**Unit 7: Investigation 5 (4 Days)**

**Modeling Data Distributions**

**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.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.

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.

**Overview**

This investigation 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 activity that introduces students to modeling sampling distributions of sample means with normal curves and using normal curves to perform statistical inference.

**Assessment Activities**

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

* Construct a discrete probability distribution from empirical data
* Calculate and interpret the mean and standard deviation of a discrete random variable
* Construct a histogram for a discrete probability distribution
* Identify the parameters of a binomial experiment
* Construct a binomial probability distribution
* Calculate and interpret the mean, variance, and standard deviation of a binomial random variable
* Fit a normal curve to a data distribution
* Use a normal curve to determine probabilities associated with continuous random variables
* Use normal models to find probabilities associated with sampling distributions of sample means

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

* **Exit Slip 7.5.1** asks students to construct a probability distribution from empirical data, use the distribution to find a probability, and calculate the mean of the distribution.
* **Exit Slip 7.5.2** asks students to find binomial probabilities given the parameters of a binomial experiment, and find the mean, variance, and standard deviation of a binomial random variable.
* **Exit Slip 7.5.3** asks students to fit a density curve to a data distribution and use the curve to estimate percentages and probabilities.
* **Exit Slip 7.5.4** asks students to find areas below normal curves.
* **Exit Slip 7.5.5** asks students to calculate *z*-scores and find areas below the standard normal curve.
* **Exit Slip 7.5.6** asks students to find percentages and probabilities associated with normal distributions.
* **Exit Slip 7.5.7** asks students to perform a hypothesis test for a population mean using a normal model of a sampling distribution of sample means.
* **Journal Prompt 1** Discuss the meaning and significance of the mean and standard deviation of a discrete probability distribution.
* **Journal Prompt 2** List the characteristics of a binomial experiment.
* **Journal Prompt 3** Describe features of a normal distribution.
* **Journal Prompt 4** Explain the meaning and significance of the standard normal distribution.
* **Activity 7.5.1 Discrete Probability Distributions** introduces students to discrete random variables and discrete probability distributions.
* **Activity 7.5.2 Binomial Distributions** introduces binomial experiments and the binomial probability formula.
* **Activity 7.5.3 Density Curves and Normal Curves** introduces students to density curves, continuous random variables, and continuous probability distributions.
* **Activity 7.5.4** **Finding Areas** **Under Normal Curves** introduces students to techniques to find areas below normal curves using technology.
* **Activity 7.5.4A** **Finding Areas** **Under Normal Curves** is an alternative activity to **Activity 7.5.4** that addresses the same learning goals.
* **Activity 7.5.5 Inference with Normal Curves** shows students how normal curves can be used to perform statistical inference on population means.

**Launch Notes**

This investigation returns students’ focus to the exploration of data distributions in the form of frequency histograms. **Activity 7.5.1** presents students a frequency histogram of the number of days a small population of students are absent from school. The activity begins with questions that ask students to use the distribution to find probabilities associated with specific values. Students then learn how to convert a frequency distribution to a discrete probability distribution and explore properties of the distribution.

**Teaching Strategies**

1. **Activity 7.5.1 Discrete Probability Distributions** introduces students to discrete random variables and discrete probability distributions. Students learn that discrete random variables are random variables that can assume a finite number of values that can be listed. Students construct probability distributions from empirical data, graph distributions using histograms, calculate and interpret the mean and standard deviation of discrete random variables, and recognize when individual random variable values in a distribution are unusual. You can assign **Exit Slip 7.5.1** after students complete **Activity 7.5.1.**

**Journal Prompt 1** Discuss the meaning and significance of the mean and standard deviation of a discrete probability distribution.

Students should state that the mean of a discrete random variable is the long-term average of values (or outcomes) of the random variable if many values are collected. The standard deviation describes how much, on average, values of the random variable differ from the mean.

1. **Activity 7.5.2 Binomial Distributions** introduces binomial experiments and the binomial probability formula. Students graph binomial distributions using histograms and calculate the mean, variance, and standard deviation of binomial random variables. Students use the properties of binomial distributions to characterize random variable values as unusual. The initial problems allow students an opportunity to extend their understanding of probability and the multiplication rule to find the probability of observing a specific value of the random variable in a binomial experiment. Ensure that students understand the binomial probability formula by discussing the meaning of each term. To ensure they understand the binomial coefficient $(n C k)$, ask students to list the number of combinations of size *k* = 3 from a set of *n* = 5 objects. Feel free to do a few more combinations and encourage students to use a tree diagram to organize the possible outcomes. You can assign **Exit Slip 7.5.2** after students complete **Activity 7.5.2.**

**Journal Prompt 2** List the characteristics of a binomial experiment.

Students should state that a binomial experiment is an experiment that has a fixed number of trials, where the experiment has two possible outcomes: success and failure, where the probability of success is the same for each trial, and where the random variable is a count of the number of successes in the fixed number of trials.

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

To solidify understanding of the binomial probability formula, demonstrate how to create tree diagrams on binomial experiments with a small number of trials (e.g. *n* = 3, *p* = 0.70). Help students connect the binomial coefficient to the number of paths that lead to the corresponding random variable outcome.

**Differentiated Instruction (Enrichment)**

Provide students opportunities to model real-world phenomena using binomial distributions, graph binomial distributions; and find the mean, variance, and standard deviation of binomial distributions.

**Differentiated Instruction (Enrichment)**

Provide students opportunities to explore Pascal’s triangle to explore combinatorial relationships and the coefficient in the binomial probability formula.

1. **Activity 7.5.3 Density Curves and Normal Curves** introduces students to density curves, continuous random variables, and continuous probability distributions. Students learn to model bell-shaped data distributions with normal curves. Students sketch regions below normal curves that correspond to intervals of random variable values, learn the association between area and probability, and use normal curves to reason about continuous random variables. You can assign **Exit Slip 7.5.3** after students complete **Activity 7.5.3.**

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

Students who have difficulty understanding the distinction between discrete and continuous random variables (or discrete and continuous probability distributions) can create concept maps or word walls to differentiate between these types of distributions.

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

Provide students opportunities to sketch normal curves with varying parameters. Pose questions to ensure students understand the role of the parameters in defining a normal curve.

**Differentiated Instruction (For Enrichment)**

Ask students to explore the probability distribution function for specific parameters. Have students explain how the normal curve changes as the mean changes and as the standard deviation changes.

1. **Activity 7.5.4** **Finding Areas** **Under Normal Curves** introduces students to techniques to find areas below normal curves using technology. Students learn to find areas to the left, to the right, and between random variable values. This activity emphasizes key relationships inherent in normal curves: (1) area equals probability; and the probability can only be determined for an interval of values (as opposed to discrete probability distributions which allow for the assignment of a probability to a single value), and (2) probability equals percentage: the probability of randomly selecting a data value within an interval is equal to the percentage of data values within the interval. Students are introduced to the standard normal distribution as an alternate means for finding probabilities associated with normal distributions. This activity involves the use of graphing calculator technology and/or computer software. You can assign **Exit Slip 7.5.4** after students complete **Activity 7.5.4.**

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

Instruct students to use the Empirical Rule to approximate the areas of regions below the normal curve and then verify that their estimates are correct by finding exact areas.

**Activity 7.5.4A** **Finding Areas** **Under Normal Curves** is an alternative activity to **Activity 7.5.4** that addresses the same learning goals. Students use *z*-scores and statistical tables to find probabilities associated with normal curves. Students were introduced to z-scores as a measure for assessing whether a data value is unusual in **Activity 7.2.3**. You can assign **Exit Slip 7.5.5** after students complete **Activity 7.5.4A.**

You can assign **Exit Slip 7.5.6** after students complete **Activity 7.5.4** or **Activity 7.5.4A.**

**Journal Prompt 3** Describe features of a normal distribution.

Students should state that normal distributions are probability distributions of continuous random variables. Normal distributions have a symmetric, bell-shape. Values in the distribution adhere to the Empirical Rule. Normal curves are centered on the mean of the distribution and the area below a normal curve and between values of the variable represents the probability of obtaining values of the variable within that interval.

**Journal Prompt 4** Explain the meaning and significance of the standard normal distribution.

(This journal prompt is most appropriate for students who use z-scores to find areas below normal curves or students who complete **Activity 7.5.4A**.) Students should state that the standard normal distribution is a distribution of z-scores, or distribution of standardized values. If we transform any normal random variable to a z-score, we can use the standard normal distribution to find probabilities associated with the normal random variable.

**Activity 7.5.5 Inference with Normal Curves** shows students how normal curves can be used to perform statistical inference on population means. Students are presented formulas for the mean and standard deviation of sampling distributions of sample means, learn how to model sample means with normal curves, and use normal curves to find *P*-values and perform statistical inference. This is an optional activity. You can assign **Exit Slip 7.5.7** after students complete **Activity 7.5.5**.

**Group Activity**

Students are encouraged to work in pairs tor triples to complete **Activity 7.5.5.**

**Differentiated Instruction (For Enrichment)**

Introduce students to statistical techniques that can be used to test for normality and provide opportunities for students to employ these techniques.

**Differentiated Instruction (For Enrichment)**

Have students compare the results of randomization hypothesis tests and normal distribution-based hypothesis tests.

**Closure Notes**

On the final day of this investigation, ask students to summarize the key differences between discrete probability distributions and continuous probability distributions. Encourage students to identify examples of discrete and continuous random variables and ask students to create questions that can be answered through the construction and exploration of probability distributions of discrete and continuous random variables. For example, a questions could be, “What percentage of high school students live in a household with 5 or more televisions?” or “What percentage of cable bills are greater than $200?” Discuss the type of data that must be collected and the statistical procedures that are necessary to answer the questions that the students develop.

**Vocabulary**

Approximately normal distribution

Bell-shaped distribution

Binomial coefficient

Binomial distribution

Binomial experiment

Binomial probability formula

Central limit theorem for sample means

Combination formula

Continuous random variable

Continuous probability distribution

Cumulative distribution function

Discrete random variable

Discrete probability distribution

Mean of a random variable

Normal distribution

Normal curve

Pascal’s triangle

Probability distribution function

Random variable

Sampling distribution of sample means

Symmetric distribution

Standard error of sample means

Standard normal distribution

Standard deviation of a random variable

*Z*-score

**Resources and Materials**

**Activity 7.5.1 – 7.5.3 should be completed and either Activity 7.5.4 or 7.5.4A should be completed. Activity 7.5.4A** **Finding Areas** **Under Normal Curves** **is an alternative activity to** **Activity 7.5.4 and addresses the same learning goals. Activity 7.5.5 Inference with Normal Curves is an optional activity that can be omitted in the interest of time.**

Activity 7.5.1 Discrete Probability Distributions

Activity 7.5.2 Binomial Distributions

Activity 7.5.3 Density Curves and Normal Curves

Activity 7.5.4 Finding Areas Under Normal Curves

Activity 7.5.4A Finding Areas Under Normal Curves

Activity 7.5.5 Inference with Normal Curves

Graphing calculators – binomial cumulative probability function

Statistical technology (TI 83, Excel, Minitab) to find areas below normal curves

Statkey