

AP Statistics - Least-Squares Regression and Residuals - Answer Key

Common Core AP STAT 2.6, AP STAT 2.7, AP STAT 2.8, AP STAT 2.9, HSS-ID.B.6, HSS-ID.C.7 · maths-pdf.fr/en/

Each solution shows the substitution into the regression equation, the residual calculation, and the interpretation in context. Equations are rounded to the precision shown in the problem.

AP Statistics · Common Core State Standards: **AP STAT 2.6, AP STAT 2.7, AP STAT 2.8, AP STAT 2.9, HSS-ID.B.6, HSS-ID.C.7**

Exercise 1 - Predict

$$\hat{y} = 40 + 0.12(1800) = 40 + 216 = 256.$$

About \$256,000.

Exercise 2 - Interpret the slope

The slope 0.12 is in thousands of dollars per square foot.

For each additional square foot, the predicted price increases by about \$120.

Exercise 3 - Interpret the intercept

When $x = 0$ classes, $\hat{y} = 3.5$.

A student with no classes is predicted to do 3.5 minutes of homework. It is not meaningful because $x = 0$ is outside the sensible range.

Exercise 4 - Compute a residual

$$\hat{y} = 12 + 3(5) = 27. e = 30 - 27 = 3.$$

Residual +3: Sofia's score was 3 points higher than predicted.

Exercise 5 - Find the slope and intercept

$$b = r \frac{s_y}{s_x} = 0.60 \times \frac{20}{3} = 4. \quad a = 150 - 4(66) = -114.$$

$$\hat{y} = -114 + 4x.$$

Exercise 6 - Point of means

$$\hat{y} = -114 + 4(66) = -114 + 264 = 150, \text{ which equals the mean of } y.$$

Yes: the line goes through (66, 150).

Exercise 7 - Interpret r squared

r squared is the proportion of variation explained.

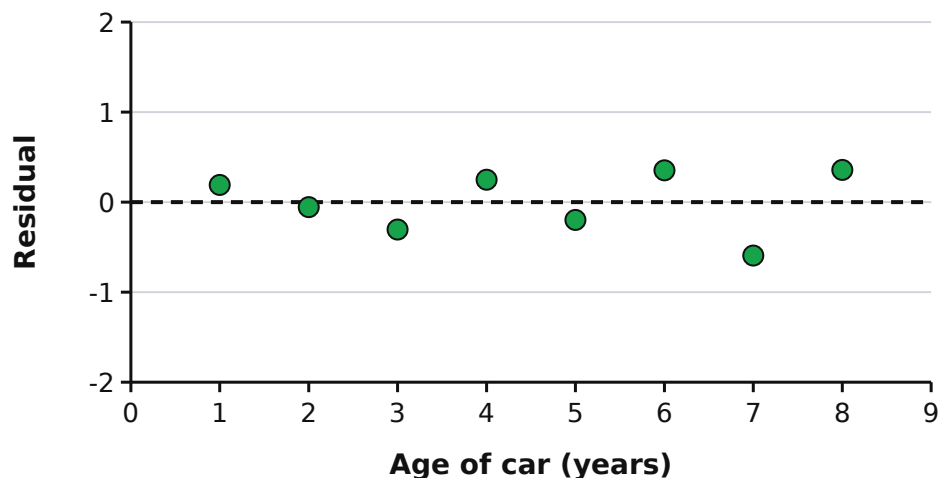
About 64% of the variation in battery life is explained by the linear relationship with screen brightness.

Exercise 8 - Get r from r squared

$$r = \pm\sqrt{0.81} = \pm 0.9. \text{ The slope is positive, so } r \text{ is positive.}$$

$$r = 0.9.$$

Exercise 9 - Residual plot: good fit

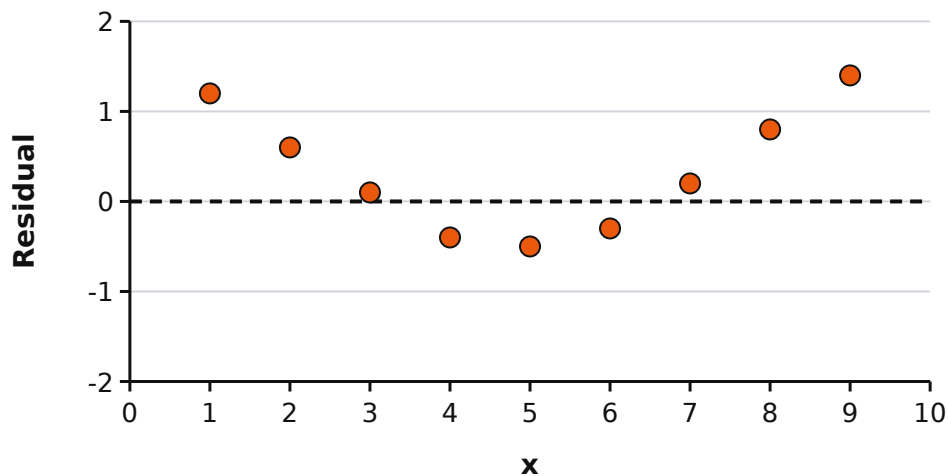


A residual plot with no pattern: a linear model is appropriate.

The residuals scatter randomly around zero with no pattern.

Yes, a linear model is appropriate.

Exercise 10 - Residual plot: curved



A curved residual pattern signals that a straight line is not the right model.

The residuals form a U shape.

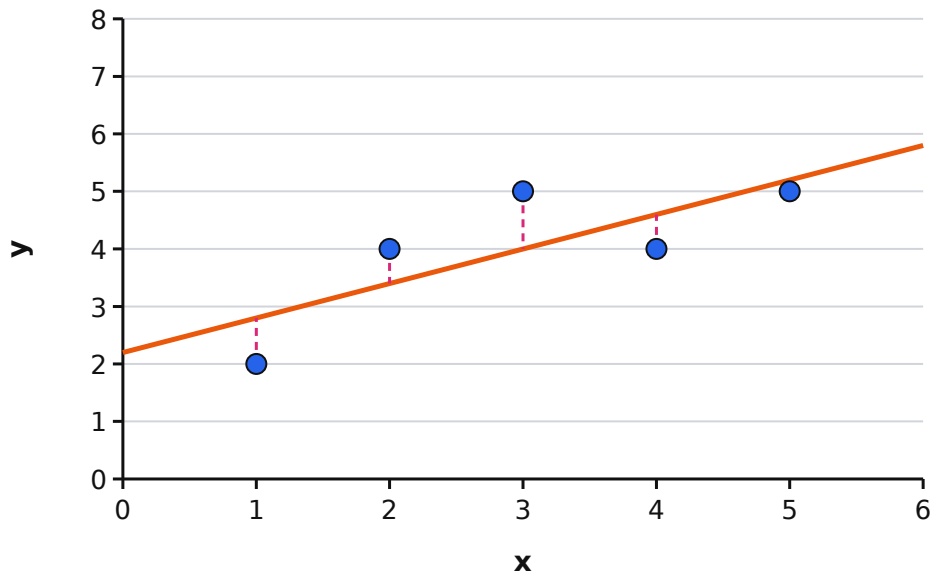
No; the relationship is curved. Try a transformation of the data or a curved model.

Exercise 11 - Extrapolation

$$\hat{y} = 26.56 - 2.25(15) = -7.22 \text{ thousand dollars.}$$

-7.22 thousand dollars is below zero, which is impossible. Age 15 is outside the data range (extrapolation), so the prediction is unreliable.

Exercise 12 - LSRL for five points



Least-squares line through five points.

$$b = 0.775 \times \frac{1.095}{1.581} \approx 0.6. \quad a = 4 - 0.6(3) = 2.2.$$

$$\hat{y} = 2.2 + 0.6x.$$

Exercise 13 - Residuals add to zero

Predicted: 2.8, 3.4, 4.0, 4.6, 5.2. Residuals: -0.8, 0.6, 1.0, -0.6, -0.2.

Sum: $-0.8 + 0.6 + 1.0 - 0.6 - 0.2 = 0$.

Residuals: -0.8, 0.6, 1.0, -0.6, -0.2; their sum is 0.

Exercise 14 - Standard deviation of residuals

Squares: 0.64, 0.36, 1.00, 0.36, 0.04, sum 2.4.

$$s = \sqrt{\frac{2.4}{5-2}} = \sqrt{0.8} \approx 0.894.$$

s = 0.894: actual y values typically differ from the predicted values by about 0.894 units.

Exercise 15 - Read computer output

Predictor	Coef
Constant	25.6
Hours	-3.2
R-Sq	78%

Output for sleep data.

Equation: $\hat{y} = 25.6 - 3.2x$.

r squared = 0.78, slope negative, so $r = -\sqrt{0.78} \approx -0.883$.

For each extra hour of screen time, the predicted score drops by 3.2; $r = -0.883$.

Exercise 16 - Influential point

Its removal substantially changes the slope.

It is an influential point, probably with an extreme x value.

Exercise 17 - Compare two predictions

10 miles: $\hat{y} = 43$, $e = 39 - 43 = -4$. 12 miles: $\hat{y} = 48$, $e = 52 - 48 = 4$.

Both residuals have size 4 (one low, one high), so the predictions are equally accurate; the 10-mile cost is overpredicted and the 12-mile cost underpredicted.

Exercise 18 - Why least squares?

For any candidate line, square each vertical distance from a point to the line and add them.

The least-squares line is the one that makes this sum of squared residuals as small as possible.

Exercise 19 - Describe a fit

High r squared, small s, and a patternless residual plot all support a linear model.

Linear regression fits well: 92% of the variation in mileage is explained by speed, and predictions typically miss by about 1.4 mpg.

Exercise 20 - Two predictions

$$\hat{y} = 15 + 0.8(40) = 47. \quad e = 52 - 47 = 5.$$

47 minutes predicted; residual +5 (the order took 5 minutes longer than predicted).