

Baseline Professional Learning Handout

M112 Modeling with Data: Two-Variable Measurement Data

Baseline Professional Learning Handout	1
Selected Common Core High School Standards for Mathematics	2
Statistics and Probability Excerpt	3
M112 Content and Practice Expectations	4
M112 Content and Practice Expectations and Indicators	5
Let's Investigate	7
Restaurant Bill and Party Size Task	8
Best Phone Charger Task	10
Let's do some math! Round 1	11
Task 1: Graduation Rates	12
Learning Principles	15
Task 1: Graduation Rates Student Work	16
Let's do some math! Round 2	26
Task 2: Heart Disease	27
Task 2: Heart Disease Student Work	29
Criteria for Success	38
References	38

Selected Common Core High School Standards for Mathematics

High School - Statistics and Probability

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

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.
6. Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
 - a. Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.
 - b. Informally assess the fit of a function by plotting and analyzing residuals.
 - c. Fit a linear function for a scatter plot that suggests a linear association.

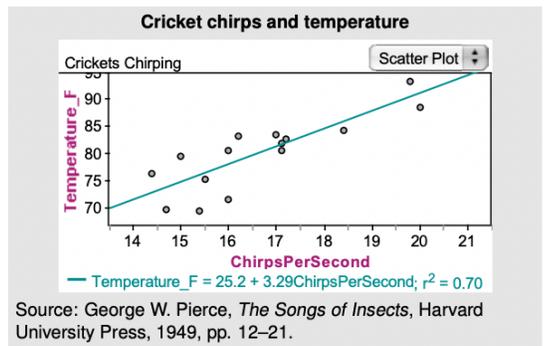
Interpret linear models

7. Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.
8. Compute (using technology) and interpret the correlation coefficient of a linear fit.
9. Distinguish between correlation and causation.

Statistics and Probability Excerpt

Students have seen scatter plots in Grade 8 and now extend that knowledge to fit mathematical models that capture key elements of the relationship between two variables and to explain what the model tells us about that relationship. Some of the data should come from science, as in the examples about cricket chirps and temperature, and tree growth and age, and some from other aspects of their everyday life, e.g., cost of pizza and calories per slice.

If you have a keen ear and some crickets, can the cricket chirps help you predict the temperature? The margin shows data modeled in a scientific investigation of that phenomenon. In this situation, the variables have been identified as chirps per second and temperature in degrees Fahrenheit. The cloud of points in the scatter plot is essentially linear with a moderately strong positive relationship. It looks like there must be something other than random behavior in this association. A model has been formulated: The least squares regression line has been fit by technology.^{S-ID.6} The model is used to draw conclusions: The line estimates that, on average, each added chirp predicts an increase of about 3.29 degrees Fahrenheit. (Common Core Standards Writing Team, 2012)



M112 Content and Practice Expectations

112.a	Engage in the modeling cycle.
112.b	Use scatter plots to represent two-variable measurement data.
112.c	Describe visible patterns between two data sets in a scatter plot.
112.d	Informally fit a straight line for scatter plots that suggest a linear association.
112.e	Fit a linear, quadratic, or exponential function to two-variable measurement data.
112.f	Use functions that have been fitted to two-variable measurement data to answer questions about the relationship being modeled.
112.g	Assess the fit of a function using the correlation coefficient, residuals, and other tools.
112.h	Explain insights gained from analyzing two-variable measurement data and limitations of curves fitted to data.

M112 Content and Practice Expectations and Indicators

Content and Practice Expectations	Indicators Choose an artifact where you...
112.a: Engage in the modeling cycle.	i. engage with the full modeling cycle (problem, formulate, compute, interpret, validate, revise as necessary, report).
112.b: Use scatter plots to represent two-variable measurement data.	i. construct a scatter plot from collected data
	ii. explain the meaning of a data point plotted on a scatter plot
112.c: Describe visible patterns between two data sets in a scatter plot.	i. identify the type of association visible in a scatter plot
	ii. give a hypothesis that might explain the association identified in a scatter plot
112.d: Informally fit a straight line for scatter plots that suggest a linear association.	i. sketch a line that follows a pattern of linear association
	ii. justify location and direction of a line fitted to data
112.e: Fit a linear, quadratic, or exponential function to two-variable measurement data.	i. graph a linear, quadratic, or exponential function that models two-variable measurement data
	ii. find the equation for a linear, quadratic, or exponential function that models two-variable measurement data
	iii. justify the choice of linear, quadratic, or exponential function
112.f: Use functions that have been fitted to two-variable measurement data to answer questions about the relationship being modeled.	i. explain the meaning of parameters for a function fitted to two-variable measurement data
	ii. use a function fitted to two-variable measurement data to predict values
112.g: Assess the fit of a function using the correlation coefficient, residuals, and other tools.	i. find the correlation coefficient for a function fitted to two-variable measurement data
	ii. explain the meaning of a correlation coefficient
	iii. plot residuals for a function fitted to two-variable measurement data
	iv. interpret a residual plot in terms of a function's goodness of fit
112.h: Explain insights gained from analyzing two-variable measurement data and limitations of curves fitted to data.	i. use statistical tools to justify a conclusion about two-variable measurement data
	ii. explain a limitation of a curve fitted to data

[Back to Table of Contents](#)

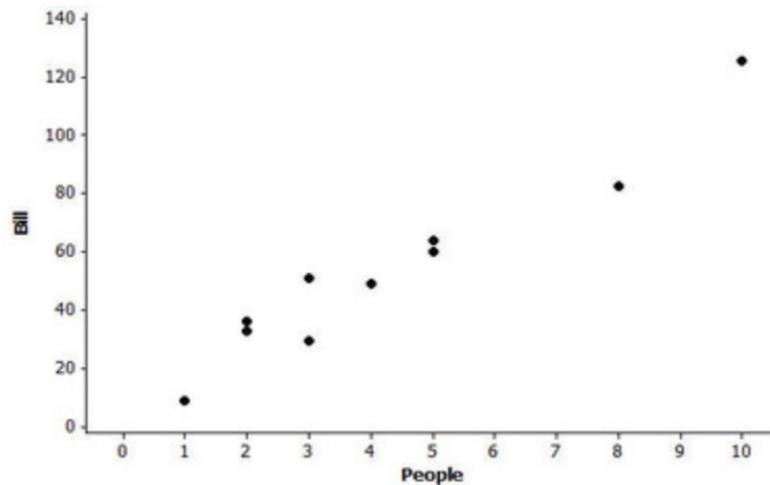
Content and Practice Expectations	Indicators Choose an artifact where you...
	iii. distinguish between correlation and causation

Let's Investigate

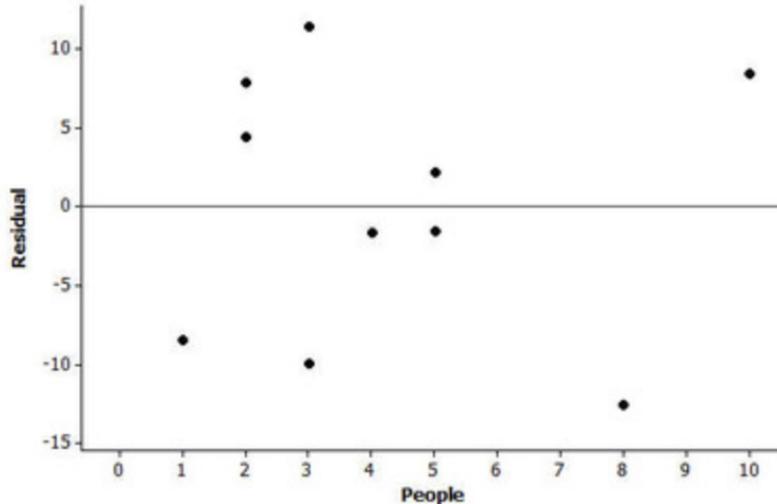
Restaurant Bill and Party Size Task

The owner of a local restaurant selected a random sample of dinner tables at his restaurant. For each table, the owner recorded the total amount of the dinner bill and the number of people at the table. The data are given in the table below. A scatterplot of the data is also shown.

People	Bill
1	\$8.50
2	\$36.00
2	\$32.55
3	\$29.30
3	\$50.65
4	\$48.75
5	\$63.75
5	\$60.00
8	\$82.50
10	\$125.75

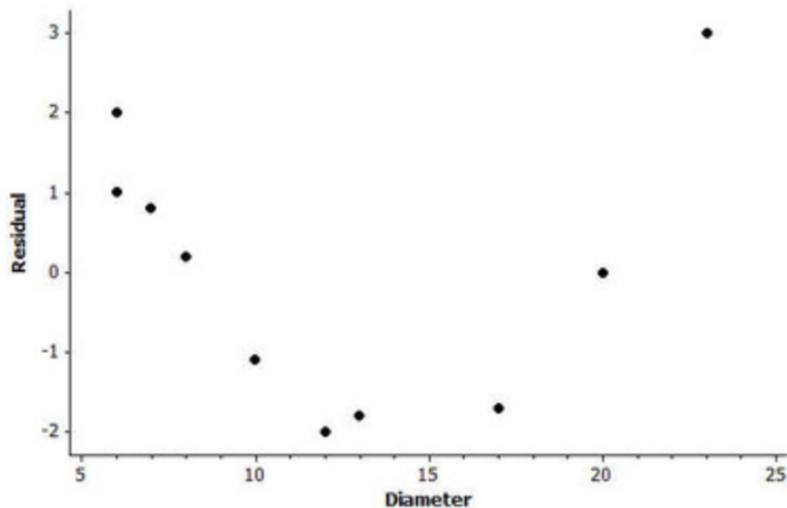


- Does the relationship between the number of people and the bill appear to be weak, moderate or strong?
- Does the relationship between the number of people and the bill appear to be linear?
- The equation of the line of best fit is $\text{Bill} = 5.80 + 11.15 \cdot \text{People}$. Sketch this line on the scatterplot.
- Interpret the slope of the line of best fit in the context of this problem.
- Note that the points in the scatterplot do not all lie on the line. What could explain this variability?
- Use the equation of the line of best fit to estimate how much would you predict the bill to be for a party of 8?
- There was one party of 8 whose bill was \$82.50. Did they pay more than or less than the predicted amount? How much more or how much less did they pay?
- The difference calculated in part (g) is called a residual. When each residual is calculated and plotted against the corresponding number of people, the resulting plot is called a residual plot. The residual plot for this data set is shown below.



Explain how one could use the residual plot in order to determine if a linear model is an appropriate way to describe the relationship between number of people and the amount of the bill.

- A random sample of several pizza restaurants in the area was selected. Data was collected on the diameter and the price of the smallest cheese pizza sold at the restaurant. The line of best fit was fit to the data and the corresponding residual plot is given below.



Does there appear to be a linear relationship between the diameter and the price of a cheese pizza in this area? Use the residual plot to explain your answer.

Source: [Illustrative Mathematics](#)

Best Phone Charger Task

Consider the following statement from [The New York Times' Wirecutter post The Best USB Phone Charger](#) by Nick Guy:

“If you’ve ever scrambled out the door with your phone’s battery in the red, you know that power is precious, and the faster you can get it, the better. Anker’s PowerPort III Nano is the best option for charging any phone fast: It’s tiny and inexpensive, and it can get a fully drained phone battery to more than 50% full in half an hour.”

In this task your job is to determine which charger is the best charger in your class.

1. Select at least two chargers to compare how quickly they can increase a phone’s battery power. What data will you need to collect?
2. Gather enough data from each charger to create a scatter plot for each one. What labels did you use for each axis?
3. Use technology to compute the best fit line for each charger’s scatter plot.
4. What are the values for the slope and y -intercept for each best fit line? What do these values mean in this situation?
 - a. Describe the relationship you see between the two quantities you used in this analysis.
 - b. Do you believe there is a causal relationship between these two quantities?
5. Based on your work, which of the chargers is the best?
 - a. What are the limitations of your mathematical analysis?

Source: Student Achievement Partners XQ HS Badging Team

Let's do some math! Round 1

[Back to Table of Contents](#)

Task 1: Graduation Rates

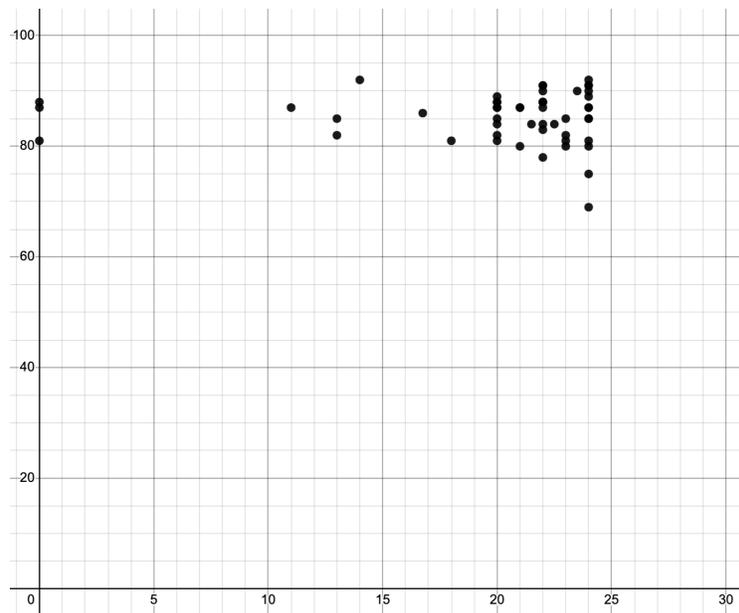
Sample Learning Experience

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; the response variable is the state's graduation rate.



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

- *What could be the context of 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 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 [Education Commission of the States' comparison of each state's high school graduation requirements](#) and the [National Center for Education Statistics' comparison of state graduation rates](#). They respond to questions such as these:

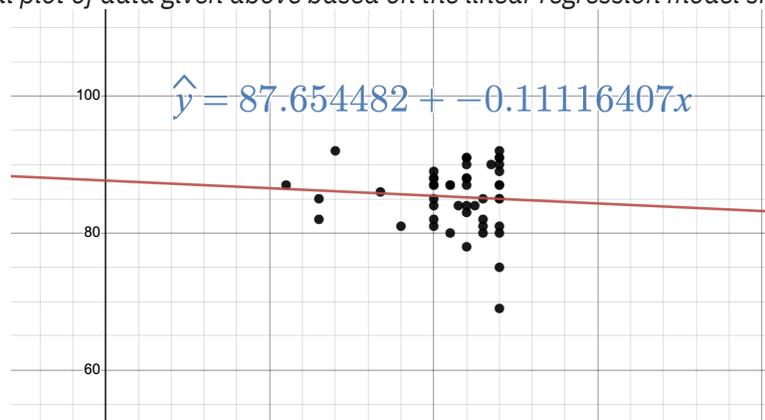
- *What are the coordinates for a state like Alabama? What do the coordinates represent?*
- *What are the coordinates of Pennsylvania? What do the coordinates represent?*
- *How does the meaning of the coordinates of these points compare? Are they tracking the same things?*

- 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?

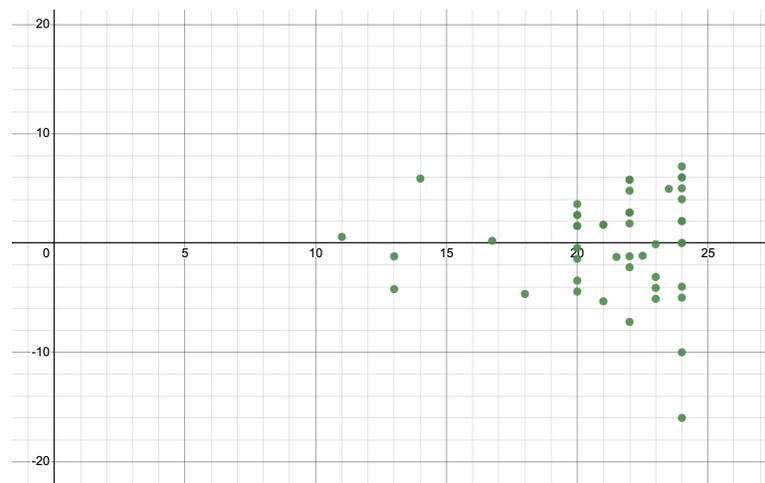
Part 2: Using a Line of Best Fit

Using technology, students are asked to estimate and draw in a line of 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 lead 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 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?*

Learning Principles

Engage with cognitively demanding tasks in heterogeneous settings (LP 1). Students should be given opportunities to grapple with multistep, non-routine tasks that promote mathematical rigor. These experiences should be differentiated so that all students engage in appropriate challenges—for example, through tasks with multiple entry points and solution pathways. These experiences should continue to integrate knowledge and skills developed in grades 6-8 at the level of sophistication of high school mathematics.

Engage in social activities (LP 2). Students should have opportunities to work independently and to communicate with one another about mathematics by engaging in collective and collaborative learning activities. Explaining and having opportunities to revise one’s thinking has excellent value in solidifying one’s knowledge.

Build conceptual understanding through reasoning (LP 3). Students should be given the opportunity to reason, justify, and problem solve with critical thinking, reading, writing, speaking, and listening. Through reasoning and work with multiple representations, students learn why procedures work and build a conceptual understanding of key mathematical ideas.

Have agency in their learning (LP 4). Students should be able to choose tasks and learning experiences that align with their interests and aspirations. All students have rich and varied experiences and home lives. Learning mathematics should bring students’ identities and interests to the fore and build on the strengths that they bring to the learning space.

View mathematics as a human endeavor across centuries (LP 5). Students should understand that mathematical ideas emanated over time from civilizations around the world and have opportunities to explore these contributions to mathematics. Students should develop an appreciation of mathematics as a human endeavor: one in which they feel a sense of belonging, where they see themselves as mathematicians, and one that offers opportunities to broaden their ideas about what mathematics is, how it is used, and who it is for.

See mathematics as relevant (LP 6). Students should engage with mathematics in ways that authentically involve real-world situations. Problem-solving contexts should allow students to perceive mathematics as a tool for addressing the questions that arise in everyday life, as well as the ways it can model our world and address global economic, social, and environmental challenges. Students should also engage with mathematics in ways that connect to academic disciplines and future careers by doing the mathematics used by artists, designers, engineers, and other professionals.

Employ technology as a tool for problem-solving and understanding (LP 7). Research indicates that technology is a powerful tool for learning deeper mathematics by improving calculation efficiency and enabling more sophisticated analysis. Students should learn to use technology, with emphasis put on widely used tools and software, such as calculators and spreadsheets, to make sense of models. Technology use should not be limited to supporting “doing mathematics,” but should also be used as a tool for displaying and communicating results to appropriate audiences.

Task 1: Graduation Rates Student Work

Example 1:

✓

Interpreting Data

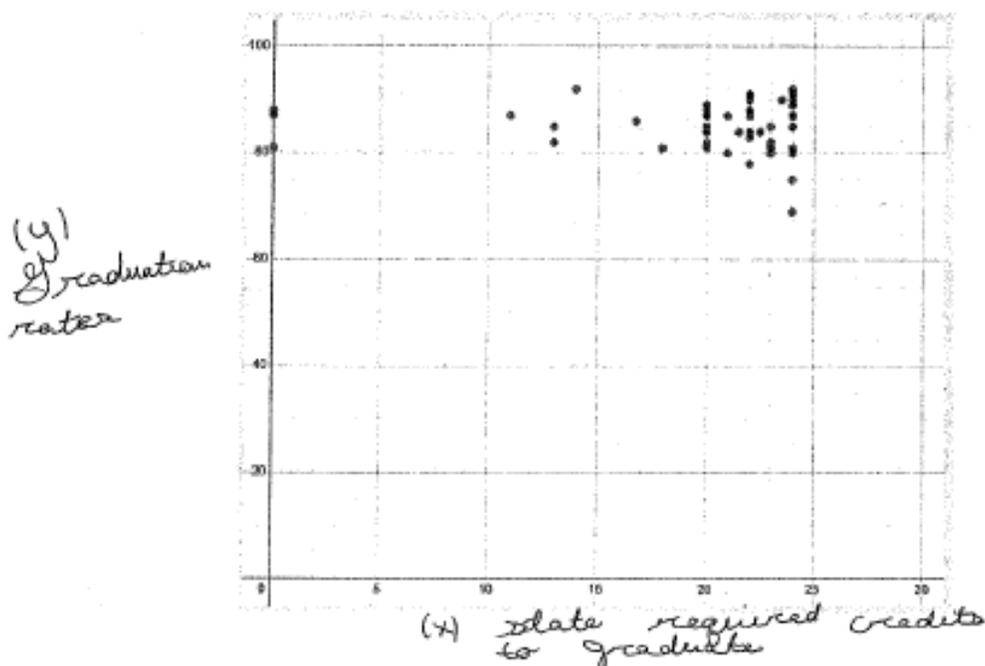
1. What do you think a state's "graduation rate" is?

82%

2. What factors might cause a state's graduation rate to be higher or lower?

students see purpose and meaning

Examine the graph below:



3. What could the x-axis represent? This can be referred to as an explanatory or independent variable.
hours required in class
4. What type of association is represented by this data (linear, quadratic, exponential, or something else)?
5. There are three data points on the y axis. What could these points represent? In other words, what do you think the y-axis represents? This can be called the response or dependent variable?
each districts came up with their own

6. Do you think the number of required courses is a good predictor for high school graduation rate?

Yes, more required less people want to do things

7. Now that you know that the response variable is high school graduation rate, what do you think the three data points on the y-axis represent?

students are interested in what they are learning

You will need to do 2 things:

- Label both axes.
- Go to the website in the Nearpod. You will need the information in order to answer the questions.

Once you have completed the two items on the list you can answer the questions below.

8. What are the coordinates for Alabama? What do the coordinates represent?

(29.5, 91)

9. What are the coordinates of Pennsylvania? What do the coordinates represent?

(0, 87)

0 = depends on country

10. How does the meaning of the coordinates of these points compare? Are they tracking the same things?

yes I think so anyway

11. 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?

they are all outliers

12. Consider High School. Determine where JHS would be on the graph AND plot the point. What do you notice? What do you wonder?

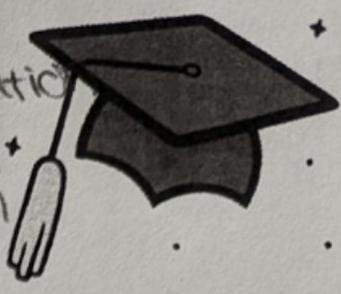
(84, 60)

Example 2:

Part 1: Let's Graduate

Consider graduation rates and the factors that might impact them, using discussion questions such as these:

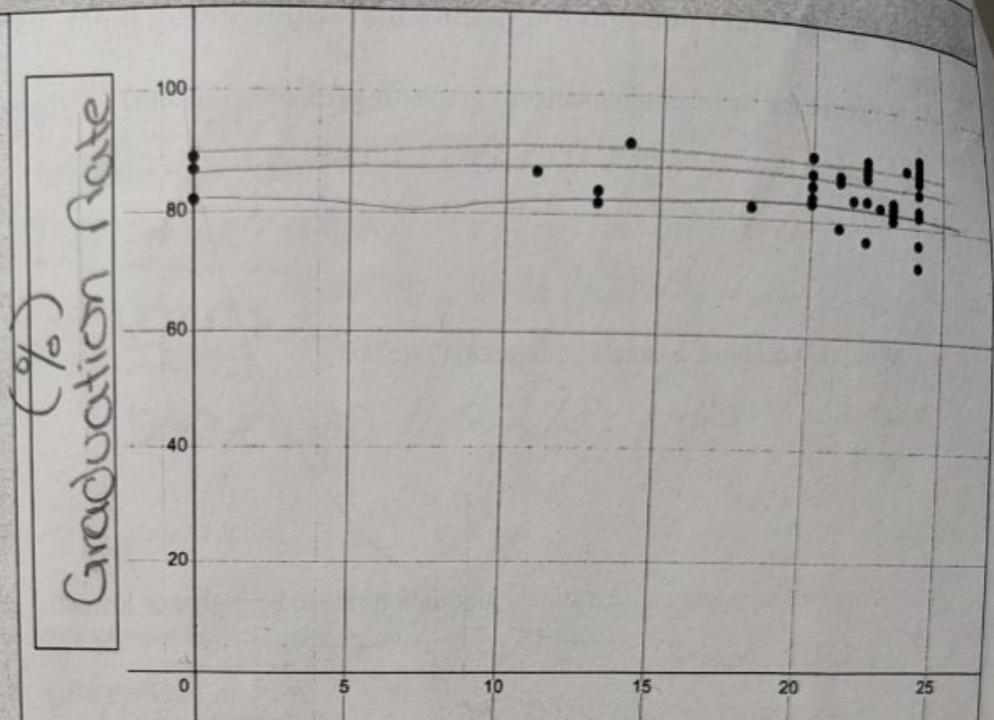
- What are the basic requirements for graduating from high school?
 - * passing certain classes/taking classes - credits
 - * sustain GPA
 - * health education
 - * 1 unit of arts/languages
 - * SATS
 - * consumers education
- What do you think a state's "graduation rate" is?
 - how many student graduate
- What factors might cause a state's graduation rate to be higher or lower?
 - * how well schools are - Teen pregnancy
 - * people priorities in community - Education
 - * how much income families make - location of school



Part 2: Making Sense of the Data Collected

The graph shows the data that has been collected. Each dot represents a state.

- Notice there are no labels on the x- or y- axis. Take a moment to make sense of the data.



Highschool Ranking

- What situation or thing might the **explanatory (independent) variable** be talking about?

Highschool ranking is independent because higher quality school will have a higher Graduation rate

- What type of relationship is represented by this data (linear, quadratic, exponential, etc.)? Justify your reasoning. Draw in your line of best fit.

This data is linear relationship between graduation rate + ranking because the dots are mostly consistent in one line.

- There are three data points on the y axis. What could these points represent?

The three dots on the y-axis represents the top 3 rankings.

An **explanatory variable** is something that helps explain or predict changes in another variable.

It's like a factor or condition that can influence the outcome of a situation.

For example, if you're studying what affects a plant's growth, the amount of sunlight it gets could be an explanatory variable because it explains why the plant might grow more or less.

Exploring the Statistics: Graduation Rates

The explanatory variable for the data above is the number of courses (credits) required by U.S. states in order to graduate high school.

- Do you think the number of required courses is a good predictor for high school graduation rate?

no, credits isn't a good graduation rate predictor because of how easy it is to fail and not be able to graduate.

- Now that you know that the response variable is high school graduation rate, what could the three data points on the y-axis represent?

The three dots represent schools with no required credits have an average amount of graduation rate.

(students with no credits)

To deepen understanding of the context being modeled, investigate the information found in the State Course Credit Requirements and the National Center for Education Statistics' comparison of state graduation rates.

- What are the coordinates for a state like Alabama? What do the coordinates represent?

24 credits
graduation rate: 90%

- What are the coordinates of Pennsylvania? What do the coordinates represent?

no state credit requirements
graduation rate: 77%

- How does the meaning of the coordinates of these points compare? Are they tracking the same things?

Both states track graduation rates but not required credits AL: Doze AA: Dozent

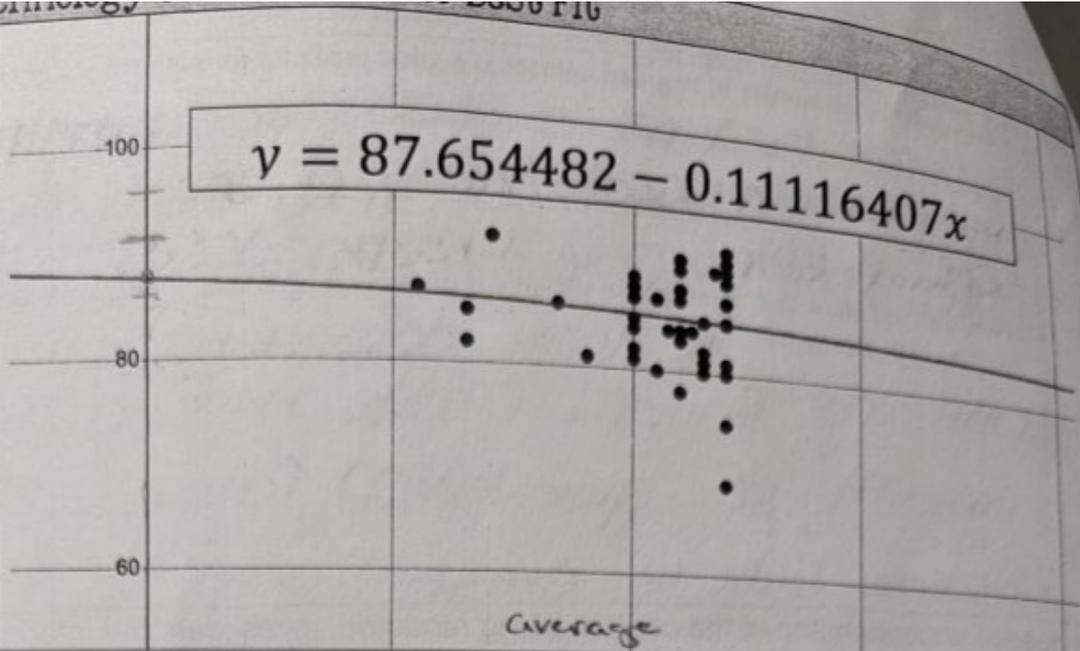
- When creating a mathematical model, it is important that the quantities being used are tracking the same things. Why would it make sense to exclude data from Pennsylvania, Colorado, and Massachusetts?

Because the state Doze not track required credits. It's up to individual schools



Part 3: Using Technology

We can use Desmos to create a line of best fit. To save time, the line of best fit has been given to you.



- What does the y-intercept and slope of your line of best fit mean for the given context?

y-intercept: 87.65 / percentage of students with no credits

- What do the dots close to the line indicate?

average % of students who graduated with required credits. High grad rate.

- What do the dots far from the line indicate?

low graduation % rate with students having required credits.

- Is a linear model a good fit for this set of data?

yes it is good fit for this data as it

represents an accurate model of graduation rates with its relations to required credits

What new insights have you gained about the number of course requirements and graduation rates?

- low required credits ↑ graduation / opposite ↓ graduation



Example 3:

Graduation Rates

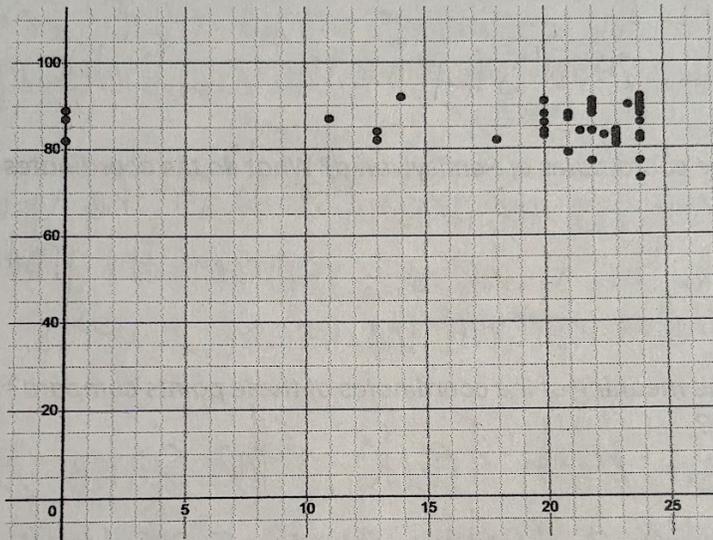
Part 1: Making Sense of a Scatter Plot

1. What do you think a state's "graduation rate" is?

How many people in a state that graduate from something like high school or college.

2. What factors might cause a state's graduation rate to be higher or lower?

The classes offered, someone's grades like if they need to be held back, people dropping out, someone going back to school, etc.



3. What could be the context of the explanatory variable?

The context could be something like the x axis is the amount of credits, and the y-axis is the graduation rate in percent.

4. What type of association is represented by this data (linear, quadratic, power, etc.)?

Quadratic, the data is different and scattered!

5. There are three data points on the y axis. What could these points represent?

The percent of graduation rates of a school with 0 credits required.

6. Do you think the number of required courses is a good predictor for high school graduation rate?

Yes, because if someone is required to take a lot of required courses, it will take longer to graduate, or someone could go to a different school with less required courses.

7. Now that you know that the response variable is high school graduation rate, what could the three data points on the y-axis represent?

The high school graduation rate with 0 required credits or required classes!

Explore the following resources: [State Course Credit Requirements](#) and the [National Center for Education Statistics' comparison of state graduation rates](#).

8. What are the coordinates for a state like Alabama? What do the coordinates represent?

Most of the coordinates would be closer to 100 on the y-axis and around 24 credits for the x. The coordinates represent the graduation rate and the amount of years.

9. What are the coordinates of Pennsylvania? What do the coordinates represent?

The coordinates are around 80-90 for the y-axis and is different locally for the x. The y coordinates represent the graduation rate, and the x coordinates represent the amount of years.

10. How does the meaning of the coordinates of these points compare? Are they tracking the same things?

They are both tracking the same point of graduation rate with a certain amount of required credits!

11. 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?

All three are below 90% graduation rates while some like Kentucky and Indiana are above 90%!

Graduation Rates

Part 2: Using a Line of Best Fit

Use technology to estimate and draw in a line of fit for the data set.

1. What does the y-intercept and slope of your line of best fit mean for the given context?

Most students graduate with 10-25 credits required with a average of around 85%.

2. 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?

$$R = -0.075$$

This means if it has a positive or negative correlation.

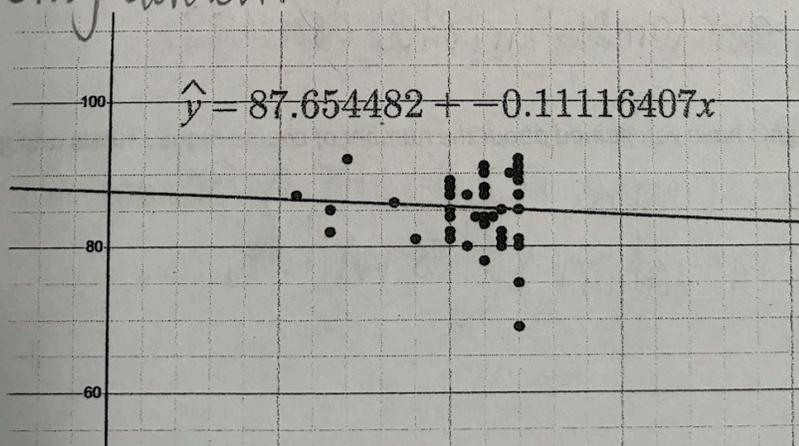
3. Is a linear model a good fit for this set of data?

Yes, these points are almost perfect, in the way that if you draw a linear line of best fit, the plots are almost perfectly even above and below the line.

4. 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 lead to higher graduation rates?

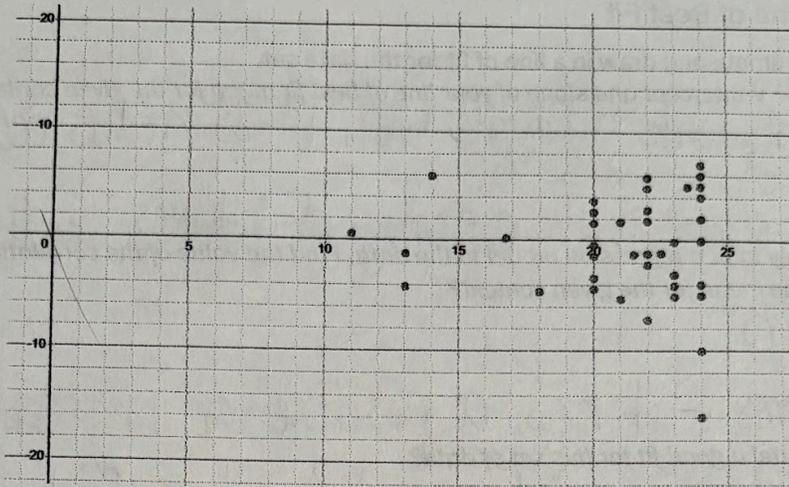
$$R^2 = 0.005$$

This disputes the claim because the line is decreasing as the amount of classes are being added.



5. 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?

Yes, if you draw a linear best fit line, it will be almost perfectly even above and below the line.

6. Which explanatory variable that you have examined impacts high school graduation rates more? Support your answer with appropriate statistics.

The amount of credits required, this is because the more the numbers on the x-axis raise (the credits), the values on the y-axis (graduation rates) decreases.

7. What new insights have you gained about the number of course requirements and graduation rates?

The more course requirements a school offers, the less their graduation rates will be.

Let's do some math! Round 2

- *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?*
- *Using technology, find the value of the coefficient of determination, r^2 . Does the value of r^2 you calculated validate or dispute the claim that a higher average speed of walking causes a decrease in lifespan due to heart disease?*

Part 4: Interpreting the Model

Students consider and discuss this question, using their modeling to support their responses:

- *Do you think that average speed of walking is a good predictor of a shortened lifespan due to heart disease?*

Part 5: Reflecting

Students consider further variables to explore, by discussing there the

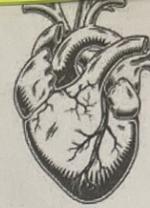
- *What other variables do you think contribute to heart disease related death?*
- *Consider your list of variables. Which ones does your group believe could be a better predictor of a shortened lifespan due to heart disease?*

Task 2: Heart Disease Student Work

Example 1:

Part 1: Develop a Statistical Question

According to a 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?

life style- diet, exercise, sleep, job, ect
~~Hereditary~~ Genetics- history of heart disease in family
STRESS

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?

Older age groups
 ↓
 Genetics

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?

old age- ~~1.5~~
 Genetic- ~~1.5~~ .1
 Researched and found these

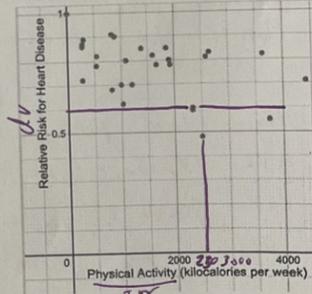
c. What are some statistical questions you can ask about heart disease?

-What groups are most subjective to heart disease?
 -What groups have a higher statistic of 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.

- 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.



1. What do you notice? What do you wonder?

This plot is kind of all over the place, it's definitely higher at the beginning but spikes again at the end why?

2. How might you use this graph to answer the statistical question, "Do higher levels of physical activity lower the risk of heart disease?"

Yes, as long as you don't go over board with it, studies show that 2500-4000 kilocalories per week in physical activity is ideal, below or above that may put you at risk.

Part 3: Manipulate the Model

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.

Line of best fit: $y = 0.0000526541 + 0.84294x$	What is the slope, what does it mean in the context of the problem? max risk = 100% Slope = 0.0000526541 The relationship between risk and the activity, it may be because of higher activity decreases risk.	What is the y-intercept? What does it mean in the context of the problem? #222 (0, 0.843) y-intercept = (4130, 0.025) It's irrelevant, people can't have 0 physical activity, breathing counts.
What is the value of the coefficient of determination, r^2 ? Does the value of r^2 you calculated validate or dispute the claim that a higher level of physical activity lowers the risk of heart disease?		
$r^2 = 0.1988$ dispute, it is very far from 1.		

Part 4: Interpreting the Model

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? Why or why not?

Yes and no, a lot of factors were left out. It seemed like the risk for heart disease only had a few ^{points} that were below the average line, even with higher kcal, it looked like there was a few with higher dots. So therefore, things like weight and age should be incorporated.

Part 5: Reflection

1. What other variables might contribute to heart disease? Make a list here.

Weight, blood pressure, cholesterol, diabetes, smoking, second-hand smoke, exposure, obesity, diet, physical activity, stress levels, and genetics. Underlying health issues also may play a part.

2. Consider your list of variables. Which ones does your group believe could be a better predictor of a shortened lifespan due to heart disease?

Obesity, the worse ones diet is, and/or physical activity, the higher all those other things may get (blood pressure, cholesterol, risk of smoking, stress, ect)

Example 2:

Task 2: Heart Disease

Students are given a problem similar to this.

Part 1: Develop a Statistical Question

According to a 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?

High blood pressure Obesity
Smoking Diabetes Stress

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?

Smokers
People with Obesity

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?

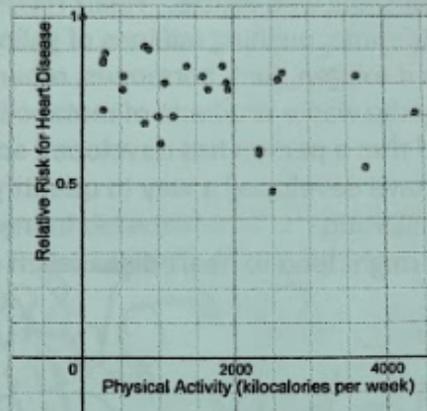
I feel like the risk ratio will be about 4 because both things are really bad and both are leading factors to heart disease.

c. What are some statistical questions you can ask about heart disease?

How many smokers have heart disease?
How many people affected with heart disease have obesity?

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?

I notice that the data is pretty bunched up.

I wonder what the line of best fit would look like?

4. How might you use this graph to answer the statistical question, "Do higher levels of physical activity lower the risk of heart disease?"

I would find the r value, the coefficient of correlation, of the graph to see if the 2 variables correlate.

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.

$$y = 0.00005x + 0.8$$

6. Find the value of the coefficient of determination, r^2 . Does the value of r^2 you calculated validate or dispute the claim that a higher level of physical activity lowers the risk of heart disease?

$$r^2 = 0.1988$$
$$r = -0.4459$$

→ This value disputes the claim that higher physical activity lowers the risk of heart disease.

Part 4: Interpreting the Model

7. 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?

I do not think that physical activity per week is good predictor of relative risk for heart disease.

Part 5: Reflection

8. What other variables do you think contribute to heart disease?

Smoke exposure, unhealthy diet, and high blood pressure.

9. Consider your list of variables. Which ones does your group believe could be a better predictor of a shortened lifespan due to heart disease?

Smoking on the back.

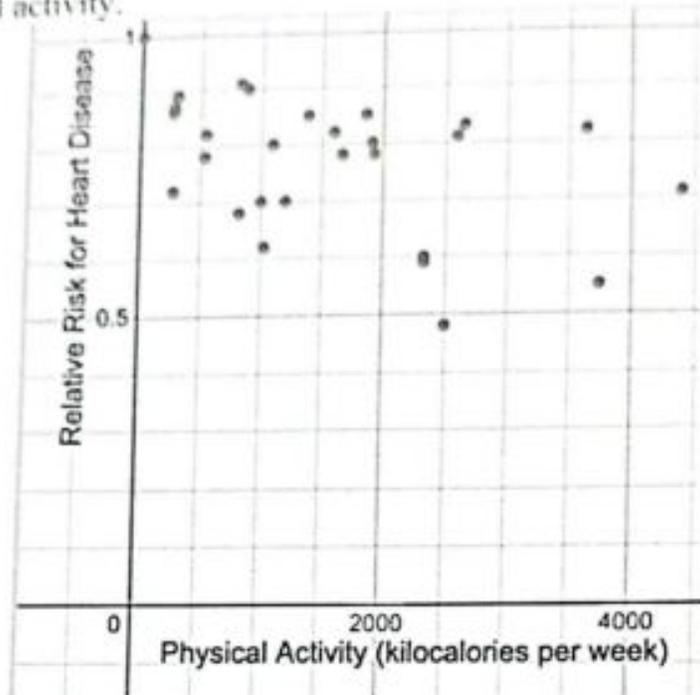
Smoking could be a better predictor for heart disease because according to CDC.gov, Smoking increases the number of plaque in your blood vessels. It can also cause blood to thicken and cause clots. These clots then block blood from reaching your heart which can lead to heart disease. The American Heart Association states, "nearly 20% of cardiovascular disease deaths are caused by cigarette smoking." This shows that smoking would be a better predictor for heart disease than physical activity levels.

Example 3:

<p>Part 1: Develop a Statistical Question</p> <p>According to a <u>Harvard Health Publishing</u> 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.</p>	
<p>1. What are some factors that might lead to heart disease?</p> <p>• Obesity - other diseases smoking</p> <p>Malnourished heredity exercise (lack of)</p> <p>Caffeine</p>	
<p>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.</p>	<p>a. Brainstorm: What are the two groups that could be involved in a statistical study regarding heart disease?</p> <p>Smokers to non smokers</p>
	<p>b. Analyze: What do you anticipate the relative risk ratio will be between the two groups you identified: about 1, greater than 1, or less than 1? How did you arrive at this conclusion?</p> <p>more than 1</p>
	<p>c. Explore: What are some statistical questions you can ask about heart disease?</p> <p>on average, how many people who smoke get heart disease</p>

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

What is the context of the explanatory (independent) variable?

Physical activity

What is the context of the response (dependent) variable?

Risk of heart disease

What type of association is represented by this data (linear, quadratic, power, etc.)?

quadratic

There are multiple data points in the set. Choose one and explain what it means.

if you burn around 2000 kilocalories your risk of heart disease goes down about 50%.

Part 3: Manipulate the Model

There is a link to a representation of the data in Schoology. You should find the link and open it. Use the graph you find to answer the questions below:

How did you decide your line of best fit?

The line that is approximately the same distance away from all the pts

What do the y-intercept and slope mean in the context of this problem?

y-int means that no exercise makes your risk 100%.
Slope - The less exercise higher risk

Using technology, find the value of the coefficient of determination, r^2 . Does the value of r^2 you calculated validate or dispute the claim that a higher average speed of walking causes a decrease in lifespan due to heart disease? Why or why not?

$$r^2 = 0.2$$

it doesn't because it is only 20%

Do you think that average speed of walking is a good predictor of a shortened lifespan due to heart disease? Explain why or why not it is a good predictor.

it is not a good one because there is not a ton of correlation between them

Part 4: Reflecting

What other variables do you think contribute to heart disease related death?

daily habits
eating choices
amount of exercise

Consider your list of variables. Which ones does your group believe could be a better predictor of a shortened lifespan due to heart disease?

heredity
daily habits/choices

Criteria for Success

Conference and Provide Revision Support	Accept with Revision	Accept
The student’s artifact shows an emerging understanding of the expectations of the indicator(s). After conferencing and additional instruction/learning, the student may provide a revised or different artifact as evidence of the indicator(s).	The student’s artifact is approaching a full understanding of the expectations of the indicator(s). The artifact may contain execution errors that should be corrected in revision. The student may revise the selected artifact or submit a different artifact.	The student’s artifact demonstrates evidence that they have met the expectations of the indicator(s).

References

Common Core Standards Writing Team. (2012, April 21). *High School Statistics and Probability*. Progressions for the Common Core State Standards in Mathematics (draft).

<https://achievethecore.org/content/upload/Draft%20High%20School%20Progression%20on%20Statistics%20and%20Probability.pdf>

Guy, N. (2023, January 20). *The Best USB Phone Charger*. Wirecutter New York Times. <https://www.nytimes.com/wirecutter/reviews/best-multiport-usb-wall-charger/>

Illustrative Mathematics. (2016, May 4). *Restaurant Bill and Party Size*. Illustrative Mathematics. <https://tasks.illustrativemathematics.org/content-standards/HSS/ID/B/6/tasks/2046>