The Least-Squares Regression Line

AP Statistics | Unit 5 | slope, intercept, r^2, computer output, extrapolation

1. The model

Least-squares regression line

$$\hat{y} = a + bx$$

$\hat{y}$ is the **predicted** value of y (the hat means predicted), x is the explanatory variable, b the slope, a the y-intercept (the predicted y when $x = 0$). Of all lines, it makes the sum of the squared residuals $\sum(y - \hat{y})^2$ as small as possible.

14 used cars of one model, age (years) and price (\$1,000s): $\widehat{\text{price}} = 22.93 - 1.40\,(\text{age})$. Write the line with the variable names, and keep the hat on the response.

2. The slope from r

$$\bar{x} = 6.21, \quad s_x = 2.75, \qquad \bar{y} = 14.25, \quad s_y = 4.16, \qquad r = -0.924$$

$$b = r\,\frac{s_y}{s_x} = -0.924 \cdot \frac{4.16}{2.75} \approx -1.40, \qquad a = \bar{y} - b\,\bar{x} = 14.25 - (-1.397)(6.21) \approx 22.93.$$

The line always passes through $(\bar{x}, \bar{y})$. In standard-deviation units the slope is r: one SD more in x moves $\hat{y}$ by r SDs. A car one SD older than average is predicted to be 0.924 SD (not a full SD) below the average price: predictions are pulled toward the mean of y.

3. Slope and intercept in context

- **Slope**: for each additional year of age, the *predicted* price decreases by about \$1,400. Say *predicted*, give the direction, use units.
- **Intercept**: a car 0 years old has a predicted price of about \$22,930. No car in the data is younger than 2 years, so this is an extrapolation: interpret with caution (an intercept may have no meaning in context at all).

4. Reading computer output

Predictor	Coef	SE Coef	T	P
Constant	**22.934**	1.125	20.38	0.000
Age	**−1.397**	0.167	−8.39	0.000

S = 1.652 R-Sq = 85.4%

- **Coef** column: the **Constant** row is the intercept a; the row named after the explanatory variable is the slope b.
- $\widehat{\text{price}} = 22.934 - 1.397\,(\text{age})$.
- R-Sq is r^2. $r = \pm\sqrt{r^2}$ with the **sign of the slope**: $r = -\sqrt{0.854} \approx -0.924$.

sad-academy.com

 © 2026 S&D Academy

5. r^2 and predictions

Coefficient of determination

$r^2 = 0.854$: about 85% of the variation in price is accounted for by the linear model with age. Not "85% of the points are on the line", not "85% accurate".

Prediction: an 8-year-old car, $\hat{y} = 22.93 - 1.40(8) \approx 11.75$, about \$11,750.

Extrapolation: a 25-year-old car gives $\hat{y} \approx -12.0$, a negative price. The data cover 2 to 11 years only, and nothing says the pattern continues: predict only inside the range of x.

6. Working backwards, and residuals in context

From $b = r\frac{s_y}{s_x}$, $r = b\frac{s_x}{s_y}$: with $b = 1.2$, $s_x = 3$, $s_y = 6$, $r = 0.6$. Doubling r multiplies r^2 by 4 ($0.3 \to 0.6$ gives $0.09 \to 0.36$).

A residual is $y - \hat{y}$, interpreted in context. A 5-year-old car that sold for 16.6 (thousand dollars), where the line predicts about 15.95, has a residual of about +0.65: it sold for about \$650 more than the line predicts for a 5-year-old car. The other 5-year-old car, at 13.5, has a residual of about −2.45: about \$2,450 less than predicted.

Where students lose points

1. No "predicted", no hat, no context in the equation or the interpretation.
2. Slope and intercept swapped in computer output (the Constant row is the intercept).
3. r^2 read as "the percent of points on the line".
4. $r = +\sqrt{r^2}$ when the slope is negative.
5. Predicting far outside the range of the data.

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.

 © 2026 S&D Academy