Georgia
Standards of Excellence
Curriculum Frameworks

Mathematics

GSE Algebra I

Unit 6: Describing Data

Georgia Department of Education
Richard Woods, Georgia's School Superintendent
"Educating Georgia's Future"

Unit 6

Describing Data

Table of Contents

OVERVIEW

In this unit students will:

- Assess how a model fits data
- Choose a summary statistic appropriate to the characteristics of the data distribution, such as the shape of the distribution or the existence of extreme data points
- Use regression techniques to describe approximately linear relationships between quantities.
- Use graphical representations and knowledge of the context to make judgments about the appropriateness of linear models

The teacher should supplement these tasks with more practice and exploration of exponential relationships.

Although the units in this instructional framework emphasize key standards and big ideas at specific times of the year, routine topics such as estimation, mental computation, and basic computation facts should be addressed on an ongoing basis. Ideas related to the eight practice standards should be addressed constantly as well. This unit provides much needed content information and excellent learning activities. However, the intent of the framework is not to provide a comprehensive resource for the implementation of all standards in the unit. A variety of resources should be utilized to supplement this unit. The tasks in this unit framework illustrate the types of learning activities that should be utilized from a variety of sources. To assure that this unit is taught with the appropriate emphasis, depth, and rigor, it is important that the **"Strategies for Teaching and Learning"** and the tasks listed under **"Evidence of Learning"** be reviewed early in the planning process.

STANDARDS ADDRESSED IN THIS UNIT

Mathematical standards are interwoven and should be addressed throughout the year in as many different units and activities as possible in order to emphasize the natural connections that exist among mathematical topics.

KEY STANDARDS

Interpreting Categorical and Quantitative Data

Summarize, represent, and interpret data on a single count or measurement variable.

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots). Choose appropriate graphs to be consistent with numerical data: dot plots, histograms, and box plots.

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers). Students will examine graphical representations to determine if data are symmetric, skewed left, or skewed right and how the shape of the data affects descriptive statistics.

Summarize, represent, and interpret data on two categorical and quantitative variables.

MGSE9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6a Decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function to best fit. Use this function to solve problems in context. Emphasize linear, quadratic, and exponential models.

MGSE9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

Interpret linear models

MGSE9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MGSE9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient "r" of a linear fit. (For instance, by looking at a scatterplot, students should be able to tell if the correlation coefficient is positive or negative and give a reasonable estimate of the "r" value.) After calculating the line of best fit using technology, students should be able to describe how strong the goodness of fit of the regression is, using "r."

MGSE9-12.S.ID.9 Distinguish between correlation and causation.

STANDARDS FOR MATHEMATICAL PRACTICE

Refer to the Comprehensive Course Overview for more detailed information about the Standards for Mathematical Practice.

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.

ENDURING UNDERSTANDINGS

- Data are gathered, displayed, summarized, examined, and interpreted to discover patterns and deviations from patterns.
- Which statistics to compare, which plots to use, and what the results of a comparison might mean, depend on the question to be investigated and the real-life actions to be taken.
- Understand and be able to use the context of the data to explain why its distribution takes on a particular shape (e.g. are there real-life limits to the values of the data that force skewness?)
- When making statistical models, technology is valuable for varying assumptions, exploring consequences and comparing predictions with data.
- Causation implies correlation yet correlation does not imply causation.

ESSENTIAL QUESTIONS

- How do I summarize, represent, and interpret data on a single count or measurement variable?
- When making decisions or comparisons, what factors are important for me to consider in determining which statistics to compare, which graphical representation to use, and how to interpret the data?
- How can I use visual representations and measures of center and spread to compare two data sets?
- How do I summarize, represent, and interpret data on two categorical and quantitative variables?
- How do I interpret relative frequencies in the context of a two-way frequency table?
- Why is technology valuable when making statistical models?
- How do you determine the regression line or line of best fit for a scatter plot of data?
- Why are linear models used to study many important real-world phenomena?
- How do I interpret linear models?
- How do I determine if linear or exponential regression is more appropriate for a scatter plot?
- How can I apply what I have learned about statistics to summarize and analyze real data?

CONCEPTS AND SKILLS TO MAINTAIN

In order for students to be successful, the following skills and concepts need to be maintained

- Know how to compute the mean, median, interquartile range, and mean standard deviation by hand in simple cases and using technology with larger data sets.
- Find the lower extreme (minimum), upper extreme (maximum), and quartiles.
- Create a graphical representation of a data set.
- Present data in a frequency table.
- Plot data on a coordinate grid and graph linear functions.
- Recognize characteristics of linear and exponential functions.
- Write an equation of a line given two points.
- Graph data in a scatter plot and determine a trend.
- Determine the slope of a line from any representation.
- Identify the y-intercept from any representation.
- Be able to use graphing technology.
- Understand the meaning of correlation.

SELECT TERMS AND SYMBOLS

The following terms and symbols are often misunderstood. These concepts are not an inclusive list and should not be taught in isolation. However, due to evidence of frequent difficulty and misunderstanding associated with these concepts, instructors should pay particular attention to them and how their students are able to explain and apply them.

The definitions below are for teacher reference only and are not to be memorized by the students. Students should explore these concepts using models and real life examples. Students should understand the concepts involved and be able to recognize and/or demonstrate them with words, models, pictures, or numbers.

The websites below are interactive and include a math glossary suitable for high school children. **Note – At the high school level, different sources use different definitions. Please preview any website for alignment to the definitions given in the frameworks**
http://www.amathsdictionaryforkids.com/

This web site has activities to help students more fully understand and retain new vocabulary.
http://intermath.coe.uga.edu/dictnary/homepg.asp

Definitions and activities for these and other terms can be found on the Intermath website. Intermath is geared towards middle and high school students.

- **Association.** A connection between data values.

- **Bivariate data.** Pairs of linked numerical observations. Example: a list of heights and weights for each player on a football team.

- **Box Plot.** A method of visually displaying a distribution of data values by using the median, quartiles, and extremes of the data set. A box shows the middle 50% of the data.

- **Box-and-Whisker Plot.** A diagram that shows the five-number summary of a distribution. (Five-number summary includes the minimum, lower quartile (25^{th} percentile), median (50^{th} percentile), upper quartile (75^{th} percentile), and the maximum. In a modified box plot, the presence of outliers can also be illustrated.

- **Categorical Variables.** Categorical variables take on values that are names or labels. The color of a ball (e.g., red, green, blue), gender (male or female), year in school (freshmen, sophomore, junior, senior). These are data that cannot be averaged or represented by a scatter plot as they have no numerical meaning.

- **Center.** Measures of center refer to the summary measures used to describe the most "typical" value in a set of data. The two most common measures of center are median and the mean.

- **Conditional Frequencies.** The relative frequencies in the body of a two-way frequency table.

- **Correlation Coefficient.** A measure of the strength of the linear relationship between two variables that is defined in terms of the (sample) covariance of the variables divided by their (sample) standard deviations.

- **Dot plot.** A method of visually displaying a distribution of data values where each data value is shown as a dot or mark above a number line.

- **First Quartile (Q_1).** The "middle value" in the *lower* half of the rank-ordered data

- **Five-Number Summary.** Minimum, lower quartile, median, upper quartile, maximum.

- **Histogram-** Graphical display that subdivides the data into class intervals and uses a rectangle to show the frequency of observations in those intervals—for example you might do intervals of 0-3, 4-7, 8-11, and 12-15

- **Interquartile Range.** A measure of variation in a set of numerical data. The interquartile range is the distance between the first and third quartiles of the data set. Example: For the data set {1, 3, 6, 7, 10, 12, 14, 15, 22, 120}, the interquartile range is 15 – 6 = 9.

- **Joint Frequencies.** Entries in the body of a two-way frequency table.

- **Line of Best Fit (**trend or regression line). A straight line that best represents the data on a scatter plot. This line may pass through some of the points, none of the points, or all of the points. Remind students that an exponential model will produce a curved fit.

- **Marginal Frequencies.** Entries in the "Total" row and "Total" column of a two-way frequency table.

- **Mean Absolute Deviation.** A measure of variation in a set of numerical data, computed by adding the distances between each data value and the mean, then dividing by the number of data values. Example: For the data set {2, 3, 6, 7, 10, 12, 14, 15, 22, 120}, the mean absolute deviation is 20.

- **Outlier.** Sometimes, distributions are characterized by extreme values that differ greatly from the other observations. These extreme values are called outliers. As a rule, an extreme value is considered to be an outlier if it is at least 1.5 interquartile ranges below the lower quartile (Q_1), or at least 1.5 interquartile ranges above the upper quartile (Q_3).

  OUTLIER if the values lie outside these specific ranges:

  $$Q_1 - 1.5 \bullet IQR$$
  $$Q_3 + 1.5 \bullet IQR$$

- **Quantitative Variables.** Numerical variables that represent a measurable quantity. For example, when we speak of the population of a city, we are talking about the number of people in the city – a measurable attribute of the city. Therefore, population would be a quantitative variable. Other examples: scores on a set of tests, height and weight, temperature at the top of each hour.

- **Scatter plot.** A graph in the coordinate plane representing a set of bivariate data. For example, the heights and weights of a group of people could be displayed on a scatter plot. If you are looking for values that fall within the range of values plotted on the scatter plot, you are interpolating. If you are looking for values that fall beyond the range of those values plotted on the scatter plot, you are extrapolating.

- **Second Quartile (**Q_2**).** The *median* value in the data set.

- **Shape**. The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity.

    - **Symmetry-** A symmetric distribution can be divided at the center so that each half is a mirror image of the other.

    - **Number of Peaks**- Distributions can have few or many peaks. Distributions with one clear peak are called unimodal and distributions with two clear peaks are called bimodal. Unimodal distributions are sometimes called bell-shaped.

    - **Direction of Skew**- Some distributions have many more observations on one side of graph than the other. Distributions with a tail on the right toward the higher values are said to be skewed right; and distributions with a tail on the left toward the lower values are said to be skewed left.

    - **Uniformity-** When observations in a set of data are equally spread across the range of the distribution, the distribution is called uniform distribution. A uniform distribution has no clear peaks.

- **Spread.** The spread of a distribution refers to the variability of the data. If the data cluster around a single central value, the spread is smaller. The further the observations fall from the center, the greater the spread or variability of the set. (range, interquartile range, Mean Absolute Deviation, and Standard Deviation measure the spread of data)

- **Third quartile.** For a data set with median M, the third quartile is the median of the data values greater than $\boldsymbol{M}$. Example: For the data set {2, 3, 6, 7, 10, 12, 14, 15, 22, 120}, the third quartile is 15.

- **Trend.** A change (positive, negative or constant) in data values over time.

- **Two-Frequency Table.** A useful tool for examining relationships between categorical variables. The entries in the cells of a two-way table can be frequency counts or relative frequencies.

EVIDENCE OF LEARNING

By the conclusion of this unit, students should be able to demonstrate the following competencies:

- Construct appropriate graphical displays (dot plots, histogram, and box plot) to describe sets of data values.
- Select the appropriate measures to describe and compare the center and spread of two or more data sets in context.
- Use the context of the data to explain why its distribution takes on a particular shape (e.g. are there real-life limits to the values of the data that force skewedness?)
- Explain the effect of any outliers on the shape, center, and spread of the data sets.
- Create a two-way frequency table from a set of data on two categorical variables.
- Calculate joint, marginal, and conditional relative frequencies and interpret in context.
- Recognize associations and trends in data from a two-way table.
- Create a scatter plot from two quantitative variables and describe the form, strength, and direction of the relationship between the two variables in context.
- Determine which type of function best models a set of data.
- Interpret constants and coefficients in the context of the data (e.g. slope and y-intercept of linear models, base/growth or decay rate and *y*-intercept of exponential models) and use the fitted function to make predictions and solve problems in the context of the data.
- Use algebraic methods and technology to fit a linear function to the data for data sets that appear to be linear.
- Interpret the slope and *y*-intercept in the context of the data.
- Compute the correlation coefficient and show that it is a measure of the strength and direction of a linear relationship between two quantities in a set of data.
- Determine if the association between two variables is a result of a cause and effect relationship.

TEACHER RESOURCES

The following pages include teacher resources that teachers may wish to use to supplement instruction.

- Web Resources
- Notes: Comparing Distributions
- Notes: Measuring
- Graphic Organizer: Measures of Center and Spread
- Graphic Organizer: Representing Data Graphically

Web Resources

The following list is provided as a sample of available resources and is for informational purposes only. It is your responsibility to investigate them to determine their value and appropriateness for your district. GaDOE does not endorse or recommend the purchase of or use of any particular resource.

- Correlation vs Causality
  https://www.khanacademy.org/math/probability/regression/regression-correlation/v/correlation-and-causality
  A video that gives a good explanation of this topic

- Fitting a Line to Data Using Excel
  https://www.khanacademy.org/math/probability/regression/regression-correlation/v/fitting-a-line-to-data
  Using a computer to find the equation of the line of best fit

- Additional tasks for this unit
  https://www.illustrativemathematics.org/content-standards/HSS

- Online boxplot maker http://www.alcula.com/calculators/statistics/box-plot/

- Interactive boxplot http://www.shodor.org/interactivate/activities/BoxPlot/
  Use one of the given data sets or enter you own data.

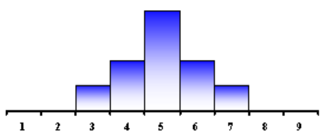
Notes: Comparing Distributions

When you compare two or more data sets, focus on four features:

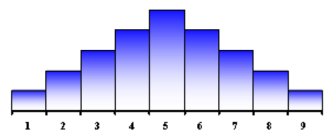
- **Center.** Graphically, the center of a distribution is the point where about half of the observations are on either side.
- **Spread**. The spread of a distribution refers to the variability of the data. If the observations cover a wide range, the spread is larger. If the observations are clustered around a single value, the spread is smaller.
- **Shape**. The shape of a distribution is described by symmetry, skewness, number of peaks, etc.
- **Unusual features**. Unusual features refer to gaps (areas of the distribution where there are no observations) and outliers.

SPREAD

The spread of a distribution refers to the variability of the data. If the data cluster around a single central value, the spread is smaller. The further the observations fall from the center, the greater the spread or variability of the set.



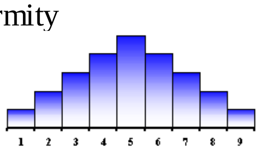
Less Spread



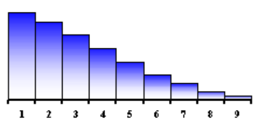
More Spread

SHAPE

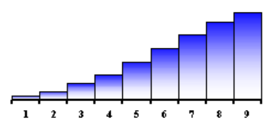
The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity



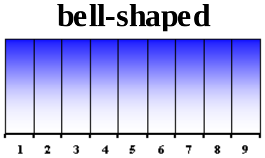
symmetric, unimodal, bell-shaped



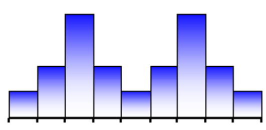
skewed right



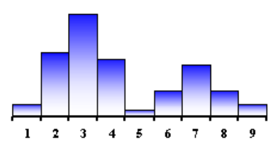
skewed left



uniform



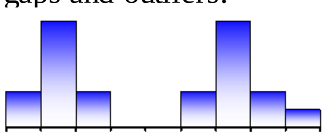
symmetric, bimodal



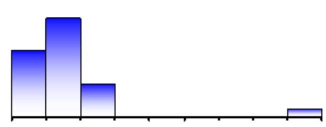
non-symmetric, bimodal

UNUSUAL FEATURES

Sometimes, statisticians refer to unusual features in a set of data. The two most common unusual features are gaps and outliers.



gap



outlier

Graphic Organizer: Measures of Center and Spread

Mean

Median

Measures of Center

Measures of Spread

Mean Absolute Deviation

Interquartile Range

Graphic Organizer: Representing Data Graphically

	Dot Plots	**Box Plots**	**Histograms**
Definition			
Example			
Characteristics			
Other Notes			

FORMATIVE ASSESSMENT LESSONS (FAL)

Formative Assessment Lessons are intended to support teachers in formative assessment. They reveal and develop students' understanding of key mathematical ideas and applications. These lessons enable teachers and students to monitor in more detail their progress towards the targets of the standards. They assess students' understanding of important concepts and problem solving performance, and help teachers and their students to work effectively together to move each student's mathematical reasoning forward.

More information on Formative Assessment Lessons may be found in the Comprehensive Course Guide.

SPOTLIGHT TASKS

A Spotlight Task has been added to each GSE mathematics unit in the Georgia resources for middle and high school. The Spotlight Tasks serve as exemplars for the use of the Standards for Mathematical Practice, appropriate unit-level Georgia Standards of Excellence, and research-based pedagogical strategies for instruction and engagement. Each task includes teacher commentary and support for classroom implementation. Some of the Spotlight Tasks are revisions of existing Georgia tasks and some are newly created. Additionally, some of the Spotlight Tasks are 3-Act Tasks based on 3-Act Problems from Dan Meyer and Problem-Based Learning from Robert Kaplinsky.

TASKS

The following tasks represent the level of depth, rigor, and complexity expected of all Algebra I students. These tasks, or tasks of similar depth and rigor, should be used to demonstrate evidence of learning. It is important that all elements of a task be addressed throughout the learning process so that students understand what is expected of them. While some tasks are identified as a performance task, they may also be used for teaching and learning (learning/scaffolding task).

Task Name *Suggested Time*	**Task Type** *Grouping Strategy*	**Content Addressed**	**Standards**
Math Class *50-60 minutes*	Homework Task *Individual*	• Represent data with plots on the real number line • Compare center and spread of two or more different data sets • Interpret differences in shape, center, and spread in the context of data sets accounting for outliers	• S.ID.1 • S.ID.2 • S.ID.3
The Basketball Star *50-60 minutes*	Performance Task *Individual*	• Represent data with plots on the real number line. • Compare center and spread of two or more different data sets. • Interpret differences in shape, center, and spread in the context of data sets, accounting for outliers. • Use frequencies and relative frequencies to compare data sets.	• S.ID.1 • S.ID.2 • S.ID.3
Representing Data 1: Using Frequency Graphs (FAL) *90-120 minutes* PDF	Formative Assessment Lesson *Individual / Small Group*	• Interpret frequency graphs. • Describe center and spread of data based on frequency graphs.	• S.ID.1 • S.ID.2 • S.ID.3
BMI Calculations *45-60 minutes* PDF / Word	Achieve CCSS-CTE Classroom Task	• Create box plots and use them to describe the center and spread of data.	• S.ID.1 • S.ID.2 • S.ID.3
Representing Data 2: Using Box Plots (FAL) *90-120 minutes* PDF	Formative Assessment Lesson *Individual / Small Group*	• Interpret box plots. • Relate box plots to frequency graphs of the same data.	• S.ID.1 • S.ID.2 • S.ID.3

Task Name *Suggested Time*	**Task Type** *Grouping Strategy*	**Content Addressed**	**Standards**
Public Opinions and Leisure Time (Spotlight Task) *50-60 minutes*	Scaffolded Learning Task *Individual / Partner*	• Practice summarizing categorical data from two-way frequency tables • Interpret relative frequencies in the context of the data • Recognize possible associations and trends in data	• S.ID.5
If the Shoe Fits! *1-2 hours*	Scaffolded Learning Task *Individual / Partner*	• Compare center and spread of two or more different data sets • Interpret differences in shape, center, and spread in the context of data sets accounting for outliers • Represent two quantitative variables on a scatter plot • Describe how two quantitative variables are related	• S.ID.1 • S.ID.2 • S.ID.3 • S.ID.6
Spaghetti Regression *50-60 minutes*	Learning Task *Small Group*	• Represent data on a scatter plot • Describe how two variables are related • Fit a linear function for a scatter plot that suggests a linear association	• S.ID.6 • S.ID.6c
Devising a Measure for Correlation (FAL) *≈ 2 hours* PDF	Formative Assessment Lesson *Individual / Small Group*	• Understand correlation as the degree of fit between two variables • Brainstorm and evaluate different approaches to measuring correlation.	• S.ID.6 • S.ID.6a • S.ID.6c
TV / Test Grades *1-2 hours*	Learning Task *Partner/Small Group*	• Represent data on a scatter plot • Describe how two variables are related • Fit a linear function for a scatter plot • Interpret the slope and the intercept of a linear model in the context of the data • Compute (using technology) and interpret the correlation coefficient of a linear fit • Distinguish between correlation and causation	• S.ID.6 • S.ID.6a • S.ID.6c • S.ID.7 • S.ID.8 • S.ID.9

Task Name ***Suggested Time***	**Task Type** ***Grouping Strategy***	**Content Addressed**	**Standards**
Equal Salaries for Equal Work? *50-60 minutes*	Performance Task *Partner/Individual*	• Represent data on a scatter plot • Describe how two variables are related • Fit a linear function to a scatter plot • Interpret the slope and the intercept of a linear model in context • Compute and interpret the correlation coefficient of a linear fit • Distinguish between correlation and causation	• S.ID.6 • S.ID.6a • S.ID.6c • S.ID.7 • S.ID.8
iRegress (FAL) *≈ 2 hours*	Formative Assessment Lesson *Individual / Small Group*	• Interpret the correlation coefficient • Interpret the slope and y-intercept of a line of best fit in the context of the data • Contrast correlation and causation	• S.ID.6a • S.ID.7 • S.ID.8 • S.ID.9
Research Design *≈ 3 hours*	Culminating Task *Individual / Small Group*	• Compare center and spread of two or more different data sets • Interpret differences in shape, center, and spread in the context of data sets, accounting for outliers • Summarize categorical data from two-way frequency tables • Interpret relative frequencies in the context of the data • Recognize possible associations and trends in data • Represent data on a scatter plot • Describe how two variables are related • Fit a linear function for a scatter plot • Interpret the slope and the intercept of a linear model in context • Compute and interpret the correlation coefficient of a linear fit • Distinguish between correlation and causation	• S.ID.1 • S.ID.2 • S.ID.3 • S.ID.5 • S.ID.6 • S.ID.6a • S.ID.6c • S.ID.7 • S.ID.8 • S.ID.9

.

Math Class (Homework Task)

Back to Task Table

Introduction

This task should be used to formatively assess students' ability to summarize, represent, and interpret data on a single count or measurement variable. The intention in this Closing or Homework Task is to adjust instruction and differentiate for students prior to engaging in the Learning Task "If the Shoe Fits." Standards are integrated within this task to help students make real world connections.

Mathematical Goals

- Represent data with plots on the real number line
- Compare center and spread of two or more different data sets
- Interpret differences in shape, center, and spread in the context of data sets, accounting for outliers

Essential Questions

- How can I use visual representations and measures of center and spread to compare two data sets?

Georgia Standards of Excellence

MGSE9-12.S.ID1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
   Students must make sense of the data given and determine what their calculations tell them about how the two classes compare.
6. Attend to precision.
   Students are working with calculations and numerical representations rather than visual representations that allow for more estimation when comparing the data sets.

Background Knowledge

- Students need to know how to find mean, median, MAD, and IQR.
- Differentiate between measures of center and measures of spread.

Common Misconceptions

- Students may believe that a bar graph and a histogram are the same. A bar graph shows frequencies of categorical data, and a histogram is for numerical data.
- Students sometimes forget to take the absolute value of the deviations for MAD. Failure to do so results in an MAD of 0, indicating no deviation. This should alert students that they have forgotten to take the absolute values.

Materials

- Calculators
- Graph or centimeter grid paper (optional)

Grouping

- Individual

Differentiation

Extension:

- Suppose Mr. Turner had a student who made a 100 on the test in each class. How would this affect the statistics for each class?
  (Solution: It would increase the mean and MAD in each class but would not affect the IQR and might only slightly increase the median.)
- Create a new list of at least ten student grades with the following properties:
  mean 70; median 75; IQR 10; MAD 16. Describe your process.
  (Possible solution: 0, 60, 70, 75, 75, 75, 80, 80, 90, 95)

Intervention:

- Students might work with smaller data sets to simplify calculations. Interpretation / comparison of statistics can still be the central purpose of this task.
- Work with extreme cases to help students understand what statistics tell us about data sets. Extreme cases include data sets where all numbers are the same, data sets where all numbers but one are the same, and data sets where numbers are clearly quite varied.
- Ask students how a low or high grade would impact their average (they're used to thinking this way) and discuss whether/how it would impact other statistics.

Formative Assessment Questions

- Describe how you determined the mean absolute deviation for each class.

Math Class – Teacher Notes

Mr. Turner has two Math 2 classes. With one class, he lectured and the students took notes. In the other class, the students worked in small groups to solve math problems. After the first test, Mr. Turner recorded the student grades to determine if his different styles of teaching might have impacted student learning.

Class 1: 80, 81, 81, 75, 70, 72, 74, 76, 77, 77, 77, 79, 84, 88, 90, 86, 80, 80, 78, 82
Class 2: 70, 90, 88, 89, 86, 86, 86, 86, 84, 82, 77, 79, 84, 84, 84, 86, 87, 88, 88, 88

1. Analyze his student grades by calculating the mean, median, mean absolute deviation and interquartile range.

Solution:

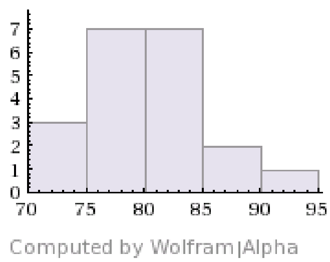
	Class 1	***Class 2***
Mean (μ):	***79.35***	***84.6***
Median:	***79.5***	***86***
Mean absolute deviation (MAD):	***3.85***	***3.28***
Interquartile range:	***5***	***4***

The grades for class 2 are higher with more consistency (less variability). Students must justify their reasoning, using the comparative statistics, as to which class they believed to be the lecture and which is the small group.

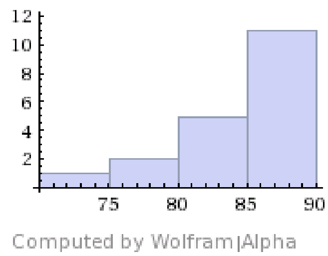
2. Draw histograms to easily compare the shapes of the distributions.

Solution:

Class 1



Class 2



3. Which measure of center and spread is more appropriate to use? Explain.

Solution:
Answers may vary. Sample answer: In Class 1, mean and median are both appropriate measures of center because the data is nearly symmetric so they are nearly the same. Similarly, either MAD or IQR would be reasonable measurements of spread. In Class 2, the fact that the data is skewed left implies that the median is a better choice as a measure of center and IQR is a better choice as a measure of spread.

Learning Task: Math Class

Name__________________________________ **Date**__________________

Mathematical Goals

- Represent data with plots on the real number line
- Compare center and spread of two or more different data sets
- Interpret differences in shape, center, and spread in the context of data sets, accounting for outliers

Essential Questions

- How can I use visual representations and measures of center and spread to compare two data sets?

Georgia Standards of Excellence

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
6. Attend to precision.

Learning Task: Math Class

Name________________________________ **Date**__________________

Mr. Turner has two Math 2 classes. With one class, he lectured and the students took notes. In the other class, the students worked in small groups to solve math problems. After the first test, Mr. Turner recorded the student grades to determine if his different styles of teaching might have impacted student learning.

Class 1: 80, 81, 81, 75, 70, 72, 74, 76, 77, 77, 77, 79, 84, 88, 90, 86, 80, 80, 78, 82

Class 2: 70, 90, 88, 89, 86, 86, 86, 86, 84, 82, 77, 79, 84, 84, 84, 86, 87, 88, 88, 88

1. Analyze his student grades by calculating the mean, median, mean absolute deviation, and interquartile range. Which class do you think was the lecture and which was the small group? Why?

2. Draw histograms to easily compare the shapes of the distributions.

3. Which measure of center and spread is more appropriate to use? Explain.

The Basketball Star (Performance Task)

Back to Task Table

Introduction

The purpose of this task is to do a final formative assessment of the standards integrated in the task prior to moving on to the next cluster of standards. Students will determine the best way to represent data with number line plots, including dot plots, histograms, or box plots. Upon determining the most appropriate graphical representation, students will use the shape of the data to compare two or more data sets with summary statistics by comparing either center (using mean and median) and/or spread (using range, interquartile range, and mean absolute deviation.)

Notes: Comparing Distributions (see Teacher Resources) may be a good supplement for this task.

Mathematical Goals

- Represent data with plots on the real number line.
- Compare center and spread of two or more different data sets.
- Interpret differences in shape, center, and spread in the context of data sets accounting for outliers.
- Use frequencies and relative frequencies to compare data sets.

Essential Questions

- How can I use visual representations and measures of center and spread to compare two data sets?

Georgia Standards of Excellence

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

MGSE9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

Standards for Mathematical Practice

3. Construct viable arguments and critique the reasoning of others.
   Students must use data to determine who is the better basketball player.
4. Model with mathematics.
   Students represent data in various forms and use each form to justify their conclusion.
6. Attend to precision.
   Calculations for 5-number summary and MAD will require precision.

Background Knowledge

- Students have practiced choosing appropriate graphs (dot plots, histograms, and box plots) to be consistent with numerical data
- Students have explained patterns in terms of center, spread, shape, and unusual features such as gaps, clusters, or outliers
- Students have described common distribution shapes (symmetric, bell-shaped, skewed).

Common Misconceptions

- Students may believe that a bar graph and a histogram are the same. A bar graph shows frequencies of categorical data, and a histogram is for numerical data.
- Students sometimes forget to take the absolute value of the deviations for MAD.

Materials

- Pencil
- paper/graph paper
- calculator or graphing calculator
- Handout: Comparing Distributions (see Teacher Resources)

Grouping

- Individual

Differentiation

Extension:

- Suppose Bob and Alan each had their best game ever, scoring 16 points in a game. How would this affect the statistics for each boy?
  Bob's mean would become 10.9; his median would not change.
  Alan's mean would become 7.5; his median would not change.

Intervention:

- Students might work with smaller data sets to simplify calculations. Interpretation / comparison of statistics can still be the central purpose of this task.

Formative Assessment Questions

- What summary statistics helped you decide who is the better basketball player?
- How can the dot plot help you find the summary statistics?

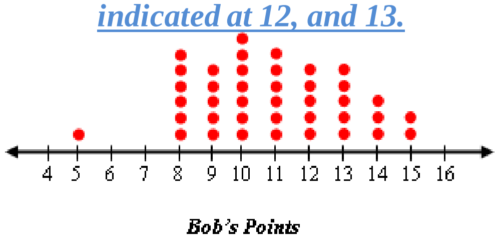
- ## <u>The Basketball Star – Teacher Notes</u>

Bob believes he is a basketball star and so does his friend Alan.

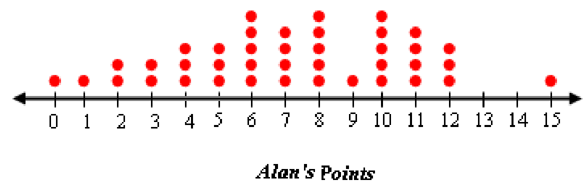
Bob's Points per Game	**Alan's Points per Game**
8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8	1, 3, 0, 2, 4, 5, 7, 7, 8, 10, 4, 4, 3, 2, 5, 6, 6, 6, 8, 8, 10, 11, 11, 10, 12, 12, 5, 6, 8, 9, 10, 15, 10, 12, 11, 11, 6, 7, 7, 8

1. Create dot plots for Bob and Alan's last forty games.

<u>Attention: The dot plot below should have three data points indicated at 7, five data points indicated at 11, and four data points indicated at 12, and 13.</u>***

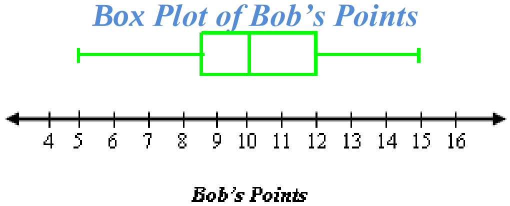


Dot Plot of Alan's Points

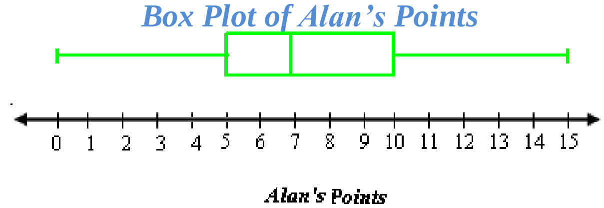


2. Create box plots for Bob and Alan's last forty games.

Box Plot of Bob's Points

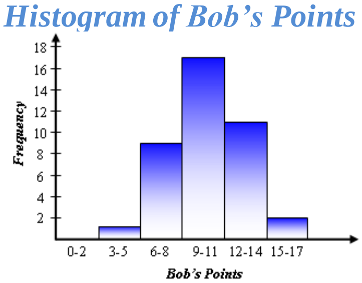


Box Plot of Alan's Points

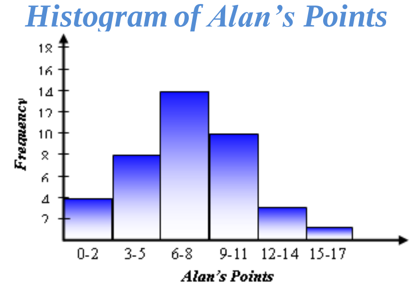


3. Create histograms of both Bob's and Alan's data.

Histogram of Bob's Points



Histogram of Alan's Points



4. Which graphical representation best displayed Bob's and Alan's data?

Answers may vary. Make sure students can defend their choice based on the comparison that they are trying to make.

5. Use summary statistics to compare Bob and Alan's points per game.

Solution:
Alan's five-number summary statistics are as follows: minimum is 0, 1st quartile is 5, median is 7, 3rd quartile is 10, and maximum is 15. The mean value is 7.25, the range of the values is 15 points, the IQR is 5 points. Any outliers must be outside of the interval (-2.5, 17.5). Comparing these two sets shows that Alan has a larger spread of data with an IQR of 5 points versus Bob's IQR of 3.5 points. In addition, Alan's overall scores are generally lower than Bob's, shown by Alan's median 7.5 points which is less that Bob's median score of 10 points. The shape of Alan's graph is not as symmetric as Bob's either, suggesting that Alan's scores might be slightly skewed right since there is a bit of a tail in that direction whereas Bob's scores appear to have a symmetric distribution.

6. Describe each person's data in terms of center, spread, and shape.

Solution:
The shape of Bob's data is fairly symmetric. The mean is 10.4 and the median is 10. The five-number summary is 5, 8.5, 10, 12, 15. The range is 10 points, the IQR is 3.5 points, and the mean absolute deviation is 2.0 points. Any outliers must lie outside the interval (3.25, 17.25). This information helps us determine that most of the values are very close to the center of the graph—this makes sense because of the small range of values and how many values are close to 10.
The shape of Alan's data is slightly skewed right and is more spread than Bob's. The mean is 7.25 and the median is 7. The five-number summary is 0, 5, 7, 10, 15. The range is 15 points, the IQR is 5 points. Any outliers must lie outside the interval (-2.5, 17.5). This information helps us see that the center of the data is lower than Bob's and the spread is larger.

7. After the season, the statistician did not have time to compute Bob's relative frequency. Complete the table by determining the relative frequency for Bob.

Points Scored	***Frequency for Bob***	***Relative Frequency for Bob***	***Frequency for Alan***	***Relative Frequency for Alan***
0	0	***0.000***	1	0.025
1	0	***0.000***	1	0.025
2	0	***0.000***	2	0.050
3	0	***0.000***	2	0.050
4	0	***0.000***	3	0.075
5	1	***0.025***	3	0.075
6	0	***0.000***	5	0.125
7	3	***0.075***	4	0.100
8	6	***0.150***	5	0.125
9	5	***0.125***	1	0.025
10	7	***0.175***	5	0.125
11	5	***0.125***	4	0.100
12	4	***0.100***	3	0.075
13	4	***0.100***	0	0.000
14	3	***0.075***	0	0.000
15	2	***0.050***	1	0.025
TOTALS	**40**	**1.000**	**40**	**1.000**

8. Discuss any trends or associations from the table below concerning points scored by the two basketball players.

Answers will vary. Sample answer: The frequency and relative frequency tables above show that Bob scored under 8 points in only 2.5% of his games, whereas Alan scored under 8 points in 52.5% of his games (over half). Bob scored 10 points or more in 70% of his games, whereas Alan scored 10 points or more in only 32.5% of his games.

9. Based on your representations and your calculations, is either friend a basketball star? Justify your answer, based on your graphical representations (#1-3), your summary statistics (#5), *and* the frequency / relative frequency table (#7).

Answers may vary as long as students use each representation to justify their answers. Most students will more than likely believe Bob to be the basketball star because Alan's data is less symmetric and slightly skewed right, with Bob having many more high-scoring games and fewer low-scoring games.

Performance Task: The Basketball Star

Name_________________________________ **Date**__________________

Mathematical Goals

- Represent data with plots on the real number line.
- Compare center and spread of two or more different data sets.
- Interpret differences in shape, center, and spread in the context of data sets accounting for outliers.
- Use frequencies and relative frequencies to compare data sets.

Essential Questions

- How can I use visual representations and measures of center and spread to compare two data sets?

Georgia Standards of Excellence

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

MGSE9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

Standards for Mathematical Practice

3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
6. Attend to precision.

Performance Task: The Basketball Star

Name___________________________________ **Date**___________________

Bob believes he is a basketball star and so does his friend Alan.

Bob's Points per Game	**Alan's Points per Game**
8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8	1, 3, 0, 2, 4, 5, 7, 7, 8, 10, 4, 4, 3, 2, 5, 6, 6, 6, 8, 8, 10, 11, 11, 10, 12, 12, 5, 6, 8, 9, 10, 15, 10, 12, 11, 11, 6, 7, 7, 8

1. Create dot plots for Bob and Alan's last forty games.

2. Create box plots for Bob and Alan's last forty games.

3. Create histograms of both Bob's and Alan's data.

4. Which graphical representation best displayed Bob's and Alan's data?

5. Use summary statistics to compare Bob and Alan's points per game.

6. Describe each person's data in terms of center, spread, and shape.

7. After the season, the statistician did not have time to compute Bob's relative frequency. Complete the table by determining the relative frequency for Bob.

Points Scored	***Frequency for Bob***	***Relative Frequency for Bob***	***Frequency for Alan***	***Relative Frequency for Alan***
0	0		1	0.025
1	0		1	0.025
2	0		2	0.050
3	0		2	0.050
4	0		3	0.075
5	1		3	0.075
6	0		5	0.125
7	3		4	0.100
8	6		5	0.125
9	5		1	0.025
10	7		5	0.125
11	5		4	0.100
12	4		3	0.075
13	4		0	0.000
14	3		0	0.000
15	2		1	0.025
TOTALS	**40**		**40**	**1.000**

8. Discuss any trends or associations from the table below concerning points scored by the two basketball players.

9. Based on your representations and your calculations, is either friend a basketball star? Justify your answer, based on your graphical representations (#1-3), your summary statistics (#5), *and* the frequency / relative frequency table (#7-8).

Formative Assessment Lesson: Representing Data 1: Using Frequency Graphs

Back to Task Table

Source: Formative Assessment Lesson Materials from Mathematics Assessment Project
http://map.mathshell.org/materials/download.php?fileid=1230

Task Comments and Introduction

Tasks and lessons from the Mathematics Assessment Project are specifically designed to help teachers effectively formatively assess their students. The way the tasks and lessons are designed gives the teacher a clear understanding of what the students are able to do and not do. Within the lesson, teachers will find suggestions and question prompts that will help guide students towards understanding. For more information access the MAP website:
http://www.map.mathshell.org/materials/background.php?subpage=formative

The task, *Representing Data 1: Using Frequency Graphs*, is a Formative Assessment Lesson (FAL) that can be found at the website:
http://map.mathshell.org/materials/lessons.php?taskid=404&subpage=concept

The FAL document provides a clear lesson design, from the opening of the lesson to the closing of the lesson.

The PDF version of the task can be found at the link below:
http://map.mathshell.org/materials/download.php?fileid=1230

Mathematical Goals

- Use frequency graphs to identify a range of measures and make sense of this data in a real-world context
- Understand that a large number of data points allow a frequency graph to be approximated by a continuous distribution.

Essential Questions

- What do frequency graphs tell me about the center and spread of data?

Georgia Standards of Excellence

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
   Students must interpret and describe the situation based on the frequency graph.
2. Reason abstractly and quantitatively.
   Students must recognize what the frequency graph represents.
3. Construct viable arguments and critique the reasoning of others.
   Students must defend their choices mathematically.

Background Knowledge

- Students can read frequency graphs.
- Students understand the meaning of range, mode, and median.

Common Misconceptions

- Students may think that a uniform distribution indicates that most values are similar. In fact, a uniform distribution tells us that the values have similar *frequencies*, while the values themselves may be different.

Materials

- See FAL website

Grouping

- Individual / small group

BMI Calculations (Career and Technical Education Task) Back to Task Table

Source: National Association of State Directors of Career Technical Education Consortium
PDF: **http://www.achieve.org/files/CCSS-CTE-Task-BMI-Calculations-FINAL.pdf**
Word: **http://www.achieve.org/files/CCSS-CTE-Task-BMI-Calculations-FINAL.doc**

Introduction

Students collect data about Body Mass Index (BMI), then they create and analyze a box plot representing this data.

Mathematical Goals

- Create box plots and use them to describe the center and spread of data.

Essential Questions

- What do box plots tell me about the center and spread of data?

Georgia Standards of Excellence

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots)

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
Students must interpret the mathematical formula for BMI and understand what it means.

3. Construct viable arguments and critique the reasoning of others.
Students must defend their descriptions of the distributions.

4. Model with mathematics.
Students model a large set of data using box plots.

5. Use appropriate tools strategically.
Students may measure their own height and weight to perform calculations.

6. Attend to precision.
Students must perform calculations carefully to accurately determine BMI.

Background Knowledge

- Students can convert units (inches to meters, pounds to kilograms).
- Students know how to find and interpret the five number summary and create a box plot.

Common Misconceptions

- Students sometimes believe a narrower box in a box plot indicates that there are fewer data points in that set. In fact, box plots split the data into four equal parts, so a narrower box indicates that there is a higher concentration of data points in that range.

Materials

- If taking real data, meter stick and scale will be necessary. (Alternatively, teacher can provide hypothetical data for students.)

Grouping

- Partner / small group

Differentiation

- See extensions in task.

Formative Assessment Lesson: Representing Data 2: Using Box Plots

Back to Task Table

Source: Formative Assessment Lesson Materials from Mathematics Assessment Project
http://map.mathshell.org/materials/download.php?fileid=1243

Task Comments and Introduction

Tasks and lessons from the Mathematics Assessment Project are specifically designed to help teachers effectively formatively assess their students. The way the tasks and lessons are designed gives the teacher a clear understanding of what the students are able to do and not do. Within the lesson, teachers will find suggestions and question prompts that will help guide students towards understanding. For more information access the MAP website: http://www.map.mathshell.org/materials/background.php?subpage=formative

The task, *Representing Data 2: Using Box Plots*, is a Formative Assessment Lesson (FAL) that can be found at the website: http://map.mathshell.org/materials/lessons.php?taskid=423&subpage=concept

The FAL document provides a clear lesson design, from the opening of the lesson to the closing of the lesson.

The PDF version of the task can be found at the link below:
http://map.mathshell.org/materials/download.php?fileid=1243

Mathematical Goals

- Interpret data represented in frequency graphs and in boxplots.
- Connect frequency graphs and boxplots.

Essential Questions

- What do box plots tell me about the center and spread of data?
- How are box plots and frequency graphs related?

Georgia Standards of Excellence

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
   Students must interpret and describe the situation based on the frequency graph and box plot.
2. Reason abstractly and quantitatively.
   Students must recognize what the frequency graphs and box plots represent and how they are related.
3. Construct viable arguments and critique the reasoning of others.
   Students must defend their choices mathematically.

Background Knowledge

- Students can read and interpret a box plot.
- Students can read and interpret a frequency graph.

Common Misconceptions

- Students sometimes believe a narrower box in a box plot indicates that there are fewer data points in that set. In fact, box plots split the data into four equal parts, so a narrower box indicates that there is a higher concentration of data points in that range.

Students may think that a uniform distribution on a frequency table indicates that most values are similar. In fact, a uniform distribution tells us that the values have similar *frequencies*, while the values themselves may be different.

Materials

- See FAL website

Grouping

- Individual / small group

Public Opinions and Leisure Time (Spotlight Task)

Back to Task Table

Mathematical Goals

- Practice summarizing categorical data from two-way frequency tables
- Interpret relative frequencies in the context of the data
- Recognize possible associations and trends in data

Essential Questions

- How do you interpret relative frequencies in the context of a two-way frequency table?

Georgia Standards of Excellence

MGSE9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

Standards of Mathematical Practice

2. Reason abstractly and quantitatively**.**
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
6. Attend to precision.
7. Look for and make sense of structure.

Background Knowledge

- Students should know the meaning of joint, marginal, and conditional relative frequencies and how to calculate them.

Common Misconceptions

- Students may confuse the meanings of the terms.
- Students should be careful about whether the question asks for frequency or percentage.

Materials

- Pencil
- Paper
- Calculator

Grouping

- Individual

Differentiation

Extension:

- Students can research actual data about this topic (or another controversial topic) in their community.

Intervention:

- Students may need remediation in determining percentages. Since the total number of respondents is 200, students can fairly easily use proportional reasoning to determine the equivalent ratio of people out of 100 ("per cent").

Formative Assessment Questions

- What is the difference between relative frequency and percentage?
- What are joint, marginal, and conditional relative frequencies?
- Why should the relative frequencies sum to 1?

TEACHER COMMENTARY

Working with contingency tables is an ideal way to introduce students to the concepts of conditional percentages and marginal percentages. So often it is percentages reported in the media as the numerical summaries (statistics) for categorical data. Traditionally, the concept of conditional percentage has been taught using mathematical formulas and students in turn struggles with understanding the concept of a conditional percentage as the student is challenged to get through a mathematical formula. In the media, it is important for students to distinguish when a marginal percentage is being reported versus a conditional percentage. Also understanding the concept of a conditional percentage will allow students to more easily understand the concept of a P-value in later high school math courses.

Working is two-way contingency tables as summary representation of two categorical variables is our most common way to explore the bivariate association between these two variables. Can we use one categorical variable to predict another categorical variable?

In part 1, the student is guided through finding and learning the terminology for the different types of percentages (joint, conditional, and marginal). A joint percentage is simply the intersection of the two variables for categories of those variables out of the entire sample. A conditional percentage is found where you interested in only one category for a categorical variable in relation to the other categories of the second categorical variable. As we say, instead of using all of the sampled individuals (as we do for joint percentages), we condition or restrict ourselves to only one category of a variable. Since we are interested in comparing the different age groups with respect to the participant's opinion preference, we restrict ourselves to each age group and then find the percentage in that age group that are for, against, or no opinion. In this case, we call age group the explanatory variable and opinion the response variable. For a marginal, we are interested in finding percentages for the different categories of one variable at a time out of all the participants sampled. Note that the frequencies for the categories of one variable at a time come from the margins of the table. Thus, the percentage of all sampled individuals who are against is 70/200 = 35%. This is a marginal. For the 21-40 age group, the percentage who are against is 20/50 = 40%. This is a conditional percentage since we restricted ourselves to the 21-40 age group. The frequency count for that age group determines the denominator of the relative frequency. For the 41-60 age group, the percentage against is 30/75 = 67% and for the over 60 age group, 20/75 = 27%.
Since all three conditional percentages are different, there appears to be an association between age group and opinion. If there were no association, we would expect these three conditional

percentages to equal the marginal of all participants being against that is 35%. NOTE: many students may want to average the three conditional percentages together to find the expected value if there is no association. Mathematically, this is not correct since the sample sizes for the three age groups are not the same. Taking the average of the conditional percentages only works if the categories of the variable being conditioned upon have the same sample size. Students should always use the correct marginal percentage and then they don't have to worry about sample size.

At this stage, students are exploring the summary frequencies of the categorical variables to see if there is an association between two variables for sampled individuals. Depending upon how the participants were selected for the sample determines the scope of inference for the possible association; i.e., can we infer an association for a larger population? Also, is the association we observe from the sample 'statistically significant"? Later in high school, students will learn how to make a decision about whether the association would be considered statistical significant by using simulation.

Part I. A public opinion survey explored the association between age and support for increasing the minimum wage. Although age is a quantitative variable, the researchers create three categories of age groups and treated the age data as categorical. The results are found in the following two-way frequency table

	For	**Against**	**No Opinion**	**TOTAL**
Ages 21-40	25	20	5	50
Ages 41-60	30	30	15	75
Over 60	50	20	5	75
TOTAL	105	70	25	200

Frequency Count

1. What percentage of those individuals surveyed were in the 21 – 40 age group and for increasing the minimum wage? Show your work for finding this percentage. This percentage is called a joint percentage.

25/200 = 12.5%

2. For the 21 to 40 age group, what percentage supports increasing the minimum wage? Explain how you arrived at your percentage. This percentage is called a conditional percentage.

25/50 = 50%

3. For the 41 to 60 age group, what percentage supports increasing the minimum wage? Show your work.

30/75=40%

4. For the over 60 age group, what percentage supports increasing the minimum wage? Show your work.

   50/75 = 67%

5. What percentage of all individuals surveyed favored increasing the minimum wage? Explain how you arrived at your percentage. This percentage is called a marginal percentage.

   105/200 = 52.5%

6. If there is no association between age group and opinion about increasing the minimum wage, we would expect the conditional percentages found in parts 2-4 to all be the same. We observe that the three percentages are different. What numerical value for the conditional percentage of each age group that favor increasing the minimum wage would we expect if there is no association between age group and opinion? Hint: Consider the percentage who favor increasing the minimum wage ignoring the age group classification of the individuals surveyed.

   We would expect all three conditional percentages to be the same as the marginal percentage of 52.5%.

Part II. The table below gives the responses of 50 teachers' responses to a survey asking which activity they enjoyed most: dancing, playing/watching sports, or seeing movies. Is there an association between gender of the teacher and type of activity enjoyed the most?

	Dance	**Sports**	**Movies**	**TOTAL**
Female	16	6	8	**30**
Male	2	10	8	**20**
TOTAL	**18**	**16**	**16**	50

7. In exploring whether an association exists between gender and type of activity, we are interested in knowing if gender of the teacher helps predict the type of activity the teacher enjoys the most. Gender is the explanatory variable and type of activity is the response variable.

Finish constructing the table below displaying conditional percentages for type of activity conditioned upon gender.

	Dance	Sports	Movies	TOTAL
Female	16/30 = 53%	***6/30=20%***	***8/30=27%***	**100%**
Male	2/20=10%	***10/20=50%***	***8/20=40%***	**100%**

8. Construct a side-by-side bar graph displaying the information from the table in part 5.

9. If a teacher is female, what percentage enjoys sports the most?

6/30 = 20%

10. If a teacher is male, what percentage enjoys sports the most?

10/20 = 50%

11. Of all the teachers surveyed, what percentage enjoys sports the most?

16/50 = 32%

12. Based on your answers to parts 6-8, does there appear to be an association between gender of teacher and type of activity enjoyed? Explain using your three percentages from parts 6-8.

Yes there appears to be some association between gender and type of activity for this group of teachers. If there were no association, we would expect the two conditional percentages in parts 9 and 10 to be the same (they are not at 20% and 50%). If there were no association, we would expect these two conditional percentages to be the marginal percentage of 32% in part 11.

13. Referring to parts 5 and 6, comment on how genders differ with respect to type of activity. Comment on any similarities.

Of the three activities, females overall prefer dance at 53% (males only prefer 10%) while overall males prefer sports at 50% (females only prefer 20%). Movies is the second most popular activity for both males and females but males prefer 40% to 27% for females.

Public Opinions and Leisure Time (Spotlight Task)

Name__________________________________ **Date**__________________

Mathematical Goals
- Practice summarizing categorical data from two-way frequency tables
- Interpret relative frequencies in the context of the data
- Recognize possible associations and trends in data

Essential Questions
- How do you interpret relative frequencies in the context of a two-way frequency table?

Georgia Standards of Excellence
MGSE9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

Standards of Mathematical Practice
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
6. Attend to precision.
7. Look for and make sense of structure.

Part I. A public opinion survey explored the association between age and support for increasing the minimum wage. Although age is a quantitative variable, the researchers create three categories of age groups and treated the age data as categorical. The results are found in the following two-way frequency table

	For	**Against**	**No Opinion**	**TOTAL**
Ages 21-40	25	20	5	50
Ages 41-60	30	30	15	75
Over 60	50	20	5	75
TOTAL	105	70	25	200

Frequency Count

1. What percentage of those individuals surveyed were in the 21 – 40 age group and for increasing the minimum wage? Show your work for finding this percentage. This percentage is called a joint percentage.

2. For the 21 to 40 age group, what percentage supports increasing the minimum wage? Explain how you arrived at your percentage. This percentage is called a conditional percentage.

3. For the 41 to 60 age group, what percentage supports increasing the minimum wage? Show your work.

4. For the over 60 age group, what percentage supports increasing the minimum wage? Show your work.

5. What percentage of all individuals surveyed favored increasing the minimum wage? Explain how you arrived at your percentage. This percentage is called a marginal percentage.

6. If there is no association between age group and opinion about increasing the minimum wage, we would expect the conditional percentages found in parts 2-4 to all be the same. We observe that the three percentages are different. What numerical value for the conditional percentage of each age group that favor increasing the minimum wage would we expect if there is no association between age group and opinion? Hint: Consider the percentage who favor increasing the minimum wage ignoring the age group classification of the individuals surveyed.

Part II. The table below gives the responses of 50 teachers' responses to a survey asking which activity they enjoyed most: dancing, playing/watching sports, or seeing movies. Is there an association between gender of the teacher and type of activity enjoyed the most?

	Dance	**Sports**	**Movies**	**TOTAL**
Female	16	6	8	**30**
Male	2	10	8	**20**
TOTAL	**18**	**16**	**16**	50

7. In exploring whether an association exists between gender and type of activity, we are interested in knowing if gender of the teacher helps predict the type of activity the teacher enjoys the most. Gender is the explanatory variable and type of activity is the response variable.

Finish constructing the table below displaying conditional percentages for type of activity conditioned upon gender.

	Dance	**Sports**	**Movies**	**TOTAL**
Female	16/30 = 53%			**100%**
Male	2/20=10%			**100%**

8. Construct a side-by-side bar graph displaying the information from the table in part 5.

9. If a teacher is female, what percentage enjoys sports the most?

10. If a teacher is male, what percentage enjoys sports the most?

11. Of all the teachers surveyed, what percentage enjoys sports the most?

12. Based on your answers to parts 6-8, does there appear to be an association between gender of teacher and type of activity enjoyed? Explain using your three percentages from parts 6-8.

13. Referring to parts 5 and 6, comment on how genders differ with respect to type of activity. Comment on any similarities.

If the Shoe Fits! (Scaffolding Task)

Back to Task Table

Introduction

In this activity, students explore and use hypothetical data collected on student shoe print lengths, height, and gender in order to help develop a tentative description of a person who entered a school's grounds over a weekend without permission. Graphs such as comparative box plots and scatter plots are drawn to illustrate the data. Measures of center (median, mean) and spread (range, Interquartile Range (IQR), Mean Absolute Deviation (MAD) are computed. Conclusions are drawn based upon the data analysis in the context of question(s) asked. Students will be able to calculate numerical summaries and use them to compare two data sets. Students will be able to determine if any data values are outliers, display data in comparative box plot and use the plot to compare two data sets, and display the relationship between two variables on a scatter plot and interpret the resulting plot.

Mathematical Goals

- Represent data with plots on the real number line
- Compare center and spread of two or more different data sets
- Interpret differences in shape, center, and spread in the context of data sets accounting for outliers
- Represent two quantitative variables on a scatter plot
- Describe how two quantitative variables are related

Essential Questions

- How do I summarize, represent, and interpret data on two categorical and quantitative variables?

Georgia Standards of Excellence

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

Standards for Mathematical Practice

4. Model with mathematics.
   Students use mathematical representations to compare men and women's shoe size and use them to make informed predictions about the likely "culprit."
6. Attend to precision.

The calculations for 5-number summary and MAD require students to be precise.

Background Knowledge

- Students will have knowledge of calculating numerical summaries for one variable: mean, median, Mean Absolute Deviation (MAD), five-number summary.
- Students will have knowledge of how to construct box plot and scatter plots. Students will have some familiarity with general concepts from linear regression.

Common Misconceptions

- Students sometimes forget to take the absolute value of the deviations for MAD.

Materials

- pencil
- graphing paper
- graphing calculator or statistical software package

Grouping

- Individual / Partner

Differentiation

Extension:

- Have students compare their own shoe size / height to the predicted value based on the line of best fit. Students can also research famous people's data (tallest man in the world, shortest woman in the world, etc.) to see how they compare to the prediction.

Intervention:

- Plotting men's and women's data on separate scatterplots (rather than plotting them on the same graph with different colors) may help students more easily identify the differences between their lines of best fit.
- Students might work with smaller data sets to simplify calculations. Interpretation / comparison of statistics can still be the central purpose of this task.

Formative Assessment Questions

- How did you determine what to label the horizontal and vertical axes of the scatterplot?
- Give another example of a situation where a random sample could be used to represent a population. Explain why a *random* sample is the best choice.

If the Shoe Fits – Teacher Notes

I. Formulate Question(s)
Begin the lesson by discussing a hypothetical background:
Welcome to CSI at School! Over the weekend, a student entered the school grounds without permission. Even though it appears that the culprit was just looking for a quiet place to study undisturbed by friends, school administrators are anxious to identify the offender and have asked for your help. The only available evidence is a suspicious footprint found outside the library door.
Ask students to write some questions that they would be interested in investigating about students' shoeprint lengths. Some possible questions might be:

1. ***What kinds of lengths do shoeprints have? What is the shortest shoe print length expected to be? What is the longest shoe print length expected to be?***
2. ***Are there differences in the shoe print lengths for males and females? If so, what are the differences?***
3. ***Are shoe print lengths related to any other variables?***

The investigation that follows is based upon some of the questions that might be posed by students and on attempting to answer the overall question of identifying the offender.

II. Design and Implement a Plan to Collect the Data
Explain the following hypothetical scenario for data collection: After the incident, school administrators arranged for data to be obtained from a random sample of this high school's students. The data table shows the shoe print length (in cm), height (in inches), and gender for each individual in the sample.

Shoe Print Length	Height	Gender	Shoe Print Length	Height	Gender
24	71	F	24.5	68.5	F
32	74	M	22.5	59	F
27	65	F	29	74	M
26	64	F	24.5	61	F
25.5	64	F	25	66	F
30	65	M	37	72	M
31	71	M	27	67	F
29.5	67	M	32.5	70	M
29	72	F	27	66	F
25	63	F	27.5	65	F
27.5	72	F	25	62	F
25.5	64	F	31	69	M
27	67	F	32	72	M
31	69	M	27.4	67	F
26	64	F	30	71	M
27	67	F	25	67	F
28	67	F	26.5	65.5	F
26.5	64	F	30	70	F
22.5	61	F	31	66	F
			27.25	67	F

1. Explain why this study was an observational study and not an experiment.

Students should note that in this context, nothing has been done deliberately to the students in order to measure their responses. From direct observation and measurement, data values were recorded for each student's height, shoe print length, and gender.

2. Why do you think the school's administrators chose to collect data on a random sample of students from the school? What benefit might a random sample offer?

What benefit might a random sample offer? Students should note that whenever possible, random selection should be used to choose samples for an observational study. By using random selection, chance determines which individuals are included in the sample. This helps ensure that a sample is representative of the population from which it was chosen. Random selection allows the researchers to generalize sample results to a larger population of interest.

Have students begin the investigation by performing data analyses to help determine the gender of the offender. By using appropriate graphs and numerical calculations students can compare shoe print lengths for males and females.

3. Suggest a graph that might be used to use to compare the shoe print length data distributions for females and males.

Comparative graphs such as dot plot or box plot are appropriate for displaying these data. Students describe one advantage of using comparative dot plot instead of comparative box plot to display these data.

4. Describe one advantage of using comparative box plots instead of comparative dot plots to display these data.

Comparative dot plot have the advantage of showing each individual data value while comparative box plot are useful for comparing the percentiles of the two distributions.

5. For each gender calculate the five-number summary for the shoe print lengths. Additionally, for each gender, determine if there are any outlying shoe print length values.

The five-number summaries of the shoe print length data are shown in the Table below.

	Minimum	*Quartile 1 (Q1)*	*Median (Q2)*	*Quartile 3 (Q3)*	*Maximum*
Male	*29*	*30*	*31*	*32*	*37*
Female	*22.5*	*25*	*26.5*	*27.4*	*31*

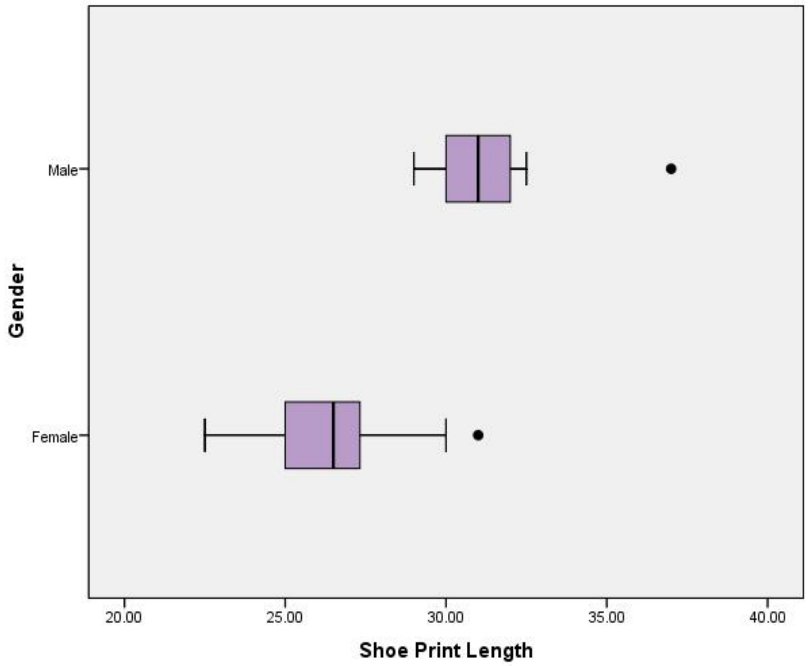
Note that the median is the halfway point in the data set. Also note that the first quartile is the median of the data points strictly below the median of the data set, and the third quartile is the median of the data points strictly above the median of the data set.
Demonstrate to students that in order to check for outlying shoe print lengths for the males:
(1) the interquartile range (IQR) is calculated: Q3 – Q1 = 32 – 30 = 2 cm
(2) the IQR is multiplied by 1.5: 1.5(2) = 3 cm
(3) 3 is subtracted from Q1: 30 – 3 = 27 cm
(4) 3 is added to Q3: 32 + 3 = 35 cm

Any shoe print length values smaller than 27 cm or greater than 35 cm are outlying values. There is one male student with an outlying shoe print length of 37 cm.
In order to check for outlying shoe print lengths for the females:
(1) the interquartile range (IQR) is calculated: Q3 – Q1 = 27.4 – 25 = 2.4 cm
(2) 1.5(IQR) is 1.5(2.4) = 3.6 cm
(3) 25 – 3.6 = 21.4 cm
(4) 27.4 + 3.6 = 31 cm

Any shoe print length values smaller than 21.4 cm or greater than 31 cm are outlying values. There is one female student with an outlying shoe print length of 31 cm.

6. Construct comparative box plots for the shoe print lengths of males and females. Discuss the similarities and differences in the shoe print length distributions for the males and females in this sample.

The Figure below illustrates the comparative box plot.



Discuss with students how to interpret the box plot. Students should understand that there are about the same number of shoe print lengths between the minimum and Q1, Q1 to Q2, Q2 to Q3, and Q3 to the maximum; or approximately 25% of the data will lie in each of these four intervals.

Ask students to describe the similarities and differences in the shoe print length distributions for the males and females in the sample. The box plots show that the females tended to have shorter shoe print lengths than the males. The median shoe print length for females was much lower than for males. The interquartile range for females is much lower than the interquartile range for males. All shoe print lengths for males are longer than at least 75% of the shoe print lengths for females. There is one high outlier in the female group and one high outlier in the male group.

7. For each gender calculate the mean shoe print length. What information does the mean shoe print length provide?

The mean and median shoe print lengths for males are 31.36 and 31 cm; respectively. For females, the mean and median lengths were 26.31 and 26.5 cm; respectively. As expected, it seems that a typical male shoe print length was longer than a typical female shoe print length by about 5 cm.

8. The mean will give us an indication of a typical shoe print length. In addition to knowing a typical length we would also like to know how much variability to expect around this length. For each gender calculate the Range; Interquartile Range; and Mean Absolute Deviation of the shoe print lengths. Interpret each of the calculated values.

For males the data cover the interval from the minimum of 29 cm to the maximum of 37 cm so the range in lengths is 8 cm. For females, the range in lengths is 8.5 cm. Thus, the ranges of shoe print lengths for males and females are very close.
In addition to the range the box plot suggests another measure of spread, the interquartile range (IQR). The IQR provides a measure of spread of the middle 50% of the shoe print lengths. The interquartile ranges for males and females are 2 cm and 2.4 cm, respectively. These IQRs are very comparable.
Yet another measure of spread can be calculated by incorporating how far the data are from the mean, on average. This measure, called the mean absolute deviation (MAD), is the arithmetic average of the absolute deviations of the data values from their mean. Students find the MAD using the following steps:

1. ***Find the deviations from the mean (by subtracting the mean from each shoe print length).***
2. ***Find the absolute value of each deviation.***
3. ***Find the mean of the absolute values.***

The table below shows the calculation of the MAD for the male shoe print lengths. For the males the MAD = 16.08/11 = 1.46 cm. In words, on average the male shoe print lengths are 1.46 cm away from the mean male shoe print length of 31.36 cm.

Shoe Print Length	***Length – Man***	***|Length – Mean|***
29	*29 – 31.36*	*2.36*
29.5	*29.5 – 31.36*	*1.86*
30	*30 – 31.36*	*1.36*
30	*30 – 31.36*	*1.36*
31	*31 – 31.36*	*.36*
31	*31 – 31.36*	*.36*
31	*31 – 31.36*	*.36*
32	*32 – 31.36*	*.64*
32	*32 – 31.36*	*.64*
32.5	*32.5 – 31.36*	*1.14*
37	*37 – 31.36*	*5.64*
	Sum ≈ 0	*Sum = 16.08*

Mention to students that the sum of the Length – Mean column is approximately zero and that the sum of the deviations from the mean for any data set will be zero. Ask students to explain why this sum is zero. The negative and positive deviations will 'cancel each other out'. This is the motivation for finding the mean absolute deviation.

A similar calculation for the female shoe print lengths gives the MAD = 42.03/28 = 1.50 cm. The female shoe print lengths are, on average, 1.50 cm away from the mean female shoe print length of 26.31 cm.

9. If the length of a student's shoe print was 32 cm, would you think that the print was made by a male or a female? How sure are you that you are correct? Explain your reasoning. Use results from Questions 5 through 8 in your explanation.

Students will say that a shoe print length of 32 cm is beyond the maximum shoe print length for the 28 females in the sample. Further the maximum female shoe print length of 31 cm was determined to be an outlier. It is highly unlikely that a shoe print 32 cm long was left by a female. Thus the offending student is most likely not a female. Note that the mean shoe print length for the 11 sampled males is 31.36 cm so a shoe print length of 32 cm long would not be unexpected for a male.

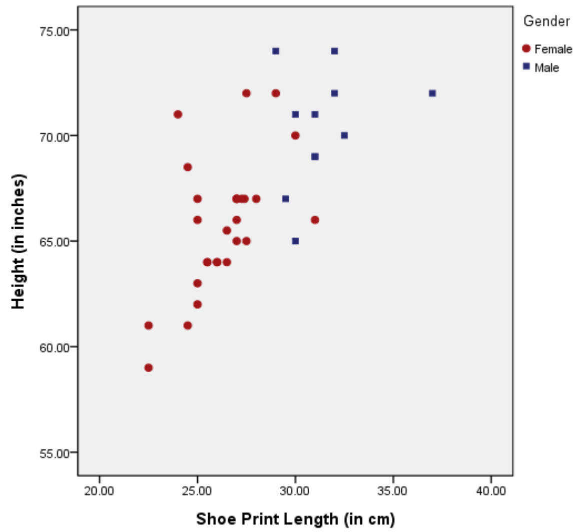
10. How would you answer Question 9 if the suspect's shoe print length was 27 cm?

Students will note that a shoe print length of 27 cm is 2 cm below the minimum shoe print length for the 11 sampled males. It is highly unlikely that a shoe print 27 cm long was left by a male. The offending student is most likely not a male. The mean shoe print length for the 28 sampled females is 26.31 cm with a MAD of 1.50 cm. Further; students will note that a shoe print length of 27 cm is only .4 cm below the third quartile of the shoe print lengths for females. We would expect close to 25% of female students to have a shoe print length of 27 cm or longer. It is much more likely for a female student to leave a shoe print 27 cm long than it is for a male student.

In the next phase of analysis have students explore the relationship between height (in inches) and length of shoe print (in cm). Discuss which variable should be the explanatory (independent) variable and which variable should be the response (dependent) variable. Here we would expect a student's shoe print length to help explain the height of the student (since the shoe print length of the offender is assumed to be known).

11. Construct a scatter plot of height (vertical scale) versus shoe print length (horizontal scale) using different colors or different plotting symbols to represent the data for males and females.

The Figure below illustrates the scatter plot.



(a) Interpret the scatter plot. Does it look like there is a linear relationship between height and shoe print length? Explain.

The scatter plot does indicate a somewhat strong, positive, linear association between shoe print length and height. We can see that as the shoe print length increases the height also tends to increase. Ask students to draw a line through the "center of the data".

(b) Does it look like the same straight line could be used to summarize the relationship between shoe print length and height for both males and females? Explain.

On the scatter plot the male heights and shoe print lengths appear to both have larger values than the values for the females. For both genders as was mentioned above the height tends to increase as the shoe print length increases. As far as the linear association within each gender it appears that the females have a larger increase in height for a unit (1 cm) increase in the shoe print length. For females it appears as though the height increases by about 1 inch for every 1 cm increase in shoe print length. For males it appears as though the height increases by about .5 inches for every 1 cm increase in shoe print length. Additionally, if students were asked to draw a line through the "center of the data" for males and then do the same for females; the y-intercept of the female line would be roughly 20 inches lower than the y-intercept of the male line.

(c) Based on the scatter plot, if a student's shoe print length was 30 cm, approximately what height would you predict for the person who made the shoe print? Explain how you arrived at your prediction.

Looking at the scatter plot if we locate 30 cm on the horizontal scale and we visualize having drawn a line through the center of the data, we see that for a 30 cm shoe print length the predicted student height is between 65 and 70 inches.

References

1. Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report, ASA, Franklin et al., ASA, 2007 http://www.amstat.org/education/gaise/

2. Adapted from an activity appearing in Making Sense of Statistical Studies by Peck and Starnes (with Kranendonk and Morita), ASA, 2009 http://www.amstat.org/education/msss

3. Adapted from the Investigation: How Long are our Shoes? In Bridging the Gap Between the Common Core State Standards and Teaching Statistics (2012, in press).
Authors: Pat Hopfensperger, Tim Jacobbe, Deborah Lurie, and Jerry Moreno.

4. CSI Stats: Helping Students Become Data Detectives with the GAISE Framework presented by Daren Starnes. Recorded web-based seminar.
http://www.amstat.org/education/webinars/

5. Assessment questions from: Mind on Statistics, Third Edition by Utts/Heckard, 2006. Cengage Learning.

Learning Task: If the Shoe Fits!

Name____________________________________ **Date**____________________

Mathematical Goals
- Represent data with plots on the real number line
- Compare center and spread of two or more different data sets
- Interpret differences in shape, center, and spread in the context of data sets accounting for outliers
- Represent two quantitative variables on a scatter plot
- Describe how two quantitative variables are related

Essential Questions
- How do I summarize, represent, and interpret data on two categorical and quantitative variables?

Georgia Standards of Excellence
MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

Standards for Mathematical Practice
4. Model with mathematics.
6. Attend to precision.

Learning Task: If the Shoe Fits!

Name________________________________ **Date**__________________

Shoe Print Length	Height	Gender
24	71	F
32	74	M
27	65	F
26	64	F
25.5	64	F
30	65	M
31	71	M
29.5	67	M
29	72	F
25	63	F
27.5	72	F
25.5	64	F
27	67	F
31	69	M
26	64	F
27	67	F
28	67	F
26.5	64	F
22.5	61	F
24.5	68.5	F
22.5	59	F
29	74	M
24.5	61	F
25	66	F
37	72	M
27	67	F
32.5	70	M
27	66	F
27.5	65	F
25	62	F
31	69	M
32	72	M
27.4	67	F
30	71	M
25	67	F
26.5	65.5	F
30	70	F
31	66	F
27.25	67	F

Welcome to CSI at School! Over the weekend, a student entered the school grounds without permission. Even though it appears that the culprit was just looking for a quiet place to study undisturbed by friends, school administrators are anxious to identify the offender and have asked for your help. The only available evidence is a suspicious footprint outside the library door.

After the incident, school administrators arranged for the data in the table below to be obtained from a random sample of this high school's students. The table shows the shoe print length (in cm), height (in inches), and gender for each individual in the sample.

1. Explain why this study was an observational study and not an experiment.

2. Why do you think the school's administrators chose to collect data on a random sample of students from the school? What benefit might a random sample offer?

3. Suggest a graph that might be used to use to compare the shoe print length data distributions for females and males.

4. Describe one advantage of using comparative box plots instead of comparative dot plots to display these data.

5. For each gender calculate the five-number summary for the shoe print lengths. Additionally, for each gender, determine if there are any outlying shoe print length values.

6. Construct comparative box plots for the shoe print lengths of males and females. Discuss the similarities and differences in the shoe print length distributions for the males and females in this sample.

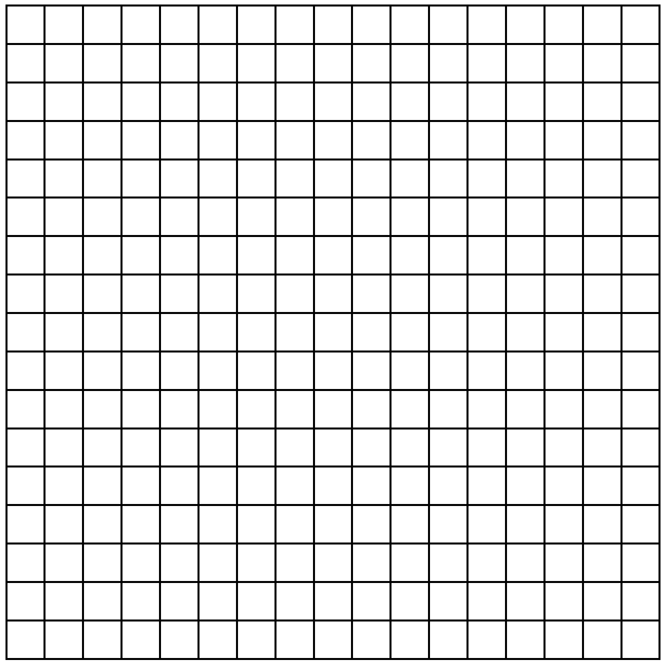
7. For each gender calculate the mean shoe print length. What information does the mean shoe print length provide?

8. The mean will give us an indication of a typical shoe print length. In addition to knowing a typical length we would also like to know how much variability to expect around this length. For each gender calculate the Range; Interquartile Range; and Mean Absolute Deviation of the shoe print lengths. Interpret each of the calculated values.

9. If the length of a student's shoe print was 32 cm, would you think that the print was made by a male or a female? How sure are you that you are correct? Explain your reasoning. Use results from Questions 5 through 8 in your explanation.

10. How would you answer Question 9 if the suspect's shoe print length was 27 cm?

11. Construct a scatter plot of height (vertical scale) versus shoe print length (horizontal scale) using different colors or different plotting symbols to represent the data for males and females.



a. Interpret the scatter plot. Does it look like there is a linear relationship between height and shoe print length? Explain.

b. Does it look like the same straight line could be used to summarize the relationship between shoe print length and height for both males and females? Explain.

c. Based on the scatter plot, if a student's shoe print length was 30 cm, approximately what height would you predict for the person who made the shoe print? Explain how you arrived at your prediction.

Spaghetti Regression (Learning Task):

Back to Task Table

Adapted from: *txcc.sedl.org/events/previous/092806/10ApplyingStrategies/math-teks-alg1.pdf*

Introduction
Students will investigate the concept of the "goodness-of-fit" and its significance in determining the regression line or best-fit line for the data. This is the first exploration in a series of three activities to explore a best- fit line. Fitting the graph of an equation to a data set is covered in all mathematics courses from Algebra to Calculus and beyond. The objective of this activity is to explore the concept in-depth.

In real life, functions arise from data gathered through observations or experiments. This data rarely falls neatly into a straight line or along a curve. There is variability in real data, and it is up to the student to find the function that best 'fits' the data. Regression, in its many facets, is probably the most widely used statistical methodology in existence. It is the basis of almost all modeling.

Students create scatter plots to develop an understanding of the relationships of bivariate data; this includes studying correlations and creating models from which they will predict and make critical judgments. As always, it is beneficial for students to generate their own data. This gives them ownership of the data and gives them insight into the process of collecting reliable data. Teachers should naturally encourage the students to discuss important concepts such as goodness-of fit. Using the graphing calculator facilitates this understanding. Students will be curious about how the linear functions are created, and this activity should help students develop this understanding.

Mathematical Goals
- To investigate the concept of goodness of fit in determining a line of best-fit

Essential Questions
- How do you determine the regression line or line of best fit for a scatter plot of data?

Georgia Standards of Excellence
MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

Standards for Mathematical Practice

2. Reason abstractly and quantitatively.
Students must determine where to place their spaghetti and compare their method to the methods of others.

Background Knowledge

- Students understand the idea of correlation and "eyeballing" lines of best fit.

Common Misconceptions

- Students may believe that a 45-degree line in a scatterplot always indicates a slope of 1, which is the case only when the two axes have the same scaling. Similarly, students may "count gridlines" for rise-over-run, but they should be careful about the scaling of each axis.
- Students often believe that a line of best fit must go *through* as many points as possible. The scatterplot students are working with is drawn intentionally to address this misconception.

Materials

- Spaghetti or linguine (3 pieces of spaghetti per student)
- Transparent tape (roll for each group)
- Transparencies of Overhead 1 Handouts – copy for each student of the Scatter plot,
- Student Activity: Spaghetti Regression

Grouping

- Small group

Formative Assessment Questions

- How did you decide where to put your line of best fit on the scatter plot?

Spaghetti Regression – Teacher Notes

Procedures	*Notes*
1. Activity 1 ***Introduce the topic of goodness of fit with Overhead 1.*** ***Ask: Why do we say that the line in the top graph fits the points better than the line in the bottom graph?*** ***Can we say that some other line might fit them better still?*** ***Say: Usually we think of a close fit as a good fit. But, what do we mean by close?***	***Discuss the importance of modeling and lead student discussions of concepts such as goodness-of-fit.***
2. Give each student 3 pieces of spaghetti, the Scatter plot handout, and Student Activity: Spaghetti Regression.	
3. Have the students examine the plots and visually determine a line of best-fit (or trend line) using a piece of spaghetti. They then tape the spaghetti line onto each of their graphs as described on the Student Activity handout.	***This should be done individually so that there is variation in the choice of lines within each group.***
4. Ask: Must the line of best fit go through any of the actual data points?	***Students often will think that the line of best fit must go through some (or most) of the actual data points.***
5. Ask: Can you write the equation of the line of best fit for each of your graphs? (Each axis has a scale of 1) ***Ask: What values do you need to know?*** ***Ask: How can you find those values?***	***Guide students to recognize that an equation in the form of y = mx + b will identify the line of best fit.*** ***Slope Formula:*** $m = \frac{y_2 - y_1}{x_2 - x_1}$ ***Use two points ON THE LINE (not necessarily a data point) to find the slope and then to solve for the y–intercept.***

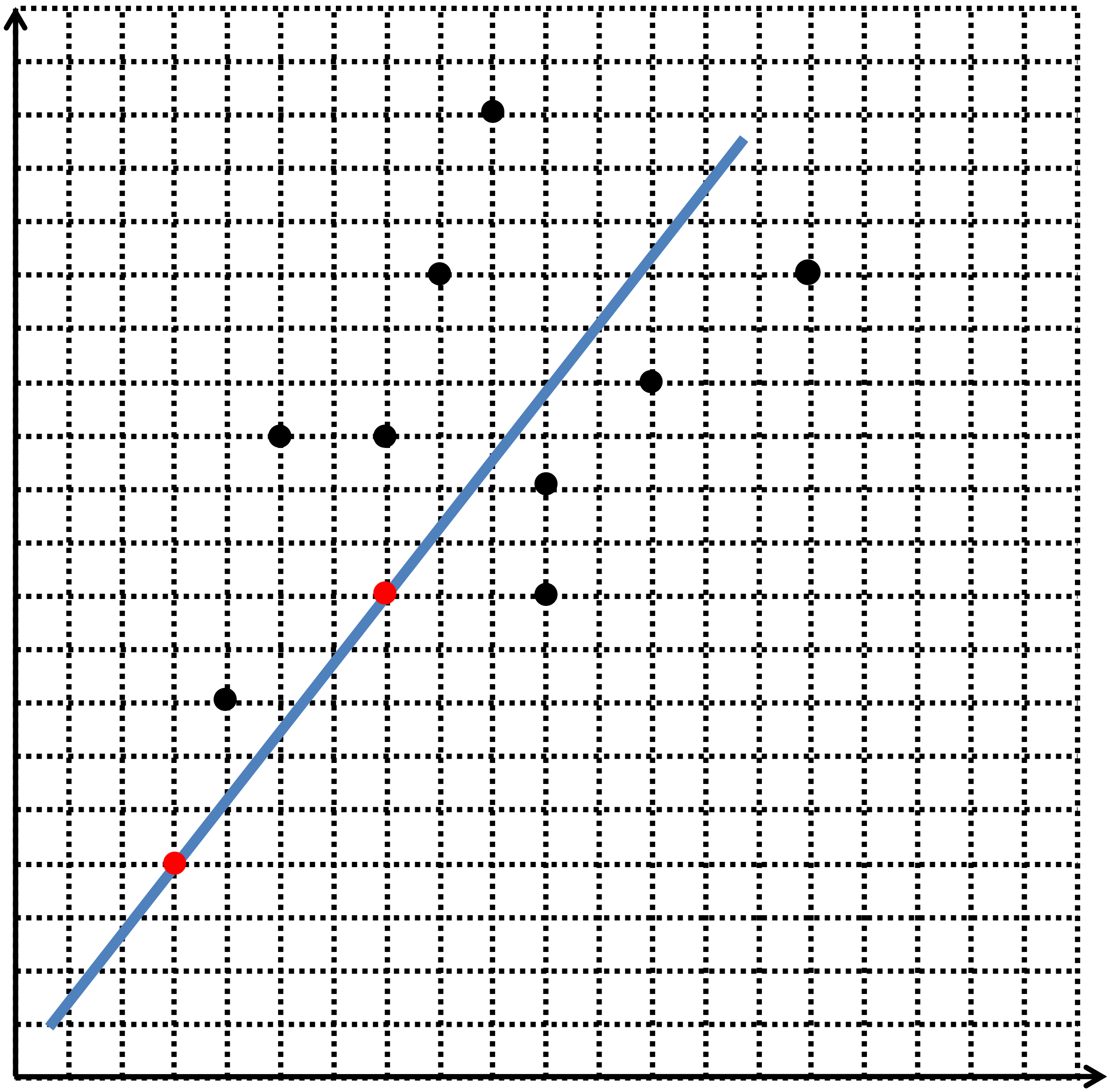
1. Examine the plot provided and visually determine a line of best-fit (or trend line) using a piece of spaghetti. Tape each spaghetti line onto your graph.

Answers will vary; one option is shown.

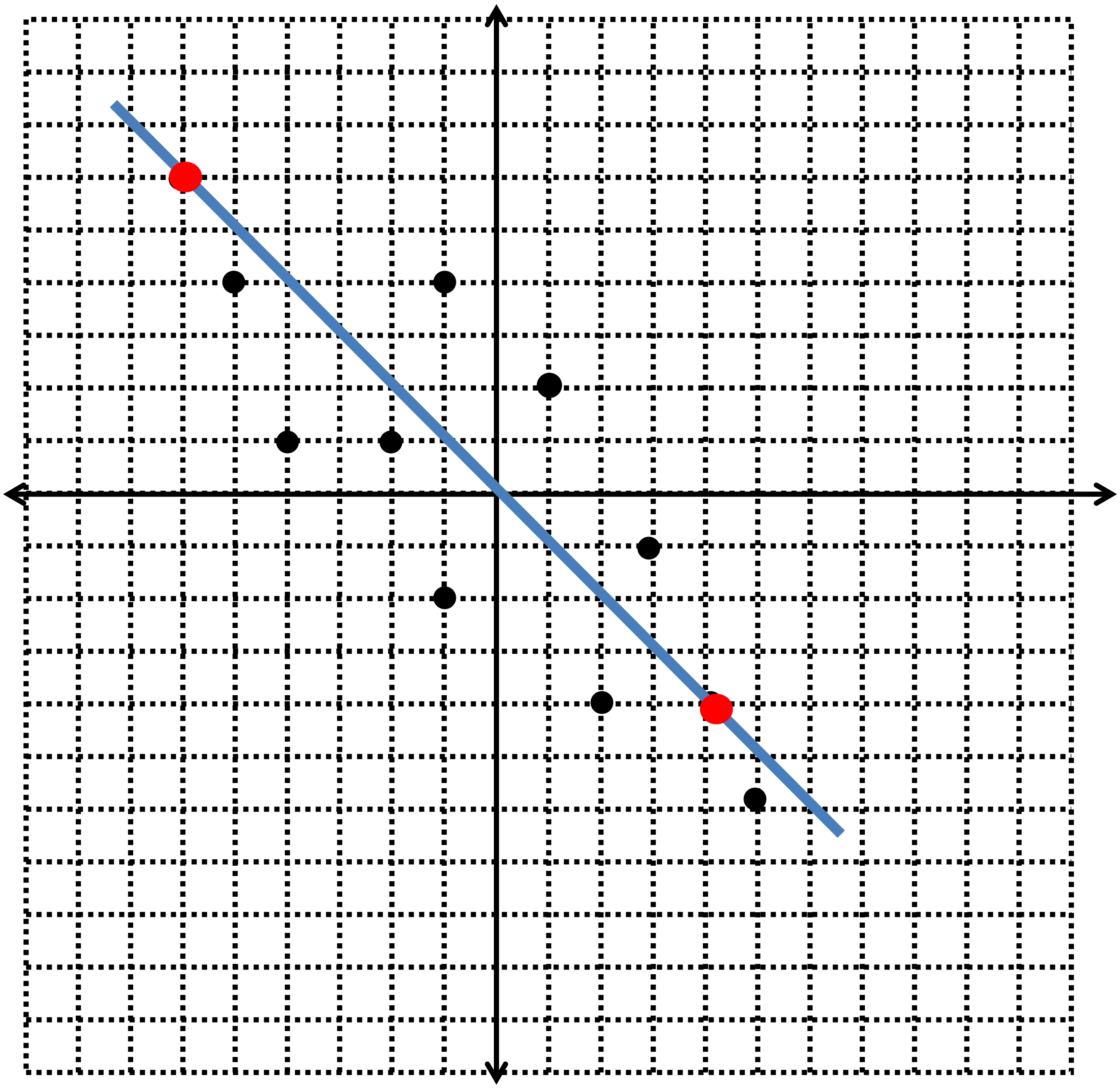
2. Write the equation of the line of best fit as identified by your spaghetti line.

Answers will vary; one option: $y = 1.25x - 0.25$

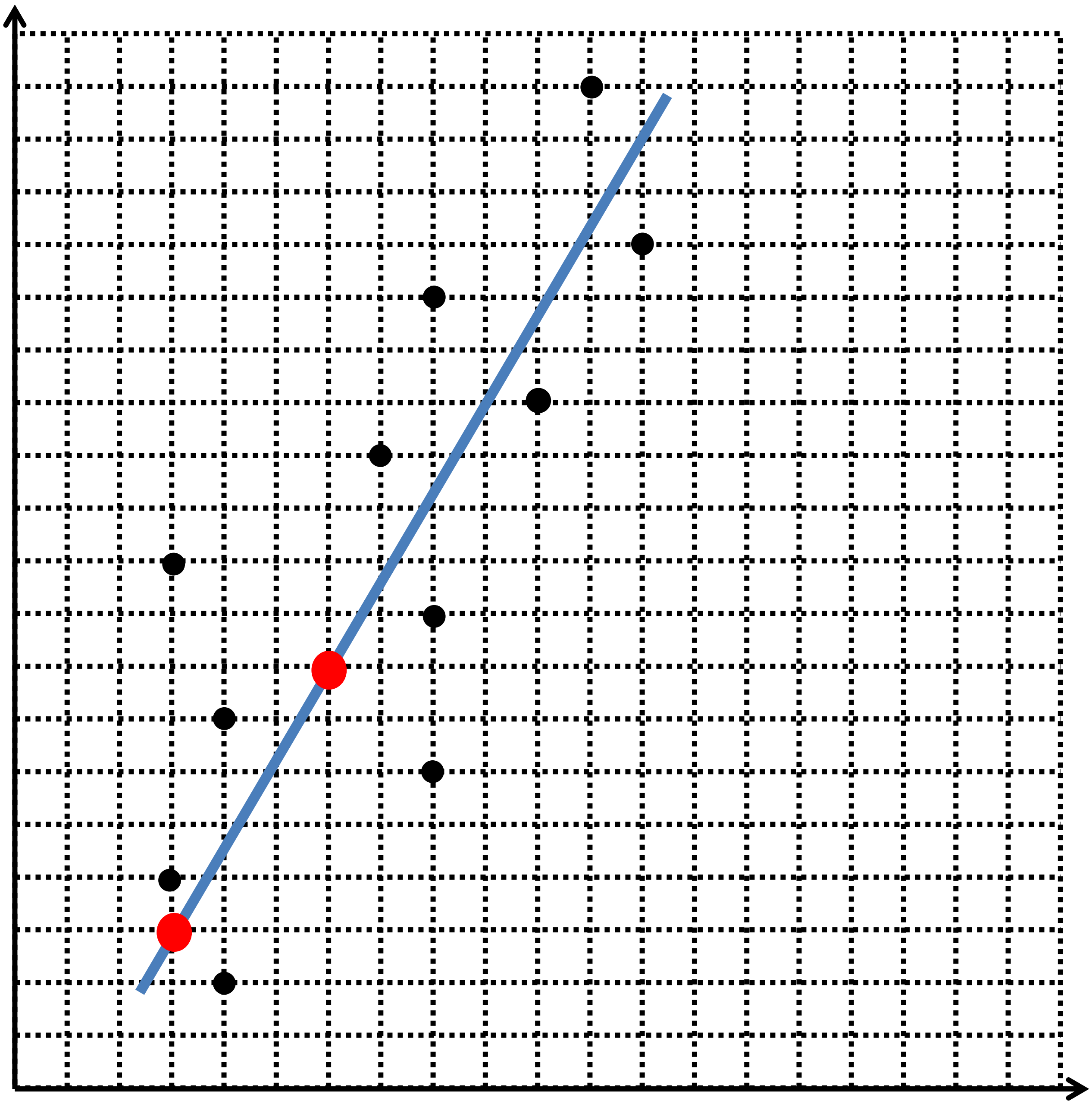
Scatter Plot 1



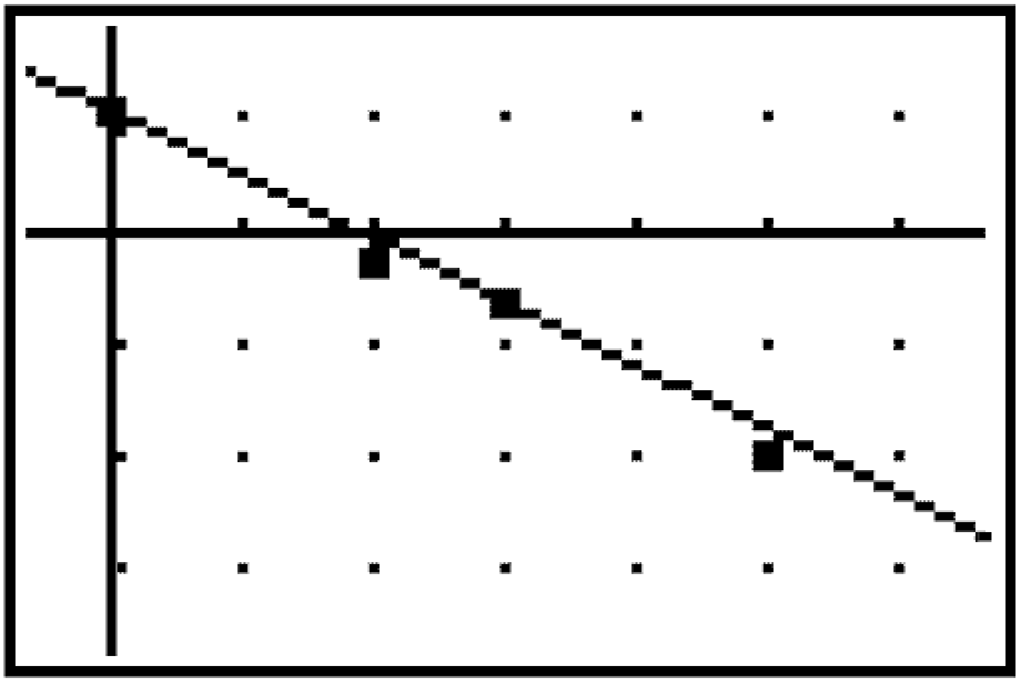
Scatter Plot 2 *Answers vary; one option is shown.* $y = -x$

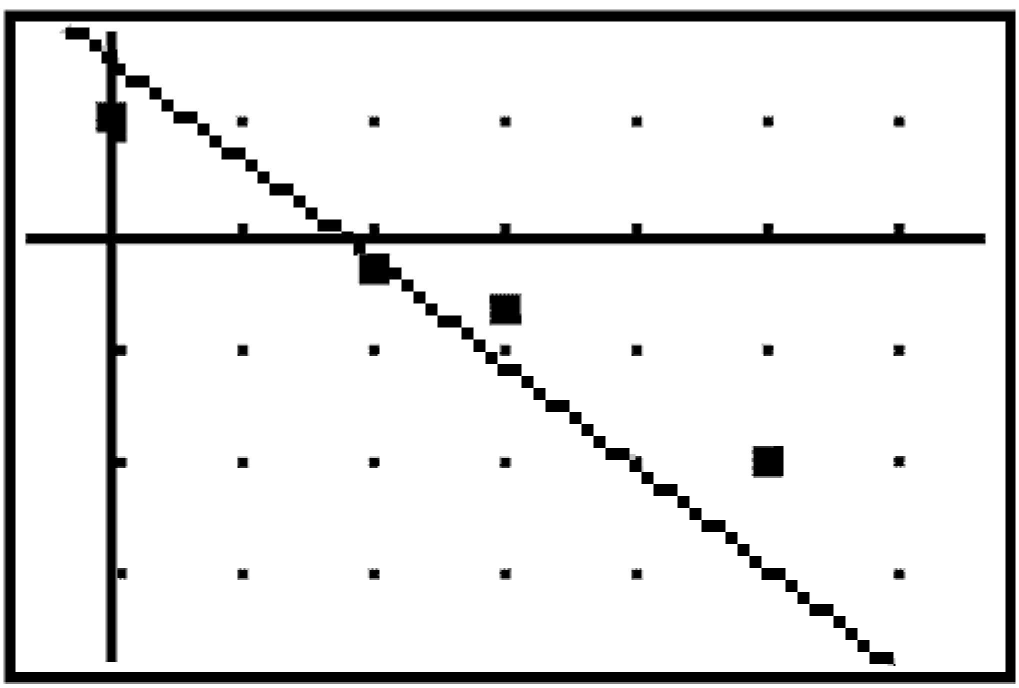


Scatter Plot 3 ***Answers vary; one option is shown.*** $y = \frac{5}{3}x - 2$



Overhead 1





Group Learning Task: Spaghetti Regression

Name__________________________________ **Date**___________________

Mathematical Goals

- To investigate the concept of goodness of fit in determining a line of best-fit

Essential Questions

- How do you determine the regression line or line of best fit for a scatter plot of data?

Georgia Standards of Excellence

MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

Standards for Mathematical Practice

2. Reason abstractly and quantitatively.
5. Use appropriate tools strategically.

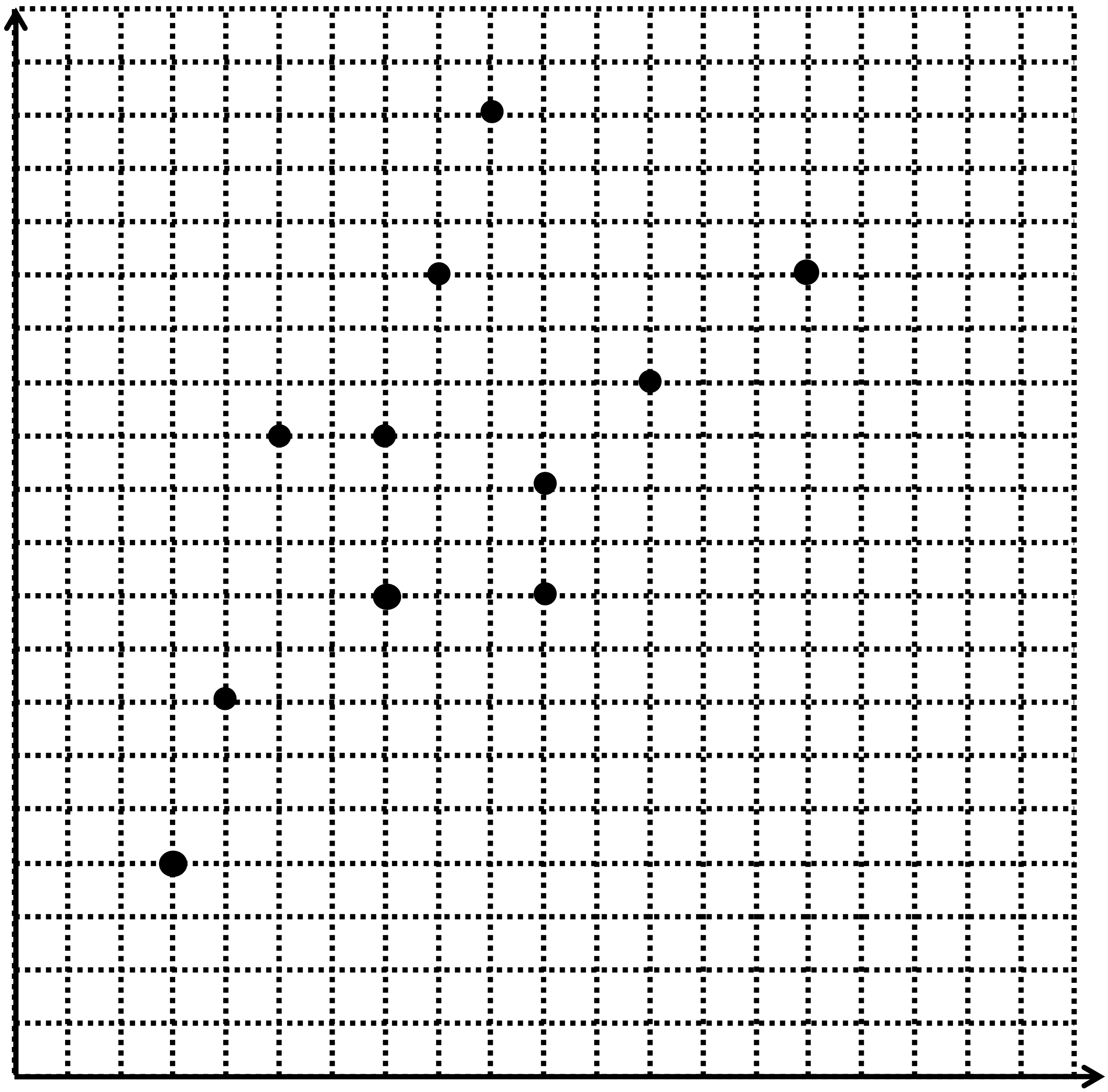
Group Learning Task: Spaghetti Regression

Name__________________________________ **Date**___________________

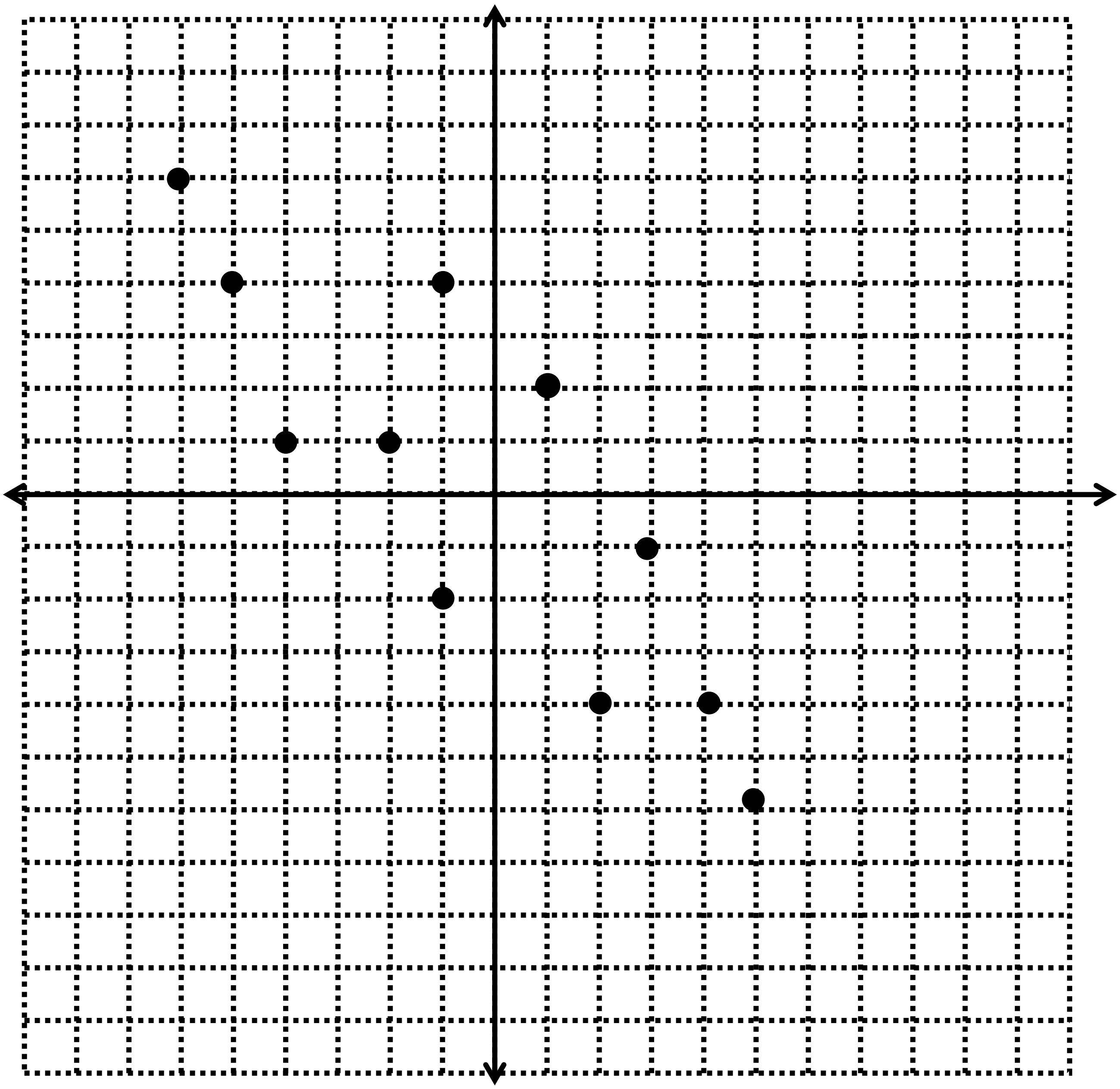
1. Examine the plots provided and visually determine a line of best-fit (or trend line) using a piece of spaghetti. Tape your spaghetti line onto each graph.

2. Write the equation of the line of best fit as identified by your spaghetti line.

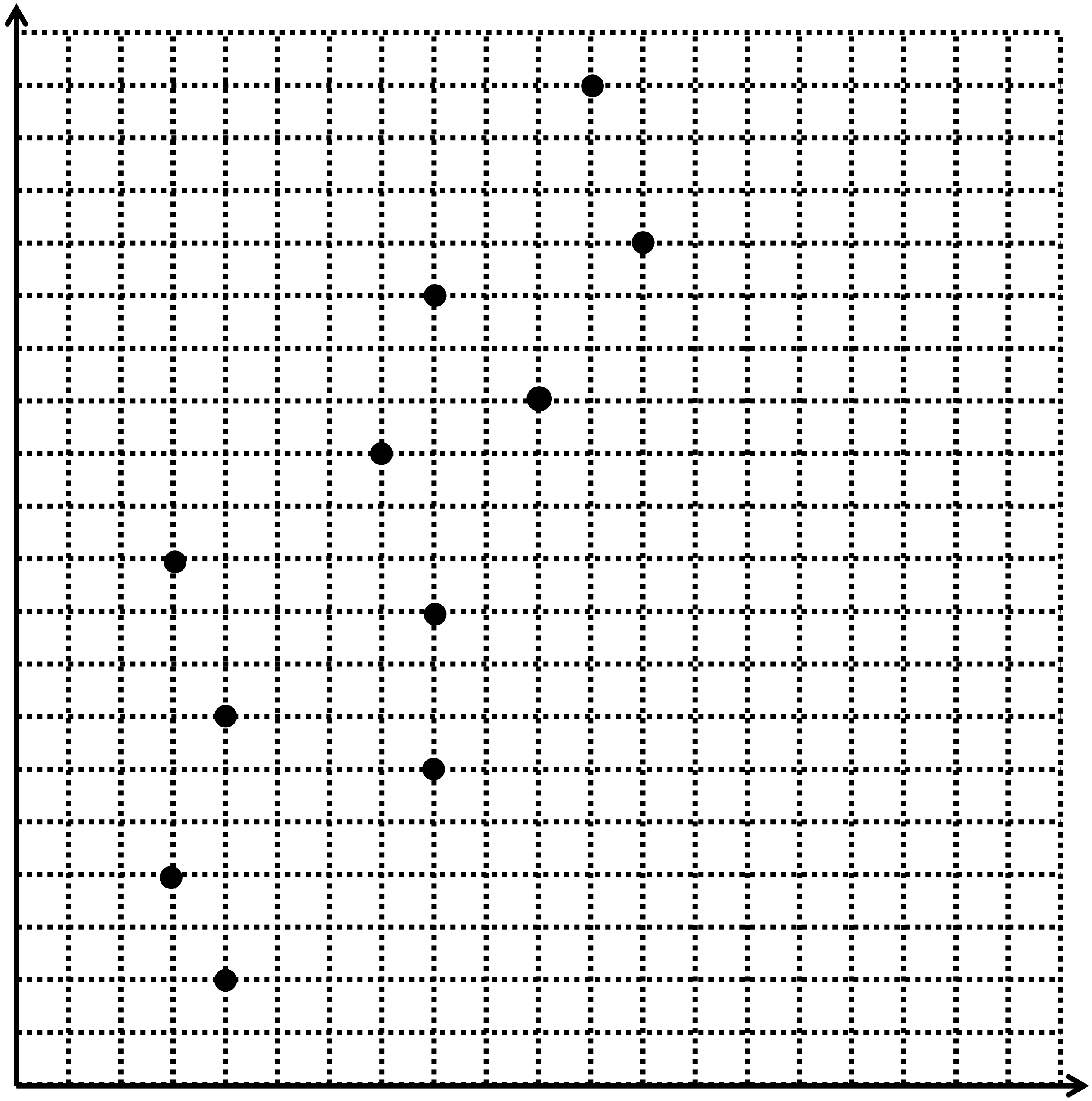
Scatter Plot 1



Scatter Plot 2



Scatter Plot 3



Formative Assessment Lesson: Devising a Measure for Correlation

Back to Task Table

Source: Formative Assessment Lesson Materials from Mathematics Assessment Project
http://map.mathshell.org/materials/download.php?fileid=1234

Task Comments and Introduction
Tasks and lessons from the Mathematics Assessment Project are specifically designed to help teachers effectively formatively assess their students. The way the tasks and lessons are designed gives the teacher a clear understanding of what the students are able to do and not do. Within the lesson, teachers will find suggestions and question prompts that will help guide students towards understanding. For more information access the MAP website:
http://www.map.mathshell.org/materials/background.php?subpage=formative

The task, *Devising a Measure for Correlation*, is a Formative Assessment Lesson (FAL) that can be found at the website:
http://map.mathshell.org/materials/lessons.php?taskid=420&subpage=problem

The FAL document provides a clear lesson design, from the opening of the lesson to the closing of the lesson.

The PDF version of the task can be found at the link below:
http://map.mathshell.org/materials/download.php?fileid=1234

Mathematical Goals
- Understand correlation as the degree of fit between two variables.
- Test and improve a mathematical model, and evaluate alternative models.

Essential Questions
- What is correlation, and what are reasonable ways to measure it?

Georgia Standards of Excellence
MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6a Decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function of best fit. Emphasize linear, quadratic, and exponential models.

MGSE9-12.S.ID.6c Using given or collected bivariate data, fit a linear function for a scatter plot that suggests a linear association.

Standards for Mathematical Practice

2. Reason abstractly and quantitatively.
   Students must make sense of the meaning of various potential ways to measure goodness of fit and calculate using their strategy.
3. Construct viable arguments and critique the reasoning of others.
   Students devise their own method to quantify the quality of fit.
4. Model with mathematics.
   Students use mathematical models to quantify correlation.

Background Knowledge

- Students can approximate the strength and direction of correlation between two variables.

Common Misconceptions

- Students often believe that a line of best fit must go *through* as many points as possible, rather than simply being a good fit for the overall trend.
- Students may believe a negative correlation indicates weak correlation, when in fact 0 indicates no correlation and +1 and -1 each indicate perfect correlation.

Materials

- See FAL website.

Grouping

- Individual / small group

TV / Test Grades (Learning Task)

Back to Task Table

Introduction
Before beginning the task, ask the class what they know about correlation. Remind them that the correlation coefficient, a measure of how closely two variables are related, is a number between –1 and 1. Determining a relationship between two sets of data, especially from a scatter plot, may be subject to interpretation. The teacher will likely want to have students use a graphing calculator with statistical capabilities to do this task, determining ahead of time which features on the calculator are appropriate.

Lines of good fit may be found using paper-and-pencil techniques (such as writing the equation based on two points) or using a graphing calculator (either generating possible lines to use for guessing and checking or using the regression feature of the calculator to determine a particular function rule). Discuss correlation and causation with the group. Ask them at the end of the task to summarize television watching and test grades and if they believe there is a causal relationship. Have them defend their position based on statistical analysis.

Mathematical Goals

- Represent data on a scatter plot
- Describe how two variables are related
- Fit a linear function for a scatter plot that suggests a linear association

Essential Questions

- How do you determine the regression line or line of best fit for a scatter plot of data?

Georgia Standards of Excellence
MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6a Decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function of best fit. Emphasize linear, quadratic, and exponential models.

MGSE9-12.S.ID.6c Using given or collected bivariate data, fit a linear function for a scatter plot that suggests a linear association.

MGSE9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MGSE9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient "r" of a linear fit. (For instance, by looking at a scatterplot, students should be able to tell if the correlation coefficient is positive or negative and give a reasonable estimate of the "r" value.) After calculating the line of best fit using technology, students should be able to describe how strong the goodness of fit of the regression is, using "r."

MGSE9-12.S.ID.9 Distinguish between correlation and causation.

Standards for Mathematical Practice

3. Construct viable arguments and critique the reasoning of others.
   Students must predict whether each correlation is positive or negative, with justification.
5. Use appropriate tools strategically.
   Students may use graphing calculators to perform linear and exponential regression.

Background Knowledge

- Students must have knowledge of writing linear equations based on two points and understand correlation.

Common Misconceptions

- Students may believe that correlation implies causation.
- Asking students to graph "[A] vs. [B]" means that A is the dependent variable and B is the independent variable. In other words, "[A] vs. [B]" should be treated as "y vs. x" NOT as "x vs. y"
- Students may believe that a 45-degree line in a scatterplot always indicates a slope of 1, which is the case only when the two axes have the same scaling. Similarly, students may "count gridlines" for rise-over-run, but they should be careful about the scaling of each axis.

Materials

- pencil
- graphing paper
- graphing calculator or statistical software package

Grouping

- Partners / small group

Differentiation

Extension:

- Name another pair of variables that are likely to have a negative correlation without a reason to expect causation.
- Find real-life examples of situations where people prove correlation and expect you to believe there is causation. (Look at advertisements, political speeches, etc.)

Intervention:

- Students may want to actually graph some (or all) of the data points in #2 to decide which pairs of variables have positive/negative/no correlation. Be sure to push students to go back to the table to see how they could have made these predictions numerically without taking the time to graph points to make the prediction graphically.

Formative Assessment Questions

- When you notice that two variables are correlated, does this mean a change in one causes a change in the other?
- If you swap the *x*- and *y*-axes, would the strength of the correlation change? Would the equation of the line of best fit change?

TV / Test Grade – Teacher Notes

1. Students in Ms. Garth's Algebra II class wanted to see if there are correlations between test scores and height and between test scores and time spent watching television. Before the students began collecting data, Ms. Garth asked them to predict what the data would reveal. Answer the following questions that Ms. Garth asked her class.

   a. Do you think students' heights will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation?

   Answers may vary, but a possible answer could be: "I do not think there will be correlation between height and test grades, since it is not reasonable to think a person's height is related to his or her intelligence or effort level."

   b. Do you think the average number of hours students watch television per week will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation? Do watching TV and low test grades have a cause and effect relationship?

   Answers may vary, but a possible answer could be: "I think the average number of hours a student watches television will be negatively correlated with the student's test grades. It is reasonable to think that the more TV you watch, the less time you spend studying, resulting in low test grades. However, it does not seem like these variables will be strongly correlated, since some people do not watch TV but do not spend time studying either. On the other hand, some students may watch a lot of TV and still study a lot." Discuss correlation vs. causation with students. Give samples of variables that correlate and have them justify their argument.

2. The students then created a table in which they recorded each student's height, average number of hours per week spent watching television (measured over a four-week period), and scores on two tests. Use the actual data collected by the students in Ms. Garth's class, as shown in the table below, to answer the following questions.

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Height (in inches)	60	65	51	76	66	72	59	58	70	67	65	71	58
TV hrs/week (average)	30	12	30	20	10	20	15	12	15	11	16	20	19
Test 1	60	80	65	85	100	78	75	95	75	90	90	80	75
Test 2	70	85	75	85	100	88	85	90	90	90	95	85	85

a. Which pairs of variables seem to have a positive correlation? Explain.

Test 1 scores and test 2 scores appear to be positively correlated. For the most part, student performance on both tests was fairly consistent, so students who did well on test 1 also did well on test 2, while those who did not do well on test 1 didn't do very well on test 2 either.

b. Which pairs of variables seem to have a negative correlation? Explain.

Test 1 scores and hours per week watching television, and test 2 scores and hours per week watching television appear to be negatively correlated. In general, students who spent more time watching television had lower test scores than those who spent less time watching television.

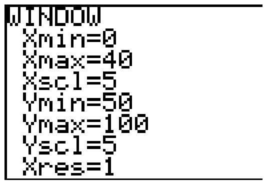
c. Which pairs of variables seem to have no correlation? Explain.

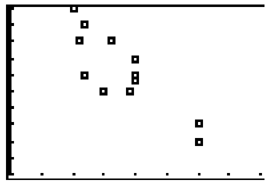
Height and hours per week watching television, test 1 scores and height, and test 2 scores and height seem to have no correlation. Height does not seem to be correlated with any of the other variables. That is, taller students do not seem to watch any more or less television or perform any better or worse on tests than shorter students.

3. For each pair of variables listed below, create a scatter plot with the first variable shown on the y-axis and the second variable on the x-axis. Are the two variables correlated positively, correlated negatively, or not correlated? Determine whether each scatter plot suggests a linear trend.

a. Score on test 1 versus hours watching television

Scatter Plot:

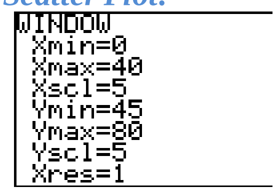


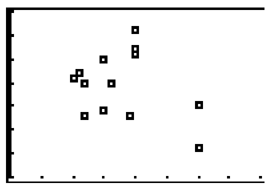


Correlation? Negative. Linear Trend? Yes.

b. Height versus hours watching television

Scatter Plot:

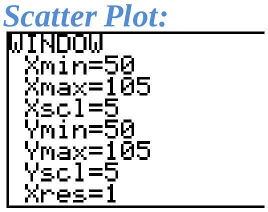


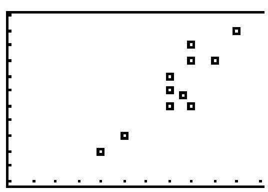


Correlation? No correlation. Linear Trend? No.

c. Score on test 1 vs. score on test 2

Scatter Plot:

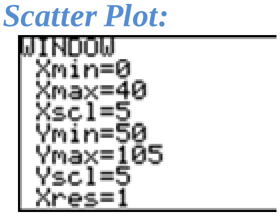


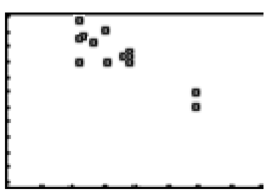


Correlation? Positive. Linear Trend? Yes.

d. Hours watching television versus score on test 2

Scatter Plot:





Correlation? Negative. Linear Trend? Yes.

4. Using the statistical functions of your graphing calculator, determine a line of good fit for each scatter plot that suggests a linear trend.

Answers may vary slightly from the ones shown here.

Using linear regression and rounding to the hundredths place:

a. Score on Test 1 versus hours watching television: ***$y = -1.43x + 105.98$***
b. Height versus hours watching television: ***no linear trend***
c. Score on test 1 versus score on test 2: ***$y = 1.32x – 33.04$***
d. Hours watching television versus score on test 2: ***$y = -0.72x + 79.64$***

Alternatively, using two points that appear to be close to a good representation of the trend in the data:

Data from score on test 1 versus hours spent watching television: (20, 78) and (11, 90)

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{90 - 78}{11 - 20} = \frac{12}{-9} = \frac{-4}{3} = -1.33$$

$$y - y_1 = m(x - x_1)$$

$$y - 90 = \frac{-4}{3}(x - 11)$$

$$y - 90 = \frac{-4}{3}x + \frac{44}{3}$$

$$y = \frac{-4}{3}x + \frac{314}{3}$$

$$y = -1.33x + 104.6$$

Guided Learning Task: TV/Test Grades

Name_________________________________ **Date**__________________

Mathematical Goals

- Represent data on a scatter plot
- Describe how two variables are related
- Fit a linear function for a scatter plot that suggests a linear association

Essential Questions

- How do you determine the regression line or line of best fit for a scatter plot of data?

Georgia Standards of Excellence

MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6a Decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function of best fit. Emphasize linear, quadratic, and exponential models.

MGSE9-12.S.ID.6c Using given or collected bivariate data, fit a linear function for a scatter plot that suggests a linear association.

MGSE9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MGSE9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient "r" of a linear fit. (For instance, by looking at a scatterplot, students should be able to tell if the correlation coefficient is positive or negative and give a reasonable estimate of the "r" value.) After calculating the line of best fit using technology, students should be able to describe how strong the goodness of fit of the regression is, using "r."

MGSE9-12.S.ID.9 Distinguish between correlation and causation.

Standards for Mathematical Practice

3. Construct viable arguments and critique the reasoning of others.
5. Use appropriate tools strategically.

Guided Learning Task: TV/Test Grades

Name__________________________________ **Date**__________________

1. Students in Ms. Garth's Algebra II class wanted to see if there are correlations between test scores and height and between test scores and time spent watching television. Before the students began collecting data, Ms. Garth asked them to predict what the data would reveal. Answer the following questions that Ms. Garth asked her class.

a. Do you think students' heights will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation?

b. Do you think the average number of hours students watch television per week will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation? Do watching TV and low test grades have a cause and effect relationship?

2. The students then created a table in which they recorded each student's height, average number of hours per week spent watching television (measured over a four-week period), and scores on two tests. Use the actual data collected by the students in Ms. Garth's class, as shown in the table below, to answer the following questions.

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Height (in inches)	60	65	51	76	66	72	59	58	70	67	65	71	58
TV hrs/week (average)	30	12	30	20	10	20	15	12	15	11	16	20	19
Test 1	60	80	65	85	100	78	75	95	75	90	90	80	75
Test 2	70	85	75	85	100	88	85	90	90	90	95	85	85

a. Which pairs of variables seem to have a positive correlation? Explain.

b. Which pairs of variables seem to have a negative correlation? Explain.

c. Which pairs of variables seem to have no correlation? Explain.

3. For each pair of variables listed below, create a scatter plot with the first variable shown on the *y*-axis and the second variable on the *x*-axis. Are the two variables correlated positively, correlated negatively, or not correlated? Determine whether each scatter plot suggests a linear trend.

a. Score on test 1 versus hours watching television

b. Height versus hours watching television

c. Score on test 1 versus score on test 2

d. Score on test 2 vs. hours watching television

4. Using the statistical functions of your graphing calculator, determine a line of good fit for each scatter plot that suggests a linear trend.

Equal Salaries for Equal Work? (Performance Task)

Back to Task Table

Introduction

This task asks students to compare additive and multiplicative growth (represented by linear and exponential models) to make predictions and solve problems within the context of gender-based salary differences. In doing this task, students analyze data sets, create scatter plots, determine the most appropriate mathematical model, and justify their model selection.

This task provides a good example of how data points can appear to be linear over a relatively small domain, but how a different type of mathematical model might be more appropriate over a larger domain. This is an opportunity for students to discuss strengths and limitations of using mathematical functions to model real data. One discussion might arise as to whether other types of mathematical functions might sometimes be used for different types of data, perhaps leading students to look for patterns in data they might gather from sources like newspapers or books of world records.

Note that students will need to make a decision about the initial value representing the year. For example, it would be reasonable to assign the year 1984 (the first year in the table) as Year 0. The sample solutions below are based on this assumption.

Mathematical Goals

- Represent data on a scatter plot
- Describe how two variables are related
- Fit a linear function for a scatter plot that suggests a linear association

Essential Questions

- How do you determine if linear or exponential regression is more appropriate for a scatter plot?

Georgia Standards of Excellence

MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6a Decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function of best fit. Emphasize linear, quadratic, and exponential models.

MGSE9-12.S.ID.6c Using given or collected bivariate data, fit a linear function for a scatter plot that suggests a linear association.

MGSE9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MGSE9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient "r" of a linear fit. (For instance, by looking at a scatterplot, students should be able to tell if the correlation coefficient is positive or negative and give a reasonable estimate of the "r" value.) After calculating the line of best fit using technology, students should be able to describe how strong the goodness of fit of the regression is, using "r."

Standards for Mathematical Practice

3. Construct viable arguments and critique the reasoning of others.
   Students use their models to make and defend informed predictions about relative pay of men vs. women, and compare the use of linear vs. exponential models.
4. Model with mathematics.
   Students use linear and exponential functions to model median pay for men and women to make informed predictions about trends in salaries.

Background Knowledge

- Students must have knowledge of using the graphing calculator to create linear and exponential models.
- Students understand how to assess the fit of a function to data.
- Students know how to choose a function suggested by context.

Common Misconceptions

- Students may believe that correlation implies causation.
- Students may believe that a 45-degree line in a scatterplot always indicates a slope of 1, which is the case only when the two axes have the same scaling. Similarly, students may "count gridlines" for rise-over-run, but they should be careful about the scaling of each axis.

Materials

- Pencil
- Paper / graphing paper
- Graphing calculator or statistical software package

Grouping

- Partner / individual

Differentiation

Extension:

- Write a letter to a representative expressing your findings mathematically. If you wish, you may take a position on the issue and use mathematical reasoning to persuade the representative. You might want to collect more recent data before writing your letter.

Intervention:

- Calculating an exponential model by hand can be challenging (finding a good estimate for the multiplier can be difficult), so graphing calculators might be helpful so students can focus on the main purpose of the task—comparing linear/exponential models and interpreting these models in the context of the data.

Formative Assessment Questions

- How do you create an exponential model for a scatter plot?
- What are the advantages and disadvantages of the linear and exponential models for these data sets? Which do you think is more reasonable?

<u>Equal Salaries for Equal Work? – Teacher Notes</u>

The data table shows the annual median earnings for female and male workers in the United States from 1984 to 2004. Use the data table to complete the task. Answer all questions in depth to show your understanding of the standards.

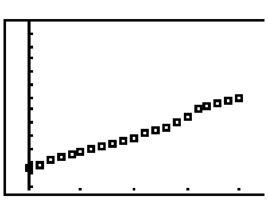
Year	Women's median earnings (in dollars)	Men's median earnings (in dollars)
1984	8,675	17,026
1985	9,328	17,779
1986	10,016	18,782
1987	10,619	19,818
1988	11,096	20,612
1989	11,736	21,376
1990	12,250	21,522
1991	12,884	21,857
1992	13,527	21,903
1993	13,896	22,443
1994	14,323	23,656
1995	15,322	25,018
1996	16,028	25,785
1997	16,716	26,843
1998	17,716	28,755
1999	18,440	30,079
2000	20,267	30,951
2001	20,851	31,364
2002	21,429	31,647
2003	22,004	32,048
2004	22,256	32,483

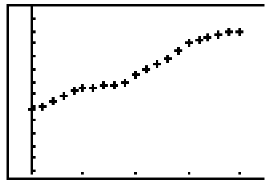
Data provided by U.S. Census Bureau

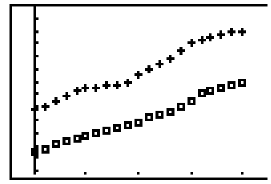
1. Create two scatter plots, one for women's median earnings over time and one for men's median earnings over time. Describe two things you notice about the scatter plots.

   Each scatter plot below is graphed with the following window:

   Window:
   Xmin=(-)2
   X max=22
   Xscl=5
   Ymin=4500
   Ymax=37000
   Yscl=2500
   Xres=1







Women's Data ***Men's Data*** ***Both Data Sets***

Answers may vary.

Possible answers: From 1984 to 2004, median earnings for both men and women increased. In each of these years, men's median earnings were greater than women's median earnings.

2. Terry and Tomas are trying to decide what type of model will most accurately represent the data. Terry thinks that a linear model might be most appropriate for each scatter plot. Help Terry find reasonable linear function rules for each scatter plot. Explain how you found these.

Answers may vary
One solution method:

To find a linear model of women's median earnings, use the starting earnings figure for women, \$8675, and the average rate of change of \$680 per year. (To find the average rate of change, find successive differences and then find the average of the successive differences.) The linear model is $m(x) = 680x + 8675$, where x represents years and $m(x)$ gives the median earnings. To find a linear model of men's median earnings, use the starting earnings figure for men, \$17,026, and the average rate of change of \$773 per year. The linear model is $m(x)= 773x + 17026$, where x represents years and $m(x)$ gives the median earnings.

Another solution method:

Using a graphing calculator to determine a regression line, women's median earnings could be represented by the function $y =703x + 8181$.

Using a graphing calculator to determine a regression line, men's median earnings could be represented by the function $y = 814x + 16709$.

3. Using the linear models, will women's annual median earnings ever equal those of men? Why or why not?

Using the linear models created from the data provided, women's annual median earnings will never equal men's annual median earnings. The men's linear model has a larger y- intercept and a larger slope, meaning the men start out earning more money and also experience a faster rate of increase in earnings.

4. Tomas thinks that an exponential model might be most appropriate for each scatter plot. Help Tomas find reasonable exponential function rules for each scatter plot. Explain how you found these.

Answers may vary.

One solution method:
To find an exponential model of women's median earnings, use the starting income for women, \$8675, and the multiplier, 1.048. (To estimate the multiplier, one possible—though imperfect—method is to find successive quotients then find the average of the successive quotients.) The exponential model would be $m(x) = 8675(1.048)^x$, where x represents years and m(x) gives the median salary. To find an exponential model of men's median earnings, use the starting earnings figure for men, \$17,026, and the average quotient, 1.033. The exponential model is $m(x) = 17026(1.033)^x$, where x represents years and m(x) gives the median earnings.

Another solution method:

Calculating an exponential regression function on a graphing calculator, women's median earnings could be represented by the function $y = 9087(1.049)^x$ and men's median earnings could be represented by the function $y = 17479(1.034)^x$

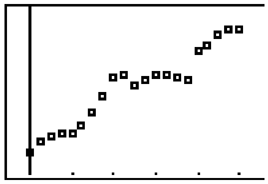
5. Using the exponential models, will women's annual median earnings ever equal those of men? Why or why not?

   Using the exponential models, women's annual median earnings will eventually equal those of men. The exponential model of men's earnings has a base of 1.034, and the exponential model of women's earnings has a base of 1.049. Since the women's model has a higher base, their earnings are increasing at a faster rate and will eventually surpass men's earnings. These functions can also be graphed to determine their intersection (45.7, 79533.88), demonstrating that at some point during the year 2029, women's annual median earnings will overtake men's annual median earnings.

6. If you answered yes to either question 3 or question 5, use that model to determine the first year women will have higher median earnings than men. Explain how you found your answer.

   Using the exponential models, women's annual median earnings will eventually equal those of men. The exponential model of men's earnings has a base of 1.034, and the exponential model of women's earnings has a base of 1.049. Since the women's model has a higher base, their earnings are increasing at a faster rate and will eventually surpass men's earnings. These functions can also be graphed to determine their intersection (45.7, 79533.88), demonstrating that at some point during the year 2029, women's annual median earnings will overtake men's annual median earnings.

7. For each year listed in the table, find the ratio of women's to men's annual median earnings expressed as a percentage. Use the data to create a scatter plot of percentage versus year. Based on this graph, do you think women's annual median earnings will ever equal those of men? Why or why not?



The scatter plot has a positive correlation. This means that women's annual earnings are approaching those of men and (if the trend continues) will eventually catch up to men's annual median earnings.

8. Considering the results of the scatter plot in question 7 above, do you think the linear model or exponential model makes more sense? Why?

Answers will vary. Generally speaking, the exponential model makes more sense because the gap between men's earnings and women's earnings is decreasing, as shown in the percentage-versus-time scatter plot. This more closely represents the real situation. The linear model shows the gap widening — an inaccurate representation of what is actually happening.

Data on earnings by gender provided by:

U.S. Census Bureau. "Table P-41. Work Experience—All Workers by Median Earnings and Sex: 1967 to2005." *Historical Income Tables—People.* www.census.gov/hhes/www/income/histinc/p41ar.html. (Date retrieved: July 24, 2007.)

Performance Task: Equal Salaries for Equal Work?

Name__________________________________ **Date**___________________

Mathematical Goals
- Represent data on a scatter plot
- Describe how two variables are related
- Fit a linear function for a scatter plot that suggests a linear association

Essential Questions
- How do you determine if linear or exponential regression is more appropriate for a scatter plot?

Georgia Standards of Excellence
MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6a Decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function of best fit. Emphasize linear, quadratic, and exponential models.

MGSE9-12.S.ID.6c Using given or collected bivariate data, fit a linear function for a scatter plot that suggests a linear association.

MGSE9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MGSE9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient "r" of a linear fit. (For instance, by looking at a scatterplot, students should be able to tell if the correlation coefficient is positive or negative and give a reasonable estimate of the "r" value.) After calculating the line of best fit using technology, students should be able to describe how strong the goodness of fit of the regression is, using "r."

Standards for Mathematical Practice
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.

Performance Task: Equal Salaries for Equal Work?

Name___________________________________ **Date**___________________

The data table shows the annual median earnings for female and male workers in the United States from 1984 to 2004. Use the data table to complete the task. Answer all questions in depth to show your understanding of the standards.

Year	Women's median earnings (in dollars)	Men's median earnings (in dollars)
1984	8,675	17,026
1985	9,328	17,779
1986	10,016	18,782
1987	10,619	19,818
1988	11,096	20,612
1989	11,736	21,376
1990	12,250	21,522
1991	12,884	21,857
1992	13,527	21,903
1993	13,896	22,443
1994	14,323	23,656
1995	15,322	25,018
1996	16,028	25,785
1997	16,716	26,843
1998	17,716	28,755
1999	18,440	30,079
2000	20,267	30,951
2001	20,851	31,364
2002	21,429	31,647
2003	22,004	32,048
2004	22,256	32,483

Data provided by U.S. Census Bureau

1. Create two scatter plots, one for women's median earnings over time and one for men's median earnings over time. Describe two things you notice about the scatter plots.

2. Terry and Tomas are trying to decide what type of model will most accurately represent the data. Terry thinks that a linear model might be most appropriate for each scatter plot. Help Terry find reasonable linear function rules for each scatter plot. Explain how you found these.

3. Using the linear models, will women's annual median earnings ever equal those of men? Why or why not?

4. Tomas thinks that an exponential model might be most appropriate for each scatter plot. Help Tomas find reasonable exponential function rules for each scatter plot. Explain how you found these.

5. Using the exponential models, will women's annual median earnings ever equal those of men? Why or why not?

6. If you answered yes to either question 3 or question 5, use that model to determine the first year women will have higher median earnings than men. Explain how you found your answer.

7. For each year listed in the table, find the ratio of women's to men's annual median earnings expressed as a percentage. Use the data to create a scatter plot of percentage versus year. Based on this graph, do you think women's annual median earnings will ever equal those of men? Why or why not?

8. Considering the results of the scatter plot in question 7 above, do you think the linear model or exponential model makes more sense? Why?

Data on earnings by gender provided by:

U.S. Census Bureau. "Table P-41. Work Experience—All Workers by Median Earnings and Sex: 1967 to2005." *Historical Income Tables—People.* www.census.gov/hhes/www/income/histinc/p41ar.html. (Date retrieved: July 24, 2007.)

Formative Assessment Lesson: iRegress!

Back to Task Table

Source: Georgia Mathematics Design Collaborative

This lesson is intended to help you assess how well students are able to:

- Interpret the correlation coefficient
- Interpret the slope and y-intercept of a line of best fit in the context of the data
- Contrast correlation and causation

Georgia Standards of Excellence

MGSE9-12.S.ID.6a Decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function of best fit. Use this function to solve problems in context. Emphasize linear, quadratic and exponential models.

MGSE9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MGSE9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient "r" of a linear fit. (For instance, by looking at a scatterplot, students should be able to tell if the correlation coefficient is positive or negative and give a reasonable estimate of the "r" value.) After calculating the line of best fit using technology, students should be able to describe how strong the goodness of fit of the regression is, using "r".

MGSE9-12.S.ID.9 Distinguish between correlation and causation.

Standards for Mathematical Practice

This lesson uses all of the practices with emphasis on:

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
6. Attend to precision

TASK DESCRIPTION, DEVELOPMENT AND DISCUSSION:

Tasks and lessons from the Georgia Mathematics Design Collaborative are specifically designed to help teachers effectively formatively assess their students. The way the tasks and lessons are designed gives the teacher a clear understanding of what the students are able to do and not do. Within the lesson, teachers will find suggestions and question prompts that will help guide students towards understanding.

The task, *iRegress!*, is a Formative Assessment Lesson (FAL) that can be found on the Georgia Mathematics online professional learning community, edWeb.net. Once logged in, the task can be found directly at: https://www.edweb.net/?14@@.5ad26830

Research Design (Culminating Task)

Back to Task Table

Introduction

Students will synthesize what they've learned in this unit and design their own study. They should work individually or in pairs to formulate research questions, use appropriate statistical methods to analyze the data, and develop and analyze predictions. Work with students on the four step research design provided and create a rubric for evaluating the study with the students collectively. Use the sample research study provided to make sure students understand the standards to be addressed as they design their own study. Facilitate the evaluation of the task to ensure content appropriateness and rigor as the student work with you to have ownership in the evaluation of their content and product.

Mathematical Goals

- Summarize, represent, and interpret data on a single count or measurement variable
- Summarize, represent, and interpret data on two categorical and quantitative variables
- Interpret linear models

Essential Questions

- How can I apply what I have learned about statistics to summarize and analyze real data?

Georgia Standards of Excellence

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots). Choose appropriate graphs to be consistent with numerical data: dot plots, histograms, and box plots.

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers). Students will examine graphical representations to determine if data are symmetric, skewed left, or skewed right and how the shape of the data affects descriptive statistics.

MGSE9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6a Decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function to best fit. Use this function to solve problems in context. Emphasize linear, quadratic, and exponential models.

MGSE9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

MGSE9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MGSE9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient "r" of a linear fit. (For instance, by looking at a scatterplot, students should be able to tell if the correlation coefficient is positive or negative and give a reasonable estimate of the "r" value.) After calculating the line of best fit using technology, students should be able to describe how strong the goodness of fit of the regression is, using "r."

MGSE9-12.S.ID.9 Distinguish between correlation and causation.

Standards for Mathematical Practice

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.

Students should use all eight SMPs when exploring this task.

Background Knowledge

- Students will apply everything they have learned in this unit.

Common Misconceptions

- Address misconceptions brought to light during the rest of the unit.

Materials

- Pencil
- Paper / graphing paper
- Graphing calculator or statistical software package
- Portfolio of work from the unit
- (Possibly) internet access to help brainstorm question design

Grouping

- Individual / Small group

Culminating Task: Research Design – Teacher Notes

You will synthesize your learning from the entire unit in this task. Follow the design process below to collect, represent, analyze, and interpret your data.

I. Formulate Questions

→ Pose your own questions (include one question that is categorical in nature and include questions that produce bivariate data in which you predict a possible relationship prior to analyzing the data)

→ Address questions involving a group larger than your classroom and begin to recognize the distinction among a population, a census, and a sample.

II. Collect Data

→ Conduct censuses of two or more classrooms or design and conduct nonrandom sample surveys

III. Analyze Data

→ Expand and demonstrate understanding of a data distribution.

→ Quantify variability within a group.

→ Compare two or more distributions using graphical displays and numerical summaries.

→ Be sure to use more sophisticated tools for summarizing and comparing distributions, including:

- Histograms
- The IQR (Interquartile Range) and MAD (Mean Absolute Deviation)
- Five-Number Summaries and box plots

→ Quantify the strength of association between two variables, develop simple models for association between two numerical variables, and use expanded tools for exploring association, including:

- Two-way frequency/contingency tables for two categorical variables (displaying conditional, marginal, joint, frequencies)
- Simple lines for modeling association between two numerical variables

IV. Interpret Results

→ Describe differences between two or more groups with respect to center, spread, and shape.

→ Explain basic interpretations of measures of association.

→ Distinguish between "association" and "cause and effect."

Sample:

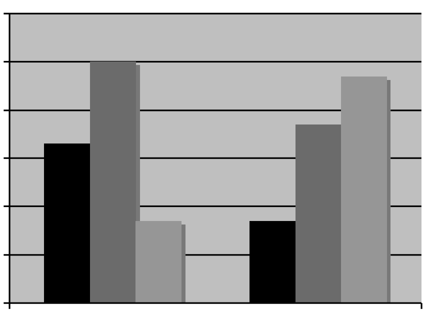
How do the favorite types of music differ between classes?

As class sizes may be different, results should be summarized with relative frequencies or percents in order to make comparisons. Percentages are useful in that they allow us to think of having comparable results for groups of size 100. Students should be comfortable summarizing and interpreting data in terms of percents or fractions. The results from two classes are summarized in Table 3 using both frequency and relative frequency (percents).

	Class 1	
Favorite	Frequency	Relative Frequency Percentage
Country	8	33%
Rap	12	50%
Rock	4	17%
Total	**24**	**100%**
	Class 2	
Favorite	Frequency	Relative Frequency Percentage
Country	4	13%
Rap	12	40%
Rock	14	47%
Total	**30**	**100%**

The bar graph below compares the percent of each favorite music category for the two classes. Students should begin to recognize that there is not only variability from one individual to another within a group, but also in results from one group to another. This second type of variability is illustrated by the fact that the most popular music is rap music in Class 1, while it is rock music in Class 2. That is, the mode for Class 1 is rap music, while the mode for Class 2 is rock music. The results from the two samples might be combined in order to have a larger sample of the entire school. The combined results indicate rap music is the favorite type of music for 44% of the students.

Comparative bar graph for music preferences



Connecting Two Categorical Variables

As rap was the most popular music for the two combined classes, the students might argue for a rap group for the dance. However, more than half of those surveyed preferred either rock or country music. Will these students be unhappy if a rap band is chosen? Not necessarily, as many students who like rock music also may like rap music. To investigate this problem, students might explore two additional questions:

Do students who like rock music tend to like or dislike rap music?
Do students who like country music tend to like or dislike rap music?

To address these questions, the survey should ask students not only their favorite type of music, but also whether they like rap, rock, and country music.

The two-way frequency table (or contingency table) below provides a way to investigate possible connections between two categorical variables.

Table 4: Two-Way Frequency Table

Like Rap Music?

		Yes	No	Row Totals
Like Rock Music?	Yes	27	6	33
	No	4	17	21
Column Totals		31	23	54

According to these results, of the 33 students who liked rock music, 27 also liked rap music. That is, 82% (27/33) of the students who like rock music also like rap music. This indicates that students who like rock music tend to like rap music as well. Notice the use of proportional reasoning in interpreting these results. A similar analysis could be performed to determine if students who like country tend to like or dislike rap music. A more detailed discussion of this example concerning a measure of association between two categorical should take place.

Practice Collecting Bivariate Quantitative Data with the Class

Come up with questions in which the responses produce bivariate data. Poll the class quickly to gather data and discuss how you would analyze the data and interpret the results.

For example:

- *What is the typical height of students in this class?*
- *What is the typical length of the feet of students in this class?*
- *What differences and similarities do we notice between these two data sets?*

Before beginning discussion, have students make a prediction about the relationship between height and length of feet. Discuss organizing and representing the data on a real number line. Also make sure students know how to describe, summarize, and compare the data in terms of center, shape, spread, and outliers as well as using summary statistics.

Take the same question and have them compare the variables (in this case height and feet). If there is a correlation, make sure they describe the type of association in terms of strength. Also ensure that cause and effect is a part of the discussion.

After modeling the research design process, work with students to create a rubric for evaluating their culminating project. Require students to get their questions approved prior to beginning their data collection process.

Culminating Task: Research Design

Name__________________________________ **Date**___________________

Mathematical Goals

- Summarize, represent, and interpret data on a single count or measurement variable
- Summarize, represent, and interpret data on two categorical and quantitative variables
- Interpret linear models

Essential Questions

- How can I apply what I have learned about statistics to summarize and analyze real data?

Georgia Standards of Excellence

MGSE9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots). Choose appropriate graphs to be consistent with numerical data: dot plots, histograms, and box plots.

MGSE9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation, ~~standard deviation~~) of two or more different data sets.

MGSE9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers). Students will examine graphical representations to determine if data are symmetric, skewed left, or skewed right and how the shape of the data affects descriptive statistics.

MGSE9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

MGSE9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MGSE9-12.S.ID.6a Decide which type of function is most appropriate by observing graphed data, charted data, or by analysis of context to generate a viable (rough) function to best fit. Use this function to solve problems in context. Emphasize linear, quadratic, and exponential models.

MGSE9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

MGSE9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MGSE9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient "r" of a linear fit. (For instance, by looking at a scatterplot, students should be able to tell if the correlation coefficient is positive or negative and give a reasonable estimate of the “r” value.) After calculating the line of best fit using technology, students should be able to describe how strong the goodness of fit of the regression is, using “r.”

MGSE9-12.S.ID.9 Distinguish between correlation and causation.

Standards for Mathematical Practice

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.

Culminating Task: Research Design

Name________________________________ **Date**__________________

You will synthesize your learning from the entire unit in this task.
Follow the design process below to collect, represent, analyze, and interpret your data.

I. Formulate Questions

→ Pose your own questions (include one question that is categorical in nature and include questions that produce bivariate data in which you predict a possible relationship prior to analyzing the data)
→ Address questions involving a group larger than your classroom and begin to recognize the distinction among a population, a census, and a sample.

II. Collect Data

→ Conduct censuses of two or more classrooms or design and conduct nonrandom sample surveys

III. Analyze Data

→ Expand and demonstrate understanding of a data distribution.
→ Quantify variability within a group.
→ Compare two or more distributions using graphical displays and numerical summaries.
→ Be sure to use more sophisticated tools for summarizing and comparing distributions, including:

- Histograms
- The IQR (Interquartile Range) and MAD (Mean Absolute Deviation)
- Five-Number Summaries and box plots

→ Quantify the strength of association between two variables, develop simple models for association between two numerical variables, and use expanded tools for exploring association, including:

- Two-way frequency/contingency tables for two categorical variables (displaying conditional, marginal, joint, frequencies)
- Simple lines for modeling association between two numerical variables

IV. Interpret Results

→ Describe differences between two or more groups with respect to center, spread, and shape.
→ Explain basic interpretations of measures of association.
→ Distinguish between "association" and "cause and effect."

Rubric for Culminating Task: Research Design

Category	4	3	2	1
Understanding	Analysis shows complete understanding of the problem	Analysis shows intermediate understanding of the problem	Analysis shows a basic understanding of the problem	No analysis or analysis shows no understanding of the problem
Analysis	Student completely analyzes information to develop an organized approach to a solution	Student substantially analyzes information to develop an organized approach to a solution	Student partially analyzes information which leads to an incomplete approach to a solution	Student demonstrates minimal analysis of the information
Strategy	Logical strategy that supports a correct solution	Logical strategy that somewhat supports a correct solution	Strategy that may or may not support a correct solution but lacks some logic	No logical strategy
Completeness	Solution completely addresses all mathematical components of the task	Solution mostly addresses all the mathematical components of the task	Solution addresses some of the mathematical components of the task	Solution does not address the mathematical components of the task
Terminology	Precise and accurate use of mathematical terminology	Effective use of mathematical terminology	Some use of mathematical terminology	No use of mathematical terminology

4=Excellent; 3=Adequate; 2=Fair; 1=Minimal