**Unit 7: Investigation 2 (3 Days)**

**Collecting and Examining Data**

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

IC.B.6 Evaluate reports based on data.

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

**Overview**

This investigation 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 as they are introduced to measures of variability: mean absolute deviation and standard deviation, shapes of data distributions, the Empirical Rule, and z-scores.

**Assessment Activities**

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

* Decide whether a study used an observational study, experiment, or sample survey
* Decide whether a conclusion of a statistical study is appropriate based on the design of the study
* Calculate and interpret the mean absolute deviation and standard deviation of a sample data set
* Apply the Empirical Rule to solve real-world problems
* Calculate and interpret the z-score of a data value

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

* **Exit Slip 7.2.1** asks students to pose statistical questions based on a list of categorical and quantitative variables.
* **Exit Slip 7.2.1** and **Exit Slip 7.2.1A** ask students to identify the type of statistical study that was used and determine whether a conclusion is appropriate based on the type of study.
* **Exit Slip 7.2.2** asks students to calculate and interpret the mean absolute deviation and standard deviation of a small sample data set.
* **Exit Slip 7.2.3** asks students to calculate and interpret z-scores to evaluate and describe data values in bell-shaped distributions.
* **Journal Prompt 1** Explain the difference between an observational study and an experiment.
* **Journal Prompt 2** Design a statistical study to address a specific research question.
* **Journal Prompt 3** Explain what the standard deviation measures.
* **Activity 7.2.1 Types of Statistical Studies** introduces students to sample surveys, observational studies, and experiments.
* **Activity 7.2.2 Measuring Variability in Data Distributions** introduces students to two measures of variability: mean absolute deviation and standard deviation.
* **Activity 7.2.3 Examining Values in Data Distributions** introduces students to the Empirical Rule and the concept of z-score for bell-shaped distributions.

**Launch Notes**

Initiate this investigation by having a class discussion on the three main types of statistical studies: sample surveys, observational studies, and experiments. You could briefly present an example of a sample survey and experiment and contrast the methods and conclusions of each study. A few examples are provided below.

**Sample Survey:** Millennials and Political News

<http://www.journalism.org/2015/06/01/millennials-political-news/>

**Randomized Experiments**: Impact of Caffeine on Alertness and Irritability in Adults

<http://users.auth.gr/~haidich/MSc/MedStatII/7.%20REVMAN/Caffeine_studies_for_data_extraction_.pdf>

**Observational Studies:** Living Near Power Lines and the Risk of Acquiring Leukemia

<https://introductorystats.wordpress.com/2011/03/30/design-of-observational-studies/>

<http://www.nejm.org/doi/full/10.1056/NEJM199707033370101>

This discussion will focus student attention on the key ideas and concepts introduced in **Activity 7.2.1**.

**Teaching Strategies**

1. **Activity 7.2.1 Types of Statistical Studies** introduces students to sample surveys, observational studies, and experiments. Students explore the key properties and the types of conclusions that are appropriate for each method. For sample surveys and observational studies, students learn the concepts of random selection and generalizing from sample to population. For experiments, students learn the concepts of random assignment, control and treatment groups, explanatory and response variables, and cause-and-effect conclusions. You can assign **Exit Slip 7.2.1** and **Exit Slip** **7.2.1A** after students complete **Activity 7.2.1.**

**Group Activity**

Students are encouraged to work in pairs or triples to complete **Activity 7.2.1**. Students should provide detailed answers to Question 15 and should discuss ways to establish a clear data collection plan and study methodology.

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

Provide students additional opportunities to distinguish between observational studies and experiments, and to distinguish between correlation and causation.

**Differentiated Instruction (Enrichment)**

Ask students to research instances of biased samples in newspapers or websites. Students should detail the flaw(s) in the study’s design that resulted in a biased sample.

**Journal Prompt 1** Explain the difference between an observational study and an experiment.

Students should state that an observational study is performed by collecting data on a sample or population without interfering with subjects’ behavior or changing subjects’ behavior. An experiment is different because researchers actively manipulate the behavior of subjects by assigning them treatments.

**Journal Prompt 2** Design a statistical study to address a specific research question.

Students should create research questions that can be answered using sample surveys, observational studies, and experiments. Examples include, “Do high school students have favorable or unfavorable views on affirmative action at universities?”, “Do NBA teams with older players perform better than NBA teams with younger players?”, and “Does reading with music playing in the background decrease memory retention?” Students should describe the features of their study that address the research question. They should detail the type of study, population, sample, and the variables. The purpose of this activity is to get students to understand and appreciate the type of design decisions that must go into designing a statistical study.

1. **Activity 7.2.2 Measuring Variability in Data Distributions** introduces students to two measures of variability: mean absolute deviation and standard deviation. Students learn how to calculate and interpret these measures of variation, learn the notation for population standard deviation and sample standard deviation, and are introduced to the range rule of thumb. You can assign **Exit Slip 7.2.2** after students complete **Activity 7.2.2.**

*Note about Calculation of Sample Standard Deviation* – You may want to briefly explain why the sample standard deviation calculation has a divisor of *n* – 1 rather than *n*. The key is that the sample standard deviation is the square root of the *sample variance* (sum of squared deviations divided by *n* – 1), and the sample variance is designed to satisfy a particular property. The sample variance is a statistic (based on a sample), the population variance is a parameter (based on a population), so the sample variance can be thought of as an *estimate* of the population variance. Since there are many samples of the same size, there are many estimates. Statisticians prefer that estimates be calculated so that the mean of all the estimates (in this case, the mean of all the sample variances) equals the population parameter (in this case, the population variance). This is the definition of an *unbiased* *estimate*. It happens to be that the sample variance is only an unbiased estimate of the population variance if the sample variance is calculated with a divisor of *n* – 1. Dividing the sum of squared deviations by *n* – 1, rather than *n*, yields *greater* sample variance values, and empirically we see that these greater values collectively are better estimates of the population variance. The sample standard deviation is the square root of the sample variance and follows this standard.

The link below provides an illustration of this result:

<http://nebula.deanza.edu/~bloom/math10/m10divideby_nminus1.pdf>

You could also show the following Khan Academy video that discusses these concepts:

<https://www.khanacademy.org/math/probability/descriptive-statistics/variance-std-deviation/v/review-and-intuition-why-we-divide-by-n-1-for-the-unbiased-sample-variance>

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

Focus learner attention on the foundational characteristics of graphical representations of univariate data by asking students to interpret the basic features of a frequency distribution or frequency graph (class width, midpoints, frequencies, cumulative frequencies, relative frequencies, etc.)

**Journal Prompt 3** Explain what the standard deviation measures.

Students should state that the standard deviation measures the typical or average deviation from the mean of all values in a data set.

1. **Activity 7.2.3 Examining Values in Data Distributions** introduces students to the Empirical Rule and the concept of z-score for bell-shaped distributions. Students apply the Empirical Rule in contextual situations to find the intervals corresponding to the middle 68%, 95%, and 99.7% of a bell-shaped distribution. Students calculate and interpret z-scores in contextual situations, use z-scores to compare values in a distribution, and use z-scores to compares values from different distributions. You may assign **Exit Slip 7.2.3** after students complete **Activity 7.2.3**.

**Differentiated Instruction (Enrichment)**

Provide students an opportunity to reason about the distribution of z-scores. Given the parameters (mean and standard deviation) of a bell-shaped distribution, ask students to generate a random sample of data values, transform each value to a z-score, construct a frequency histogram of the z-scores, and discuss how the z-scores are distributed.

**Closure Notes**

On the final day of this investigation, have students share their answers to **Journal Entry 2**. Ask students to present their research questions and statistical studies, and request that they discuss the features of their studies (type of study, population, sample, variables). You may also consider asking students to describe what they feel are the most important concepts presented in this investigation and the most challenging concepts presented in this investigation.

**Vocabulary**

Approximately normal distribution

Bell-shaped distribution

Categorical variable

Cause-and-effect conclusion

Control group

Empirical Rule

Experiment

Explanatory variable

Mean

Mean absolute deviation

Observational study

Quantitative variable

Random assignment

Random selection

Range

Range rule of thumb

Sample survey

Skewed left distribution

Skewed right distribution

Spread

Standard deviation

Statistical question

Treatment group

Uniform distribution

*Z*-score

**Resources and Materials**

**All activity sheets 7.2.1 – 7.2.3 should be completed**

Activity 7.2.1 Types of Statistical Studies

Activity 7.2.2 Measuring Variability in Data Distributions

Activity 7.2.3 Examining Values in Data Distributions

Statistical software (e.g. Excel, Minitab) to develop histogram of class-generated data

Census at School: <http://www.amstat.org/censusatschool/>

Gallup.com and Pew Research – Sources for sample surveys