

Regression Analysis: Lines, Residuals and r

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

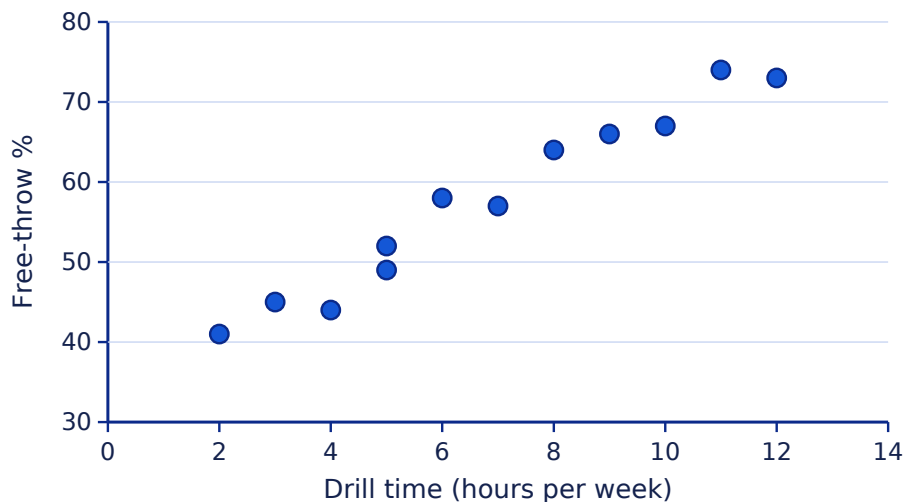
Name: _____ Date: _____ Class: _____

These twenty problems train the Unit 5 skills of AP Statistics: describing scatterplots, using correlation, finding and interpreting least-squares lines, reading residual plots, recognizing influential points and avoiding extrapolation and false causation. Later problems use raw data and computer output.

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

Problem 1 Describing a Scatterplot (easy)

The scatterplot shows weekly drill hours and free-throw percentage for 12 players.



Drill time and free-throw percentage.

Describe the direction, form, strength and any unusual features.

Problem 2 Explanatory and Response Variables (easy)

For each pair, choose the explanatory variable and the response variable. (a) A student's hours of sleep and the quiz score the next day. (b) The outdoor temperature and the number of customers at a coffee kiosk.

Problem 3 Correlation Facts (medium)

Four values of r are -0.92 , -0.15 , 0.35 and 0.88 . (a) Which describes the strongest linear relationship? (b) Which describes a weak negative relationship? (c) If the heights are converted from inches to centimeters, what happens to r between height and weight?

Problem 4 Slope and Intercept from Summary Statistics (medium)

For adult women, height has mean 64.5 inches and standard deviation 3.1 inches; weight has mean 152 pounds and standard deviation 18.4 pounds; the correlation is 0.78. Find the least-squares line for predicting weight from height.

Problem 5 Interpreting the Slope (medium)

The line for predicting weight (pounds) from height (inches) is $\hat{y} = -146.61 + 4.63x$. Interpret the slope in context.

Problem 6 Interpreting the Intercept (medium)

Interpret the y-intercept of $\hat{y} = -146.61 + 4.63x$ and decide whether it is meaningful.

Problem 7 Making a Prediction (easy)

Use $\hat{y} = -146.61 + 4.63x$ to predict the weight of a woman who is 70 inches tall.

Problem 8 Computing a Residual (medium)

A woman 70 inches tall weighs 190 pounds. Using $\hat{y} = -146.61 + 4.63x$, find her residual and interpret it.

Problem 9 Interpreting r Squared (easy)

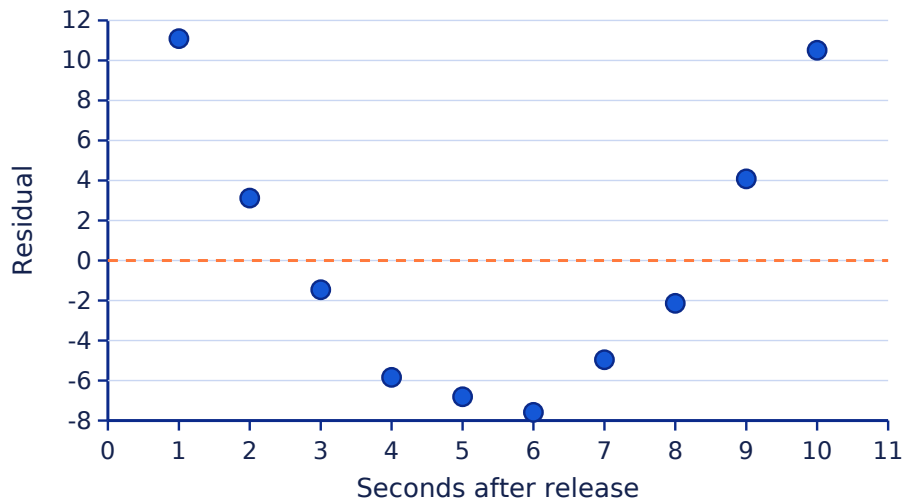
For the height and weight data the correlation is $r = 0.78$. Find r^2 and interpret it.

Problem 10 A Regression from Raw Data (hard)

A company recorded years of experience and salary (thousands of dollars) for six employees: (1, 38), (3, 44), (4, 47), (6, 55), (8, 60), (10, 68). Find the correlation and the least-squares line, and predict the salary for 5 years of experience.

Problem 11 A Curved Residual Pattern (medium)

A line was fitted to data on the speed of a ball (x , seconds after release) and its height. The residual plot is shown.

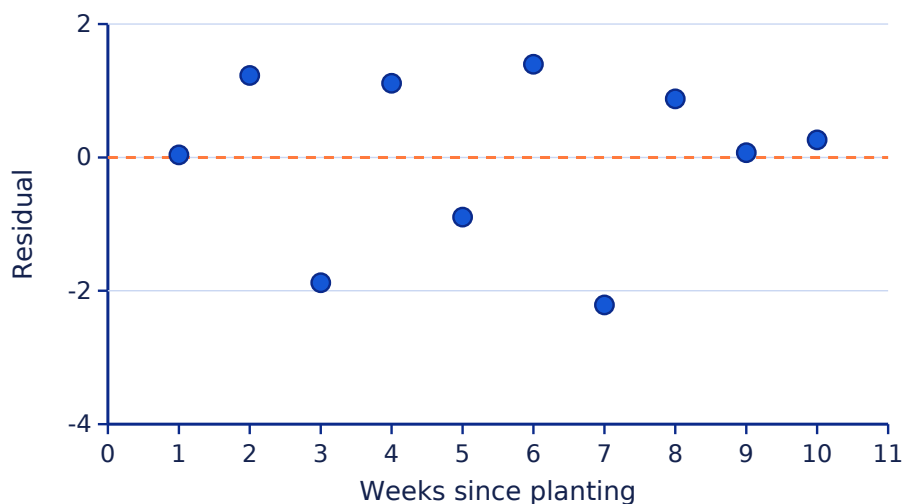


Residual plot for a linear model.

Is a linear model appropriate? Explain.

Problem 12 A Random Residual Pattern (medium)

A line was fitted to data on a plant's height (y) and weeks since planting (x). The residual plot is shown.

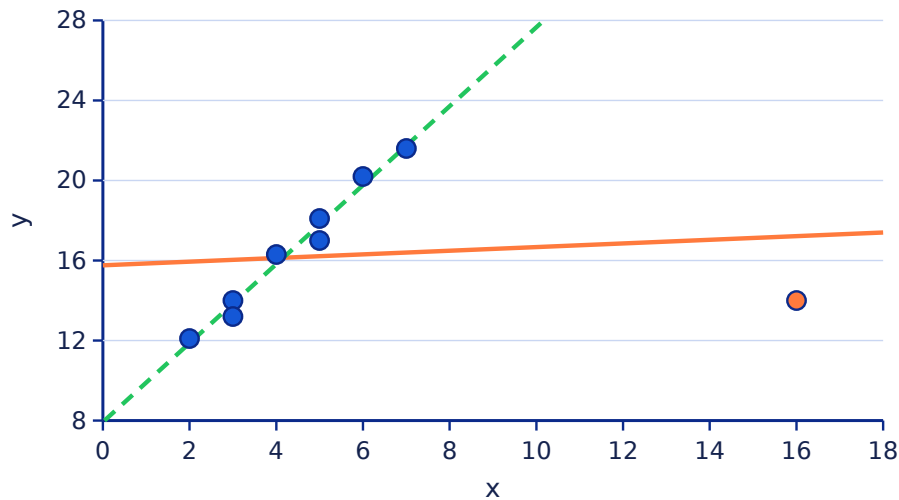


Residual plot for plant growth.

Is a linear model appropriate? Explain.

Problem 13 An Influential Point (hard)

Eight points follow a clear linear trend. A ninth point at (16, 14) is added. The solid line fits all nine points; the dashed line fits the first eight.



Solid: all nine points; dashed: first eight.

Compare the slopes and the correlations, and state whether the point at (16, 14) is influential.

Problem 14 Outlier or Influential (medium)

Three points are added to a scatterplot that has a strong positive linear pattern for x from 10 to 30. Point A is at $x = 20$ with a y -value far above the line. Point B is at $x = 60$, on the line extended. Point C is at $x = 60$ and far below the line. Which point is an outlier, which is most influential, and which has little effect on the slope?

Problem 15 Extrapolation (medium)

A streaming service finds that, for the 9 years from 2015 ($x = 0$) to 2023 ($x = 8$), subscribers (thousands) follow $\hat{y} = 12.4 + 3.8x$. A manager predicts the subscribers for 2035 ($x = 20$). Make the prediction and comment on its reliability.

Problem 16 Standard Deviation of the Residuals (hard)

For the experience and salary data ($x = 1, 3, 4, 6, 8, 10$; $y = 38, 44, 47, 55, 60, 68$) the line is $\hat{y} = 34.27 + 3.33x$. Find the residuals and the standard deviation of the residuals $s = \sqrt{(\sum e^2 / (n - 2))}$. Interpret s .

Problem 17 Finding r from r Squared (medium)

A regression of fuel use on vehicle weight has $r^2 = 0.64$ and a negative slope. What is the correlation r ? Could the correlation be $+0.8$?

Problem 18 Correlation and Causation (medium)

Over many summer months, a coastal town finds a strong positive correlation between ice cream sales and the number of swimmers rescued. A council member proposes limiting ice cream sales. Explain the flaw in this reasoning.

Problem 19 Reading Computer Output (hard)

A regression of cups sold on temperature for a juice stand gives the following output.

Predictor	Coef
Constant	-38.5
Temperature	1.38

$S = 2.4$, $R\text{-sq} = 96.0\%$. (a) Write the regression equation. (b) Interpret the slope. (c) Interpret $R\text{-sq}$. (d) Interpret S . (e) Find r .

Problem 20 Complete Regression Analysis (hard)

For used cars, age in years has mean 5.2 and standard deviation 2.1; price in thousands of dollars has mean 14.8 and standard deviation 4.5; the correlation is -0.91 . (a) Find the line for predicting price from age. (b) Interpret the slope. (c) Predict the price of an 8-year-old car. (d) A particular 8-year-old car sold for \$4,100. Find its residual. (e) What percent of the variation in price is explained by age?