

AP Statistics - Inference for the Slope of a Regression Line - Lesson

Common Core AP STAT 9.1, AP STAT 9.2, AP STAT 9.3, AP STAT 9.4, AP STAT 9.5, AP STAT 9.6, HSS-ID.B.6 · maths-pdf.fr/en/

Inference about the slope of a regression line estimates or tests the true slope β . A t interval uses $b \pm t^* SE_b$ with $df = n - 2$, and the test of $H_0: \beta = 0$ uses $t = b/SE_b$. Check the LINER conditions. Computer output gives b , SE_b , t , and the two-sided p -value.

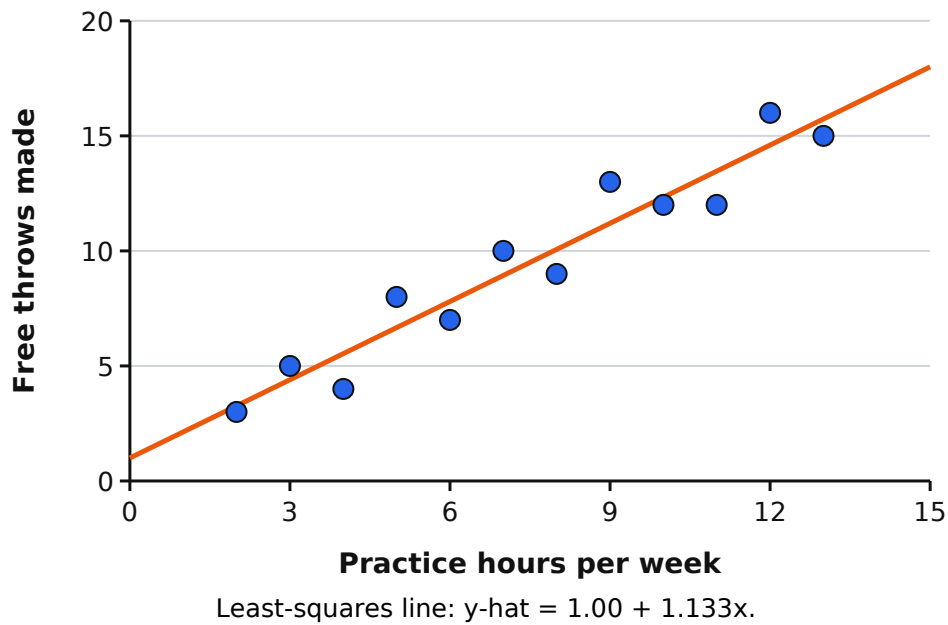
AP Statistics · Common Core State Standards: **AP STAT 9.1, AP STAT 9.2, AP STAT 9.3, AP STAT 9.4, AP STAT 9.5, AP STAT 9.6, HSS-ID.B.6**

Key facts

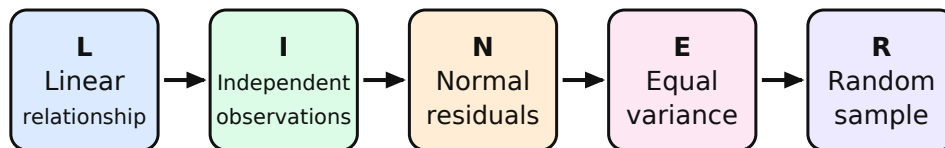
- Parameter: true slope β ; statistic: sample slope b ; $df = n - 2$.
- Interval: $b \pm t^* SE_b$; test: $t = (b - 0)/SE_b$ for $H_0: \beta = 0$.
- LINER: Linear, Independent, Normal residuals, Equal variance, Random.
- Computer output gives b , SE_b , t , and the two-sided P in the row for the explanatory variable.
- If the interval contains 0, there is no convincing evidence of a linear relationship.

What does inference about a slope tell us?

The least-squares line from a sample, $\hat{y} = a + bx$, estimates the true population line $\mu_y = \alpha + \beta x$. The sample slope b varies from sample to sample, so we use it to make inferences about the true slope β : how much the mean response changes per one-unit increase in x .

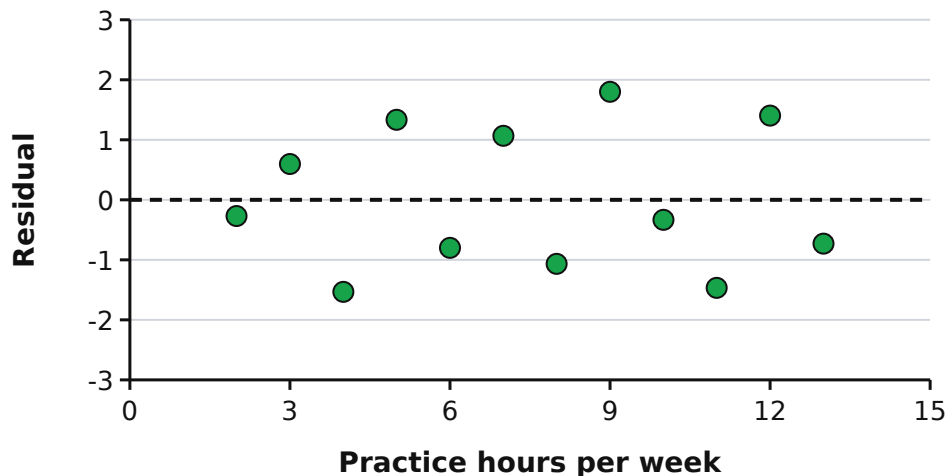


What conditions are needed?



The LINER conditions for inference about a slope.

- **Linear:** the scatterplot is roughly linear and the residual plot has no pattern.
- **Independent:** observations are independent (10% condition when sampling).
- **Normal:** residuals are approximately normal (check a histogram or dotplot of the residuals).
- **Equal variance:** the spread of the residuals is about the same for all x .
- **Random:** data come from a random sample or randomized experiment.



No pattern in the residuals and roughly constant spread: a linear model is appropriate.

How do you build a t interval for the slope?

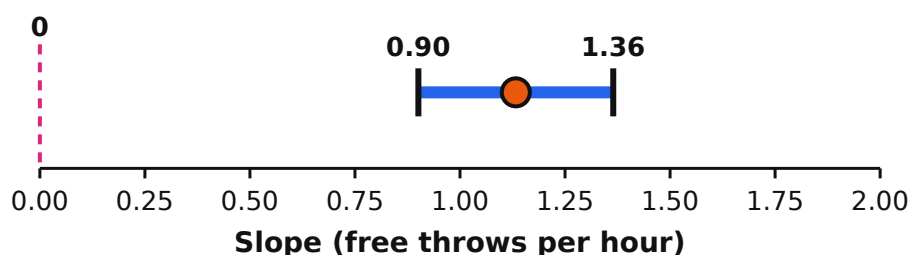
The statistic b has standard error $SE_b = \frac{s}{s_x \sqrt{n-1}}$, which computer output reports. The interval uses $df = n-2$:

$$b \pm t^* SE_b$$

Predictor	Coef	SE Coef	T	P
Constant	1.003	0.859	1.17	0.270
Hours	1.133	0.104	10.89	0.0000
S = 1.244	R-Sq = 92.2%			

Regression output: the Hours row gives the slope, its standard error, t, and the two-sided P.

From the output, $b = 1.133$ and $SE_b = 0.104$. With $df = 12 - 2 = 10$, $t^* = 2.228$, so $ME = 2.228(0.104) = 0.232$ and the 95% interval is (0.901, 1.365).



95% interval for the true slope: (0.901, 1.365); 0 is not included.

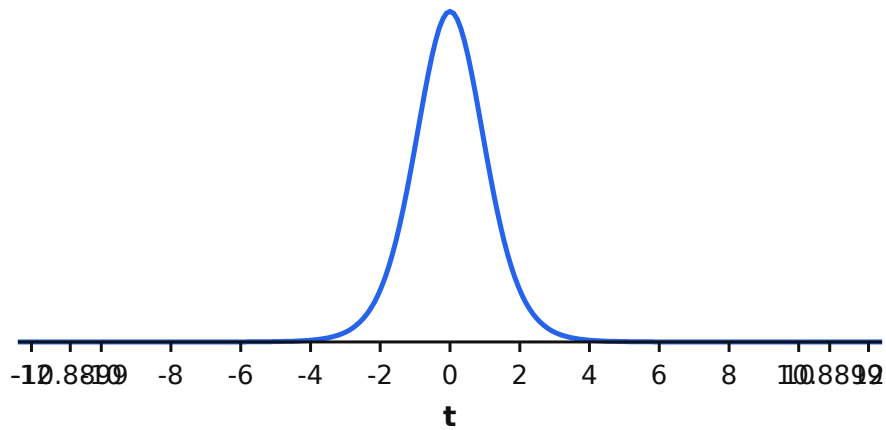
We are 95% confident that the true slope is between 0.901 and 1.365: each extra hour of weekly practice is associated with an increase of between 0.90 and 1.36 in the mean number of free throws made.

How do you test whether there is a linear relationship?

The usual test is $H_0: \beta = 0$ against $H_a: \beta \neq 0$ (no linear relationship versus a linear relationship). The test statistic is:

$$t = \frac{b-0}{SE_b}, \quad df = n-2$$

Here $t = \frac{1.133}{0.104} = 10.89$, and the two-sided p-value in the output is 0.0000. Because this is far below 0.05, we reject H_0 : there is convincing evidence of a linear relationship between weekly practice hours and free throws made. The output's P value is already two-sided; halve it for a one-sided alternative in the correct direction.



Two-sided p-value for the slope: 0.0000.

Questions about inference for the slope of a regression line

What does the true slope beta represent?

It is the change in the mean response for each one-unit increase in the explanatory variable in the whole population. The sample slope b only estimates it.

Why are the degrees of freedom $n - 2$?

Two parameters, the slope and the intercept, are estimated from the data before estimating the variation about the line. That leaves $n - 2$ degrees of freedom.

How do I read the P value in computer output?

The P value in the explanatory variable's row is two-sided for $H_0: \beta = 0$. For a one-sided alternative in the direction of the data, divide it by 2.

What if the residual plot shows a curved pattern?

Then the Linear condition fails and the inference is not valid. Try a transformation of the variables or a different model before doing a t procedure.

What does a confidence interval for beta that includes 0 mean?

It means a slope of 0 is plausible, so there is not convincing evidence of a linear relationship between the variables.