

Regression Analysis: Lines, Residuals and r

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

The scoring guide shows how points are earned. Interpretations must mention the variables and units, and conclusions about causation or extrapolation need a reason to earn full credit.

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

Question 1 Describing Association (2 points)

The association is negative, linear and strong ($r \approx -0.996$), with no outliers. No: this is observational data, so a cause-and-effect claim is not justified; lurking variables (for example, stress) could affect both.

Negative, linear, strong; association is not causation

Scoring: 1 point for describing direction, form and strength, 1 point for the causation comment.

Question 2 Least-Squares Line (3 points)

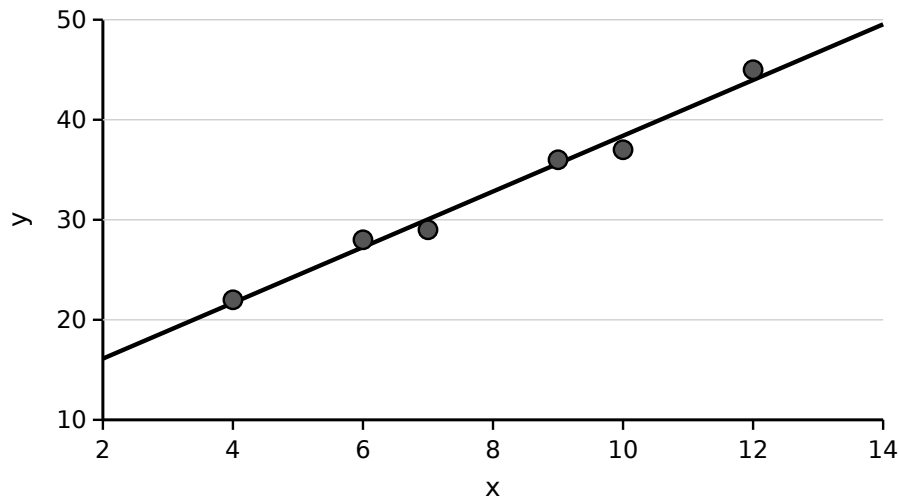
(a) $b = 0.85(6.2/2.0) = 2.635$; $a = 71 - 2.635(10.5) = 43.33$; $\hat{y} = 43.33 + 2.63x$. (b) Each extra hour of study per week is associated with a predicted score increase of about 2.63 points. (c) At 0 hours of study the predicted score is 43.3; this is outside the data range (about 5 to 15 hours), so it should not be taken literally.

$$\hat{y} = 43.33 + 2.63x$$

Scoring: 1 point each part.

Question 3 Prediction and Residual (2 points)

(a) $\hat{y} = 10.55 + 2.786(9) = 35.62$. (b) Residual = $36 - 35.62 = 0.38$. The actual value is 0.38 above the predicted value.



Points and fitted line.

$\hat{y} \approx 35.6$; residual ≈ 0.38

Scoring: 1 point each part.

Question 4 Interpreting r (2 points)

(a) $r^2 = 0.5184$, so about 51.8% of the variation in fitness scores is explained by the linear relationship with weekly TV hours. (b) The sign of r is negative: people who watch more TV tend to have lower fitness scores.

$r^2 \approx 0.518$: about 51.8% of the variation

Scoring: 1 point each part.

Question 5 Residual Plot (3 points)

(a) The U shape shows that the relationship is curved, so a linear model is not appropriate. (b) The residuals would be randomly scattered about 0 with no pattern and roughly constant spread. (c) Yes: r^2 measures how much variation a line explains overall, and a curved relationship can still have a large value; the residual plot is needed to detect the curve.

Curved relationship; the linear model is not appropriate

Scoring: 1 point each part.

Question 6 Influential Point (2 points)

(a) It is far from the other points in the x direction and far below the pattern, so it is both an outlier and an influential point. (b) The slope will decrease (the line is pulled down at the right) and r will become weaker.

Outlier and influential; slope and r decrease

Scoring: 1 point each part.

Question 7 Computer Output (3 points)

(a) $\hat{y} = 412.5 - 58.2 \times \text{age}$. (b) For each additional year of age, the predicted price decreases by about \$58.20, on average. The value $S = 31.4$ means a typical prediction is off by about \$31.40. (c) $r = -\sqrt{0.884} = -0.940$, negative because the slope is negative.

$\hat{y} = 412.5 - 58.2x; r \approx -0.940$

Scoring: 1 point each part.

Question 8 Extrapolation and Causation (3 points)

(a) $\hat{y} = 8.1 + 0.62(30) = 26.7$. Since 30 is far above the data range (2 to 14), this is extrapolation and may be unreliable. (b) No. The data are observational, so the association may be due to lurking variables such as city wealth; a cause-and-effect conclusion would need a randomized experiment.

26.7, unreliable; causation is not justified

Scoring: 1 point for the prediction, 1 point for the extrapolation comment, 1 point for the causation explanation.