

Annotated Examples: Modeling with Data:

Two-Variable Measurement Data

The examples that follow are intended to illustrate how the learning principles are used to support students' engagement with the content and practices outlined in this badge. These examples do not provide comprehensive coverage of those expectations, but rather elevate some of the learning principles that are less likely to be part of published curricular materials for mathematics instruction. The examples that follow were developed by the Math Badging writing team, unless otherwise specified. These are a small sample of types of learning activities that can be done with students, both in and out of a traditional classroom setting.

Example 1: Understanding Graduation Rates

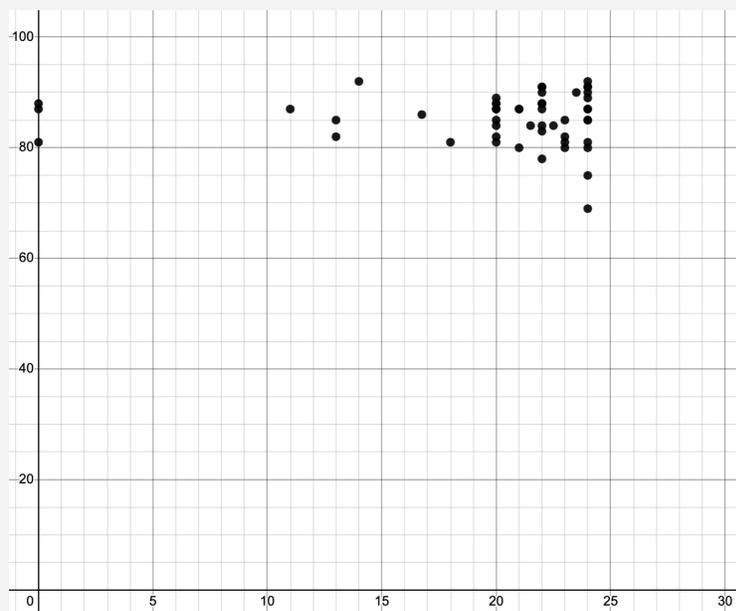
Sample Learning Activity

Part 1: Making Sense of a Scatter Plot

Students are invited to consider graduation rates and the factors that might impact them, using discussion questions such as these:

- *What do you think a state's "graduation rate" is?*
- *What factors might cause a state's graduation rate to be higher or lower?*

After discussing, students are provided this scatterplot and told that each point represents a state's data on two variables. The response variable is the state's graduation rate.



To ensure students have a secure understanding of the scatter plot and the meaning of the response variable, share with students that Mississippi has a graduation rate of 85%. Ask students to examine the graph and identify a point that could represent Mississippi.

Students are then asked to brainstorm, first individually, then with a peer group, the answers to these questions:

- *What could be the explanatory variable?*
- *What type of association is represented by this data (linear, quadratic, power, etc.)?*
- *There are three data points on the y axis. What could these points represent?*

After sharing, students are told that the explanatory variable for the data above is the number of courses required by U.S. states in order to graduate from high school.

Students are then asked to brainstorm, first individually, then with a peer group, the answers to these questions:

- *Do you think the number of required courses is a good predictor for high school graduation rate?*
- *Now that you know that the response variable is high school graduation rate, what could the three data points on the y-axis represent?*

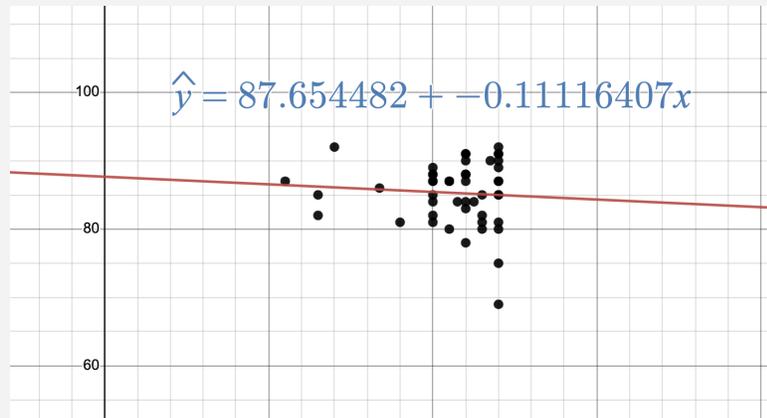
To deepen understanding of the context being modeled, students investigate the information found in the [National Center for Education Statistics' comparison of state graduation rates](#). They respond to questions such as these:

- *What are the coordinates for a state like Alabama? What do the coordinates represent?*
- *What are the coordinates of Pennsylvania? What do the coordinates represent?*
- *How does the meaning of the coordinates of these points compare? Are they tracking the same things?*
- *When creating a mathematical model, it is important that the quantities being used are tracking the same things. Why would it make sense to exclude data from Pennsylvania, Colorado, and Massachusetts?*

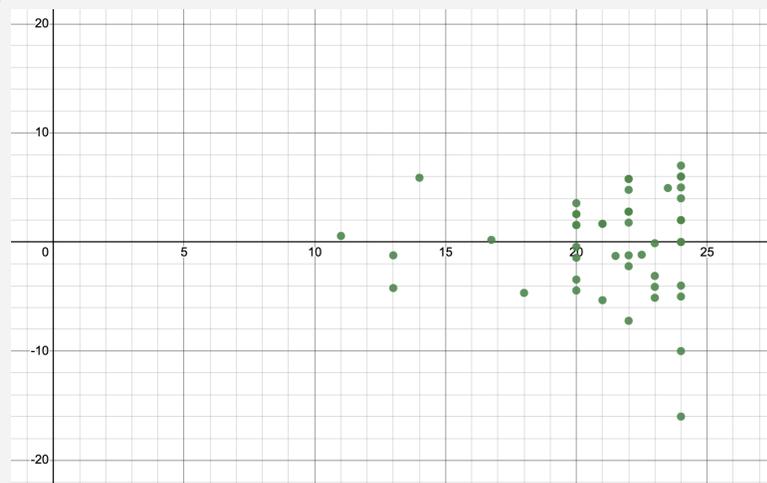
Part 2: Using a Line of Best Fit

Using technology, students are asked to estimate and draw in a line of best fit for the data set. (See, for example, [this Desmos calculator link](#).) Students discuss and respond to these questions:

- *What does the y-intercept and slope of your line of best fit mean for the given context?*
- *Use technology to fit a regression model to the data. Find the value of the correlation coefficient, r . What does this value mean for the given context? (See for example, [this Desmos calculator link](#).)*
- *Is a linear model a good fit for this set of data?*
- *Using technology, find the value of the coefficient of determination, r^2 . The coefficient of determination shows the amount of variability accounted for in the model. Does the value of r^2 you calculated validate or dispute the claim that a higher number of required courses in high school leads to higher graduation rates?*
- *Here is the residual plot of data given above based on the linear regression model shown below:*



Residual plot:



Is a linear model appropriate for the given data?

Students are then asked to consider the other potential factors influencing graduation rates brainstormed initially in Part 1. Students find state level data and repeat the process to examine the effect of their chosen explanatory variable on graduation rates and evaluate whether it is a better predictor.

In closing, students are asked to consider and respond to these questions:

- Which explanatory variable that you have examined impacts high school graduation rates more? Support your answer with appropriate statistics.
- What new insights have you gained about the number of course requirements and graduation rates?
- What recommendations would you offer or new questions surface for you?

In this example, students are:

- given a cognitively demanding task. It requires that students make sense of a situation and a statistical model, in addition to generating their own model (LP 1).
- set up to have social interactions, as they are invited to brainstorm, discuss, and share ideas with one another. These interactions may happen live in a physical classroom or asynchronously in a remote setting (LP 2).

- building conceptual understanding through reasoning, as they make and justify conjectures about the explanatory variable in a scatter plot (LP 3).
- engaging with a task that is relevant to their lives, focusing on graduation rates and requirements (LP 6).
- encouraged to use technology as a tool for understanding two-variable statistics (LP 7).

Through the implementation of this learning activity, look for opportunities to support students' growth and development on this XQ competency:



- Making Meaning: [Synthesizing information](#) (FL.ID.1a)

Example 2: Risk Ratio

Students are given a problem similar to this:

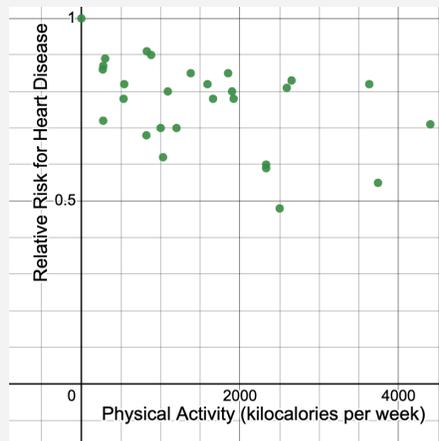
Part 1: Develop a Statistical Question

According to [Harvard Health Publishing web article, "Heart Health"](#), a healthy heart "beats about 2.5 billion times over the average lifetime, pushing millions of gallons of blood to every part of the body. This steady flow carries with it oxygen, fuel, hormones, other compounds, and a host of essential cells. It also whisks away the waste products of metabolism." When the heart can't perform at optimal health, it is said that a person has developed some form of cardiovascular disease or heart disease. Doctors have developed a way to quantify a person's risk for developing heart disease; it is called relative risk ratio.

1. What are some factors that might lead to heart disease?
2. A relative risk ratio is a number that compares the possibility of getting heart disease among one group with the possibility of getting that disease in another group. If the risk ratio is 1, that means both groups have the same possibility of getting heart disease. Based on your current understanding, answer the following questions using a factor from your answer to Question 1:
 - a. What do you believe are the two groups involved?
 - b. What do you anticipate the relative risk ratio will be: about 1, greater than 1, or less than 1? How did you arrive at this conclusion?
 - c. What are some statistical questions you can ask about heart disease?

Part 2: Understand a Mathematical Representation

In one study, researchers set out to quantify the specific amounts of physical activity required for lower risks of heart disease. In this scatter plot, the explanatory (independent) variable is the physical activity measured in kilocalories burned per week by participants in various studies. The response (dependent) variable in the data set below is the relative risk of heart disease, when comparing physical activity during leisure-time with a group who has no leisure-time physical activity.



Based on: [Dose response between physical activity and risk of coronary heart disease: a meta-analysis](#)

3. *What do you notice? What do you wonder?*
4. *How might you use this graph to answer the statistical question, “Do higher levels of physical activity lower the risk of heart disease?”*

Part 3: Manipulate the Model

5. *Work with this representation of the data on Desmos Calculator: [Physical Activity and Relative Risk for Heart Disease](#) to determine the line of best fit.*
6. *Find and interpret the value of the coefficient of determination, r^2 .*
7. *Based on the value of r^2 and the data, do you believe the claim that a higher level of physical activity lowers the risk of heart disease is valid?*

Part 4: Interpreting the Model

8. *Do you think that a level of physical activity, quantified by kilocalories per week, is a good predictor of a relative risk for heart disease?*

Part 5: Reflection

9. *What other variables do you think contribute to heart disease?*
10. *Consider your list of variables. Which ones does your group believe could be a better predictor of a shortened lifespan due to heart disease?*

Sample Learning Activity

Part 1: Develop a Statistical Question

Students are primed with information about heart disease, including what it is and possible contributing factors. Students are asked to brainstorm factors that might decrease a person’s risk for heart disease. Give students an opportunity to learn more about terms that surface. One term students may be wanting to explore is relative risk ratio; they can learn more about this at the website [Principles of Epidemiology | Lesson 3 - Section 5](#).

Some examples of factors that might contribute to a person's risk for heart disease might include:

- personal habits such as smoking, exercise, and eating
- genetic factors
- environmental factors

Students then collaboratively develop a statistical question to explore heart disease. Some may wonder about the relationship between heart disease and walking speed, which may lead to a question like: Is a higher average walking speed associated with a lower rate of heart disease?

Consider having students post their statistical questions in a visible space for all to reflect on. Invite students to give each other feedback on the questions they developed in order to spark possibilities for moving forward.

Part 2: Understand a Mathematical Representation

For this part, students will work on making sense of the given data set and model. Time permitting, you may offer students time to do research and examine another factor related to the statistical questions they developed. Highlight statistical questions that lend themselves to the data given.

Students are asked to consider how they could use a scatter plot to explore the statistical question: “Do higher levels of physical activity lower the risk of heart disease?” First, give students time to think individually for two to three minutes. Next, have them turn and talk or collaborate within small groups.

Consider posing the following questions as students discuss their insights:

- *What is the explanatory (independent) variable? Is it a reasonable variable to consider for this context?*
- *What does the response (dependent) variable mean?*
- *What type of association is represented by this data (linear, quadratic, power, etc.)?*
- *There are multiple data points in the set. Choose one and explain what it means.*

Part 3: Manipulate the Model

Students use technology to draw a line of best fit for the scatter plot. Have students navigate independently to the Desmos Calculator: [Physical Activity and Relative Risk for Heart Disease](#). Students respond to and discuss questions such as these:

- *How did you decide your line of best fit?*
- *What do the y-intercept and slope mean in the context of this problem?*
- *What type of model is best for this data set?*
- *How do you feel about using your model to explain the relationship between physical activity and risk of heart disease? What new questions do you have?*

Part 4: Interpreting the Model

Students consider and discuss this question, using their modeling, data, and analysis to support their responses. Encourage students to ask clarifying questions of each other, so that arguments are understood. Some questions to consider are as follows:

- *What about this reasoning makes sense to you?*

- *Is there anything you disagree with?*
- *What questions surface for you?*

Part 5: Reflection

Students consider further variables to explore by discussing the questions posed. Offer space for various ideas to surface. Provide time for students to reflect and revise their work throughout.

In this example, students are:

- given a cognitively demanding task. It requires that students go through the full modeling cycle to investigate the relationship between weekly physical activity and relative risk rate of heart disease (LP 1).
- set up to have social interactions, as they collaborate to develop a statistical question (LP 2).
- building conceptual understanding as they reason and reflect throughout this iterative process (LP 3).
- given opportunities to explore questions that surface as they make their way through this process, building a sense of agency (LP 4).
- applying mathematics in a relevant, real-world situation. Students are asked to consider a matter of public health. As they do, they have an opportunity to learn how researchers work to develop mathematical models to explain possible relationships between factors present in a phenomena (LP 6).
- encouraged to use technology to create a scatter plot and generate a line of best fit (LP 7).

Through the implementation of this learning activity, look for opportunities to support students' growth and development on this XQ competency:



- **Wayfinding:** [Iterative thinking](#) (LL.SD.2.b)

References

CDC. "Section 5: Measures of Association." *CDC Web Archive, Lesson 3: Measures of Risk*, CDC, 18 May 2012, <https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section5.html>.

Common Core Standards Writing Team. (2019). *Progressions for the Common Core State Standards for Mathematics, Modeling, K-12*. (draft February 8, 2019). Tucson, AZ: Institute for Mathematics and Education, University of Arizona. From *Mathematical Musings*: http://mathematicalmusings.org/wp-content/uploads/2022/05/modeling_02082019.pdf

Desmos. "AveWalkHeartDiseas." *Desmos*, <https://www.desmos.com/calculator/np3fiqspl6>. Accessed 15 November 2022.

Desmos. "Required High School Course vs. High School Graduation Rate (estimated line)." *Desmos*, <https://www.desmos.com/calculator/u86paauydj>. Accessed 15 November 2022.

Education Commission of the States. "High School Graduation Requirements: 50 State Comparison." Education Commission of the States, February 2019, <https://reports.ecs.org/comparisons/high-school-graduation-requirements-01>. Accessed 15 November 2022.

Hamblyn, C. "Data Sets for Bivariate Investigations." *CensusAtSchool*, CensusAtSchool New Zealand, 22 June 2022, <https://new.censusatschool.org.nz/resource/multivariate-data-sets/>. Accessed 15 November 2022.

Harvard Medical School. "Heart Health." *Harvard Health Publishing*, Harvard Medical School, <https://www.health.harvard.edu/topics/heart-health>. Accessed 16 December 2022.

National Center for Education Statistics. (2022). Public High School Graduation Rates. Condition of Education. U.S. Department of Education, Institute of Education Sciences. Retrieved [date], from <https://nces.ed.gov/programs/coe/indicator/coi>.

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). High School: Modeling. From Common Core State Standards Initiative: <http://www.corestandards.org/Math/Content/HSM/>

Safir, S., & Dugan, J. (2021). Building A Performance Assessment System. In *Street Data* (pp. 131-137). Thousand Oaks, CA: Corwin, A SAGE Company.

Sattelmair J, Pertman J, Ding EL, Kohl HW 3rd, Haskell W, Lee IM. Dose response between physical activity and risk of coronary heart disease: a meta-analysis. *Circulation*. 2011 Aug 16;124(7):789-95. doi: 10.1161/CIRCULATIONAHA.110.010710. Epub 2011 Aug 1. PMID: 21810663; PMCID: PMC3158733.