

TABLE OF CONTENTS

◆ KEY CONCEPTS

01 Probability	006
02 Data Analysis	010
03 Statistical Inference and Prediction	018
04 Modeling and Interpreting Data Trends	022

◆ CHALLENGE QUESTIONS

Difficulty Level 4–5 Questions	027
--------------------------------------	-----

◆ ANSWERS & EXPLANATIONS

Detailed Step-by-Step Solutions	053
---------------------------------------	-----

VOLUME 4

PROBLEM SOLVING & DATA ANALYSIS

This volume develops the analytical and quantitative reasoning skills required for success on the SAT Math section.

- Probability
- Data Analysis
- Statistical Inference
- Data Interpretation

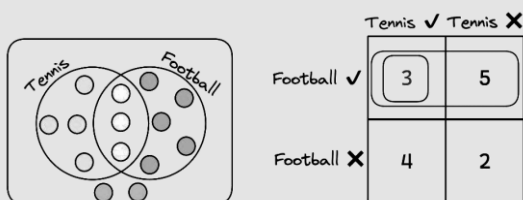
KEY CONCEPTS

1

Key Concepts 2 Conditional Probability

- ① Conditional Probability **Formula**:
The probability of **event A** occurring **given** that **event B** has occurred is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



- ② Independence Check:
Two events **A** and **B** are **independent** if:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

012.

The table summarizes the distribution of color and type for 100 clothes, categorized into shirts and pants.

	blue	black	yellow	total
shirt	30	20	15	65
pant	15	10	10	35
total	45	35	25	100

What is the probability of selecting a black pant if one of these clothes is selected at random? (Express your answer as a decimal or fraction, not as a percent.)

013.

Employees working at a retail store recorded all customer visits last Friday and noted whether the customer purchased an item and whether they used a discount coupon. The results are summarized in the table below.

	used discount coupon	did not use discount coupon	total
purchased an item	62	514	576
did not purchase an item	150	124	274
total	212	638	850

If a customer last Friday who purchased an item is selected at random, which of the following is closest to the probability that the customer also used a discount coupon?

- A) 0.05
B) 0.08
C) 0.11
D) 0.14

014.

activity	responded in person	responded online
movie night	40%	70%
concert	25%	15%
art exhibition	20%	10%
charity event	15%	5%
total	100%	100%

A community group conducted a survey asking participants to select their preferred event for the group's next gathering. Responses were collected either in person or online. The table above shows the distribution of preferences, in percent, for each type of response. For the survey, the number of online responses was three times the number of in-person responses. If a person who preferred an art exhibition is selected at random, what is the probability that the person responded online?

02. Data Analysis

Key Concepts 1 Measures of Central Tendency

Measures of central tendency summarize a data set by **identifying a central value**. The three most common measures are **mean**, **median**, and **mode**, each providing unique insights into the data.

① Mean:

Sum of all data values divided by the number of data points:

$$\text{mean} = \frac{\text{sum of data}}{\text{number of data}}$$

② Median:

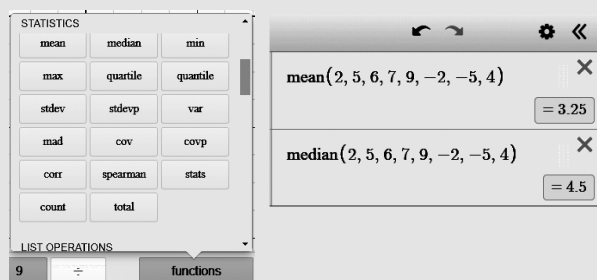
The **middle value** of an **ordered data set**:
If the **number of data points is odd**,
the median is the **middle value**.

If **even**, the median is the **average of the two middle values**.

③ Mode:

The **most frequently** occurring value in a data set.

Or use DESMOS:



015.

5, 9, 11, 11, 15, 18, 20

Which of the following lists represents a data set that has the same median as the data set shown?

- A) 6, 10, 11, 12, 17, 19
- B) 5, 9, 10, 11, 14, 20
- C) 6, 9, 10, 12, 14, 16
- D) 6, 9, 12, 14, 14, 18

016.

weights (grams)

Ella	12.2	13.1	14.3	12.8	13.0	13.2
Liam	x	13.0	12.9	13.3	13.5	13.2

Ella and Liam each collected six fruit samples, and the weights of the samples are shown in the table above. The mean weight of the samples Liam collected is 0.8 grams greater than the mean weight of the samples Ella collected. What is the value of x ?

017.

percent of residents who participated in a community activity

city	A	B	C	D	E	F	G
percent of participants	15.3	19.8	22.7	18.2	21.5	24.0	23.1

A survey was conducted to determine the percentage of residents who participated in a community activity across 100 cities. The median percentage of participants across all 100 cities was 22.0%. What is the difference between the median percentage of participants in these 7 cities and the median for all 100 cities?

- A) 0.25%
- B) 0.50%
- C) 0.75%
- D) 1.00%

Key Concepts 2 Measures of Spread

Measures of spread **describe** how much **variability exists in a data set**. They complement measures of central tendency by showing the range, distribution, and consistency of the data.

Common measures include **range**, **interquartile range (IQR)**, and **standard deviation**.

① Range:

Difference between the largest and smallest values in a data set:

Range = Maximum Value - Minimum Value

② Interquartile Range (IQR):

The spread of the middle 50% of the data:

$$\text{IQR} = Q_3 - Q_1$$

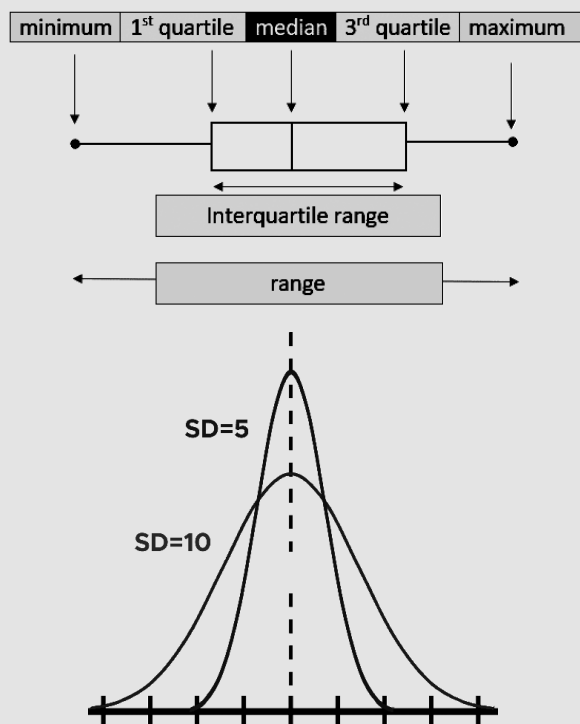
Q_1 : First quartile (25th percentile).

Q_3 : Third quartile (75th percentile).

③ Standard Deviation (SD):

Measures **how spread out** the data values are **around the mean**.

A **smaller SD** means the data points are **closer to the mean**, while a **larger SD** indicates **greater variability**.



026.

28, 35, 35, 38, 39, 42, 49, 61, 73

What is the range of the 9 scores shown?

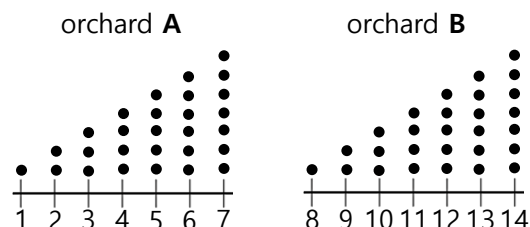
027.

data value	4	5	6	7	8	9	10	11
frequency	2	4	7	7	8	13	0	5

The frequency table summarizes the 46 data values in a data set. What is the maximum data value in the data set?

028.

The following dot plots represent the number of apples picked by each worker in two different orchards, orchard A and orchard B:



Which statement best compares the standard deviations of the number of apples picked in orchard A and orchard B?

- A) There is not enough information to compare these standard deviations.
- B) The standard deviation of the number of apples picked in orchard A is less than the standard deviation of the number of apples picked in orchard B.
- C) The standard deviation of the number of apples picked in orchard A is equal to the standard deviation of the number of apples picked in orchard B.
- D) The standard deviation of the number of apples picked in orchard A is greater than the standard deviation of the number of apples picked in orchard B.

Key Concepts 3 Data Interpretation

Data interpretation involves analyzing visual representations of data to **extract meaningful information, identify trends, and make predictions** or decisions based on the data provided.

① **Graphs and Charts:**

Bar Graphs:

Compare quantities across categories.

Line Graphs:

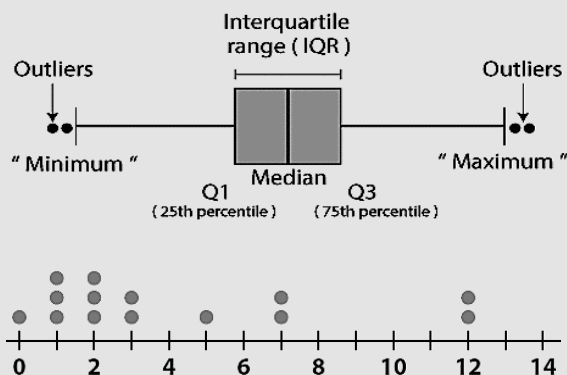
Show trends over time or continuous data changes.

Box Plots:

Summarize data distribution, including **median, quartiles, and outliers**.

Dot Plots:

Display individual data points and frequencies.



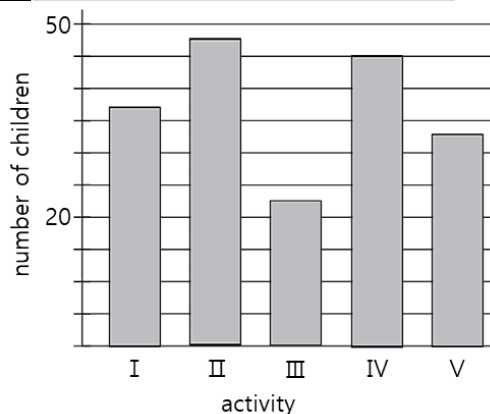
② **Proportional Reasoning:**

Use **percentages or ratios** to estimate totals or proportions in a larger population.

estimated total

$$= \text{sample value} \times \frac{\text{population size}}{\text{sample size}}$$

029.



A group of children voted on five different after-school activities. The bar graph shows the number of students who voted for each of the five activities. How many children chose activity III?

- A) 23
- B) 32
- C) 48
- D) 47

030.

In District Y, a science class consisting of 30 students was surveyed, and 40% of the students reported that they had participated in a science fair. The average class size in the district is 30 students. If the students in the science class are representative of students in the district's science classes, and there are 2,000 science classes in the district, which of the following best estimates the number of students in the district who have participated in a science fair?

- A) 22,000
- B) 24,000
- C) 26,000
- D) 28,000

04. Modeling and Interpreting Data Trends

Key Concepts 1 Scatterplots and Best Fit Lines

① Key Concepts and Principles

Scatterplot:

A **graphical representation** of data points showing the **relationship between two variables**.

Line of Best Fit (Regression Line):

A straight line that best represents the data on a scatterplot.

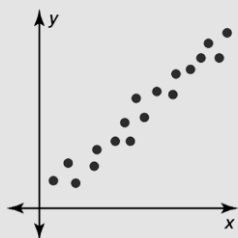
Used to **predict values and analyze trends**.

② **Correlation:**

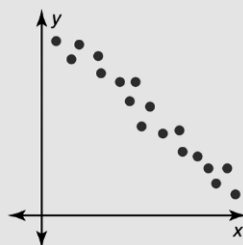
Describes the strength and direction of the linear relationship between variables.

Positive correlation:

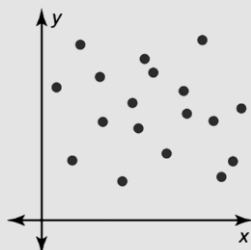
Both variables **increase together**.

**Negative correlation:**

One variable **decreases** as the **other** increases.

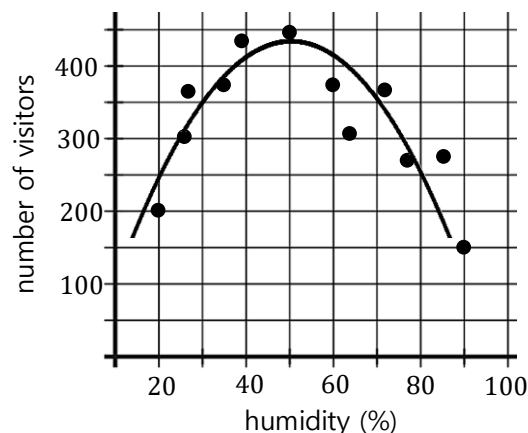
**No correlation:**

No discernible relationship.



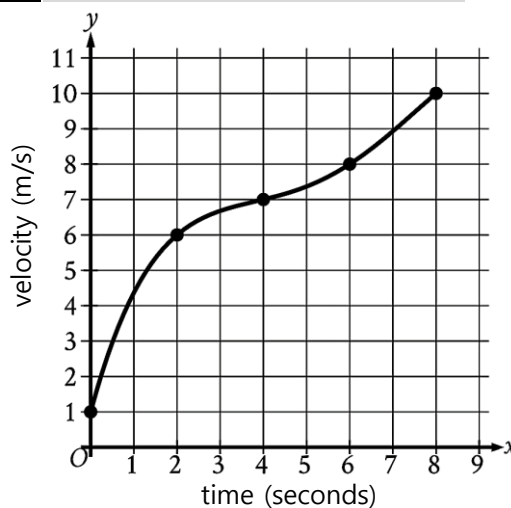
052.

The scatterplot below shows the number of visitors y , to a wildlife sanctuary on different days with varying average humidity levels x . The line of best fit for the data is also shown.



According to the line of best fit, how many visitors, rounded to the nearest 50, are predicted to visit the sanctuary on a day with an average humidity of 30%?

053.



The graph displays the velocity y , in meters per second, of a car x seconds after it started accelerating, for $0 \leq x \leq 9$. What is the average rate of change, in meters per second squared, in the velocity of the car from $x = 4$ to $x = 8$?

Key Concepts 2 Modeling

① Linear Models

A linear relationship is one where the **rate of change** between variables **is constant**.

Represented by: $y = mx + b$

Graph is a **straight line**.

The slope m indicates the change in y for **every unit increase in x** .

② Exponential Models

An exponential relationship involves a quantity **increasing or decreasing by a fixed percentage or ratio over equal intervals of time**.

Represented by: $y = a \cdot b^x$

a : Initial value

b : **Growth** factor ($b > 1$) or
Decay factor ($0 < b < 1$)

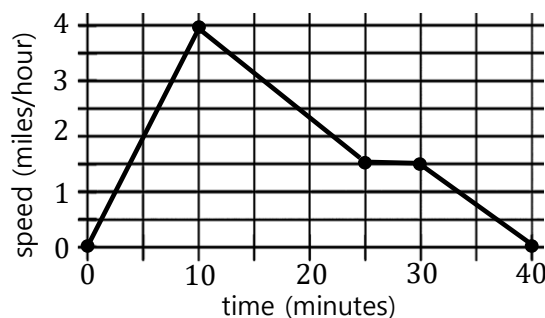
Graph has a curve, **not a straight line**.

057.

A chef begins preparing a large batch of pies to bake. The chef removes one pie from the oven every 5 minutes at a steady pace. What type of model best represents the number of pies remaining to be removed over time?

- A) an exponential decay model
- B) an exponential growth model
- C) a linear model with a positive slope
- D) a linear model with a negative slope

058.



James rode a bicycle for forty minutes, and his time and speed are displayed on the graph above. Based on the graph, which of the following statements is accurate about James's ride?

- A) James maintained a constant speed for ten minutes.
- B) James's speed increased steadily during the final ten minutes.
- C) James's speed was increasing for a shorter duration than it was decreasing.
- D) James achieved his highest speed during the last fifteen minutes of the ride.

059.

Which of the following is true about the values of 3^x and $3x + 5$ for $x > 0$?

- A) for all $x > 0$, it is true that $3^x < 3x + 5$
- B) for all $x > 0$, it is true that $3^x > 3x + 5$
- C) There is a constant c such that if $0 < x < c$, then $3^x < 3x + 5$, but if $x > c$, then $3^x > 3x + 5$.
- D) There is a constant c such that if $0 < x < c$, then $3^x > 3x + 5$, but if $x > c$, then $3^x < 3x + 5$.

CHALLENGE QUESTIONS



CHALLENGE QUESTIONS

072.

The table gives the distribution of participants for a coding competition and their experience levels.

	beginner	intermediate	advanced	total
Python	5	7	6	18
Java	4	6	5	15
C++	3	8	6	17
JavaScript	8	5	7	20
total	20	26	24	70

If one of these participants is selected at random, what is the probability of selecting a participant who registered for Python?

- A) $\frac{2}{7}$
- B) $\frac{3}{35}$
- C) $\frac{9}{35}$
- D) $\frac{5}{14}$

073.

The table gives the distribution of votes for preferred subjects and grade level for 80 students.

subject	grade level			
	6th	7th	8th	total
math	4	9	9	22
chemistry	9	2	9	20
biology	4	6	4	14
physics	6	9	9	24
total	23	26	31	80

If one of these students is selected at random, what is the probability of selecting a student whose vote for preferred subject was for biology?

- A) $\frac{7}{40}$
- B) $\frac{11}{40}$
- C) $\frac{1}{4}$
- D) $\frac{1}{20}$

074.

At a baseball stadium, there are a total of 4,800 spectators. Each spectator is seated in either section A, section B, or section C. If one of these spectators is selected at random, the probability of selecting a customer who is located in section A is 0.37, and the probability of selecting a spectators who is located in section B is 0.41. How many spectators are located in section C?

- A) 872
- B) 988
- C) 1,056
- D) 1,134

075.

animal shelter exhibit

type of animal	male	female	total
dogs	52	63	115
cats	23	31	54
rabbits	41	40	81
birds	30	25	55
total	146	159	305

The table above shows the number of animals in an animal shelter exhibit, sorted by gender and type. Which type of animal has a female population closest to being $\frac{1}{4}$ of the total female animal population in the shelter?

- A) cats
- B) dogs
- C) birds
- D) rabbits

CHALLENGE QUESTIONS

119.

Two surveys were conducted to determine the level of satisfaction among customers at two different stores. The results are shown in the table below:

Sample	Percent Satisfied	Margin of error
Store A	65%	5.2%
Store B	60%	2.8%

The samples were selected from the same population, and the margins of error were calculated using the same method. Which of the following is the most appropriate reason that the margin of error for Store **A** is greater than the margin of error for Store **B**?

- A) Store **A** had a lower number of satisfied customers.
- B) Store **A** had a larger percentage of satisfied customers.
- C) Store **A** had a smaller sample size.
- D) Store **A** had a larger sample size.

120.

To estimate the proportion of a population that has a certain characteristic, a random sample was selected from the population. According to the sample, the proportion of the population possessing the characteristic is estimated to be 6.73, with a margin of error of 0.07. Given this estimate and margin of error, which of the following is the most appropriate conclusion about the proportion of the population that has the characteristic?

- A) The proportion is exactly 6.73.
- B) It is plausible that the proportion is less than 6.66.
- C) It is plausible that the proportion is between 6.66 and 6.80.
- D) It is plausible that the proportion is greater than 6.80.

121.

Views on Renewable Energy Use in Europe

Response	Frequency
Strongly favor	75
Somewhat favor	220
Somewhat oppose	130
Strongly oppose	45

A researcher interviewed 470 randomly selected residents of Germany and asked about their views on the use of renewable energy. The table above summarizes the responses of the interviewees. If the population of Germany was 83 million when the survey was given, based on the sample data for the 470 residents, what is the best estimate, in millions, of the difference between the number of German residents who somewhat favor or strongly favor the use of renewable energy and the number of those who somewhat oppose or strongly oppose it? (Round your answer to the nearest whole number.)

122.

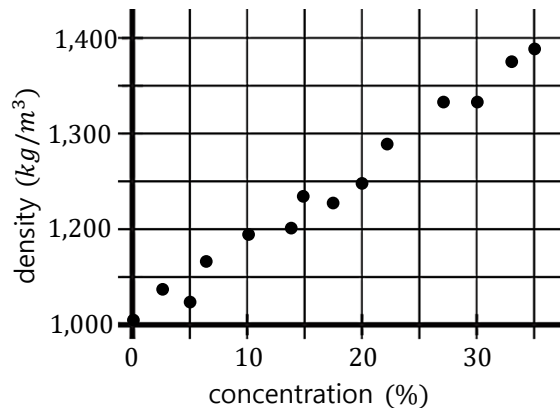
Favorite Types of Food in France

Response	Frequency
Strongly prefer	40
Somewhat prefer	100
Somewhat dislike	70
Strongly dislike	20

A researcher interviewed 230 randomly selected French residents and asked about their preferences for Italian food. The table above summarizes the responses. If the population of France is 67 million, what is the best estimate, in millions, of the difference between the number of French residents who somewhat or strongly prefer Italian food and the number of those who somewhat or strongly dislike it? (Round your answer to the nearest whole number.)

CHALLENGE QUESTIONS

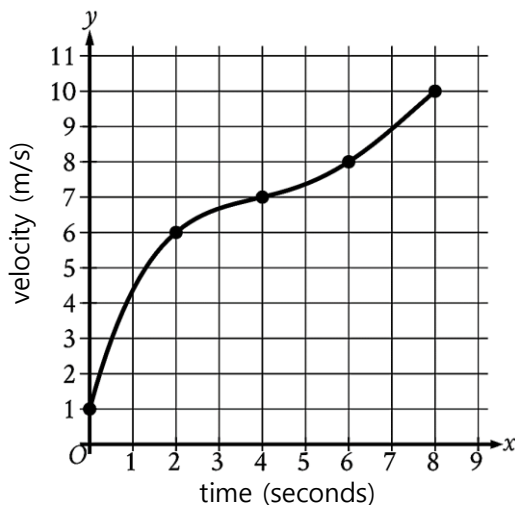
125.



The scatterplot above shows the densities of different concentrations of saltwater. According to the trend shown by the data, which of the following is closest to the predicted density, in kilograms per cubic meter (kg/m^3), for saltwater with a concentration of 25%?

- A) 1,250
- B) 1,275
- C) 1,300
- D) 1,325

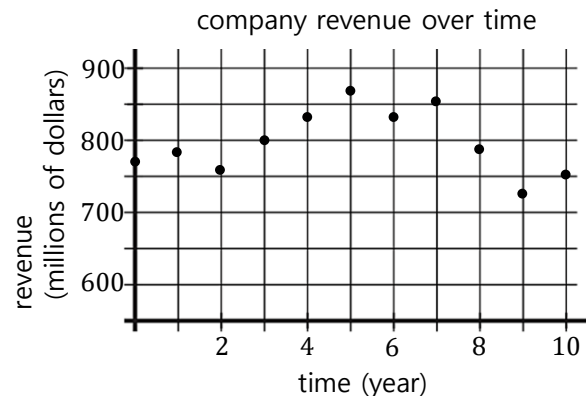
126.



The graph displays the velocity y , in meters per second, of a car x seconds after it started accelerating, for $0 \leq x \leq 10$. What is the average rate of change, in meters per second squared, in the velocity of the car from $x = 4$ to $x = 8$?

127.

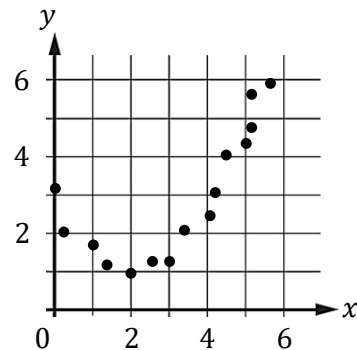
The scatterplot below shows the revenue y , in millions of dollars, generated by a company over a 10-year period.



Which of the following equations best represents the data in the scatterplot?

- A) $y = 1.584x^2 + 16.32x - 790.8$
- B) $y = 1.584x^2 + 23.54x + 790.8$
- C) $y = -2.134x^2 + 18.64x - 790.8$
- D) $y = -2.134x^2 + 21.54x + 790.8$

128.



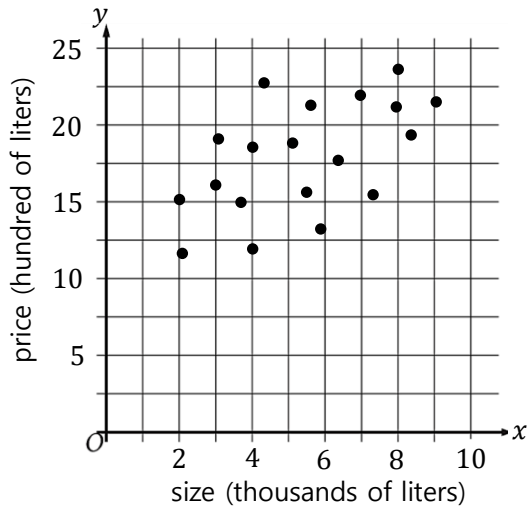
The scatterplot above shows the distances traveled by six rockets, in miles, with respect to the time of travel in hours. Which of the following equations is the best model for the data in the scatterplot?

- A) $y = 5x - 3$
- B) $y = 0.2x^2 - 6x + 3$
- C) $y = 0.3x^2 + 6x - 3$
- D) $y = -4x + 3$

CHALLENGE QUESTIONS

129.

The scatterplot above shows the relationship between the size (x) in thousands of liters and the price (y) in hundreds of dollars of 30 water tanks sold in City P .

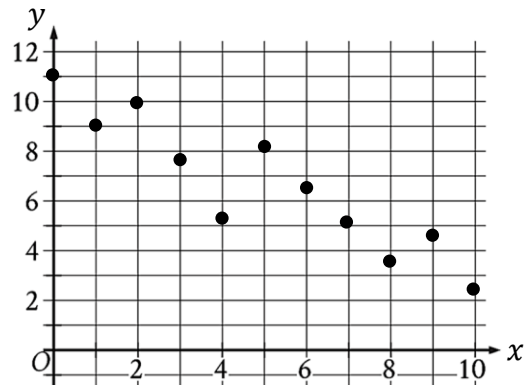


Which of the following could be an equation for the line of best fit for the data?

- A) $y = 1.54x - 5.92$
- B) $y = 1.67x + 8.32$
- C) $y = 241x - 6.03$
- D) $y = 241x + 6.03$

130.

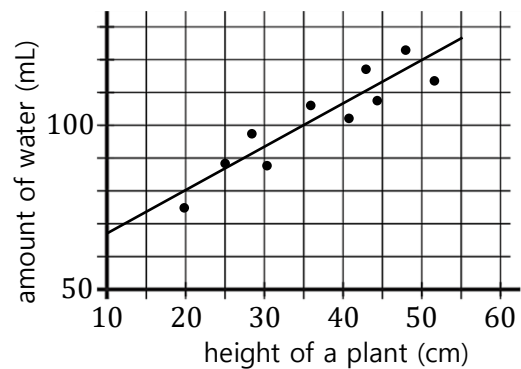
The scatterplot shows the relationship between two variables, x and y .



Which of the following equations best represents the linear relationship suggested by the data in the scatterplot?

- A) $y = 6.5 - 0.35x$
- B) $y = 11 - 0.35x$
- C) $y = 6.5 - 0.82x$
- D) $y = 11 - 0.82x$

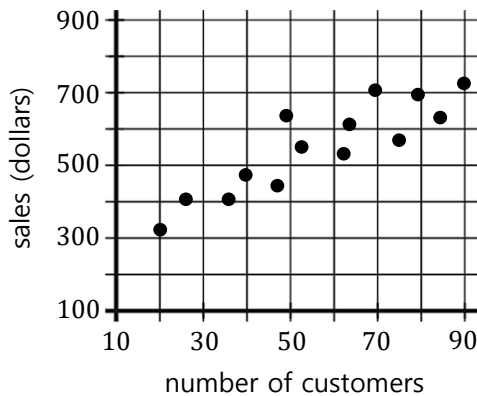
131.



Each point in the scatterplot above represents the height of a plant (in cm) and the amount of water (in mL) given to it daily over a 10-day period. The line of best fit for the data is also shown. The line of best fit for the data has a slope of approximately 0.8. According to this estimate, how many additional milliliters of water are predicted to be needed per day for each 100 cm increase in the plant's height?

CHALLENGE QUESTIONS

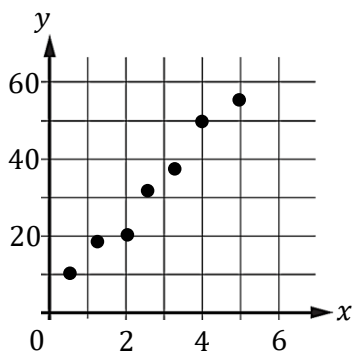
135.



The scatterplot above shows a café's coffee sales c , in dollars, and the number of customers n , on 10 different days. A line of best fit for the data is also shown. Which of the following could be an equation of the line of best fit?

- A) $c = 0.5n + 195$
- B) $c = 0.75n + 250$
- C) $c = 3n + 175$
- D) $c = 5n + 250$

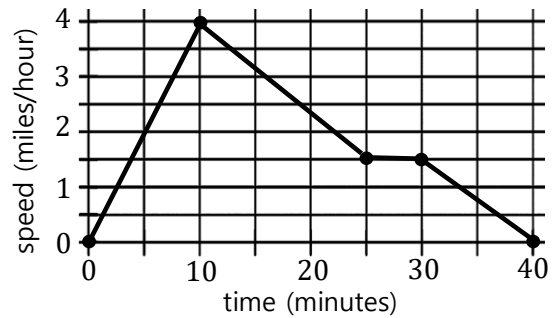
136.



Which of the following equations best represents the line of best fit for the data shown in the scatterplot?

- A) $y = 1.5x + 3.2$
- B) $y = 15x + 3.2$
- C) $y = -1.5x + 2.5$
- D) $y = 15x - 25$

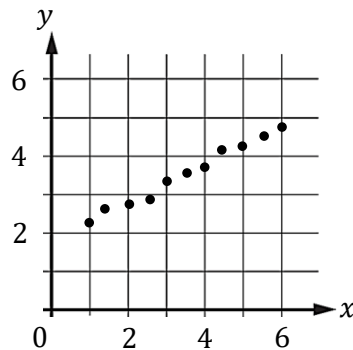
137.



James rode a bicycle for forty minutes, and his time and speed are displayed on the graph above. Based on the graph, which of the following statements is NOT accurate about James's ride?

- A) James maintained a constant speed for ten minutes.
- B) James's speed increased steadily during the final ten minutes.
- C) James's speed was increasing for a shorter duration than it was decreasing.
- D) James achieved his highest speed during the last fifteen minutes of the ride.

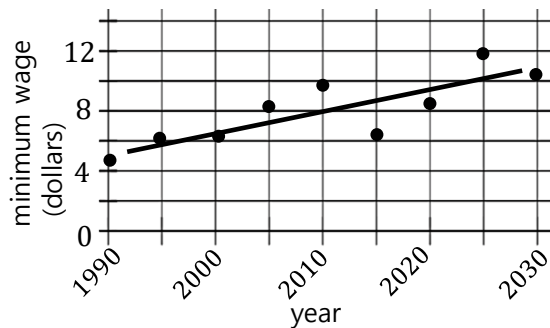
138.



Which of the following could represent the equation of a line of best fit for the relationship between the number of hours a student studies (x) and their score on a test (y)?

- A) $y = 0.75x + 1.2$
- B) $y = -0.75x + 1.2$
- C) $y = 2.5x - 1.7$
- D) $y = -2.5x + 1.7$

139.

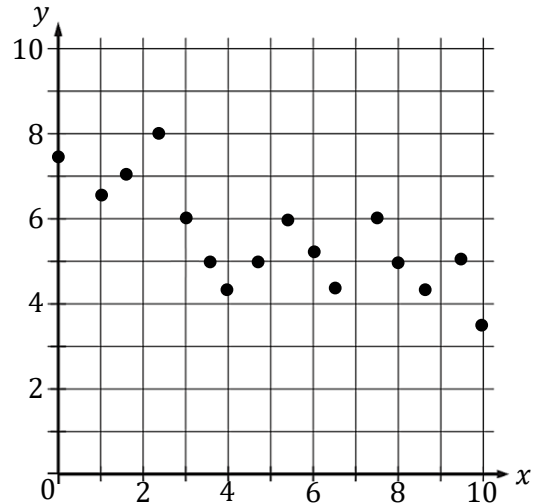


The scatterplot above shows the federal-mandated minimum wage every 5 years between 1990 and 2030. A line of best fit is shown, and its equation is $y = 0.63x + 4.71$. What does the line of best fit predict about the increase in the minimum wage over the 40-year period?

- A) Each year between 1990 and 2030, the average increase in minimum wage was 3.15 dollars.
- B) Each year between 1990 and 2030, the average increase in minimum wage was 0.63 dollars.
- C) Every 5 years between 1990 and 2030, the average increase in minimum wage was 3.15 dollars.
- D) Every 5 years between 1990 and 2030, the average increase in minimum wage was 0.63 dollars.

140.

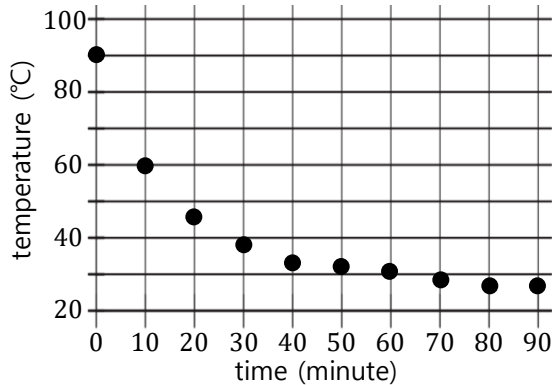
The scatterplot shows the relationship between two variables, x and y .



What is the most appropriate linear model for the data represented by the following equations?

- A) $y = 0.4x + 7.6$
- B) $y = -0.4x + 7.6$
- C) $y = -7.6x + 0.4$
- D) $y = 7.6x + 0.4$

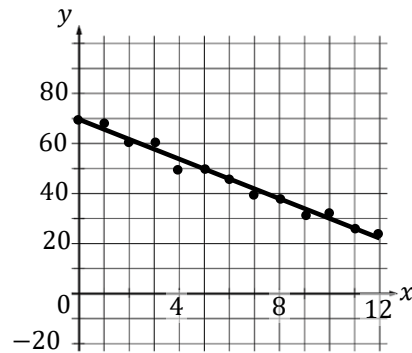
141.



In an experiment, a heated pot of soup is removed from a stove, and the soup is placed in a room maintained at a constant temperature. The graph above shows the temperature, in degrees Celsius ($^{\circ}\text{C}$), of the soup immediately after being removed from the stove and at 10-minute intervals thereafter. During which of the following 10-minute intervals does the soup's temperature decrease at the greatest average rate?

- A) between 0 and 10 minutes
- B) between 20 and 30 minutes
- C) between 40 and 50 minutes
- D) between 80 and 90 minutes

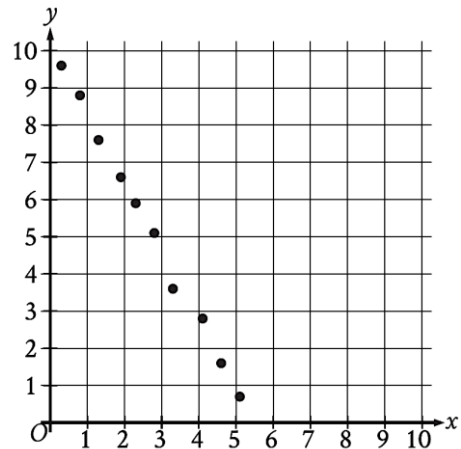
142.



The scatterplot shows the relationship between two variables, x and y , for data set A . A line of best fit is shown. Data set B is created by multiplying the y -coordinate of each data point from data set A by 0.25. Which of the following could be an equation of a line of best fit for data set B ?

- A) $y = 17.5 - x$
- B) $y = 17.5 - 0.25x$
- C) $y = 70 - x$
- D) $y = 70 - 4x$

143.



Which of the following equations is the most appropriate linear model for the data shown in the scatterplot?

- A) $y = 1.9x + 10.1$
- B) $y = 1.9x - 10.1$
- C) $y = -1.9x + 10.1$
- D) $y = -1.9x - 10.1$

SAT Math Question Bank ANSWER EXPLANATIONS

015. Explanation and Answer

The median is the middle value of a data set when arranged in ascending order:

if the number of data points is **odd**, the median is the **exact middle value**

if the number of data points is **even**, the median is the **average of the two middle values**

Determine the median of the original data set:

The number of data points is **7 (odd)**, so the median is the **4th value**: median=11

Analyze each choice:

The **number of data points is 6 (even)**, so the **median is the average of the two middle values (3rd and 4th)**.

$$\text{for A: } \frac{11+12}{2} = 11.5 \quad \text{for B: } \frac{10+11}{2} = 10.5 \quad \text{for C: } \frac{10+12}{2} = 11 \quad \text{for D: } \frac{12+14}{2} = 13$$

Therefore, the correct answer is **C**.

016. Explanation and Answer

The mean of a set of numbers is calculated by dividing the total sum of the numbers by the count of numbers in the set.

Calculate the **mean of Ella's samples**:

The mean of Ella's samples are:

$$\frac{12.2 + 13.1 + 14.3 + 12.8 + 13.0 + 13.2}{6} = \frac{78.6}{6} = 13.1$$

Find the **mean of Liam's samples**:

It's given that the mean of Liam's samples is **0.8 grams greater than** Ella's: $13.1 + 0.8 = 13.9$

Set up an equation for Liam's mean:

the sum of Liam's samples: $x + 13.0 + 12.9 + 13.3 + 13.5 + 13.2 = x + 78.9$

the mean of Liam's samples is given as:

$$\frac{x + 78.9}{6} = 13.9 \rightarrow x + 78.9 = 83.4 \rightarrow x = 4.5$$

Therefore, **the value of x is 4.5** and the correct answer is **4.5**.

017. Explanation and Answer

The median of a data set is the middle value when the data is ordered from least to greatest. If the number of data points is odd, the median is the middle value; if even, the median is the average of the two middle values.

Order the percentages **from least to greatest**: 15.3%, 18.2%, 19.8%, **21.5%**, 22.7%, 23.1%, 24.0%

Since there are 7 cities (odd number), the median is the **4th value** in the ordered list:

median of these 7 cities = 21.5%

Median for all 100 cities is given as 22.0%.

difference: $|21.5\% - 22.0\%| = 0.5\%$

Therefore, the correct answer is **B**.

018. Explanation and Answer

To find the **median** in a data set, **arrange** the values

from least to greatest and **locate the middle value**.

For an odd number of values, the median is the middle value.

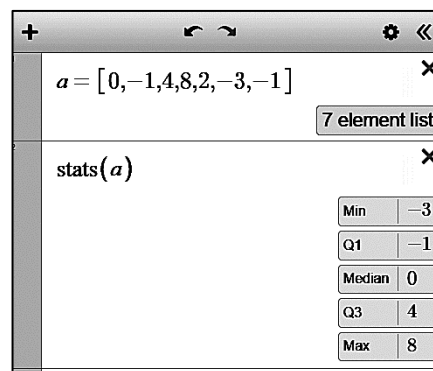
Arrange the data in order: **-3, -1, -1, 0, 2, 4, 8**.

Since there are **seven values**, the median is the **fourth value** in the ordered list.

The **fourth value in the list is 0**, so the **median is 0**.

Or use DESMOS

Therefore, the correct answer is **C**.



SAT Math Question Bank ANSWER EXPLANATIONS

093. Explanation and Answer

The median of a data set is the middle value when the data is ordered from least to greatest. If the number of data points is odd, the median is the middle value; if even, the median is the average of the two middle values.

Order the percentages **from least to greatest**: 15.3%, 18.2%, 19.8%, **21.5%**, 22.7%, 23.1%, 24.0%

Since there are 7 cities (odd number), the median is the **4th value** in the ordered list:

median of these 7 cities = 21.5%

Median for all 100 cities is given as 22.0%.

difference: $|21.5\% - 22.0\%| = 0.5\%$

Therefore, the correct answer is **B**.

094. Explanation and Answer

The median of a data set is the middle value when the data is ordered from least to greatest. For a data set with an **odd number of values**, the **median is the $\frac{n+1}{2}$ -th value**, where n is the total number of values.

Determine the total number of fruits:

enclosure X contains **15** fruits, enclosure Y contains **18** fruits, and enclosure Z contains **20** fruits

So, **total number of fruits is: $15 + 18 + 20 = 53$** .

Since **there are 53 fruits** (odd number), the **median is the $\frac{53+1}{2} = 27$ -th value**.

The **first 15 fruits** (1st to **15th**) weigh less than **2 kilograms** (from enclosure X). The next **18 fruits** (**16th to 33rd**) weigh **between 2 kilograms and 5 kilograms** (from enclosure Y). The final 20 fruits (34th to 53rd) weigh more than 5 kilograms (from enclosure Z).

The **27th fruit lies** in the range of weights for enclosure Y (**between 2 kilograms and 5 kilograms**).

Only **3.3 kilograms falls within the range $2 \leq \text{weight} \leq 5$** .

Therefore, the correct answer is **B**.

095. Explanation and Answer

To determine the mean of data set C , we need to first understand how means are calculated and how to combine two data sets. We know that the **mean of a data set is the sum of all the values in the data set divided by the number of values**.

Data Set A : The mean is 52kg , and there are 85 students. So, the **total sum** of body weight in Data Set A is:

$$\text{sum of } A = 52 \times 85 = 4,420 \text{ kg}$$

Data Set B : The mean is 58kg , and there are 60 students. So, the **total sum** of body weights in Data Set B is:

$$\text{sum of } B = 58 \times 60 = 3,480 \text{ kg}$$

Data set C consists of all students from Data Set A and Data Set B combined. So, the total number of students in Data Set C is:

$$\text{number of students in } C = 85 + 60 = 145$$

The total sum of body weights in Data Set C is:

$$\text{sum of } C = 4,420 + 3,480 = 7,900 \text{ kg}$$

Now, to find the **mean of Data Set C** , divide the total sum of body weights by the total number of students:

$$\frac{7,900}{145} \approx 54.48275 \dots \approx 54.5 \text{ kg}$$

Therefore, the correct answer is **54.5**

096. Explanation and Answer

Standard deviation measures the **spread of data around the mean**. If **two distributions have identical shapes and frequency distributions**, their **standard deviations will also be identical**, regardless of the specific values.

Analyze the shapes of the distributions for orchard A and orchard B :

Both distributions are identical in shape and frequency. Orchard B 's data values are simply shifted higher (all values are larger by 7 compared to Orchard A). **Standard deviation is not affected by shifting all data points by the same amount because the relative spread of the data remains the same**.

Therefore, the **standard deviations** of orchard A and orchard B **are identical** and the correct answer is **C**.

SAT Math Question Bank ANSWER EXPLANATIONS

128. Explanation and Answer

To determine the best model for a dataset displayed in a scatterplot, the chosen model must approximate the data points closely, with roughly equal numbers of points above and below the curve or line. For quadratic models, key considerations include the vertex, axis of symmetry, and shape of the parabola.

The scatterplot shows a **parabolic trend** where the data form an **upward-facing curve**. This indicates that the **best-fit model is likely quadratic**.

Examine the Equations:

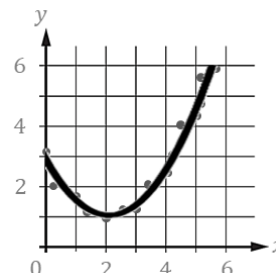
choice A: A linear equation. This cannot capture the parabolic trend.

choice B: $y = 0.2x^2 - 6x + 3$, an **upward-facing parabola** with a vertex and **y-intercept in reasonable alignment with the scatterplot points**.

choice C: $y = 0.3x^2 + 6x - 3$, another upward-facing parabola, but its vertex and **y-intercept do not align** closely with the scatterplot.

choice D: A linear equation. Again, this is unsuitable for modeling the parabolic trend.

Therefore, the correct answer is **B**.



129. Explanation and Answer

This problem involves determining the equation of the line of best fit based on a scatterplot. A linear equation is represented as $y = mx + b$, where: m is the slope of the line, and b is the y-intercept.

The scatterplot shows the relationship between the size (x) and the price (y) of water tanks. The two points **(4, 15) and (7, 20) are chosen** because they are representative **points through which the best fit line passes**. These points are either directly indicated in the problem or are visually closest to the trend of the data cluster.

The slope of the line passing through (x_1, y_1) and (x_2, y_2) is given by:

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{20 - 15}{7 - 4} = \frac{5}{3} \approx 1.666 \dots$$

Write the equation in slope-intercept form ($y = mx + b$):

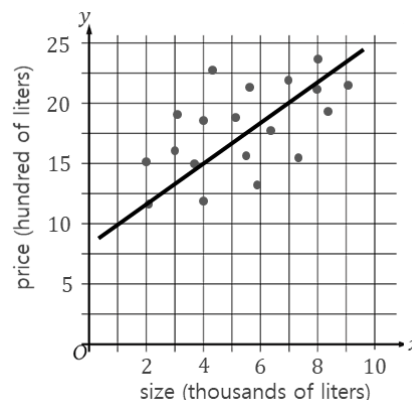
$$y = mx + b \rightarrow y = 1.67x + b$$

Use one of the **given points, such as (4, 15)**, and substitute $x = 4$ and $y = 15$:

$$y = 1.67x + b \rightarrow 15 = 1.67(4) + b \rightarrow b = 8.32$$

So, the equation of the line of best fit is: $y = 1.67x + 8.32$

Therefore, **the equation for the line of best fit for the data is $y = 1.67x + 8.32$** and the correct answer is **B**.



130. Explanation and Answer

This question involves **linear models**, where the goal is to determine the most appropriate linear equation that fits the data in the scatterplot. A linear equation has the form $y = mx + b$, where: **m is the slope** (rate of change of y with respect to x), and **b is the y-intercept** (the value of y when $x = 0$). Key observations in a scatterplot, such as the direction of the **slope (positive or negative)** and approximate values for b , help determine the best fit model.

Understand the problem:

The **scatterplot shows a negative relationship** between x and y :
as x increases, y decreases.

This indicates that the **slope (m) of the equation must be negative**.

Analyze the slope (m):

For **every increase of 1 in x , y decreases by approximately 0.8**.

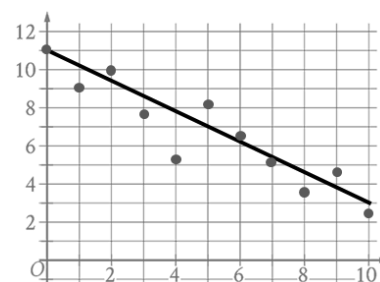
This suggests that the **slope (m) is close to -0.8** .

Analyze the y-Intercept (b):

From the **scatterplot, when x is near 11**.

This suggests that the **y-intercept (b) should be close to 11**.

Therefore, the **most appropriate linear model for the data is $y = 11 - 0.82x$** , and the correct answer is **D**.



SAT Math Question Bank ANSWER EXPLANATIONS

142. Explanation and Answer

The problem involves interpreting data from a scatterplot and **finding the line of best fit for the original data set A**, and then **determining the transformed equation for data set B**.

To find the line of best fit for data set A, we need to determine the equation in the form: $y = mx + b$, where **b is the y-intercept** of the line and **m is the slope** of the line.

To estimate the equation, we **need at least two points** that lie on or very close to the line of best fit. Let's suppose from the scatterplot, we observe the line passing through two approximate points: **(0, 70)** and **(5, 50)**.

The slope m is calculated using the formula:

$$m = \frac{y_2 - y_1}{x_2 - x_1} \rightarrow m = \frac{50 - 70}{5 - 0} = -4.$$

The **y-intercept b** is the value of y when $x = 0$. From point **(0, 70)**, we can directly see that: **$b = 70$** . Substituting the values of m and b into the equation: $y = -4x + 70 \rightarrow y = 70 - 4x$.

Data set B is created by **multiplying the y-coordinates of each data point in data set A by 0.25**. This scaling affects both the slope and the y-intercept of the line of **best fit proportionally**. The new line of best fit for data set B will be:

$$y = 0.25 \times 70 - 0.25 \times 4x \rightarrow y = 17.5 - x$$

Therefore, the correct answer is **A**.

143. Explanation and Answer

To determine the most appropriate linear model for the data shown in the scatterplot, we need to identify the equation that **best fits the pattern of the data points**. The general form of a linear equation is $y = mx + b$, where **m is the slope** and b is the **y-intercept**.

Analyzing the Scatterplot:

The **scatterplot shows** data points that appear to have a **negative slope**, meaning **as x increases, y decreases**. The **y-intercept** appears to be a **positive value**, as the data points start above the origin on the y-axis.

We need to find an equation where the **slope m is negative** and the **y-intercept b is positive**.

The correct equation must have a negative slope and a positive y-intercept. Hence, the equation **$y = -1.9x + 10.1$ is the most appropriate** linear model for the data.

Therefore, the correct answer is **C**.

144. Explanation and Answer

This question pertains to exponential models and linear models, focusing on the difference between compounding interest (exponential growth) and fixed annual increase (or simple interest: linear growth).

Compound Interest:

The amount grows **by a percentage of the current balance**, which can be modeled using an exponential function:

$$A = P(1 + r)^t$$

where **A is the balance**, **P is the principal**, **r is the annual interest rate**, and **t is time in years**.

Simple interest (fixed annual increase):

The amount grows **by a constant value each year**, which can be represented by a linear function:

$$A = P + rt$$

where A is the balance, P is the principal, r is the annual interest rate, and t is time in years.

Compound interest formula for **account X**:

$$A_X = P(1 + r)^t \rightarrow A_X = 700(1 + 0.05)^t$$

the interest earned in t years, I_X , is:

$$I_X = A_X - P \rightarrow I_X = 700(1.05)^t - 700$$

Simple interest formula for **account Y**:

$$A_Y = P + I_Y \rightarrow A_X = 1200 + 40t$$

the interest earned in t years, I_Y , is:

$$I_Y = A_Y - P \rightarrow I_Y = 1200 + 40t - 1200 \\ \rightarrow I_Y = 40t$$

Use DESMOS:

Therefore, due to compounding interest, **account X initially earns less but surpasses account Y's fixed earnings over time**, and the correct answer is **D**.

