

Regression Analysis: Lines, Residuals and r

Grade 12 – AP Statistics math · Worksheet Answer Key · Standards: AP Statistics Unit 5 (2026-27)

The answer key computes every slope, intercept, correlation and residual with the exact summary statistics, rounds only at the end, and states each interpretation in context with units. Raw-data lines were verified with software.

Standards: AP Statistics Unit 5 (2026-27) · **Level:** Grade 12 AP Statistics

Problem 1 Describing a Scatterplot

Direction: positive (more drill time goes with higher percentages). Form: linear. Strength: strong ($r \approx 0.98$). Unusual features: no outliers or gaps.

Positive, linear, strong, no outliers

Problem 2 Explanatory and Response Variables

(a) Explanatory: hours of sleep; response: quiz score. (b) Explanatory: temperature; response: number of customers.

(a) sleep → quiz score; (b) temperature → customers

Problem 3 Correlation Facts

(a) The value closest to -1 or 1 in size is -0.92 . (b) -0.15 . (c) r is unchanged, because correlation does not depend on units.

(a) -0.92 ; (b) -0.15 ; (c) unchanged

Problem 4 Slope and Intercept from Summary Statistics

$$b = 0.78(18.4/3.1) = 4.630. \quad a = 152 - 4.630(64.5) = -146.61.$$

$$\hat{y} = -146.61 + 4.63x \text{ (weight in pounds, height in inches)}$$

Problem 5 Interpreting the Slope

For each additional inch of height, the predicted weight increases by about 4.63 pounds, on

average.

About 4.63 pounds more per inch

Problem 6 Interpreting the Intercept

The intercept -146.6 is the predicted weight of a woman 0 inches tall. A negative weight is impossible, and 0 inches is far outside the data (heights near 64.5 inches), so the intercept has no practical meaning; it only positions the line.

Not meaningful

Problem 7 Making a Prediction

$$\hat{y} = -146.61 + 4.630(70) = 177.5.$$

About 177 pounds

Problem 8 Computing a Residual

Residual = actual – predicted = $190 - 177.5 = 12.5$. She weighs about 12.5 pounds more than the line predicts for her height.

Residual ≈ 12.5 pounds (above the line)

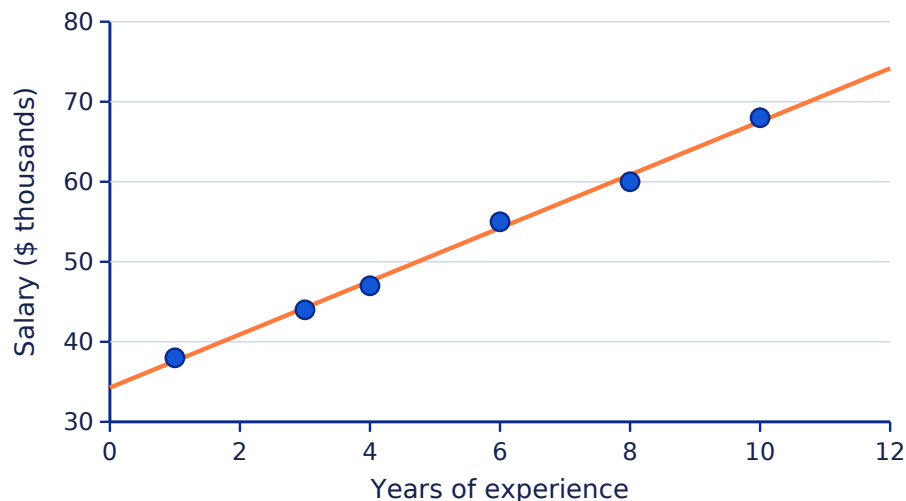
Problem 9 Interpreting r Squared

$r^2 = 0.78^2 = 0.6084$. About 60.8% of the variation in weight is explained by the linear relationship with height.

$r^2 \approx 0.608$: about 60.8% of the variation

Problem 10 A Regression from Raw Data

$\bar{x} = 5.33$, $s_x = 3.327$, $\bar{y} = 52.00$, $s_y = 11.082$, $r = 0.9983$. $b = 3.325$, $a = 34.27$, so $\hat{y} = 34.27 + 3.33x$. At $x = 5$: $\hat{y} = 50.9$.



The six points and the line.

$r \approx 0.998$; $\hat{y} = 34.27 + 3.33x$; predicted salary $\approx$ \$50.9 thousand

Problem 11 A Curved Residual Pattern

The residuals form a U-shaped curve: positive at both ends and negative in the middle. A random scatter would show no pattern, so the pattern shows that the relationship is curved and a linear model is not appropriate.

Not appropriate: curved residual pattern

Problem 12 A Random Residual Pattern

The residuals are scattered randomly above and below 0, with no curve and no changing spread. This supports using a linear model.

Appropriate: no pattern in the residuals

Problem 13 An Influential Point

Without the point: slope = 1.973, $r = 0.989$. With the point: slope = 0.091, $r = 0.117$. One point changed the slope from strongly positive to nearly flat and weakened r a great deal. It is far from the others in the x direction, so it is an influential point.

Yes, influential: slope 1.97 $\rightarrow$ 0.09

Problem 14 Outlier or Influential

Point A has a large residual but is in the middle of the x -values, so it is an outlier with little pull on the slope. Point C is far from the other points in x and far below the line, so it is the most

influential point. Point B is far in x but on the line, so it has little effect on the slope (though it raises r).

A: outlier; C: most influential; B: little effect on the slope

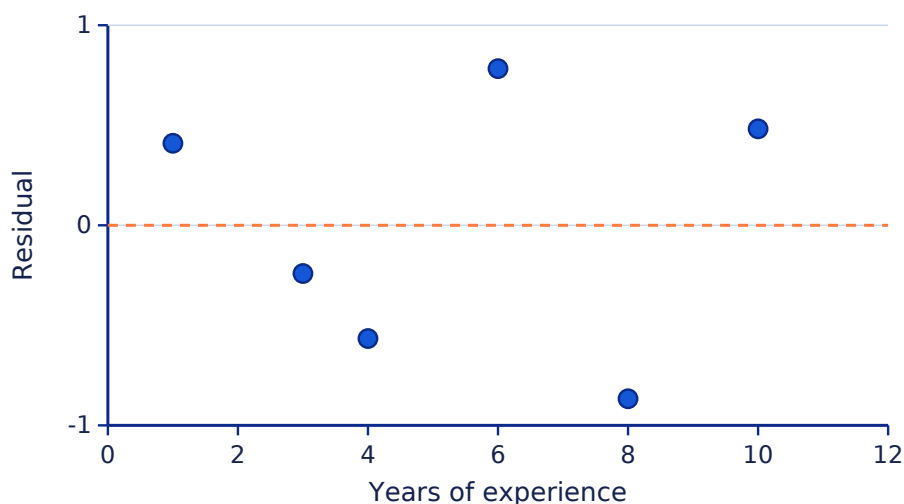
Problem 15 Extrapolation

$\hat{y} = 12.4 + 3.8(20) = 88.4$ thousand. The data cover only $x = 0$ to 8, so $x = 20$ is extrapolation, and the linear trend may not continue (the market could level off). The prediction is unreliable.

88.4 thousand, but unreliable extrapolation

Problem 16 Standard Deviation of the Residuals

Residuals: 0.41, -0.24 , -0.57 , 0.78, -0.87 , 0.48. Sum of squares = 2.145. $s = \sqrt{(2.145/4)} = 0.732$. A typical prediction of salary is off by about 0.73 thousand dollars.



Residuals for the salary model.

$s \approx 0.73$ thousand dollars

Problem 17 Finding r from r Squared

$r = \pm\sqrt{0.64} = \pm 0.8$. The sign of r matches the sign of the slope, so $r = -0.8$. It cannot be $+0.8$ with a negative slope.

$r = -0.8$

Problem 18 Correlation and Causation

The correlation does not show that ice cream causes rescues. A lurking variable, hot weather,

increases both ice cream sales and the number of swimmers, so the two variables are associated without one causing the other. Limiting sales would not reduce rescues.

Lurking variable: hot weather; association is not causation

Problem 19 Reading Computer Output

(a) $\hat{y} = -38.5 + 1.38 \times \text{temperature}$. (b) Each additional degree is associated with about 1.38 more cups sold, on average. (c) About 96.0% of the variation in cups sold is explained by the linear relationship with temperature. (d) A typical prediction is off by about 2.4 cups. (e) $r = +\sqrt{0.960} = 0.980$, positive because the slope is positive.

$$\hat{y} = -38.5 + 1.38x; r \approx 0.980$$

Problem 20 Complete Regression Analysis

(a) $b = -0.91(4.5/2.1) = -1.950$; $a = 14.8 - (-1.950)(5.2) = 24.94$; $\hat{y} = 24.94 - 1.95x$. (b) Each additional year of age is associated with a predicted price decrease of about \$1.95 thousand, on average. (c) $\hat{y} = 24.94 - 1.950(8) = 9.34$, about \$9,340. (d) Residual = $4.1 - 9.34 = -5.24$, so the car sold for about \$5,240 below the prediction. (e) $r^2 = 0.8281$, about 82.8%.

$$\hat{y} = 24.94 - 1.95x; \hat{y}(8) \approx 9.34; \text{residual} \approx -5.24; r^2 \approx 82.8\%$$