

Residuals & Influential Points

AP Statistics | Unit 2 | residual plots, r^2 , s , unusual points

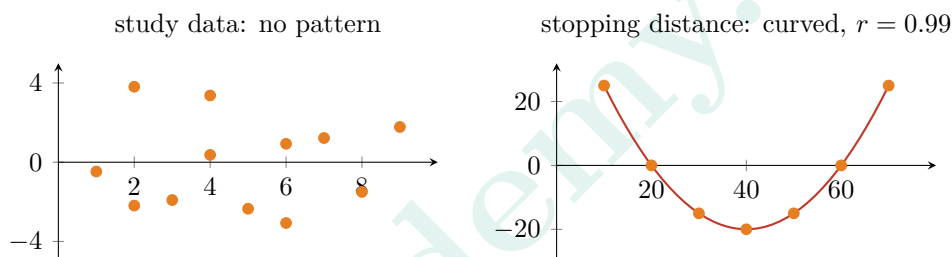
1. Residuals

$$\text{residual} = y - \hat{y} = \text{actual} - \text{predicted}.$$

For the 12 students with $\hat{y} = 54.75 + 3.72x$ (hours studied, exam score; hypothetical data): a student who studied 4 hours and scored 73 has $\hat{y} = 69.63$ and residual $73 - 69.63 = 3.37$. *This student scored about 3.4 points higher than predicted for someone who studied 4 hours.* Positive residual: above the line. Negative: below.

2. The residual plot

Plot the residuals against x . **No pattern** (points scattered randomly above and below 0) means a linear model is appropriate. A **curved pattern** means it is not, however high r is.



A high r says the points are close to a line; only the residual plot says a line is the right *shape*.

3. r^2 and s

Interpretations AP readers look for

r^2 (coefficient of determination): $r^2 = 1 - \frac{57}{1029} \approx 0.944$. *About 94.4% of the variation in exam scores is explained by the linear relationship with hours studied.*

s (standard deviation of the residuals): $s = \sqrt{\frac{57.4}{10}} \approx 2.40$. *When we use this line, our predictions of exam score are typically off by about 2.4 points.*

4. Outliers, leverage, influence

- **Outlier:** a point with a large residual. Adding (4.5, 95) shifts the line up but the slope barely changes ($3.72 \rightarrow 3.64$).
- **High-leverage point:** an extreme x value. Adding a student at 16 hours exactly on the line leaves the slope at 3.72.
- **Influential point:** removing it would substantially change the slope, intercept or r . Adding (16, 70) drops the slope from 3.72 to 1.26 and r from 0.97 to 0.54.

High-leverage points that do not follow the pattern are the dangerous ones.

Where students lose points

1. The residual backwards: it is actual – predicted.
2. Judging the model by r alone instead of checking the residual plot.
3. Reading r^2 as “94% of the points are on the line”.
4. Interpreting s without units or context.
5. Calling every unusual point influential.

Remember

- Residual = actual – predicted.
- Residual plot: no pattern → a line is appropriate; a curve → it is not.
- r^2 : % of variation in y explained; s : typical prediction error.
- Outlier: big residual; leverage: extreme x ; influential: changes the line.

Now practise it

Theory sticks once you use it. On **sad-academy.com** you'll find practice questions for this exact subtopic in the AP format, and an AI that grades your free-response answers against the scoring guidelines, the way an AP reader would.

sad-academy.com

Data in this guide are hypothetical. AP® is a registered trademark of College Board, which is not affiliated with and does not endorse this material.