Algebra I

Investigate Data

Option #1 Performance Task |   
Student Document

Authors: Initiative Team

April 30, 2024

Directions

Please review the task below and answer the various questions within the task to the best of your ability. If needed, you may have an adult or peer read the task out loud to aid your understanding. Additionally, feel free to use the following resources when answering each item:

* text-to-speech software
* speech-to-text software
* scratch paper

Key Words

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

Part 1. Investigate Data - Comparing Numerical Data Sets

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 |  |  |

Item 1 – Directions

Item 1 has no sub-items.

Item 1 Question

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

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 |  |  |

Item 2 – Directions

Item 2 has four sub-items. Please complete each sub-item.

Represent and analyze the data.

Item 2 Questions

1. Using the minimum, maximum, quartiles, and median, create two side-by-side box plots to compare your numerical variable between the two groups.
2. How would you describe the difference in the variable between the two groups? Be sure you discuss differences or similarities in shape, center, and spread.
3. 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.
4. Is the median or mean a more appropriate choice for describing the “centers” of these two distributions? Explain your reasoning.

PART 2. Investigate Data

Comparing Categorical Data Sets

Item 1 – Directions

Item 1 has no sub-items. You will use the same data set for these items to understand and organize data.

Item 1 Question

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

Figure 1. Two-Way Frequency Table for Collected Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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:** |

Item 2 – Directions

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.

Item 2 Questions

1. Calculate three joint relative frequencies from your table and explain their significance in the context of the question you’ve asked.
2. Calculate two marginal relative frequencies from your table and explain their significance in context.
3. 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?

PART 3. Investigate Data

Analyzing the Relationship Between Two Numerical Variables

Item 1 – Directions

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.

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.

Item 2

Item 2 has four sub-items. Analyze the data illustrated by your scatter plot to complete each item.

Item 2 Questions

1. 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.
2. If you concluded in Item 2(A) that a linear model is not a good fit for your data, please use the following data, scatter plot, and model to answer this question instead. 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 3. Optional Scatter Plot Example (if linear model is not a good fit)

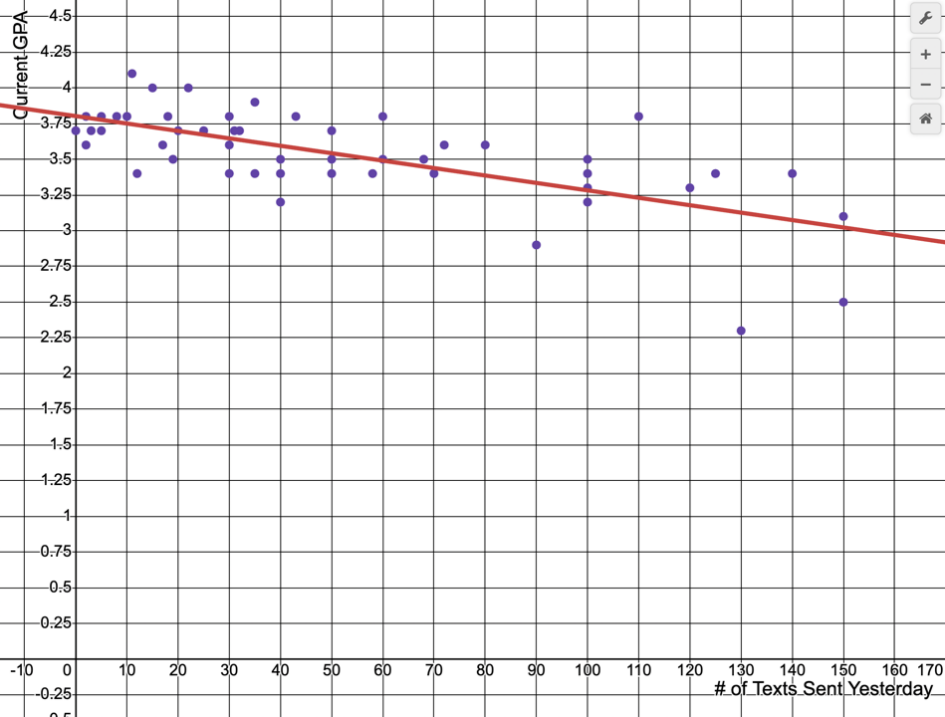


Table 4. Optional Data Tied to Scatter Plot (if linear model is not a good fit)

| Student | Number of texts sent yesterday | 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 |

1. 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?
2. 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.