Algebra I

Investigate Data

Option #1 Performance Task |   
Teacher Document

Authors: Initiative Team

April 30, 2024

Table of Contents

[Overview of the Performance Task 1](#_Toc171516697)

[Administering the Performance Task and Embedding   
Resources for Students 2](#_Toc171516698)

[Guidance for Task Entry Activities 2](#_Toc171516699)

[Key Vocabulary Associated with the Standards 3](#_Toc171516700)

[Strategies for Supporting Students 3](#_Toc171516701)

[Potential Alternative Means of Expression 4](#_Toc171516702)

[PART 1. Comparing Numerical Data Sets 6](#_Toc171516703)

[Task Alignment to Key Elements of Big Ideas and Standards 6](#_Toc171516704)

[Optional Task Entry Activity for Part 1 8](#_Toc171516705)

[Part 1. Items 12](#_Toc171516706)

[Part 1. Sample Student Responses 18](#_Toc171516707)

[PART 2. Comparing Categorical Data Sets 22](#_Toc171516708)

[Task Alignment to Key Elements of Big Ideas and Standards 22](#_Toc171516709)

[Task Entry Activity for Part 2 24](#_Toc171516710)

[Part 2. Items 25](#_Toc171516711)

[Part 2. Sample Student Responses 29](#_Toc171516712)

[PART 3. Analyzing the Relationship Between Two   
Numerical Variables 32](#_Toc171516713)

[Task Alignment to Key Elements of Big Ideas and Standards 32](#_Toc171516714)

[Task Entry Activity for Part 3 33](#_Toc171516715)

[Part 3. Items 35](#_Toc171516716)

[Part 3. Sample Student Responses 44](#_Toc171516717)

Overview of the Performance Task

This performance task evaluates students’ understanding of key concepts within the Algebra I Investigate Data Big Idea. It is divided into parts, each targeting a specific component of the Big Idea. Each part offers accessible strategies and examples of how students can demonstrate proficiency with the concepts. Various tools, mediums, and connections are provided for teachers to customize the task to the unique needs, cultures, interests, and abilities of their students, promoting an inclusive and relevant educational experience.

When preparing this performance task, distinguish between the flexible and fixed elements to ensure students have multiple ways to demonstrate their knowledge without compromising the concepts’ depth and rigor within the standards. Furthermore, educators should always consult the student’s Individualized Education Program (IEP) to ensure that all required accommodations and supplementary aids are provided during the assessment.

Additional information on providing alternative means of expression can be found in the best practice guides and content-aligned practice briefs defined as part of the California Alternative Means to a Diploma Project.

Administering the Performance Task and  
Embedding Resources for Students

Each part of this task is broken into a series of Items for administration. This section provides guidance to the educator on how to administer each part of the task and support the student in demonstrating their understanding of the Big Idea. As you are planning to administer this performance task, it is suggested to review these recommendations as they offer associated key vocabulary, appropriate and inappropriate resources, and potential methods and means of expression.

Guidance for Task Entry Activities

For each part of this performance task, there are task entry activities designed to create a more relevant and engaging assessment opportunity for students, where they orient themselves to the task by doing their own data investigations. These entry activities are not required for completion of the task, nor does the student need to engage in these activities alone or without support. Completing these activities can be done within a group or with the support of a teacher. The activities center on identifying variables, developing research questions, or collecting data each of which is not being formally assessed in this Big Idea. Task entry activities **do not have to be completed** **to accomplish this task**. It is allowable for the educator to provide students with a pre-existing dataset. Students can either choose to collect their own data (for example, polling friends about their height in inches and shoe size) or use a pre-existing dataset. The table below provides an **optional** data set that could be used in the event a teacher is forgoing the entry activity.

Key Vocabulary Associated with the Standards

The key vocabulary terms provided are essential to the concepts within the Big Ideas, therefore unless otherwise noted, the vocabulary cannot be taught during completion of the task.

* box plot, distribution, skew/skewed, maximum, minimum, tail, center, spread, median, variability, interquartile range

Strategies for Supporting Students

Whether completed after gathering their own data or through analyzing a teacher-created dataset that has been provided as part of the assessment, when working on the actual items of this task, students will engage in the mathematical work focused on the core areas of the Big Idea and therefore should complete this work independently other than with the appropriate resources listed below.

Appropriate Resources

Appropriate resources maintain the rigor of the standards while also accommodating any student difficulties such as confusion or anxiety while providing a resource the student could use to complete the task. If the teacher proceeds with the student using their own data set, the following guidance and support are allowable.

* clarifying terms in the table or how the data are organized (for example, clarifying what is meant by Group 1, Group 2, and Number of Observations; Clarifying Maximum signifies greatest value; explaining the meaning of side-by-side or population)
* supporting students in **accessing resources** to remind them of the terms “median,” “mean,” first quartile,” and “third quartile”

Inappropriate Resources

The inappropriate resources identify what assistance should be avoided as it may alter the rigor of the standards and negatively impact the student’s ability to independently demonstrate proficiency and be objectively scored on that task.

* explaining the definitions of key vocabulary like median or mode is not allowable as understanding these are required to demonstrate proficiency
* giving instructions on each item is also not allowable because it is reteaching the concepts during the assessment and negatively impacts the student’s ability to objectively demonstrate proficiency

Potential Alternative Means of Expression

Potential methods and means of expression show the various ways students can demonstrate their knowledge of the standards being assessed in this part of the task.

* using paper and pen/pencil
* using a spreadsheet
* describing the contents of the table verbally to the teacher
* dictating to a scribe[[1]](#footnote-1)

Students can create box plots by

* using paper and pen/pencil
* using graphing software such as Desmos or Geogebra
* dictating to a scribe

Students can respond to the narrative items in Item 5 (items b, c, and d) by

* using paper and pen/pencil
* typing a response using a word processor
* using speech-to-text software
* dictating to a scribe
* explaining in words to the teacher

PART 1. Comparing Numerical Data Sets

Part 1 of the Investigate Data performance task outlines the following:

* associated standards that will be assessed
* student task requirements
* rubrics that assess each item
* sample student responses

Teachers should familiarize themselves with the related standards, review the student task, explore each item’s rubric, and view the sample student responses to sufficiently prepare students to use this performance task to show proficiency with the Investigate Data Big Idea in Algebra I. Additionally, teachers must be careful to incorporate any IEP-defined supplementary aids and services specific to individual students with disabilities taking this performance task.

As a reminder, task entry activities allow students to engage with and orient to the task by gathering their own data sets based on their interests or values. Entry activities do not have to be completed for students to carry out the items in this part. If the task entry activity is skipped, the teacher must provide a data set having a minimum of twenty-five data points across two different categories.

Task Alignment to Key Elements of Big Ideas and Standards

Clusters of content standards exist within the Big Ideas allowing the Big Ideas to demonstrate the central concepts and key understandings of the course content. The indicator statements provide the teacher with the key concepts being evaluated in each Big Idea as well as the associated content standards centered within the Big Idea of this task and come from the *Mathematics Framework for California Public Schools: Kindergarten Through Grade Twelve* (*Mathematics Framework*) and are aligned to California adopted mathematics state standards.

Investigate Data: Big Ideas Indicator 1

Represent data from two or more data sets with plots, dot plots, histograms, and box plots—comparing and analyzing the center and spread, using technology, and interpreting the results.

Related Standards

The following are standards that align with the indicator statements above:

* **Summarize, represent, and interpret data on a single count or measurement variable.**
  + (*Item 2a*) **S-ID.1** Represent data with plots on the real number line (dot plots, histograms, and box plots).

Investigate Data: Big Idea Indicator 2

Interpret and compare data distributions using center (median, mean) and spread (interquartile range, standard deviation) through the use of technology.

Related Standards

The following are standards that align with the indicator statements above.

* **Summarize, represent, and interpret data on a single count or measurement variable.**
  + (*Items 1, 2b, 2c*) **S-ID.2** Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
  + (*Items 2b, 2c, 2d)* **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).

Optional Task Entry Activity for Part 1

For this task entry activity, students will create a research question pertaining to a numerical value across two different categories. Then, they create a plan for gathering these data and carry out their plan resulting in their gathering the needed data for answering the items of this task.

Step 1 – Task Entry Activity

Decide on a question you would like to answer about a numerical variable across two different categories.

You will need to gather a minimum of twenty-five data points for each group.

Some examples of such questions include

* How does the cost of girls’ haircuts compare to the cost of boys’ haircuts? (numerical value = cost of a haircut; two categories = boys and girls)
* Who tends to own more pairs of shoes, adults or kids? (numerical value = number of pairs of shoes owned; two categories = adults and kids)

As part of this performance task, your teacher may facilitate you working through a series of activities that allow you to create a research question, formulate a plan for gathering data related to this question, and organize the data you gather in productive ways. If you engage with the task entry activities, then the data sets you produce will be what you use for the items in this task. Your teacher also has the option of giving you a pre-created data set.

Use the data set you have gathered yourself or the one you have been given to complete each part of this performance task. Below you will find a table to use as an **optional** tool.

Table 1. Optional Table for Gathering Data – Student Entry Activity

| Observation Number | Category | Value |
| --- | --- | --- |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |
| 11 |  |  |
| 12 |  |  |
| 13 |  |  |
| 14 |  |  |
| 15 |  |  |
| 16 |  |  |
| 17 |  |  |
| 18 |  |  |
| 19 |  |  |
| 20 |  |  |
| 21 |  |  |
| 22 |  |  |
| 23 |  |  |
| 24 |  |  |
| 25 |  |  |

Criteria for Progressing to the Next Item

A student can proceed to Item 2 when they have identified a question considering how a numerical variable varies across two different categories and populations while aligning with the variable on feasibly collected data.

Step 2 – Task Entry Activity

Come up with a plan for how you will gather and record your data.

You will need to gather a minimum of twenty-five data points for each group.

Explain your plan, including

* Who will you survey?
* How will you contact them?
* How will you record and store data?
* How will you organize your data?
* Do you foresee any problems gathering this type of data?

Criteria for Progressing to the Next Item

A student is ready to go on to Item 3 when (with support from a teacher or a small group of peers) the student has clear, easy-to-understand, and contextually relevant survey questions and describes a reasonable and coherent plan for gathering and recording data, including

* Who will they survey?
* What means or tools will they use the conduct the survey?
* How will they record and store the data?
* Do they identify any potential problems or challenges related to their data collection plan?

If the student would prefer not to collect their own data, they can use the data in the table below. The survey question used to generate this data was, “How much did you pay for your most recent haircut?” with the two categories being girls and boys.

Table 2. Sample Student-Collected Data

| Observation number | Category  (boy or girl) | Value  (cost of most recent haircut) |
| --- | --- | --- |
| 1 | Girl | 20 |
| 2 | Boy | 30 |
| 3 | Boy | 10 |
| 4 | Girl | 30 |
| 5 | Boy | 25 |
| 6 | Boy | 10 |
| 7 | Girl | 30 |
| 8 | Girl | 0 |
| 9 | Boy | 35 |
| 10 | Girl | 25 |
| 11 | Girl | 10 |
| 12 | Girl | 55 |
| 13 | Boy | 0 |
| 14 | Girl | 20 |
| 15 | Girl | 90 |
| 16 | Girl | 65 |
| 17 | Girl | 125 |
| 18 | Boy | 10 |
| 19 | Boy | 17 |
| 20 | Girl | 75 |
| 21 | Girl | 15 |
| 22 | Boy | 25 |
| 23 | Boy | 25 |
| 24 | Girl | 150 |
| 25 | Girl | 75 |

Step 3 – Task Entry Activity

Gather and record your data.

Criteria for Progressing to the Next Item

Unless a data set is already provided, the student is ready to go on to start the task when they have collected at least twenty-five data points that align with the numerical variable and populations they selected in Item 1 and the data are organized in a clear and usable way.

Part 1. Items

The items below must be completed independently by the student. They can use a given data set **or** one gathered during their task entry activity. Remember, for data sets to be viable, they must have at least two data points for each category.

Item 1

Item 1 has no sub-items.

Item 1 Task

You will use the same data set for these items to understand and organize data. Use your data to complete the student document table 1 below (**S-ID.2**).

Use your data to complete the table below.

Student Document Table 2. Observational Data for Group 1 and Group 2

|  | Group 1 | Group 2 |
| --- | --- | --- |
| Number of observations |  |  |
| Minimum |  |  |
| Maximum |  |  |
| First quartile |  |  |
| Median |  |  |
| Third quartile |  |  |
| Mean |  |  |

A Rubric for Assessing a Response to Item 1

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

Rubric for Task Item 1

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student correctly identifies or calculates fewer than six of the minimum, maximum, median, and mean for each population.  The student correctly identifies fewer than half of the remaining statistics in the table. | The student correctly identifies or calculates at least six values out of the minimum, maximum, median, and mean for each population.  The student correctly identifies or calculates at least half of the remaining statistics in the table (number of observations, first and third quartiles) for both populations. | The student correctly identifies or calculates the minimum, maximum, median, and mean for both populations.  The student correctly identifies or calculates four of the six remaining statistics in the table (number of observations, first and third quartiles) for both populations. |

Item 2

Item 2 has four sub-items. Please be sure the student completes each. Represent and analyze data (**S-ID.1, S-ID.2, S-ID.3**).

Item 2a [Student Document (A)]

Using the minimum, maximum, quartiles, and median, create two side-by-side box plots to compare your numerical variable between the two groups.

A Rubric for Assessing a Response to Item 2a

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

Rubric for Item 2a

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student does not generate two plausibly correct box plots. | The student generates two plausibly correct box plots with statistics labeled but statistics, variables, units, or scale are not labeled or are incorrect —OR— the student correctly generates one of the two box plots with the numerical axes correctly labeled with variables, units, and scale —OR— the student correctly generates one of the two box plots with minimum, maximum, quartiles, and median all accurately indicated. | The student correctly generates two box plots, one for each of the two populations, including   * the numerical axes are correctly labeled with variable, units, and scale, and * the minimum, maximum, quartiles, and median are all accurately indicated for both populations.   There are no major conceptual errors and no more than 2–3 minor procedural errors (that is, via calculation, transcription, or placement). |

Item 2b [Student Document (B)]

How would you describe the difference in the variable between the two groups? Be sure you discuss differences/similarities in shape, center, and spread.

A Rubric for Assessing a Response to Item 2b

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

**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).

Rubric for Item 2b

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student attempts to describe similarities and differences between the two box plots but there are multiple major conceptual errors. | The student correctly describes at least one way that the two box plots are statistically similar.  The student correctly describes at least one way that the two box plots are statistically different.  If there are no similarities or differences, the student acknowledges this.  The student may include one major conceptual error, but overall, the response is generally conceptually correct. | The student correctly describes any ways that the two box plots are statistically similar by indicating, for example,   * center, * spread, * median, and/or * interquartile range.   The student correctly describes any ways that the two box plots are statistically different.  Any significant statistical similarities or differences between the two populations for the variable are addressed.  There are no major conceptual errors and no more than 2–3 minor procedural errors. |

Item 2c [Student Document (C)]

For each group, which is greater—the mean or median? Why? Explain your reasoning. If the mean and median are the same for each population, explain why.

A Rubric for Assessing a Response to Item 2c

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

**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).

Rubric for Item 2c

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student attempts to describe the relationship between the mean and median for at least one of the two populations but there are multiple major conceptual errors —OR— the explanation is so unclear and imprecise that it is not possible to determine the student’s level of conceptual understanding. | The student correctly describes the relationship between mean and median for at least one of the two populations.  The explanation may lack clarity, specificity, or precise language regarding the connection to skew, extreme values, or the difference between median and mean.  The student may include one major conceptual error, but overall, the response is generally conceptually correct. | The student correctly identifies for each population whether the mean or median is greater (or that they are equal).  The student correctly explains the connection between mean/median and skew, for example,   * if the data is skewed to the right, the mean will be greater than the median and vice versa * extreme values “pull” the mean further toward those extremes * because the median is the fiftieth percentile, it is not “pulled” toward extreme values the way the mean is   If the two populations both have equal mean and median, the student correctly explains the connection between this fact and the symmetry of the data.  There are no major conceptual errors and no more than 2–3 minor procedural errors. |

Item 2d [Student Document (D)]

Is the median or mean a more appropriate choice for describing the “centers” of these two distributions? Explain your reasoning.

A Rubric for Assessing a Response to Item 2d

**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).

Rubric for Item 2d

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student attempts to describe whether mean or median is more appropriate for at least one of the two populations but there are multiple major conceptual errors —OR— the explanation is so unclear and imprecise that it is not possible to determine the student’s level of conceptual understanding. | The student correctly identifies for at least one population whether the mean or median is more appropriate for describing the center of that distribution.  The explanation may lack clarity, specificity, or precise language regarding the connection to skew, extreme values, or the difference between median and mean.  The student may include one major conceptual error, but overall, the response is generally conceptually correct. | The student correctly identifies for each population whether the mean or median is more appropriate for describing the center of that distribution.  The explanation includes some reference to extreme values (for example, if values are more extreme in one direction than the other, then the median will generally be more appropriate).  If both populations have mean and median equal,   * the student acknowledges that it does not matter which to use * the student refers to the symmetric nature of the data. |

Part 1. Sample Student Responses

The content below provides a sample of proficient responses from a student. The text that leads with “Student Voice” is an example of how a student might respond to each item. This section should only serve as a model—different students will arrive at solutions in different ways.

Task Entry Activity

The task entry activity is not required, and therefore, should not be scored. Data sets either produced through this activity or directly provided to the student for the completion of task items should incorporate at least twenty-five data points for each of the two data categories.

Item 1

Use your data to complete the table below.

Table 3. [Student Document Table 1.] Sample Student-Generated Table

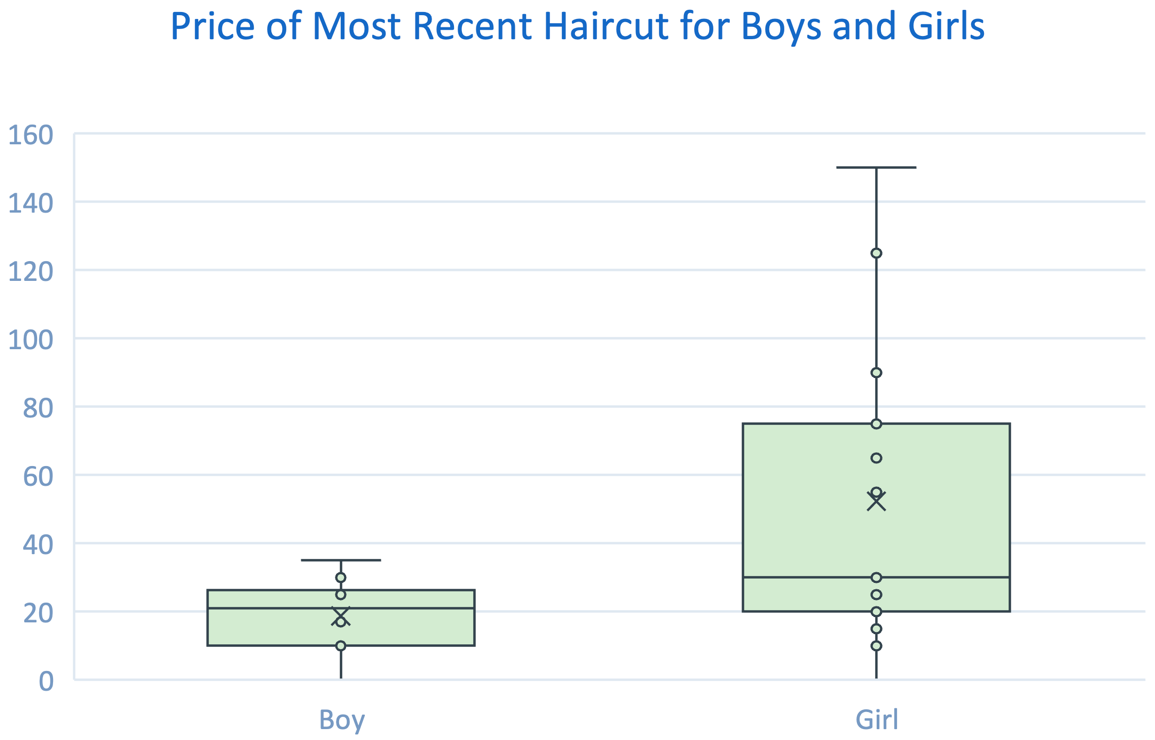
|  |  |  |
| --- | --- | --- |
|  | Group 1: Girls | Group 2: Boys |
| Number of observations | 75 | 24 |
| Minimum | 0 | 0 |
| Maximum | 150 | 35 |
| First quartile | 20 | 10 |
| Median | 30 | 21 |
| Third quartile | 75 | 25 |
| Mean | 52.33 | 18.7 |

Item 2

Item 2a [Student Document (A)]

Using the minimum, maximum, quartiles, and median, create two side-by-side box plots to compare your numerical variable between the two groups.

Figure 1 Sample Student-Generated Box Plot



Item 2b [Student Document (B)]

How would you describe the difference in the variable between the two groups? Be sure you discuss differences/similarities in shape, center, and spread.

Student Voice: The box plot for girls is significantly skewed to the right (upward). It makes sense that most haircuts will not cost too much, but a few students will spend a large amount. This is why the average is so much higher than the median. Since the cost will always be a positive number, the minimum cannot be less than 0 and there is a long right tail. The box plot for boys is slightly skewed to the left (downward) meaning there are a few more extreme lower values (0 and 10). For boys, the mean and median are close to the average just a little less than the median. This tells us that the distribution is closer to normal/symmetrical than the distribution for girls. The centers and spreads are quite different. The median cost for girls’ haircuts is 50% more than for boys, and there is much more variability in the haircut costs for girls. The interquartile range (IQR) for girls is $55, while for boys it is only $15.

Item 2c [Student Document (C)]

For each group, which is greater—the mean or median? Why? Explain your reasoning. If the mean and median are the same for each population, explain why.

Student Voice: We should not be surprised that the mean is larger than the median for girls because the distribution appears to be skewed to the right. The mean averages all the values in the data, so is “pulled” toward the high ones. The median is the 50th percentile and is resistant to extreme values. For boys, there are not as many extreme values in either direction, so the mean and median are not that different, but the mean is just slightly less than the median.

Item 2d [Student Document (D)]

Is the median or mean a more appropriate choice for describing the “centers” of these two distributions? Explain your reasoning

Student Voice: Since the median gives a better description of the center, or a “typical” haircut cost, it is more appropriate. For girls, the median is $30, indicating that half of women spent $3o or less, but the mean haircut cost for girls is $52.33. So the mean doesn’t give us a good idea of what we could expect for a typical girl’s haircut cost. It is best to only use the mean when the data distribution is reasonably symmetric in shape. With the boys in this population, there were not that many extreme values, so it doesn’t hugely matter which you use, though the median still gives a better idea of “center.”

PART 2. Comparing Categorical Data Sets

Part 2 of the Investigate Data performance task outlines the following:

* associated standards that will be assessed
* student task requirements
* rubrics that assess each item
* sample student responses

Teachers should familiarize themselves with the related standards, review the student task, explore each item’s rubric, and view the sample student responses to sufficiently prepare students to use this performance task to show proficiency with the Investigate Data Big Idea in Algebra I. As a reminder, task entry activities allow students to engage with and orient to the task by gathering their own data sets based on their interests or values. Entry activities do not have to be completed for students to carry out the items in this part. If the task entry activity is skipped, the teacher must provide a data set having a minimum of twenty-five data points across two different categories.

Task Alignment to Key Elements of Big Ideas and Standards

Clusters of content standards exist within the Big Ideas allowing the Big Ideas to demonstrate the central concepts and key understandings of the course content. The indicator statements provide the teacher with the key concepts being evaluated in each Big Idea as well as the associated content standards centered within the Big Ideas of this task and come from the 2023 California Mathematics Framework and are aligned to California adopted mathematics state standards.

Investigate Data: Big Ideas Indicator 1

Represent data from two or more data sets with plots, dot plots, histograms, and box plots, comparing and analyzing the center and spread, using technology, and interpreting the results.

Related Standards

The following are standards that align with the indicator statement above:

* **Summarize, represent, and interpret data on a single count or measurement variable.**
  + (*Items 1, 2a, 2b, 2c*) **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.

Investigate Data: Big Idea Indicator 2

Interpret and compare data distributions using center (median, mean) and spread (interquartile range, standard deviation) through the use of technology.

Related Standards

The following are standards that align with the indicator statement above:

* **Summarize, represent, and interpret data on a single count or measurement variable.**
  + (*Items 1, 2a, 2b, 2c*) **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.

Task Entry Activity for Part 2

For this task entry activity, students will create a research question pertaining to a numerical value across two different categories. Then they create a plan for gathering these data and carry out their plan resulting in their gathering the needed data for answering the items of this task.

Step 1 – Task Entry Activity

Select a question about which you’d like to poll two different populations. The question should be something that people can answer “Yes/Support,” “No/Don’t Support,” or “No Opinion.” You will need to gather a minimum of twenty-five data points for each group.

* Do you think we should lengthen the school day by 30 minutes? (girls versus boys, students versus teachers, athletes versus non-athletes)
* Do you think kids should be able to vote starting at age 16? (girls versus boys, kids versus adults, parents versus non-parents)

Try to select a question and population where the responses of each group aren’t very predictable.

Step 2 – Task Entry Activity

Come up with a plan for how you will gather and record your data. You will need to gather a minimum of twenty-five data points for each group.

* Who will you survey?
* How will you contact them?
* How will you record and store data?
* How will you organize your data?
* Do you foresee any problems gathering this type of data?

Step 3 – Task Entry Activity

Follow your plan and gather your data.

Part 2. Items

The items below must be completed independently by the student. They can use a given data set **or** one gathered during their task entry activity. Remember, for data sets to be viable, they must have at least two data points for each category.

Item 1

Item 1 has no sub-items.

Item 1 Task

Use your data to complete the two-way frequency table below (**S-ID.5**).

Figure 2. [Student Document Figure 1] Two-Way Frequency Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Opinion on Question:** No/Oppose | **Opinion on Question:** Yes/Support | **Opinion on Question:** No Opinion | **Totals** |
| **Population:** Group 1 |  |  |  |  |
| **Population:** Group 2 |  |  |  |  |
| **Totals** |  |  |  | **Grand Total:** |

A Rubric for Assessing a Response to Item 1

**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.

Rubric for Item 1

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student attempts to correctly complete a two-way frequency table using the data they collected, but there are significant errors that suggest deeper conceptual misconceptions. | The student ***mostly*** correctly completes a two-way frequency table; some data may be missing or there may be 2–3 significant mismatches with the collected data that need to be corrected.  While there may be errors, it is clear that the student has a conceptual understanding of how to represent data in a two-way frequency table. | The student correctly completes a two-way frequency table.  There are no major errors but a few minor transcription errors are acceptable. |

Item 2

Item 2 has three sub-items. Answer each by analyzing the data from your two-way table and discussing what conclusions you can draw from your data and table. The standard does not require students to calculate these values by hand; using a calculator or spreadsheet is acceptable (**S-ID.5**).

Item 2a [Student Document (A)]

Calculate three joint relative frequencies from your table and explain their significance in the context of the question you’ve asked.

A Rubric for Assessing a Response to Item 2a

**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.

Rubric for Item 2a

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student correctly calculates fewer than two joint relative frequencies from the table.  The explanations are missing or it is not clear from the explanations that the student understands the significance of the joint relative frequencies within the context of the original question. | The student correctly calculates **two** of the three joint relative frequencies from the table.  The student provides ***generally***correct explanations of the significance of the joint relative frequencies they calculated, though the explanations may lack clarity, precision, or thoroughness. | The student correctly calculates **three** joint relative frequencies from the table.  The student correctly explains the significance of each of the joint relative frequencies they calculated within the context of the original question. |

Item 2b [Student Document (B)]

Calculate two marginal relative frequencies from your table and explain their significance in context.

A Rubric for Assessing a Response to Item 2b

**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.

Rubric for Item 2b

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student does not correctly calculate a marginal relative frequency from the table.  The explanations are missing, or it is not clear from the explanations that the student understands the significance of the marginal relative frequencies within the context of the original question. | The student correctly calculates **one** of the marginal relative frequencies from the table.  The student provides ***generally***correct explanations of the significance of the marginal relative frequencies they calculated, though the explanations may lack clarity, precision, or thoroughness. | The student correctly calculates **two** marginal relative frequencies from the table.  The student correctly explains the significance of each of the marginal relative frequencies they calculated within the context of the original question. |

Item 2c [Student Document (C)]

Discuss the conditional frequencies in your table and their significance. Which ones surprised you? Which ones did not? Why? What conclusions can you draw based on this data?

A Rubric for Assessing a Response to Item 2c

**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.

Rubric for Item 2c

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student does not correctly calculate a conditional frequency from the table.  The explanations are missing, or it is not clear from the explanations that the student understands the significance of the conditional frequencies within the context of the original question. | The student correctly calculates **one** of the conditional frequencies from the table.  The student provides ***generally***correct explanations of the significance of the conditional frequencies they calculated, though the explanations may lack clarity, precision, or thoroughness. | The student correctly calculates **both** conditional frequencies from the table.  The student correctly explains the significance of each of the conditional frequencies they calculated within the context of the original question. |

Part 2. Sample Student Responses

The content below provides a sample of proficient responses from a student. The text that leads with “student voice” is an example of how a student might respond to each item. This section should only serve as a model—different students will arrive at solutions in different ways.

Task Entry Activity

The task entry activity is not required, and therefore, it should not be scored. Data sets either produced through this activity or directly provided to the student for the completion of task items should incorporate at least twenty-five data points for each of the two data categories.

Item 1

Use your data to complete the two-way frequency table below.

Figure 3. [Student Document Figure 1.] Sample Student-Generated   
Two-Way Frequency Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Opinion on Question:** No/Oppose | **Opinion on Question:** Yes/Support | **Opinion on Question:** No Opinion | **Totals** |
| **Population:**  **Group 1: Students** | **42** | **79** | **12** | **133** |
| **Population:**  **Group 2: Teachers** | **23** | **14** | **3** | **40** |
| **Totals** | **65** | **93** | **15** | **Grand Total:**  **173** |

Item 2

Analyze the data from your two-way table and discuss what conclusions you can draw from your data and table.

Item 2a [Student Document (A)]

Calculate three joint relative frequencies from your table and explain their significance in the context of the question you’ve asked.

Student Voice: Examples of joint relative frequencies from the table include (a) 42 of the 173 respondents were students who oppose shifting the school day; (b) 12 of the 173 respondents were students who had no opinion on shifting the school day; and (c) *f* the 173 respondents were teachers who support shifting the school day.

Item 2b [Student Document (B)]

Calculate two marginal relative frequencies from your table and explain their significance in context.

Student Voice: Examples of marginal relative frequencies from the table include:

* Students polled: 133/173. You can tell that most of the respondents were students, so the responses of students may be more representative of the entire population of students.

Respondents who support shifting the school day: 93/173. You can see that a little more than half of everyone who responded said they would support shifting the school day.

Item 2c [Student Document (C)]

Discuss the conditional frequencies in your table and their significance. Which ones surprised you? Which ones did not? Why? What conclusions can you draw based on this data?

Student Voice: Examples of conditional relative frequencies from the table include:

* Students: 79/133 support, 42/133 do not support, 12/133 not sure/no opinion à Over half of the students would support shifting the school day one hour later. I am not surprised because many kids feel that school starts too early and think they would be more rested and alert if it started later.
* Teachers: 14/40 support, 23/40 do not support, 3/40 not sure/no opinion à Less than half of teachers support shifting the school day one hour later. I am not surprised by this because adults probably don’t mind getting up early as much as kids and they are more used to this schedule. They also might like getting home earlier and would not want to work later. I am surprised only slightly over half of teachers do not support shifting the school day, I thought it would be much higher.

PART 3. Analyzing the Relationship Between Two Numerical Variables

Part 3 of the Investigate Data performance task outlines the following:

* associated standards that will be assessed
* student task requirements
* rubrics that assess each item
* sample student responses

Teachers should familiarize themselves with the related standards, review the student task, explore each item’s rubric, and view the sample student responses to sufficiently prepare students to use this performance task to show proficiency with the Investigate Data Big Idea in Algebra I. As a reminder, task entry activities allow students to engage with and orient to the task by gathering their own data sets based on their interests or values. Entry activities do not have to be completed for students to carry out the items in this part. If the task entry activity is skipped, the teacher must provide a data set having a minimum of twenty-five data points across two different categories.

Task Alignment to Key Elements of Big Ideas and Standards

Clusters of content standards exist within the Big Ideas allowing the Big Ideas to demonstrate the central concepts and key understandings of the course content. The indicator statements provide the teacher with the key concepts being evaluated in each Big Idea as well as the associated content standards centered within the Big Ideas of this task and come from the 2023 California Mathematics Framework and are aligned to California adopted mathematics state standards.

Investigate Data: Big Idea Indicator 1

Represent data from two or more data sets with plots, dot plots, histograms, and box plots, comparing and analyzing the center and spread, using technology, and interpreting the results.

Related Standards

The following are standards that align with the indicator statement above:

* **Summarize, represent, and interpret data on a single count or measurement variable.** 
  + (*Items 1, 2a, 2b, 2c, 2d*) **S-ID.6** Represent data on two quantitative variables on a scatter plot and describe how the variables are related.
    - (*Items 2a, 2b, 2c, 2d*) **S-ID.6.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.
    - (*Item 2a*) **S-ID.6.b** Informally assess the fit of a function by plotting and analyzing residuals.
    - (*Item 2a*) **S-ID.6.c** Fit a linear function for a scatter plot that suggests a linear association.

Task Entry Activity for Part 3

For this task entry activity, students will create a research question pertaining to a numerical value across two different categories. Then they create a plan for gathering these data and carry out their plan resulting in their gathering the needed data for answering the items of this task.

Step 1 – Task Entry Activity

Select two numerical variables that you suspect are related to one another, and generate a hypothesis (something you think might be true) about the relationship between the two variables.

You will need to gather a minimum of twenty-five data points for each group.

Some examples include

* “I think that as the years go on, the world record for [the marathon; the 100-meter dash] keeps steadily decreasing” (numerical variables = “year” and “world record [in minutes; in seconds]).”
* “I wonder if there is a clear relationship between the year and global temperature” (numerical variables = “year” and “average global temperature [in Celsius; in Fahrenheit]).”
* “I predict that as pollution levels increase, so does the global temperature.” (numerical variables = “parts per million of [carbon, methane] in the atmosphere” and “average global temperature [in Celsius; in Fahrenheit]).”

Table 5. Sample Table for Step 1 Task Entry Activity

|  |  |
| --- | --- |
| Variables | Units |
|  |  |
|  |  |

Step 2 – Task Entry Activity

Come up with a plan for how you will gather and record your data.

* Will you need to survey people? If so, how will you contact them?
* Can you gather your data from the internet? If so, where will you find it? How will you ensure that it is accurate?
* How will you record and store data?
* How will you organize your data?
* Do you foresee any problems gathering this type of data?

You will need to gather a minimum of twenty-five data points for each group.

Step 3 – Task Entry Activity

Gather your data and organize it in a table like the one below.

Table 6. Sample Table for Step 3

|  |  |  |
| --- | --- | --- |
| Observation # | Variable 1 | Variable 2 |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| (more) |  |  |

Part 3. Items

The items below must be completed independently by the student. They can use a given data set **or** one gathered during their task entry activity. Remember, for data sets to be viable, they must have at least two data points for each category.

Item 1

Item has no sub-items. This item should be completed using data sets gathered in the task entry activity or that have been provided to you by your teacher. (**S-ID.6**).

Item 1 Task

Use the data gathered through the entry activity for this task or the data set provided by your teacher to create a scatter plot. Be sure to label your axes with variables and units.

A Rubric for Assessing a Response to Item 1

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

Rubric for Item 1

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The scatter plot cannot be said to be ***mostly***correct;there may be a major conceptual error or more than three minor errors in labeling axes, scales, or variables.  More than 30% of the points are plotted incorrectly. | The student generates a scatter plot from the data that is ***mostly***correct.  There are no major conceptual errors, though there may be up to three minor errors in labeling axes, scales, or variables.  Up to 30% of the points may be plotted incorrectly due to data entry or transcription errors. | The student generates a scatter plot from the data and correctly labels axes, scales, and variables.  Up to 20% of the points may be plotted incorrectly due to data entry or transcription errors. |

Item 2

Item 2 has four sub-items. Analyze the data illustrated by your scatter plot to complete each item (**S-ID.6a, b, c**).

Item 2a [Student Document (A)]

Is a linear model a good fit for your data? Explain, commenting on the strength and direction of the association. Use residuals as part of your explanation.

A Rubric for Assessing a Response to Item 2a

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

* **S-ID.6.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.
* **S-ID.6.b** Informally assess the fit of a function by plotting and analyzing residuals.
* **S-ID.6.c** Fit a linear function for a scatter plot that suggests a linear association.

Rubric for Item 2a

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student does not accurately identify whether a linear model is a good fit for their data.  The student does not generate a generally accurate residuals plot; there is at least one major conceptual error and/or more than a small number of minor procedural errors.  The student does not correctly use the residual plot to justify their conclusion about whether a linear model is a good fit, or it is not clear from the explanation that they understand how to use a residual plot in that way.  The student does not accurately comment on the strength of the association, or it is not clear from the explanation that they understand how to determine the strength of association from the scatter plot.  The student does not accurately comment on the direction of the association, or it is not clear from the explanation that they understand how to determine the direction of association from the scatter plot. | The student accurately identifies whether a linear model is a good fit for their data or not.  The student generates a ***generally*** accurate residual plot with no major conceptual errors, though a small number of minor procedural errors may be present.  The student uses the residual plot to justify their conclusion about whether a linear model is a good fit, though the explanation may lack clarity, specificity, or thoroughness.  The student accurately comments on the strength of the association (weak, moderate, strong), though the explanation may lack clarity or specificity.  The student accurately comments on the direction of the association (positive, negative), though the explanation may lack clarity or specificity. | The student accurately identifies whether a linear model is a good fit for their data or not.  The student generates an accurate residuals plot, including labeling axes, variables, and scale, and uses it to clearly justify their conclusion about whether a linear model is a good fit.  The student clearly and accurately comments on the strength of the association (weak, moderate, strong).  The student clearly and accurately comments on the direction of the association (positive, negative). |

Item 2b [Student Document (B)]

If you concluded in Item 2a [Student Document (A)] that a linear model is **not** a good fit for your data, please use the following data, scatter plot, and model instead to answer this question. If it is a good fit continue to use the provided data to answer this question.

What does the slope of your linear model mean in the context of the dataset you are using?

Figure 4. [Student Document Figure 2. Optional Scatter Plot Example (if linear model is not a good fit)] Sample Slope of Linear Model

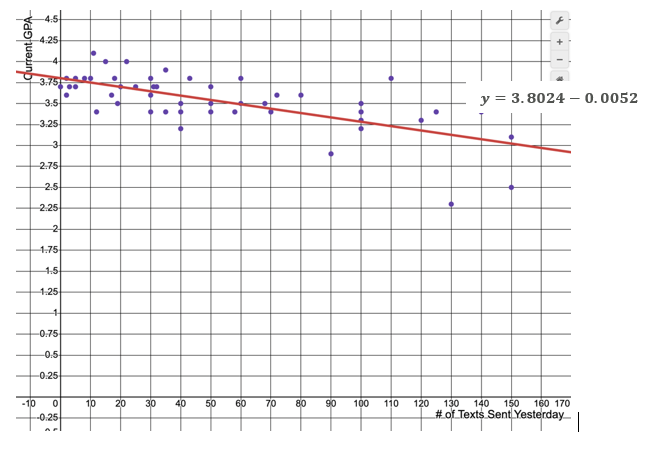


Table 7. [Student Document Table 2. Optional Data Tied to Scatter Plot (if linear model is not a good fit)] Sample Data Set

| Students | Number of texts sent | Current GPA |
| --- | --- | --- |
| 1 | 0 | 3.7 |
| 2 | 2 | 3.8 |
| 3 | 2 | 3.6 |
| 4 | 3 | 3.7 |
| 5 | 5 | 3.8 |
| 6 | 5 | 3.7 |
| 7 | 8 | 3.8 |
| 8 | 20 | 3.7 |
| 9 | 10 | 3.8 |
| 10 | 11 | 4.1 |
| 11 | 15 | 4 |
| 12 | 17 | 3.6 |
| 13 | 18 | 3.8 |
| 14 | 19 | 3.5 |
| 15 | 20 | 3.7 |
| 16 | 22 | 4 |
| 17 | 25 | 3.7 |
| 18 | 130 | 2.3 |
| 19 | 30 | 3.6 |
| 20 | 30 | 3.8 |
| 21 | 30 | 3.4 |
| 22 | 31 | 3.7 |
| 23 | 32 | 3.7 |
| 24 | 35 | 3.9 |
| 25 | 35 | 3.4 |
| 26 | 40 | 3.5 |
| 27 | 40 | 3.2 |
| 28 | 43 | 3.8 |
| 29 | 12 | 3.4 |
| 30 | 50 | 3.5 |
| 31 | 50 | 3.7 |
| 32 | 125 | 3.4 |
| 33 | 50 | 3.4 |
| 34 | 58 | 3.4 |
| 35 | 60 | 3.5 |
| 36 | 68 | 3.5 |
| 37 | 70 | 3.4 |
| 38 | 72 | 3.6 |
| 39 | 80 | 3.6 |
| 40 | 90 | 2.9 |
| 41 | 100 | 3.5 |
| 42 | 100 | 3.4 |
| 43 | 100 | 3.3 |
| 44 | 100 | 3.2 |
| 45 | 110 | 3.8 |
| 46 | 120 | 3.3 |
| 47 | 140 | 3.4 |
| 48 | 150 | 2.5 |
| 49 | 60 | 3.8 |
| 50 | 30 | 3.4 |
| 51 | 150 | 3.1 |
| 52 | 40 | 3.4 |

A Rubric for Assessing a Response to Item 2b

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

* **S-ID.6.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.

Rubric for Item 2b

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| It is not clear that the student understands the meaning of the slope of the linear model in the context of the dataset. | The student ***generally***correctly identifies the meaning of the slope of the linear model, but the explanation may lack clarity, precision, or specificity (for example, “It means that another text loses -0.005”). | The student clearly and correctly identifies the meaning of the slope of the linear model in the context of the dataset (for example, *m* = -0.005 means that for every additional text a student sent yesterday, the model predicts a decrease in their GPA of -0.005). |

Item 2c [Student Document (C)]

What is the vertical intercept of the function's graph? What does it mean in the context of the variable you’ve plotted on the *y*-axis?

A Rubric for Assessing a Response to Item 2c

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

* **S-ID.6.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.

Rubric for Item 2c

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| The student does not correctly identify the  *y*-intercept  It is not clear that the student understands the meaning of the slope of the linear model in the context of the dataset. | The student correctly identifies the *y*-intercept of the linear model.  The student ***generally*** correctly identifies the meaning of the slope of the linear model, but the explanation may lack clarity, precision, or specificity (for example, “It means that another text loses -0.005”). | The student correctly identifies the *y*-intercept of the linear model.  The student clearly and correctly identifies the meaning of the *y*-intercept of the linear model in the context of the dataset (for example, *b* = 3.8024, so the *y*-intercept is (0, 3.8024); this means that for someone who sent zero tests yesterday, the model predicts a GPA of 3.8024 (or about 3.8)). |

Item 2d [Student Document (D)]

Suppose your dataset was missing a particular *x*-value. How could you use your model to predict the *y*-value we should expect for that *x*-value? Give an example either from your data set or from the “Text Messages versus GPA” data above.

A Rubric for Assessing a Response to Item 2d

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

* **S-ID.6.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.

Rubric for Item 2d

| Attempted | Approaching | Proficient |
| --- | --- | --- |
| It is not clear that the student understands the meaning of their explanation —OR— the student provides an example that they misunderstand how the model could be used to predict the  *y-*value that corresponds to a missing *x*-value. | The student gives a ***generally*** correct explanation for how the model could be used to predict the *y*-value for a missing  *x*-value (for example, “You can put the *x*-value in the equation”).  The student includes an example in which the strategy is executed with no more than 1–2 minor procedural errors —OR the student gives a clear, mathematically sound explanation for how the model could be used to predict the  *y-*value for a missing  *x-*value but does not give an example —OR— the student gives a fully correct example but does not provide an explanation. | The student gives a clear, mathematically sound explanation for how the model could be used to predict the *y-*value for a missing *x-*value and includes an example in which the strategy is executed with no more 1–2 minor procedural errors (for example, “You can substitute the missing *x*-value into the linear model to find the corresponding *y*-value; no one in the text messaging data reported sending 160 texts but if they had, the model predicts that they would have a GPA of *y* = 3.8024 – 0.0052 \* 160 = 2.9704). |

Part 3. Sample Student Responses

The content below provides a sample of proficient responses from a student. The text that leads with “Student Voice” is an example of how a student might respond to each item. This section should only serve as a model—different students will arrive at solutions in different ways.

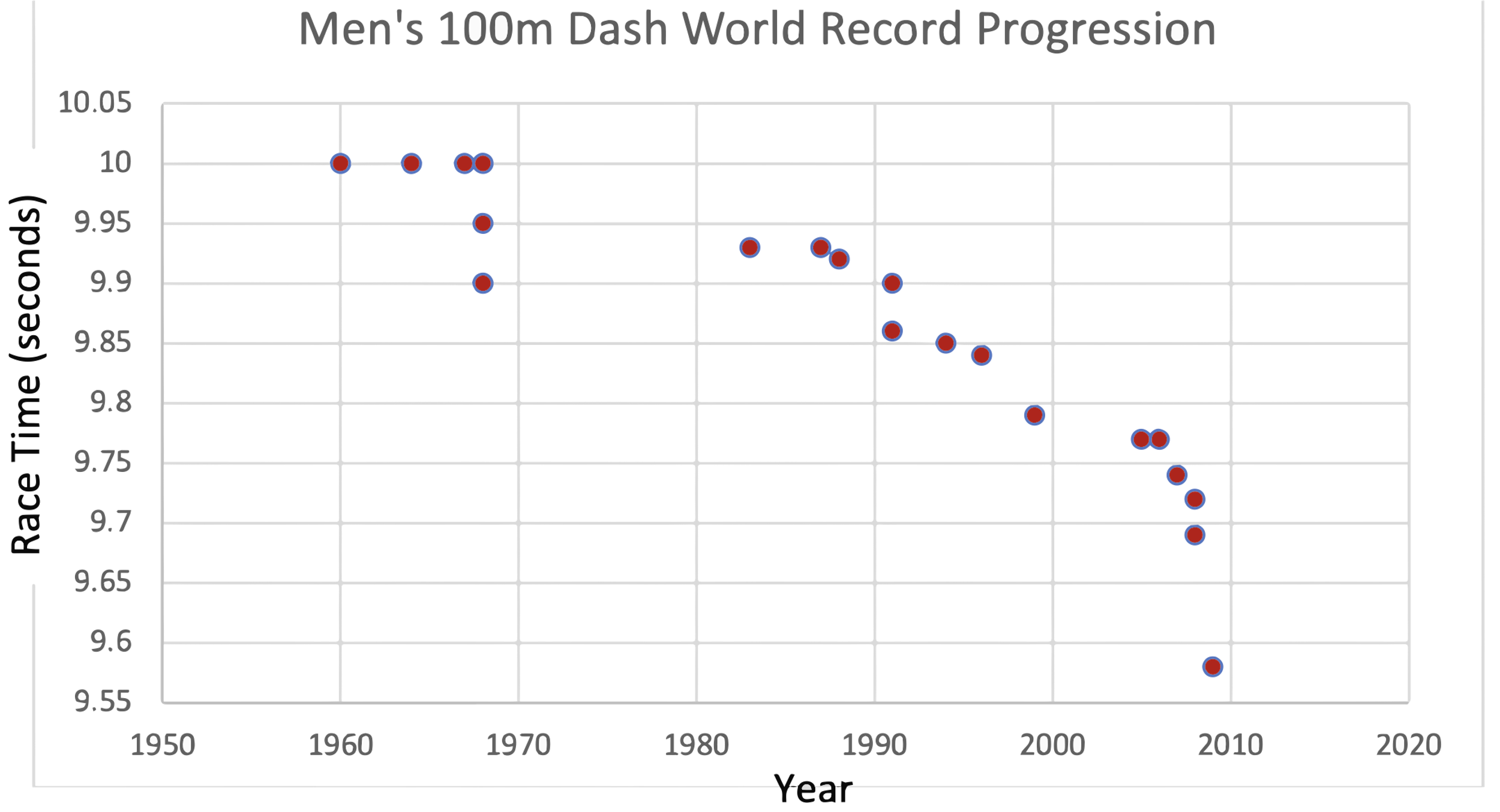
Task Entry Activity

The task entry activity is not required, and therefore, should not be scored. Data sets either produced through this activity or directly provided to the student for the completion of task items should incorporate at least twenty-five data points for each of the two data categories.

Item 1

Use the data gathered through the entry activity for this task or the data set provided by your teacher to create a scatter plot. Be sure to label your axes with variables and units.

Figure 5. Sample Student-Generated Scatter Plot



Item 2

Analyze the data illustrated by your scatter plot.

Item 2a [Student Document (A)]

Is a linear model a good fit for your data? Explain, commenting on the strength and direction of the association. Use residuals as part of your explanation.

Student Voice: “I used Excel to create a scatter plot using the most recent 29 data points. For the most part, a linear trendline (*y* = -0.0059*x* + 21.611) showing a moderately strong negative correlation between year and men’s world record 100m time looks like a pretty good (but not perfect) fit. I also used Excel to graph the residuals. Based on that graph, linear does not look like a perfect fit (because the dots are not super evenly spaced above and below the *x*-axis) but it’s also not a terrible fit, so maybe it’s an okay model at a high level.”

Figure 6. Sample Student-Generated Graph

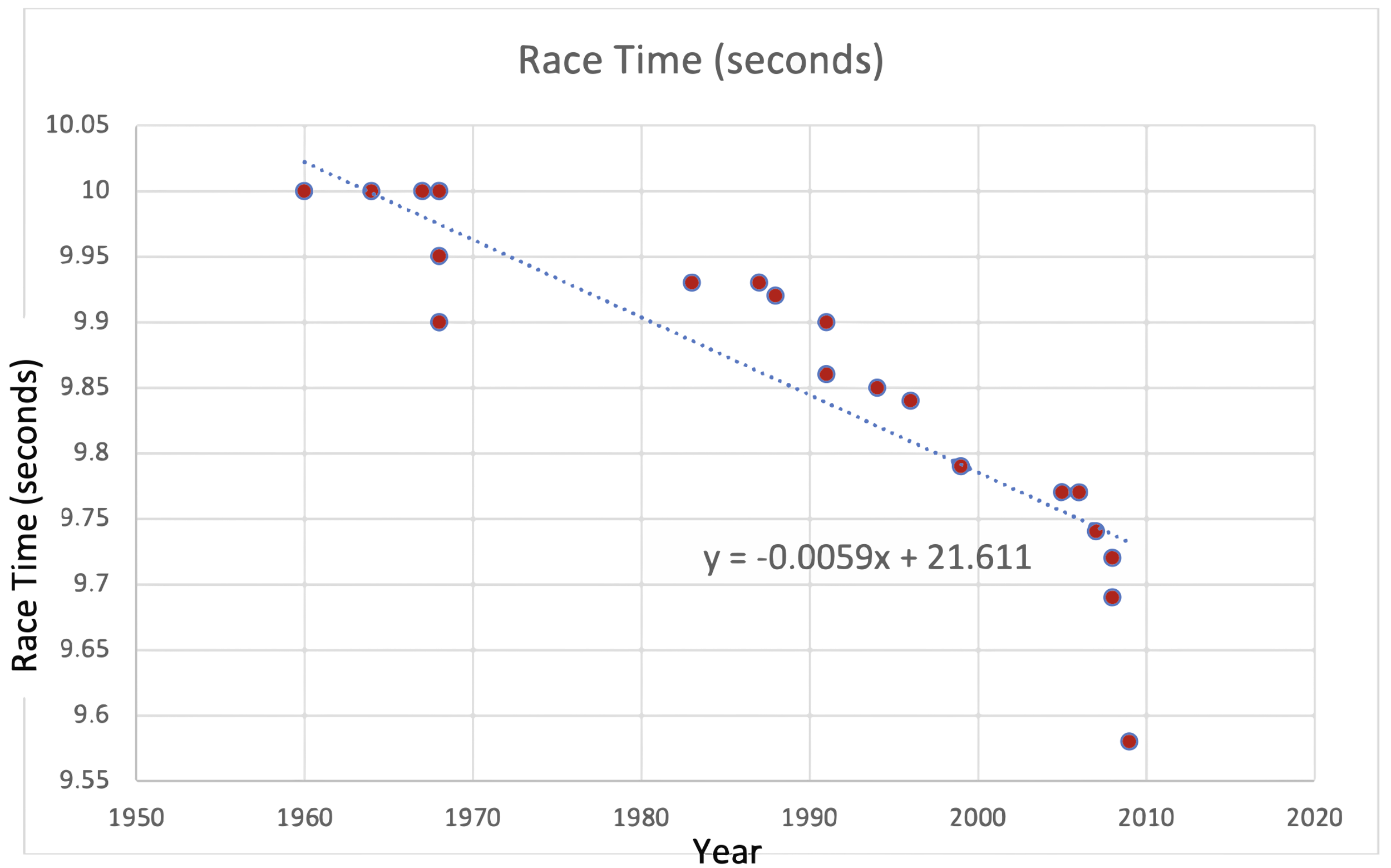


Figure 7. Sample Student-Generated Graph

Item 2b [Student Document (B)]

What does the slope of your linear model mean in the context of the dataset you are using?

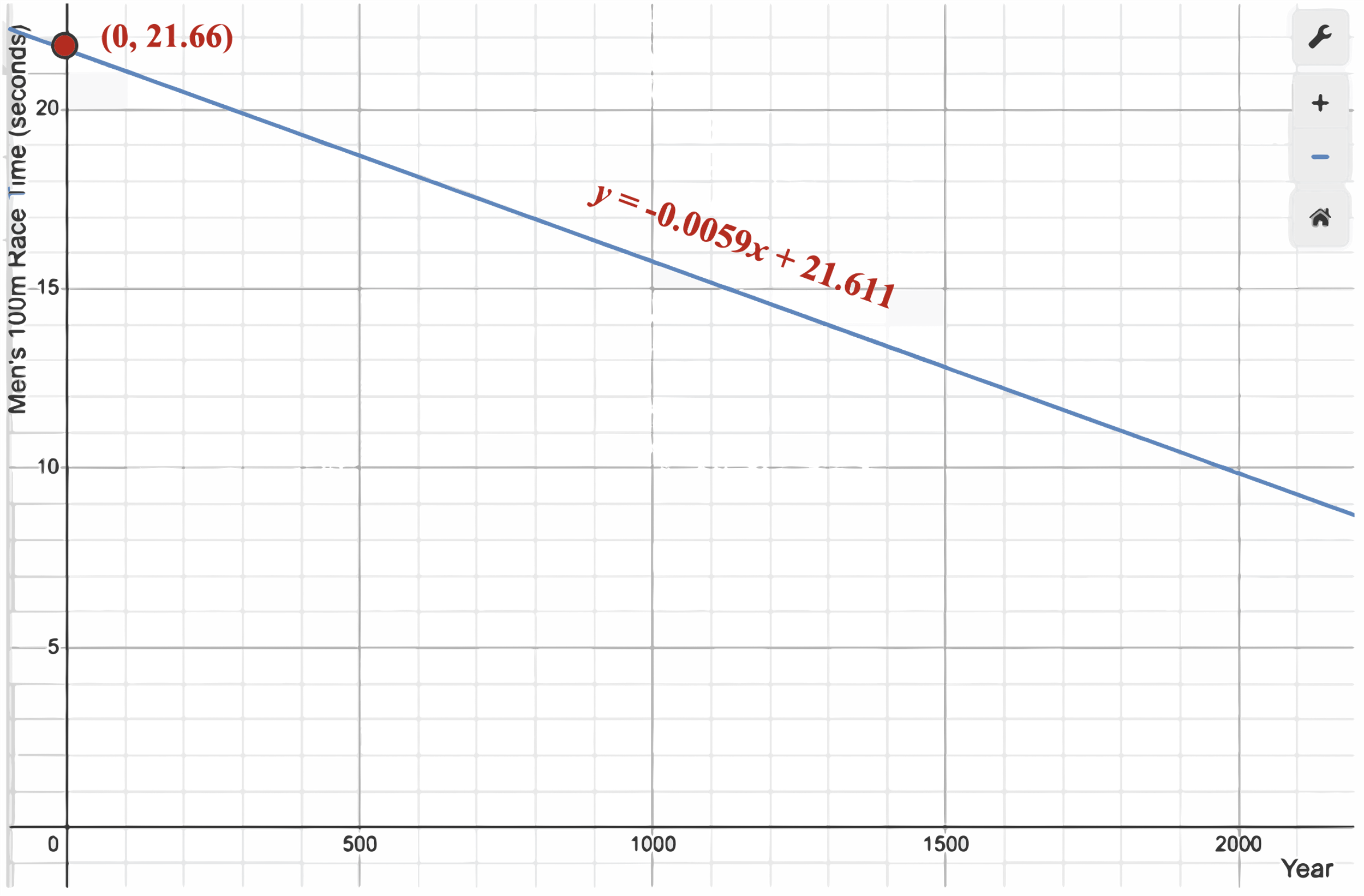
Student Voice: “With the linear model *y* = -0.0059*x* + 21.611, each time the year goes up by one, the race time in seconds goes down by 0.0059 seconds.”

Item 2c [Student Document (C)]

What is the vertical intercept of the function’s graph? What does it mean in the context of the variable you’ve plotted on the *y*-axis?

A possible response: “This is kind of a strange dataset and graph to talk about the *y*-intercept so I graphed the trendline or linear model in Desmos to see the intercept better.

Figure 8. Sample Student-Generated Graph



Student Voice: “Basically what this graph is saying is that in the year 0, the men’s 100m World Record would have been 21.66 seconds. This doesn’t really make sense because the record progression didn’t exist until the twentieth century and even if it did, it probably wouldn’t be universally linear. (Also, if it was, that means that at some point in the future, the record will be 0 seconds, which doesn’t make sense either).”

Item 2d [Student Document (D)]

Suppose your dataset was missing a particular *x*-value. How could you use your model to predict the *y*-value we should expect for that *x*-value? Give an example either from your data set or from the “Text Messages versus GPA” data above.

Student Voice: “For example, if you wanted to know what the world record would be likely to be (according to this model) in 2025, you could substitute *x* = 2025 into the equation *y* = -0.0059*x* + 21.611 à *y* = -0.0059 \* 2025 + 21.611 à -11.9475 + 21.611 = 9.6635 seconds.”

1. In this situation, it is important for the scribe to be careful to record **only** what the student explicitly communicates, rather than making interpretations and “filling in the blanks.” [↑](#footnote-ref-1)