

Student Workbook

Fitting Linear Models to Data

Learning Target: I can describe correlation, find a line of best fit, and use it to interpolate, extrapolate, and evaluate residuals.

Vocabulary

Term	Meaning
Scatter Plot	A graph of individual data points showing the relationship between two variables.
Line of Best Fit	A line that best represents the overall trend in a set of data.
Correlation Coefficient (r)	A number between -1 and 1 that measures the strength and direction of a LINEAR relationship.
Interpolation	Predicting a value using an x INSIDE the range of the original data (more reliable).
Extrapolation	Predicting a value using an x OUTSIDE the range of the original data (less reliable).
Residual	Actual value minus predicted value: shows how far off a prediction was for a specific data point.

Key Concept: Reading Correlation

r VALUES

r value	Meaning
Close to +1	Strong positive correlation (both variables increase together)
Close to -1	Strong negative correlation (one increases as the other decreases)
Close to 0	Weak or no LINEAR correlation

COMMON MISTAKE

r close to 0 does NOT mean there's no relationship at all — it only means there's no strong LINEAR relationship. There could still be a strong curved (nonlinear) pattern in the data!

Key Concept: Finding a Line of Best Fit

EXAMPLE 1

Using two representative data points (1, 15) and (5, 35), find the line of best fit.

$$\text{slope} = (35-15)/(5-1) = \underline{\hspace{2cm}}$$

$$y - 15 = \underline{\hspace{2cm}}(x - 1) \rightarrow f(x) = \underline{\hspace{2cm}}$$

Key Concept: Interpolation vs. Extrapolation

EXAMPLE 2

Using $f(x) = 5x + 10$ (built from data with x ranging from 1 to 5):

Predict $f(3)$: this is **interpolation** (3 is inside the data range) $\rightarrow f(3) = \underline{\hspace{2cm}}$

Predict $f(10)$: this is **extrapolation** (10 is outside the data range) $\rightarrow f(10) = \underline{\hspace{2cm}}$ (less reliable!)

Key Concept: Residuals

FORMULA

Residual = actual value - predicted value

EXAMPLE 3

Using $f(x) = 5x + 10$, the actual data value at $x=4$ was $y=32$. Find the residual.

$$\text{Predicted: } f(4) = \underline{\hspace{2cm}} \quad \text{Residual} = 32 - \underline{\hspace{2cm}} = \underline{\hspace{2cm}}$$

Guided Practice — Part A: Describe the Correlation

- A1** As study hours increase, test scores increase, with points tightly clustered near a line.

- A2** As car age increases, resale value decreases, with points loosely scattered.

Guided Practice — Part B: Find the Line of Best Fit

B1 Representative points: (1, 15) and (5, 35)

B2 Representative points: (0, 50) and (10, 10)

Guided Practice — Part C: Interpolate or Extrapolate?

C1 Using B1 (data range $x=1$ to 5), predict at $x=3$ and $x=10$. Classify each.

C2 Using B2 (data range $x=0$ to 10), predict at $x=5$ and $x=15$. Classify each.

Guided Practice — Part D: Calculate the Residual

D1 Using B1's model, the actual value at $x=4$ was $y=32$. Find the residual.

D2 Using B2's model, the actual value at $x=6$ was $y=20$. Find the residual.

BEFORE YOU MOVE ON

Check with a partner: for extrapolation problems, did you note that the prediction is LESS reliable, since it goes beyond the observed data?