Fitting Regression Models to Scatter Plots

Name: ____________	Date: ____________	Score: ______ / 32

Quick Review

A **scatter plot** of two variables x and y is just a cloud of points (x, y). If the cloud looks roughly linear, we can fit a straight line through it – the **line of best fit**, also called the **least-squares regression line**.

Least-squares principle. For every point, the *residual* is the vertical distance from the point to the line: $e_i = y_i - \hat{y}_i$. The best-fit line is the unique line that *minimizes the sum of the squared residuals*. Squaring kills the sign and punishes large misses more than small ones.

Equation form. $\hat{y} = mx + b$ where m is the slope and b is the y-intercept. The hat on $\hat{y}$ is a reminder that it's a *predicted* value, not an actual data value.

Correlation coefficient r. A number between -1 and 1 that measures the strength and direction of the linear relationship. $r = 1$ is perfect positive linear (the points lie exactly on a rising line). $r = -1$ is perfect negative linear (falling line). $r = 0$ means no linear association – though there could still be a nonlinear pattern.

Coefficient of determination r^2. The proportion of variation in y that the regression line explains. If $r^2 = 0.81$, then 81% of the spread in y can be attributed to the linear relationship with x; the other 19% is residual scatter.

When a line is a bad fit. If the scatter plot curves, if there's a funnel-shaped spread (variance grows with x), or if there's a clear outlier dragging the line around, a straight line is not the right model.

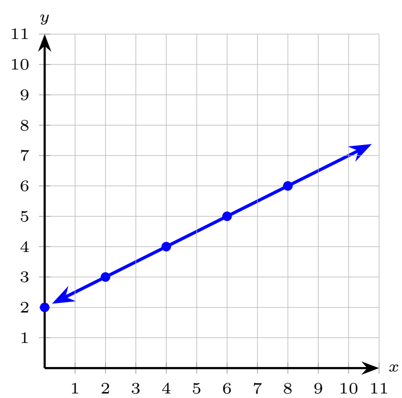
Common slips. Reporting r when r^2 was asked (or vice versa). Treating a strong correlation as proof of causation (it isn't – it's just association). Extrapolating the line far outside the data range (the relationship might break down).

PRACTICE

Use a regression line to predict; interpret slope, intercept, and correlation.

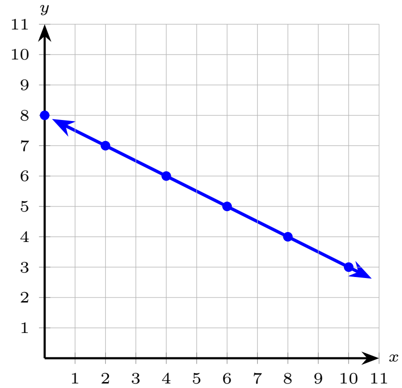
1. What does the line of best fit minimize? ________

2. The line of best fit is drawn through the scatter below. Use it to predict $\hat{y}$ when $x = 10$. ________



3. What's the range of the correlation coefficient r? ________

4. The scatter below has $r = -0.95$. Describe the relationship. ________



5. $r = 0.30$. Describe the relationship. ________

6. $r^2 = 0.64$. What fraction of the variation in y is explained? ________

7. In $\hat{y} = mx + b$, what does m represent? ________

8. $\hat{y} = 0.8x + 15$. Predict $\hat{y}$ when $x = 7$. ________

9. $\hat{y} = -1.8x + 28$. What does 28 represent (when $x = 0$)? ________

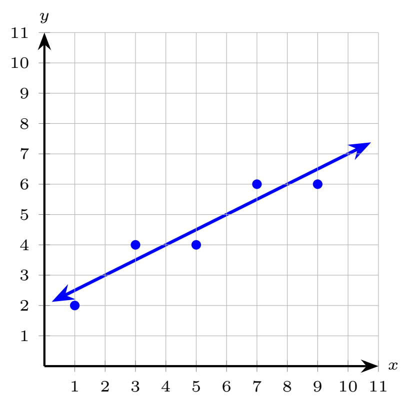
10. $r = 0$. Does that mean x and y are independent? ________

11. True or false: a strong correlation proves causation. ________

12. $\hat{y} = 4x - 2$. Find the residual at $(3, 12)$. ________

13. $r^2 = 0.49$. What is $|r|$? ________

14. From the scatter and its best-fit line below, is the slope positive or negative? ________



15. Why is extrapolating a regression line far outside the data range risky? ________

16. If the residuals show a clear curve when plotted against x, what does that mean? ________

17. $\hat{y} = 2.5x + 10$. Predict $\hat{y}$ at $x = 4$. ________

18. True or false: the regression line always passes through $(\bar{x}, \bar{y})$. ________

19. A line is fit and $r = 0.9$. Find r^2 and interpret. ________

20. In $\hat{y} = mx + b$, what does b represent? ________

Scan Me

Word Problems

21. A nutritionist fits a regression model relating daily calories (x) to weight gain over 30 days (y, in pounds): $\hat{y} = 0.0015x - 2.5$. Interpret the slope in context, and predict the weight change for someone who eats 3000 calories per day. ____

22. A scatter plot of study time (hours) vs. test score yields $r = 0.78$. What proportion of the variation in test scores does the linear model explain? Round to two decimal places. ____

23. A researcher finds a correlation of $r = 0.92$ between ice-cream sales and shark attacks at the beach. Can she conclude that ice cream causes shark attacks? Explain. ____

24. A real-estate agent fits $\hat{y} = 120x + 30{,}000$ to data on square footage (x) and home price (y, in dollars) for homes between $1{,}000$ and $3{,}000$ sq ft. Predict the price of a $2{,}000$ sq ft home. Why would using this model to predict a 200 sq ft tiny home or a $20{,}000$ sq ft mansion be risky? ____

Additional Practice

25. Find the mean of $4, 6, 8, 10$. ________

26. Find the median of $3, 9, 4, 10, 7$. ________

27. Find the range of $12, 5, 9, 20$. ________

28. Find the mode of $2, 3, 3, 5, 8$. ________

29. Find z for $x = 72$, mean 60, standard deviation 6. ________

30. Interpret $z = -1.5$. ________

31. Predicted y for $\hat{y} = 2x + 5$ at $x = 6$. ________

32. Residual if actual $= 20$ and predicted $= 17$. ________

Answer Keys

1. sum of squared residuals
2. 7
3. $[-1, 1]$
4. strong negative linear
5. weak positive linear
6. 64%
7. slope – predicted change in y per unit change in x
8. 20.6
9. $\hat{y} = 28$ when $x = 0$
10. Not necessarily
11. False
12. 2
13. 0.7
14. positive
15. the linear pattern may not hold there
16. a line is the wrong model
17. 20
18. True
19. $r^2 = 0.81;\ 81\%$ of variation explained
20. $\hat{y}$ at $x = 0$
21. $\hat{y} = 2$ lb; slope $= 0.0015$ lb/cal/day
22. $r^2 \approx 0.61$
23. No – a confounder (warm weather) drives both
24. $\hat{y} = \$270{,}000;$ extrapolation is unreliable

Additional Practice Answers

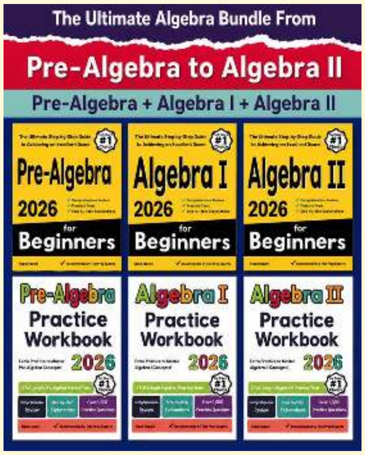
25. 7
26. 7
27. 15
28. 3
29. 2
30. 1.5 SD below mean
31. 17
32. 3

Additional Practice: Answers for all numbered items, including the added practice, are shown in the grid above.

Step-by-Step Explanations

1. Residual $= y - \hat{y}$. Squaring kills signs and weights big misses more. The best fit makes the total of those squares as small as possible.
2. The fitted line is $\hat{y} = 0.5x + 2$. At $x = 10$: $\hat{y} = 0.5(10) + 2 = 5 + 2 = 7$ – right where the line crosses $x = 10$.
3. One steady path is: r measures linear strength on a -1 to 1 scale. Anything outside that range is a calculation error. That gives a quick check on the answer.
4. The points fall along a downward line, so y goes down as x goes up (negative). They hug the line tightly and $|r| = 0.95$ is near $1 \Rightarrow$ strong.
5. Positive sign, small magnitude. y tends to rise with x but the cloud is loose.
6. Keep the rule visible: $r^2 = 0.64 = 64\%$. The rest (36%) is unexplained scatter. That gives a quick check on the answer.
7. That's the standard slope interpretation, with the caveat that it's *predicted* change, not exact.
8. Start with the key idea: $\hat{y} = 0.8(7) + 15 = 5.6 + 15 = 20.6$. Slope times x, then add the intercept. That gives a quick check on the answer.
9. A careful way to see it: 28 is the y-intercept – the predicted value of y when $x = 0$. In a car-value model where x is age in years, this is the predicted value of a brand-new car. That gives a quick check on the answer.
10. Keep the rule visible: $r = 0$ only rules out a *linear* relationship. There could still be a strong nonlinear pattern (e.g., $y = x^2$ has $r \approx 0$ on a symmetric range). That gives a quick check on the answer.
11. Correlation says the variables move together. Causation requires ruling out confounders – typically through a controlled experiment.
12. Predicted: $\hat{y} = 4(3) - 2 = 10$. Residual: $y - \hat{y} = 12 - 10 = 2$. The point sits 2 units above the line.
13. A careful way to see it: $|r| = \sqrt{r^2} = \sqrt{0.49} = 0.7$. The sign comes from the scatter plot, not from r^2. That gives a quick check on the answer.
14. The line rises from left to right, so the slope is positive – and the correlation r shares that sign.
15. The line was fit to the observed range. Outside that range, the relationship might curve, level off, or break down entirely.
16. A good linear fit has residuals that look random. A curved pattern says the data has nonlinear structure your line is missing.
17. A careful way to see it: $\hat{y} = 2.5(4) + 10 = 10 + 10 = 20$. This is the part to check before moving on, because it keeps the answer tied to the original question.
18. The least-squares line is guaranteed to pass through the point of means – a useful sanity check when computing one by hand.
19. One steady path is: $r^2 = 0.9^2 = 0.81$. The model explains 81% of the variation in y; the remaining 19% is residual. That gives a quick check on the answer.
20. Start with the key idea: b is the y-intercept – the predicted value of y when x is zero. Whether that's meaningful in context depends on whether $x = 0$ falls inside the data range. That gives a quick check on the answer.
21. Plug in: $\hat{y} = 0.0015(3000) - 2.5 = 4.5 - 2.5 = 2$ pounds gained. The slope 0.0015 means each extra calorie per day predicts about 0.0015 pounds of weight gain over the 30-day window. Equivalently, 1000 extra cal/day predicts about 1.5 lb gain over a month. (Real nutrition is messier than a line, but this is the model's prediction.)
22. Keep the rule visible: $r^2 = 0.78^2 \approx 0.6084$, so about 0.61 – or 61% – of the variation in test scores is explained by the linear relationship with study time. The other 39% is variation in scores that the model can't account for (test anxiety, prior knowledge, sleep, you name it). That gives a quick check on the answer.
23. A strong correlation says the two variables move together, but it can't tell you *why*. Here, both ice-cream sales and shark attacks shoot up in hot months – the lurking variable is temperature. This is a classic warning: correlation $\neq$ causation. To establish causation, you'd need a controlled experiment (impossible here, thankfully).
24. Predict: $\hat{y} = 120(2000) + 30{,}000 = 240{,}000 + 30{,}000 = \$270{,}000$. For tiny homes or huge mansions, we're *extrapolating* far outside the data range (1000–3000 sq ft). The relationship might curve, plateau, or behave entirely differently outside that range – think of how price per sq ft changes dramatically once a property crosses into the luxury market. The model can't tell us. Stay inside the data.

Build Algebra Confidence From Pre-Algebra Through Algebra II



The Complete Algebra Success Bundle

Pre-Algebra, Algebra I, and Algebra II in one clear path

Friendly lessons, focused practice, and full-review support for every stage.

Scan Me

Scan for the Bundle

6 Books
3 Courses
1 Path

Bundle Value: Six coordinated books help students review missing skills, learn new algebra topics, and practice until the steps feel natural.

Complete Course Path

- ✓ Starts with Pre-Algebra foundations
- ✓ Moves smoothly into Algebra I skills
- ✓ Extends learning through Algebra II topics
- ✓ Great for review, tutoring, and summer study

One bundle, one steady path.

Step-by-Step Lessons

- ✓ Plain-English explanations students can follow
- ✓ Worked examples that show every important step
- ✓ Common mistakes called out before they stick
- ✓ Skill-building practice after each lesson
- ✓ Helpful for independent study or class support

Less guessing. More understanding.

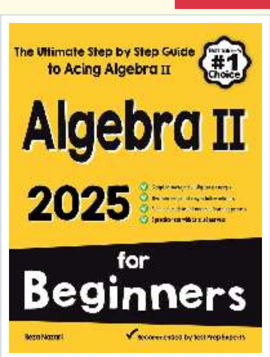
Practice That Sticks

- ✓ Matching practice workbooks for extra repetition
- ✓ Review sets to keep older skills fresh
- ✓ Answer explanations for checking thinking
- ✓ Strong support before tests and final exams
- ✓ Designed to build fluency and confidence

Practice today. Remember tomorrow.

STUDENT FAVORITE • Master Algebra II From the Ground Up

Algebra II for Beginners



✓ 100% Guaranteed
✓ Lifetime Support
✓ Trusted by Teachers

Start Your Algebra Journey Today! →

Written by a top math teacher & aligned with national and state Algebra II courses. From polynomial functions to logarithms, trigonometry, and rational expressions — explained the easy way.

- ✓ **Complete coverage** of every Algebra II concept — perfect companion to these worksheets
- ✓ **Step-by-step explanations** with worked examples on every topic
- ✓ **QR codes in every chapter** for free video lessons & bonus practice
- ✓ **2 full-length practice tests** with detailed answer keys

⋆ STUDENT'S #1 CHOICE ⋆

Teacher-recommended • 12,000+ Happy Students

↓ PDF EDITION

Scan Me

Instant download • any device

PAPERBACK

Scan Me

Paperback on Amazon

Hold it in your hands

Pair these free worksheets with *Algebra II for Beginners* and you have a complete self-paced course — concept lessons, daily practice, and full exam-style reviews, all in one path. → EffortlessMath.com/product/algebra-ii-for-beginners