**Unit 7: Investigation 3 (4 Days)**

**Inference on Population Proportions**

**Common Core State Standards**

IC.A1 Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

IC.A.2 Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation.

IC.B.4 Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.

IC.B.5 Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

**Overview**

This investigation introduces students to statistical inference on population proportions. Students explore sampling distributions of sample proportions, learn the Central Limit Theorem for sample proportions, use bootstrap distributions to construct interval estimates for population proportions, and use randomization distributions to test claims about population proportions and differences in population proportions. Students construct randomization distributions through hands-on activities and the use of technology.

**Assessment Activities**

**Evidence of Success: What Will Students Be Able to Do?**

* Construct, describe and interpret a distribution of sample proportions
* Use distributions of sample proportions to reason about individual sample proportions
* Apply the Central Limit Theorem for sample proportions
* Calculate the margin of error of a 95% confidence interval for a population proportion
* Construct a 95% confidence interval for a population proportion
* Perform a randomization test to test a claim about a population proportion
* Perform a randomization test to test a claim about the difference in population proportions

**Assessment Strategies: How Will They Show What They Know?**

* **Exit Slip 7.3.1** asks students to estimate the interval that contains the middle 95% of sample proportions and identify sample proportions that are unusual.
* **Exit Slip 7.3.2** asks students to construct and interpret a 95% confidence interval for a population proportion.
* **Exit Slip 7.3.3** asks students to interpret a randomization distribution of sample proportions, identify a *P*-value, determine statistical significance, and draw a conclusion about a population proportion.
* **Exit Slip 7.3.4** asks students to interpret a randomization distribution of differences in sample proportions and draw a conclusion about two population proportions.
* **Journal Prompt 1** Define a sampling distribution of sample proportionsand explain why its standard deviation is called the standard error.
* **Journal Prompt 2** List the steps required to construct a 95% confidence interval for a population proportion.
* **Journal Prompt 3** Explain theprocess for generating a randomization distribution of differences in sample proportions given two random samples.
* **Journal Prompt 4** Describe why small probabilities (small *P*-values) lead us to reject claims made about population parameters.
* **Activity 7.3.1 Exploring Sample Proportions** introduces students to sampling distributions of sample proportions, sampling variability.
* In **Activity 7.3.2 The Role of Sample Size** students use Statkey to explore sampling distributions of sample proportions for random samples of size 20, 50 and 100 from a population with a known population proportion.
* **Activity 7.3.3 Estimating Population Proportions** introduces students to the concepts of point estimates, interval estimates, margin of error, and 95% confidence intervals for population proportions.
* **Activity 7.3.5** **Testing Differences in Two Population Proportions** introduces students to randomization tests on difference in population proportions.

**Launch Notes**

Initiate this investigation by discussing the concepts of statistical inference, population parameters, and sample statistics. State a few numerical values and ask students to decide whether the values are parameters or statistics.

Examples of numerical values include:

* The mean height of the players on the 2015 Boston Celtics is 80.25 inches.
* A random sample of fifty 8th graders in Connecticut had a median height of 61 inches.
* Of the roughly 2.16 million registered voters in Connecticut, 36% are registered as supporting the Democratic Party.
* 54% of adults in an ESPN May 2015 national poll stated that they believed New England Patriot’s quarterback Tom Brady was involved in deflating footballs prior to the 2015 AFC Championship game.

**Teaching Strategies**

1. **Activity 7.3.1 Exploring Sample Proportions** introduces students to sampling distributions of sample proportions and sampling variability. Students generate random samples from a population using random number sequences, calculatesample proportions, create an empirical distribution of sample proportions, and examine properties of the empirical distribution. Students also examine how a distribution of sample proportions changes when the sample size increases. As students complete the activity, they are introduced to the terms statistic, parameter, sampling distribution, sampling variability, and standard error.

*Note on Random Sampling*: Students can generate random samples in two ways.

1. Distribute index cards to each student numbered 1 to 50. Students should shuffle the index cards and then select 10 cards (for problems 1 and 2), or select 20 cards (for problem 10). The numbers that they select will comprise the random sample.
2. Use the graphing calculator to generate a list of random integers. The list of integers will comprise the random sample. Distribute the **Unit 7 Technology Supplement** to provide students instructions on how to use the TI 83/84 to generate sequences of random integers.

Question 15 asks students to compare distributions of sample proportions for two different sample sizes using statistics developed in class. If time permits, show students how to simulate this process in Statkey.

To create simulated distributions of sample proportions:

* Go to <http://lock5stat.com/statkey/index.html>
* Click on Proportions in the Sampling Distributions section
* Click Edit Proportion to set the population proportion to *p* = 0.48
* Set Samples of size *n* to 10
* Click Generate 1000 Samples
* The simulation will show 1000 sample proportions. In the upper right corner, you will see the mean and standard deviation of the simulated sample proportions. Record the mean and standard deviation.
* Set Samples of size *n* to 20
* Click Generate 1000 Samples
* Record the mean and standard deviation.

You can assign **Exit Slip 7.3.1** after students complete **Activity 7.3.1**.

**Group Activity**

Students are encouraged to work in groups to complete **Activity 7.3.1**. Students could partition the responsibilities as follows: one student obtains a list of random numbers and one student records the corresponding values in the sample. All students should get practice calculating sample proportions.

**Differentiated Instruction (For Learners Needing More Assistance)**

Provide students an opportunity to generate and record results from multiple random samples and create multiple sample statistics. This will enable students to more concretely understand the concept of sampling variability.

In **Activity 7.3.2 The Role of Sample Size** students use Statkey to explore sampling distributions of sample proportions for random samples of size 20, 50 and 100 from a population with a known population proportion. Students learn that as the sample size increases, the standard error of sample proportions decreases, making the sample proportions better estimates of the population proportion. This activity is intended as an out-of-class activity.

**Journal Prompt 1** Define a sampling distribution of sample proportionsand explain why its standard deviation is called the standard error.

Students should state that a sampling distribution of sample proportionsis the distribution of all sample proportions from random samples of the same size taken from a population. The standard deviation of sample proportions is the typical or average deviation from the population proportion (mean of of all sample proportions). Sample proportions are estimates of the population proportion, so the deviations from the population proportion are errors. Since the average of these deviations is an average of the errors, the standard deviation is called the standard error.

1. **Activity 7.3.3 Estimating Population Proportions** introduces students to the concepts of point estimates, interval estimates, margin of error, and 95% confidence intervals for population proportions. Students learn the process of constructing a 95% confidence interval solely from sample statistics: calculating a sample proportion, computing an approximate standard error using the sample proportion, computing the margin of error based on the approximate standard error, and then combining the margin of error with the sample proportion to form an interval estimate. The appropriate interpretation of confidence intervals is also discussed.

On page 1 of the activity, students are presented a simulated distribution of sample proportions from a population with a population proportion *p* = 0.70. If time permits, present the creation of a simulated distribution of sample proportions using Statkey.

To create a simulated distribution of sample proportions:

* Go to Statkey (<http://lock5stat.com/statkey/index.html>)
* Click on Proportions in the Sampling Distributions section
* Click Edit Proportion to set the population proportion to *p* = 0.70
* Set Samples of size *n* to 50
* Click Generate 1 sample (this will create a random sample, calculate the sample proportion, and plot the sample proportion), repeat nine times, discuss the results
* Click Generate 10 samples and Generate 100 samples until the distribution has 500 sample proportions.

Make sure students understand what the distribution represents, how it was constructed via random sampling, and the connection between the mean of the distribution of sample proportions and the assumed population proportion.

You may assign **Exit Slip 7.3.2** after students complete **Activity 7.3.3**.

**Journal Prompt 2** List the steps required to construct a 95% confidence interval for a population proportion.

Students should state: 1) Find a sample proportion, (2) Use the sample proportion to calculate the approximate standard error, (3) Calculate the margin of error, (4) Add and subtract the margin of error from the sample proportion to construct an interval estimate.

1. **Activity 7.3.4** **Hypothesis Tests on Population Proportions** introduces students to concepts related to hypothesis testing. The activity begins with a hypothesis-testing scenario involving a teacher who supposedly is using a random process to select students to come to the board. Using the multiplication rule, students quantify the likelihood of a particular sequence of events occurring, under the assumption that a random process is being used. If the probability of the sequence of events is very low the probability challenges the assumption that the events resulted from a random process. This opening problem situation highlights key components of a hypothesis test: making an assumption about a population parameter (hypothesis), collecting sample data, calculating a conditional probability, and using the probability to make a decision about the population parameter.

The activity shifts to the construction of randomization tests using randomization distributions. Students use random number sequences to randomly sample from a population to obtain random samples, calculate sample proportions, and construct a distribution of sample proportions. Students are introduced to the concepts of *P*-value and statistical significance and learn how to use *P*-values to make decisions about population proportions.

*Notes about Randomization Tests* – A randomization test consists of four basic steps:

1. Make an assumption (hypothesis) about the value of a population parameter
2. Construct a randomization distribution of sample statistics under the assumption that the population parameter is equal to the hypothesized value
3. Find the probability of observing a sample statistic as extreme as the one found
4. Make a decision about the population parameter

*Note on Random Sampling*: Problems 6 and 7 require students to construct random samples. Students can generate random samples in two ways.

1. Distribute 50 index cards to each student. Label 24 index cards with S, label 26 with F. This models a population of size *N* = 50 with *p* = 0.48. Students should shuffle the index cards and then select 20 cards, determine the number of “successes”, and calculate sample proportions.
2. Use the graphing calculator to generate a list of random integers. The list of integers will comprise the random sample. Distribute the **Unit 7 Technology Supplement** to provide students instructions on how to use the TI 83/84 to generate sequences of random integers. Additional instructions are provided in the activity.

You may assign **Exit Slip 7.3.3** after students complete **Activity 7.3.4**.

1. **Activity 7.3.5** **Testing Differences in Two Population Proportions** introduces students to randomization tests on differences in population proportions. Students use random number sequences to randomly sample from two populations to obtain random samples, calculate the differences in sample proportions, and construct a distribution of differences in sample proportions. Students learn to interpret a randomization distribution of differences, use randomization distributions to find *P*-values, and use *P*-values to make decisions about the difference in population proportions.

*Note on Random Sampling*: Problems 3 – 6 require students to construct random samples. Students can generate random samples in two ways.

1. Distribute 50 index cards to each student. Label 30 index cards with S, label 20 with F. This models a population of size *N* = 50 with a proportion of successes, *p* = 0.60. Students should shuffle the index cards and then separate the cards into two groups of 25, one group representing males, and one group representing females. Then, students should determine the number and proportion of successes in each sample, and then determine the difference in sample proportions.
2. Use the graphing calculator to generate a list of random integers. The list of integers will comprise the random sample. Distribute the **Unit 7 Technology Supplement** to provide students instructions on how to use the TI 83/84 to generate sequences of random integers. Additional instructions are provided in the activity.

You can assign **Exit Slip 7.3.4** after students complete **Activity 7.3.5**.

 **Group Activity**

Students are encouraged to work in pairs to complete **Activity 7.3.4** and **Activity 7.3.5**. Students could partition the responsibilities as follows: one student obtains a list of random numbers and one student records the corresponding values in the sample. All students should get practice calculating sample proportions.

**Differentiated Instruction (Enrichment)**

Provide students hypothesis-testing situations that involve testing the difference in two population proportions. Ask students to use Statkey to develop randomization distributions and find *P*-values.

**Differentiated Instruction (For Learners Needing More Assistance)**

Implement activities with scaffolds that provide students explanations for each part of the hypothesis testing process.

**Activity 7.3.6** **Inference Problems on Population Proportions** provides students additional opportunities to practice constructing confidence intervals and performing hypothesis tests for population proportions and differences in population proportions.

This is intended as an out-of-class activity. Some problems require students to have access to a computer as students are prompted to use Statkey to construct and explore randomization distributions.

**Journal Prompt 3** Explain theprocess for generating a randomization distribution of differences in sample proportions given two random samples.

Students should state that given two samples we must first form a hypothetical population by combining the two samples. This yields a population with a certain population proportion. We then randomly sample from the population to create randomization samples, one for each group, calculate sample proportions for each group, and find the difference in sample proportions between the two groups. We repeat this process many times to develop a distribution of differences in sample proportions.

**Journal Prompt 4** Describe why small probabilities (small *P*-values) lead us to reject claims made about population parameters.

Students should state that small *P*-values are small probabilities that provide evidence that the assumption about the value of a population parameter is false. When an observed sample proportion, or difference in observed sample proportions, is very unlikely to occur, we surmise that the sample results (sample proportions) are typical and representative and conclude that the assumption about the unknown population parameter (population proportions) is false.

**Group Activity**

Have students work in pairs to create a brief survey (3 “Yes or No” questions), administer the survey to 50 randomly selected students outside of class, summarize the data using two-way tables, and perform a randomization hypothesis test to determine if the sample data provides evidence that a difference exists between population proportions for male and female students. For example, include questions such as: (a) Do you watch Netflix during the week? (b) Are you interested in playing sports in college? (c) Are you older than your best friend?

**Closure Notes**

On the final day of this investigation, have students discuss their results on the problems they completed in **Activity 7.3.6**. Have students discuss features of the randomization distributions that they developed using Statkey and explain the logic that they used to construct the distributions and use the distributions to make decisions about population proportions. Ask students to discuss situations that would warrant the use of a one-sample and two-sample hypothesis test for population proportions.

**Vocabulary**

Categorical variable

Distribution of difference in sample proportions

Distribution of sample proportions

Distributions of sample statistics

Empirical sampling distributions

Hypothesis

Hypothesis test

*P*-value

Parameter

Random sample

Randomization distribution

Randomization test

Sample proportion

Sampling variability

Sampling distribution

Standard error

Statistic

**Resources and Materials**

**All activities 7.3.1 – 7.3.6 should be completed. Activities 7.3.2 and 7.3.6 are intended to be out-of-class activities.**

Activity 7.3.1 Exploring Sample Proportions

Activity 7.3.2 The Role of Sample Size

Activity 7.3.3 Estimating Population Proportions

Activity 7.3.4 Hypothesis Tests on Population Proportions

Activity 7.3.5 Testing Differences in Two Population Proportions

Activity 7.3.6 Inference Problems on Population Proportions

Census at School (<http://www.amstat.org/censusatschool/index.cfm>)

Statkey (<http://lock5stat.com/statkey/>)

Graphing calculator – Random number generator

Random number tables (alternative to using graphing calculator)

Unit 7 Technology Supplement

Index cards