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

**Inference on Population Means**

**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 means. Students explore sampling distributions of sample means, apply the Central Limit Theorem for sample means, construct interval estimates for population means, and use randomization distributions to test claims about population means and differences in population means. 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 means
* Use distributions of sample means to reason about individual sample means
* Apply the Central Limit Theorem for sample means
* Calculate the margin of error of a 95% confidence interval for a population mean
* Construct a 95% confidence interval for a population mean
* Perform a randomization test to test a claim about a population mean
* Perform a randomization test to test a claim about the difference in population means

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

* **Exit Slip 7.4.1** asks students to estimate the interval that contains the middle 95% of sample means and determine when sample means are unusual.
* **Exit Slip 7.4.2** asks students to construct and interpret a 95% confidence interval for a population mean.
* **Exit Slip 7.4.3** asks students to interpret a randomization distribution of sample means, identify a *P*-value, determine statistical significance, and draw a conclusion about a population mean.
* **Exit Slip 7.4.4** asks students to interpret a randomization distribution of differences in sample means and draw a conclusion about two population means.
* **Journal Prompt 1** Define a sampling distribution of sample meansand explain how the variability of the sampling distribution changes as the sample size increases.
* **Journal Prompt 2** List the steps required to construct a 95% confidence interval for a population mean.
* **Journal Prompt 3** Explain theprocess for generating a randomization distribution of differences in sample means given two random samples.
* **Journal Prompt 4** Describe why small probabilities (small *P*-values) lead us to reject claims made about population parameters.
* **Activity 7.4.1 Exploring Distributions of Sample Means** is a hands-on activity in which students explore distributions of sample means for various sample sizes.
* **Activity 7.4.1A Exploring Distributions of Sample Means** can be used as a replacement for **Activity 7.4.1**. This activity introduces students to sampling distributions of sample means and sampling variability.
* **Activity 7.4.2 The Role of Sample Size** hasstudents use Statkey to explore sampling distributions of sample means for random samples of size 20, 50 and 100 from a population with a known population mean.
* **Activity 7.4.3 Estimating Population Means** introduces students to the concepts of point estimates, interval estimates, margin of error, and 95% confidence intervals for population means.
* **Activity 7.4.4** **Testing Claims About Population Means** introduces students to concepts related to hypothesis testing for population means.
* **Activity 7.4.5** **Testing Differences in Two Population Means** introduces students to randomization tests on differences in population means.
* **Activity 7.4.6** **Inference Problems on Population Means** provides students opportunities to practice constructing confidence intervals and performing hypothesis tests for population means and differences in population means.

**Launch Notes**

This investigation has a structure similar to that of Investigation 3 and can be completed in place of or after Investigation 3. If this investigation is completed in place of Investigation 3, begin the investigation by introducing students to the difference between sample statistics and population parameters and the definition of statistical inference. 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 a May 2015 ESPN 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.

If this investigation is completed after Investigation 3, begin the investigation by distributing **Activity 7.4.1**. Activity 7.4.1 provides students opportunities to reason about sample means to understand distributions of sample means and the manner in which sample means vary and are distributed around the population mean.

**Teaching Strategies**

1. **Activity 7.4.1 Exploring Distributions of Sample Means** is a hands-on activity in which students explore distributions of sample means for various sample sizes. Students are provided a random sample of pennies, determine the mean age of the pennies in the sample, and then construct distributions of sample means for samples of size 5, 10, and 25. As students complete the activity, they are introduced to the terms statistic, parameter, sampling distribution, sampling variability, and standard error.

Questions 11 – 13 ask students to compare distributions of sample means for three different sample sizes using statistics developed in class. If time permits, show students how to simulate this process in Statkey using a hypothetical penny population (or the actual penny population used in class). The Excel file (Penny.xls) contains the ages of 325 pennies. The distribution of ages (shown below) is a skewed-right distribution.



To create simulated distributions of sample means:

* Go to Statkey (<http://lock5stat.com/statkey/index.html>)
* Click on Mean in the Sampling Distributions section
* Click Edit Data to enter in the Penny data. Copy in the column of the ages.
* Set Samples of size *n* to 5
* Click Generate 1000 Samples
* The simulation will show 1000 sample means. In the upper right corner, you will see the mean and standard deviation of the simulated sample means. Record the mean and standard deviation.
* Set Samples of size *n* to 10
* Click Generate 1000 Samples
* Record the mean and standard deviation.
* Set Samples of size *n* to 25
* Click Generate 1000 Samples
* Record the mean and standard deviation.

**Activity 7.4.1A Exploring Distributions of Sample Means** can be used as a replacement for **Activity 7.4.1**. This activity introduces students to sampling distributions of sample means and sampling variability. Students generate random samples from a population using random number sequences, calculatesample means, create an empirical distribution of sample means, and examine properties of the empirical distribution. Students explore how the distribution of sample means changes as the sample size increases from *n* = 5 to *n* = 10. As students complete the activity, they are introduced to the terms statistic, parameter, sampling distribution, sampling variability, and standard error.

Questions 10 – 16 ask students to compare distributions of sample means for two different sample sizes using statistics developed in class. If time permits, show students how to simulate this process in Statkey using the distribution of 100 SAT Critical Reading scores. The Excel file (SATCriticalReading.xls) contains the distribution of 100 scores. The distribution of ages (shown below) is a bell-shaped distribution.



To create simulated distributions of sample means:

* Go to Statkey (<http://lock5stat.com/statkey/index.html>)
* Click on Mean in the Sampling Distributions section
* Click Edit Data to enter in the SAT score data. Copy in the column of the scores.
* Set Samples of size *n* to 5
* Click Generate 1000 Samples
* The simulation will show 1000 sample means. In the upper right corner, you will see the mean and standard deviation of the simulated sample means. Record the mean and standard deviation.
* Set Samples of size *n* to 10
* Click Generate 1000 Samples
* Record the mean and standard deviation.

You may assign **Exit Slip 7.4.1** after students complete **Activity 7.4.1 or Activity 7.4.1A.**

**Group Activity**

Students are encouraged to work in groups to complete **Activity 7.4.1**. 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 means.

**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 7.4.2 The Role of Sample Size** students use Statkey to explore sampling distributions of sample means for random samples of size 20, 50 and 100 from a population with a known population mean. This requires that students copy data from an Excel file (NBAPlayers2011) into Statkey. Students learn that as the sample size increases, the standard error of sample means decreases, making the sample means better estimates of the population mean. This activity is intended as an out-of-class activity.

**Journal Prompt 1** Define a sampling distribution of sample meansand explain how the variability of the sampling distribution changes as the sample size increases.

Students should state that a sampling distribution of sample meansis the distribution of all sample means from random samples of the same size taken from a population. As the sample size increases, the sample means have values that are more similar and closer to the population mean, which results in the standard error of the distribution becoming smaller.

1. **Activity 7.4.3 Estimating Population Means** introduces students to the concepts of point estimates, interval estimates, margin of error, and 95% confidence intervals for population means. Students learn the process of constructing a 95% confidence interval solely from sample statistics: calculating a sample mean, computing an approximate standard error using the sample mean, computing the margin of error based on the approximate standard error, and then combining the margin of error with the sample mean to form an interval estimate. The appropriate interpretation of confidence intervals is also discussed.

You can assign **Exit Slip 7.4.2** after students complete **Activity 7.4.3.**

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

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

1. **Activity 7.4.4** **Testing Claims About Population Means** introduces students to concepts related to hypothesis testing for population means. Students learn the 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 focuses on randomization tests using randomization distributions. Students use random number sequences to randomly sample from a populations to obtain random samples, calculate sample means, and construct a distribution of sample means. Students are introduced to the concepts of *P*-value and statistical significance and learn how to use *P*-values to make decisions about population means.

*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

In **Activity 7.4.4**, randomization tests for a population mean begin with a claim about a population mean and a random sample used to assess the validity of the claim. The random sample is modified so that the sample mean equals the assumed population mean. This modified sample is then treated as the population, and the randomization distribution is obtained by repeatedly sampling from this population.

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

1. Distribute index cards to each student numbered 1 to 20. Students should shuffle the index cards and select a card. Since they are sampling with replacement, students must place the card back in the deck, shuffle, and select a card again. This process must be repeated 20 times (for problems 4 and 5).
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.

You can assign **Exit Slip 7.4.3** after students complete **Activity 7.4.4.**

1. **Activity 7.4.5** **Testing Differences in Two Population Means** introduces students to randomization tests on differences in population means. Students use random number sequences to randomly sample from two populations to obtain random samples, calculate the differences in sample means, and construct a distribution of difference in sample means. 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 means.

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

1. Distribute 20 index cards to each student. Label the 20 cards with the values in the population displayed on page 2. Students should shuffle the index cards and then separate the cards into two groups of 10, one group representing the control group, and one group representing the treatment group. Then, students should determine the mean of each group, and then determine the difference in sample means.
2. Use the graphing calculator to generate a list of 10 random integers. The list of integers can be matched to values in the population and the population values will comprise the control group. The treatment group will then consist of the remaining 10 population values. 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 7.4.4** after students complete **Activity 7.4.5.**

**Group Activity**

Students are encouraged to work in pairs to complete **Activity 7.4.4** and **Activity 7.4.5**. 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 means.

**Differentiated Instruction (Enrichment)**

Introduce students to the concept of two-tailed hypothesis tests and provide them a few two-tailed hypothesis-testing problems to solve.

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

Students should state that given two random samples we must first form a hypothetical population by combining the two samples. We then randomly reallocate data values to two samples and calculate differences in the sample means. We repeat this process many times to develop a distribution of differences in sample means.

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

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

**Activity 7.4.6** **Inference Problems on Population Means** provides students opportunities to practice constructing confidence intervals and performing hypothesis tests for population means and differences in population means. Students will use Statkey to construct randomization distributions. This is intended as an out-of-class activity and some problems require students to have access to a computer.

**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 mean, or difference in observed sample means, is very unlikely to occur, we surmise that the sample results (sample means) are typical and representative and conclude that the assumption about the unknown population parameter (population means) is false.

**Closure Notes**

On the final day of this investigation, have students discuss their results on the problems they completed in **Activity 7.4.6**. 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 means. Ask students to discuss situations that would warrant the use of a one-sample and two-sample hypothesis test for population means.

**Vocabulary**

Categorical variable

Central Limit Theorem for sample means

Distribution of difference in sample means

Distribution of sample means

Distributions of sample statistics

Empirical sampling distributions

Hypothesis

Hypothesis test

*P*-value

Parameter

Random sample

Randomization distribution

Randomization test

Sample mean

Sample median

Sampling variability

Sampling distribution

Standard error

Statistic

**Resources and Materials**

**Activity 7.4.1A Exploring Distributions of Sample Means** can be used as a replacement for **Activity 7.4.1**. **Activity 7.4.2 is intended as an out-of-class activity**. **Activity 7.4.6 is intended as an out-of-class activity but does need some access to technology**.

Activity 7.4.1 Exploring Distributions of Sample Means

Activity 7.4.1A Exploring Distributions of Sample Means

Activity 7.4.2 The Role of Sample Size

Activity 7.4.3 Estimating Population Means

Activity 7.4.4 Testing Claims About Population Means

Activity 7.4.5 Testing Differences in Two Population Means

Activity 7.4.6 Inference Problems on Population Means

* Census at School (<http://www.amstat.org/censusatschool/index.cfm>)
* Statkey (<http://lock5stat.com/statkey/>)
* Graphing calculator – Random number generator
* Random number tables
* Penny Excel File
* SATCriticalReading Excel File
* NBAPlayers2011 Excel File