer "yes" to this question.

Using the Document

Page 2.1: Generate Simulated Samples and Graph

Clicking the arrow in the top work area on Page 2.1 will generate a simulated sampling distribution of the proportion of 'yeses' in random samples for g1 and g2, in sets of 10, until the proportions of 'yeses' for 100 samples are displayed. Each simulated sample of g1 and of g2 is drawn from a population where the proportion who answered 'yes' to the question is approximately 91% (0.909) and 16% (0.161) respectively.

Each point in the dotplot for g1 represents the proportion of 22 senior citizens who answered yes in that sample. Each point in the dotplot for g2 represents the proportion of the 62 college students who answered yes. Clicking on a point in either distribution will highlight the corresponding random sample from the other population. By hovering over the points, students can record the specific proportion for each sample and calculate the difference in the two sample proportions.

Page 2.3: Simulated Sampling Distribution of Differences in Proportions

Page 2.3 will generate and display a simulated sampling distribution of the differences in the proportions of 'yeses' for 100 random samples from each population on Page 2.1, first in a set of 10 and then in sets of 20. Each point represents the difference obtained from a simulated sample in g1 and a corresponding simulated sample in g2. Note that if any point in the simulated sampling distributions for g1 or g2 on Page 2.1 is highlighted, the difference in proportions for those two samples will be highlighted on Page 2.3.

### Page 3.1: Change the Known Proportion of 'yeses"

The arrows on Page 3.1 will again generate simulated samples for each generation; however, students can now click an arrow to set the value of a known proportion of 'yes' answers in each population, p1 and p2, and change the difference in the proportions accordingly.
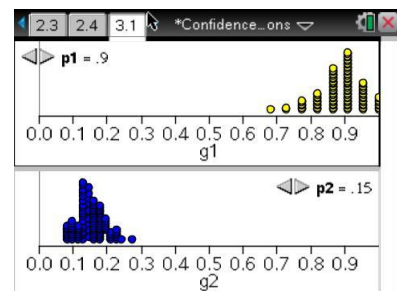
Page 3.2: Simulated Sampling Distribution of Differences in Proportions

Page 3.2 displays the simulated sampling distribution of the differences in the proportions of 'yeses' for the two populations displayed on Page 3.1. The vertical line represents the observed difference in the problem, approximately 0.75. If 0.75 seems likely to be part of the simulated sampling distribution of differences, then the known difference (identified as the actual difference at the top of this page) is a possible value for the true difference in the proportions within populations of senior citizens and college students who would answer yes to the survey.
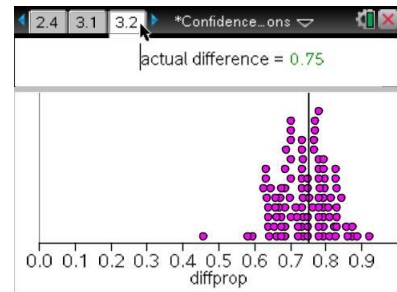
Page 4.2: Possible Population Differences

The goal of this investigation is developed on Page 4.2. Page 4.2 combines choosing p1 and p2 as on Page 3.1 and the simulated sampling distribution of the differences in the proportions as on Page 3.2. Students assign various population proportion values to p1 and p2. If the (fixed) observed difference (0.75) seems to be a likely candidate for the simulated sampling distribution of differences in
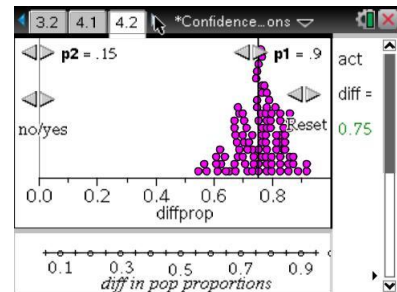
sample proportions for a given p1 and p2, students click the "yes" response, and a point is marked on the lower screen representing that difference in population proportions. Repeating the process for different values of p1 and p2 allows students to collect a set of differences in population proportions for which the proportion likely to respond yes is known and for which the observed difference of 0.75 was a likely sample difference.

Students can determine an interval of "yeses" from the bottom screen on Page 4.2 to estimate a lower and an upper bound of a confidence interval of the population differences based on the observed sample difference 0.75.

Possible Applications

This activity can be used as an informal introduction to the concept of a confidence interval for the difference between two sample proportions and the underlying reasoning from simulated sampling distributions for known population differences to population differences that lead to an observed difference being a likely outcome. Understanding what this interval represents can help students interpret the results of the formula that calculates the confidence intervals for the difference in two proportions.

Related Activities

- Statistics Nspired *Confidence Intervals for Proportions*

- Statistics Nspired *Confidence Levels for Proportions*