Inference for Quantitative Data: Slopes

AP Statistics

Do Those Points Align?

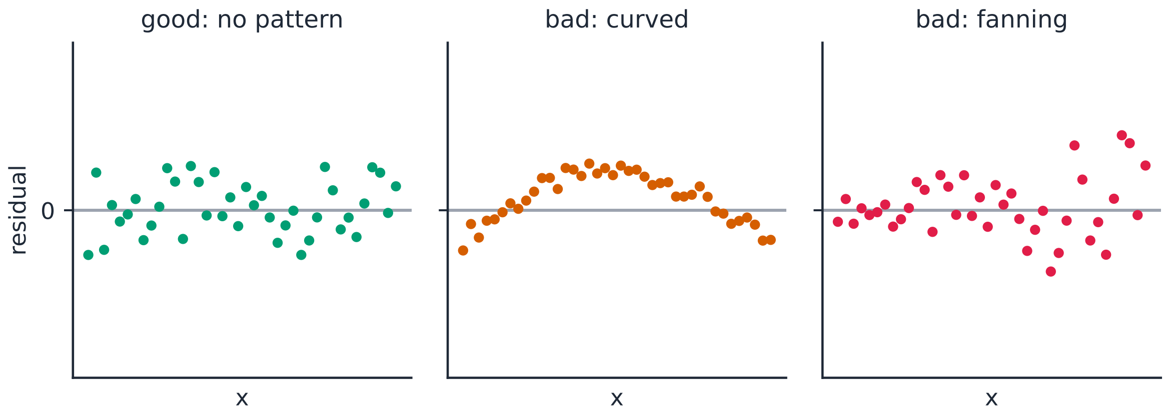
A sample **scatterplot** 散点图 gives a **sample slope** 样本斜率 b for the least-squares **regression** 回归 line –but a different sample would give a slightly different slope. So b is a statistic with **sampling variability** 抽样变异性, estimating the **true (population) slope** 总体斜率 β. This unit does **inference** 推断 for β: is there a real **linear** 线性 relationship, and how strong is it?

Confidence Interval for a Slope

A t interval for the true slope β:

$$b \pm t^* \, SE_b, \qquad df = n - 2,$$

where b is the sample slope and SE_b its **standard error** (read from computer output). Conditions (**LINER**): the true relationship is **Linear**, observations **Independent**, residuals **Normal**, and residuals have **Equal** spread (check the residual plot and a histogram of residuals), from **Random** data. Interpret the interval for β in context, with units of y per unit of x.



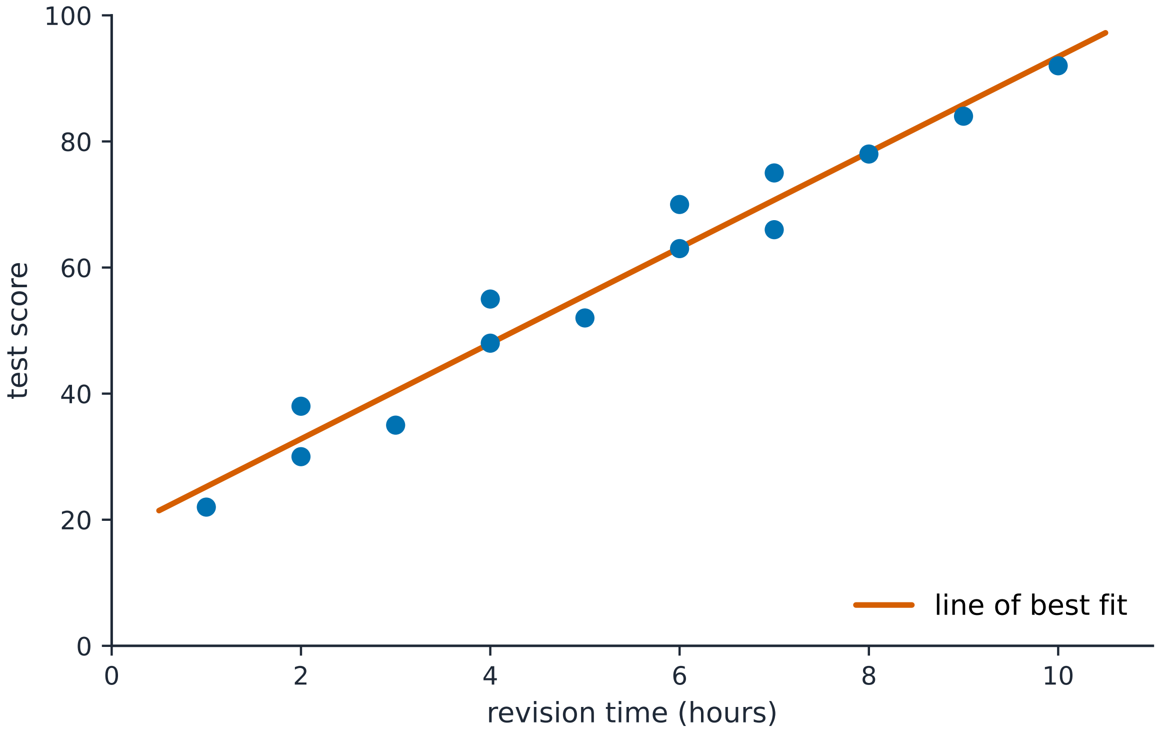
A random, patternless residual plot supports the conditions; a curve or a fan does not

The **residual plot** 残差图 is where you check Linear and Equal-spread: you want a formless cloud around zero. A **curve** means the relationship is not linear; a **fan** (spread growing with x) means the residuals do not have equal spread –both break a condition.

Worked example. Regression output gives slope $b = 2.5$ with $SE_b = 0.8$ from $n = 20$ points. For a 95% interval, $df = 18$ gives $t^* = 2.101$:

$$2.5 \pm 2.101(0.8) = 2.5 \pm 1.68 = (0.82,\ 4.18).$$

Because 0 is **not** in the interval, there is evidence of a positive linear relationship.



Slope inference is based on the least-squares regression line through the points

Justifying a Claim About a Slope

If the confidence interval for β contains 0, a slope of zero is plausible –no evidence of a linear relationship. If the interval is entirely positive or negative, there is evidence of a real (positive or negative) linear relationship. State the direction in context.

Setting Up a Test for a Slope

The usual test asks whether there is any linear relationship:

$$H_0 : \beta = 0 \quad \text{(no linear relationship)} \qquad H_a : \beta \neq 0 \text{ (or } <, >).$$

Check the LINER conditions. This is a t-test on the slope.

Carrying Out a Test for a Slope

The slope t statistic:

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

Both b and SE_b come straight from the regression output. Find the p-value from the t-distribution, compare to α, and conclude in context –evidence (or not) of a linear relationship between the two variables.

Worked example. For the same output ($b = 2.5$, $SE_b = 0.8$, $n = 20$), test $H_0 : \beta = 0$:

$$t = \frac{2.5 - 0}{0.8} = 3.13, \qquad df = 18,$$

a small p-value (< 0.01), so **reject** H_0 –convincing evidence of a linear relationship. This matches the interval, which excluded 0.

Selecting the Right Procedure

Across all of inference, identify: **what** is estimated or claimed (a proportion, a mean, a difference, a distribution of counts, or a slope), **how many** samples, and **which** design (independent or paired; sample or experiment). Then name the procedure, verify its conditions, carry it out, and communicate the conclusion with the statistic, the p-value or interval, and a plain-language answer in context. This selecting-and-communicating skill is what the investigative-task question rewards most.

Exam tips

- Inference for a **slope** tests whether the true slope is 0 (no linear relationship).
- If a slope's confidence interval **includes 0**, you cannot conclude a real relationship.
- Read the slope, standard error, t-statistic, and p-value straight from computer output.
- Check the regression conditions (linearity, independence, roughly normal residuals, equal spread) via the residual plot.
- Interpret the interval and test in context, tied to the true slope.