**Unit 7: Inferential Statistics**

**UNIT OVERVIEW**

20 days (for non-STEM intending students)

Inferential statistics is the process of using sample statistics to draw conclusions about population parameters. This unit builds on students’ prior knowledge of descriptive statistics and univariate and bivariate data distributions. The activities are designed to foster statistical reasoning and develop students’ conceptual understanding of fundamental concepts central to statistical inference: chance variation, random sampling, sampling variability and probability.

The activities in this unit are designed to be in accord with the framework for statistical problem solving outlined in the American Statistical Association’s (ASA) *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report[[1]](#footnote-1)*. The framework defines four components of any statistical investigation:

1. Formulate Questions
2. Collect Data
3. Analyze Data
4. Interpret Results

**Tasks and activities in this unit allow students opportunities to derive questions that can be answered through the collection of data, to collect data, and to analyze data via a variety of statistical tools.** Students learn to interpret and analyze data distributions, empirical distributions developed through randomizations, and theoretical probability distributions. Students conduct the two key forms of statistical inference: estimation and hypothesis testing, and explore inference on single population parameters and on comparisons between two population parameters.

Sampling distributions – distributions of sample statistics – are introduced as the fundamental tool in the inference process. Students perform inference on a variety of population parameters (correlation coefficient, regression line slope, proportion, mean, difference in proportions, difference in means) via randomization distributions of sample statistics and are introduced to the logic that underlies the inference process.

The activities in this unit reinforce student understanding and fluency with graphical representations (scatterplots, dot plots, histograms, normal curves) and provide a variety of contextual situations for students to explore. Simulations, in-class data collection activities, authentic data sets, and statistical technology are employed to provide students engaging and meaningful opportunities to perform statistical inference.

Unit 7 is intended to be flexible so teachers can traverse the unit in multiple ways. Teachers can complete a subset of the investigations, or a subset of activities within an investigation, while still providing students an opportunity to thoroughly explore statistical inference. If time is an issue, teachers may choose to complete either Investigation 3 (Proportions) or Investigation 4 (means), but not both, as both investigations introduce statistical inference using one sample and two samples. In addition, Investigation 6 is optional and can be omitted if time is not available.

**Investigation 1 - Inference on Correlation and Regression** introduces statistical inference on population correlation coefficients and population regression line slopes. Students compute and interpret sample correlation coefficients and least-square regression lines. They perform randomization tests via randomization distributions by permuting data values in the original samples (under the assumption that no correlation exists between the two variables) and using the randomization distributions to evaluate the likelihood of observing sample statistics as extreme as the ones observed. This activity involves the use of hands-on activities and simulated randomization distributions.

**Investigation 2 – Collecting and Examining Data** introduces students to the three main types of statistical studies: observational studies, experiments, and sample surveys. Students learn the types of research questions that are appropriate for observational studies and experiments and the types of conclusions that are appropriate. Students explore methods of analyzing data and they are introduced to measures of variability: mean absolute deviation and standard deviation, shapes of data distributions, the Empirical Rule, and z-scores.

**Investigation 3** **– Inference on Population Proportions** introduces students to statistical inference on population proportions. Student explore sampling distributions of sample proportions, apply the Central Limit Theorem for sample proportions, construct interval estimates for population proportions, and use randomization distributions to test claims about population proportions and claims about differences in two population proportions. This investigation addresses the difference between population parameters and sample statistics, properties of sampling distributions, sampling variability, *P*-values and statistical significance, and making decisions about a parameter. Activities support in-class random sampling and the use of statistical technology to construct simulated distributions of sample proportions.

**Investigation 4** **– Inference on Population Means** introduces students to statistical inference on population means. Student 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 claims about differences in two population means. This investigation addresses the difference between population parameters and sample statistics, properties of sampling distributions, sampling variability, *P*-values and statistical significance, and making decisions about a parameter. Activities support in-class random sampling and the use of statistical technology to construct simulated distributions of sample means.

**Investigation 5** **– Modeling Data Distributions** introduces students to probability experiments, discrete and continuous random variables, and discrete and continuous probability distributions. Students construct discrete probability distributions based on empirical data and theoretical probabilities, represent distributions graphically using histograms, calculate and interpret means and standard deviation of probability distributions, and examine binomial experiments and binomial distributions. Students also examine continuous probability distributions and normal curves, learn to associate normal curves with density curves, and recognize how normal curves are used to model data distributions. Given population parameters (mean and standard deviation), students learn to use technology to find areas below normal curves. The investigation concludes with an optional activity that introduces students to modeling sampling distributions of sample means with normal curves and using normal curves to perform statistical inference.

**Investigation 6 – Inference on Categorical Data** is an optional investigation that provides students an opportunity 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.

**Essential Questions**

* What is statistical inference?
* What is sampling variability?
* What is a sampling distribution?
* What is a randomization distribution of sample statistics?
* What is a confidence interval?
* How can we create randomization distribution of sample statistics?
* How can sample statistics be used to make inferences about population parameters?

**Enduring Understandings**

* Sample statistics are estimates of population parameters.
* Randomization distributions of sample statistics are important tools that enable us to evaluate sample statistics and make conclusions about population parameters.
* Statistical inference is the process of using sample statistics to draw a conclusion about an unknown population parameter

**Unit Understandings**

* Sample statistics vary but a population parameter is unique.
* Distributions of sample statistics can be modeled by randomization distributions and theoretical distributions.
* There are two basic types of statistical inference: confidence intervals and hypothesis tests.
* To form a confidence interval we must add and subtract a margin of error from a point estimate.
* A *P*-value is a conditional probability – the probability of observing a sample statistic as extreme as the one observed assuming the population parameter is equal to a certain value.
* A hypothesis test requires that we make an assumption about the value of a population parameter, construct a distribution of sample statistics, evaluate the probability of observing a sample statistic as extreme as the one observed, and that we make a decision about a population parameter.

**Unit Contents**

Investigation 1: Inference on Correlation and Regression (2 days)

Investigation 2: Collecting and Examining Data (3 days)

Investigation 3: Inference on Population Proportions (4 days)

Investigation 4: Inference on Population Means (4 days)

Investigation 5: Modeling Data Distributions (4 days)

Investigation 6: (Optional) Inference on Categorical Data (2 days)

Performance task (1 – 2 days)

End of unit review of unit (1 day)

End of unit assessment (1 day)

**Common Core Practice Standards**

*Mathematical Practices #1 and #3* *describe a classroom environment that encourages thinking mathematically and are critical for quality teaching and learning. Practices in bold are to be emphasized in the unit:*

1. Make sense of problems and persevere in solving them.

2. Reason abstractly and quantitatively.

3. **Construct viable arguments and critique the reasoning of others.**

4. Model with mathematics.

5. **Use appropriate tools strategically.**

6. Attend to precision.

7. Look for and make use of structure.

8. Look for and express regularity in repeated reasoning.

**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.

ID.A.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

ID.A.4 Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.

IC.B.3 Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.

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.

IC.B.6 Evaluate reports based on data.

MD.A.1 (+) Define a random variable for a quantity of interest by assigning a numerical value to each event in a sample space; graph the corresponding probability distribution using the same graphical displays as for data distributions.

MD.A.2 (+) Calculate the expected value of a random variable; interpret it as the mean of the probability distribution.

MD.A.3 (+) Develop a probability distribution for a random variable defined for a sample space in which theoretical probabilities can be calculated; find the expected value.

MD.A.4 (+) Develop a probability distribution for a random variable defined for a sample space in which probabilities are assigned empirically; find the expected value.

**Assessment Strategies**

**Performance Task**

The Unit 7 Performance Task is an open-ended assignment that requires students to conduct two forms of statistical inference: estimating a population parameter and testing a claim about a population parameter, based on a random sample of data. Students are encouraged to explore data from the *Census at School* project, but they are not limited to this data set. For each form of inference, students are required to complete the four components of a statistical investigation as outlined in the GAISE framework: (1) formulate statistical questions that can be answered using data, (2) collect a random sample of data, (3) analyze data and calculate statistics in a manner appropriate to address the statistical question, and (4) interpret the statistical results and establish appropriate conclusions. Students must create a report detailing their work and reasoning for each step in the statistical investigation process. Students could make a class presentation.

**Other Evidence (Formative and Summative Assessments)**

* Exit slips
* Class work
* Homework assignments
* Math journals
* End-of-unit test

**Vocabulary**

|  |  |
| --- | --- |
| Approximately normal  Bell-shaped  Boxplot  Categorical variable  Center  Chi-square test  Chi-square distribution  Confidence interval  Continuous random variable  Correlation coefficient  Data  Discrete random variable  Distribution of sample statistics  Distribution of sample means  Distribution of sample proportions  Dotplot  Empirical Rule  Empirical sampling distribution  Expected value  Experiment  Explanatory variable  Goodness-of-fit test  Histogram  Hypothesis  Hypothesis test  Interval estimate  Least-squares regression line  Margin of error  Mean  Mean absolute deviation  Median  Mode | Normal curve  Normal distribution  Observational study  Parameter  Probability distribution  *P*-value  Point estimate  Quantitative variable  Random sample  Randomization distribution  Randomization sample  Randomization test  Range rule of thumb  Regression line  Response variable  Sample mean  Sample survey  Sample proportion  Sampling distribution  Sampling variability  Shape  Skewed distribution  Spread  Standard deviation  Standard error  Stratified random sampling  Statistic  Statistical significance  Statistical inference  Treatment variable  Uniform distribution  *Z*-score |

1. Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K–12 Curriculum Framework, <http://www.amstat.org/education/gaise/GAISEPreK12_Intro.pdf> [↑](#footnote-ref-1)