**Unit 7: Investigation 6 (2 Days)**

**Inference on Categorical Data (Optional)**

**Common Core State Standards**

IC.A.1 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.

**Overview**

This optional investigation provides students opportunities to perform inference on categorical variables. Students are introduced to the chi-square statistic and the chi-square goodness-of-fit test for a single categorical variable. The investigation culminates with an activity in which students apply chi-square goodness-of-fit tests to investigate social-justice topics.

**Assessment Activities**

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

* Calculate a chi-square statistic
* Perform and interpret a chi-square goodness-of-fit test

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

* **Exit Slip 7.6.1** asks students to calculate a chi-square statistic given an expected (hypothesized) distribution and observed distribution.
* **Exit Slip 7.6.2** asks students to use a randomization distribution to perform a chi-square goodness-of-fit test.
* **Journal Prompt 1** Explain why a large chi-square statistic provides evidence that an observed distribution differs from a hypothesized distribution.
* **Journal Prompt 2** Explain how to construct a randomization distribution of chi-square statistics based on a hypothesized distribution.
* **Activity 7.6.1** **Evaluating the Fit of Categorical Variable** introduces students to inference situations that can be addressed by the chi-square goodness-of-fit test.
* **Activity 7.6.2 Goodness-of-Fit Test** provides students opportunities to compute chi-square statistics and perform goodness-of-fit tests using randomization distributions.
* **Activity 7.6.3 Applying Chi-Square Tests to Social-Justice Topics** introduces students to applications of chi-square goodness-of-fit tests to social justice topics.

**Launch Notes**

Begin this investigation by presenting students the opening problem in **Activity 7.6.1.** This problem presents a hypothesized probability distribution that describes high school seniors’ views on the cost of college. The hypothesis details how high school students’ views fit into three categories. Ask students to think of a way to test this hypothesis using a random sample from the population.

Remind students that the variable of interest, high school seniors’ level of concern with the cost of college, is a *categorical* variable. Responses are grouped into three categories: not concerned, somewhat concerned, and highly concerned.

**Teaching Strategies**

1. **Activity 7.6.1** **Evaluating the Fit of Categorical Variable** introduces students to inference situations that can be addressed by the chi-square goodness-of-fit test. The activity provides students an opportunity to develop a measure that evaluates the extent to which an observed frequency distribution fits an expected frequency distribution. Students are introduced to the chi-square statistic for a single categorical variable and learn to recognize why large chi-square values indicate that the observed distribution differs significantly from the expected distribution. This activity serves as a foundation for understanding the chi-square goodness-of-fit test and the chi-square distribution. You can assign **Exit Slip 7.6.1** after students complete **Activity 7.6.1.**

**Journal Prompt 1** Explain why a large chi-square statistic provides evidence that an observed distribution differs from a hypothesized distribution.

Students should state that a large chi-square statistic emerges when expected counts and observed counts are very different. The chi-square statistic is a sum of terms, and each term in the sum is based on the square of the difference between a category’s observed count and expected count.

1. **Activity 7.6.2 Goodness-of-Fit Test** provides students opportunities to compute chi-square statistics and perform goodness-of-fit tests using randomization distributions. Students learn to perform a randomization chi-square hypothesis test. Students learn how to create a randomization distribution of chi-square statistics based on randomized samples, and use the distribution to find *P*-values and state a conclusion about the population’s distribution. Students also explore a large simulated randomization distribution of chi-square statistics to find a *P*-value. You can assign **Exit Slip 7.6.2** after students complete **Activity 7.6.2.**

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

Students may need assistance understanding the meaning of the randomization distribution. To better understand randomization distributions, have students enter in the observed and expected frequencies for a categorical variable in Statkey, verify that the calculation of the chi-square statistic is correct, and generate a few re-randomized samples and corresponding chi-square statistics under the assumption of the null hypothesis.

**Differentiated Instruction (For Enrichment)**

Introduce students to theoretical chi-square distributions. Ask students to explore how theoretical chi-square distributions change as the degrees of freedom increase and use theoretical chi-square distributions to find *P*-values.

**Journal Prompt 2** Explain how to construct a randomization distribution of chi-square statistics based on a hypothesized distribution.

Students should state that a distribution of randomization statistics is formed by first using the hypothesized distribution to form a population with a specific distribution. Then, randomization samples of the same size are obtained from the population and sample chi-square statistics are computed for each randomized sample. The sample chi-square statistics are collected to form a distribution of sample chi-square statistics.

**Activity 7.6.3 Applying Chi-Square Tests to Social-Justice Topics** introduces students to applications of chi-square goodness-of-fit tests to social justice topics. This optional activity can be completed as homework. Students construct randomization distributions of chi-square statistics using Statkey, calculate the chi-square statistic for the observed sample, determine the corresponding *P*-value, and state a conclusion about the population’s distribution. This activity requires students to have access to a computer, tablet, or smartphone in order to construct a randomization distribution using Statkey. If appropriate, you could allow students to use technology or theoretical chi-square distributions to find *P*-values. After completing one or two hypothesis tests in the activity, encourage students to develop their own research question on a social-justice topic and search for data on the Internet that can be used to investigate their topic.

**Group Activity**

Encourage students to work in pairs to complete **Activity 7.6.3**. Make sure each student understands the research question being addressed, hypotheses, calculations, and interpretations of the hypothesis test.

**Closure Notes**

On the final day of this investigation, ask students to develop sample surveys that could be used to obtain observed counts for categorical variables. For example, “Are high school students very interested, somewhat interested, or not interested, in watching the U.S. Open tennis tournament in August 2016?” This question focuses on a single categorical variable (interest in watching the U.S. Open) that assumes three values (very interested, somewhat interested, not interested). Inform students that they could collect data on the categorical variable using a sample. Once observed counts from the sample are determined, students could use the observed counts to test a set of specific hypothesis about the distribution. Encourage students to explore designing and implementing chi-square goodness-of-fit tests. This can be done by: (1) establishing a research question, (2) creating a sample survey, (3) collecting data, and (4) using the sample data to test a hypothesis on the categorical variable’s distribution.

**Vocabulary**

Categorical variable

Chi-square distribution

Chi-square statistic

Chi-square goodness-of-fit test

Continuous probability distribution

Expected frequencies

Observed frequencies

Randomization distribution

Sample statistic

Skewed-right distribution

**Resources and Materials**

**This investigation is optional. If time is limited then Activity 7.6.3 can be omitted.**

Activity 7.6.1 Evaluating the Fit of Categorical Variable

Activity 7.6.2 Goodness-of-Fit Test

Activity 7.6.3 Applying Chi-Square Tests to Social-Justice Topics

Statistical technology (TI 83, Excel, Minitab) to calculate chi-square statistics and find corresponding *P*-values

Statkey