Researchers are examining the relationship between hours of sleep and athletic performance among college athletes. Athletic performance will be measured on a numeric scale, with greater numbers indicating better performance. The researchers expect that the more hours the athletes sleep, the better they will perform. Assuming all conditions for inference are met, the researchers will create a $95$ percent confidence interval for the slope of the regression line for predicting athletic performance from amount of sleep.

For which of the following would the confidence interval support the researchers' expectations?

(A) The confidence interval includes only positive values. (A)

(B) The confidence interval includes only negative values. (B)

(C) The confidence interval has a width less than $1$. (C)

(D) The confidence interval has a width greater than $1$. (D)

(E) The confidence interval includes the value $0$. (E)

A

| | |
|---|---|
| Confidence Interval | 置信区间 |
| Population Parameter | 总体参数 |
| Point Estimate | 点估计 |
| Margin of Error | 误差范围 |
| Confidence Level | 置信水平 |
| Standard Error | 标准误 |
| Critical Value | 临界值 |
| Sample Statistic | 样本统计量 |
| Precision | 精确度 |
| Variability | 变异性 |

## Confidence Intervals — Basic Review: Key Points

1. **Purpose** – A confidence interval estimates an unknown **population parameter** using sample data.
2. **General form:**

$$\text{Point Estimate} \pm \text{Margin of Error}$$

3. **Point estimate** – The sample statistic used to estimate the population parameter.
   - $\bar{x}$ estimates $\mu$
   - $\hat{p}$ estimates $p$
4. **Margin of Error (ME)** – Measures how much uncertainty there is in the estimate.

$$ME = (\text{critical value})(\text{standard error})$$

5. **Confidence level** – Common levels are **90%, 95%, and 99%**.
6. **Correct interpretation** – A 95% confidence interval means that **if we repeatedly took random samples and constructed intervals in the same way, about 95% of those intervals would contain the true population parameter**.
7. **Higher confidence → wider interval.**
   - 99% CI → wider
   - 95% CI → medium
   - 90% CI → narrower
8. **Larger sample size → smaller standard error → narrower confidence interval.**
9. **More variability → larger standard error → wider confidence interval.**
10. **Conditions matter** – Before constructing a confidence interval, check the appropriate conditions, such as **randomness, independence, and the relevant distribution/large-sample condition**.

**Question: Estimating a Population Mean**

A random sample of **36 students** is selected from a school. The students' mean daily study time is **2.8 hours**, with a sample standard deviation of **1.2 hours**.

A **95% confidence interval** for the population mean study time is constructed.

**(a)** Identify the **population parameter** being estimated.

**(b)** Identify the **point estimate**.

**(c)** Calculate the **standard error** of the sample mean.

**(d)** Using $t^* = 2.03$, calculate the **margin of error**.

**(e)** Construct the **95% confidence interval**.

**(f)** Interpret the confidence interval in context.

**(a)** $\mu$ = true mean daily study time for all students.

**(b)** $\bar{x} = 2.8$ hours.

**(c)**

$$SE = \frac{s}{\sqrt{n}} = \frac{1.2}{\sqrt{36}} = 0.20$$

**(d)**

$$ME = t^*(SE) = 2.03(0.20) = 0.406$$

**(e)**

$$2.8 \pm 0.406$$

$$\boxed{(2.394,\ 3.206)}$$

**(f)** We are **95% confident that the true mean daily study time for all students is between 2.394 and 3.206 hours.**

1. **Purpose:** A confidence interval for the slope ($\beta_1$) estimates the range of plausible values for the true relationship between $x$ and $y$.
2. **Check conditions:** Linearity, nearly normal residuals, constant variability (homoscedasticity), and independent observations.
3. **Calculate the standard error:** $SE_{b_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}}$, where $s$ is the standard deviation of residuals.
4. **Construct the interval:** $b_1 \pm t^* \cdot SE_{b_1}$, where $b_1$ is the sample slope and $t^*$ comes from $t$-distribution with $n - 2$ degrees of freedom.
5. **Interpret in context:** We are (e.g., 95%) confident that the true slope falls within this interval, meaning the estimated effect of $x$ on $y$ is likely within this range.

- Scatterplot — 散点图
- Association / relationship — 关联 / 关系
- Positive association — 正相关
- Negative association — 负相关
- Outlier — 离群点

A researcher is studying the relationship between hours of exercise per week (x) and resting heart rate (y). A random sample of 25 adults is taken, and the least-squares regression line is calculated:

$$\hat{y} = 75 - 0.8x$$

The **standard error of the slope** is $SE_{b_1} = 0.3$.

**a.** Construct a **95% confidence interval** for the **population slope ($\beta_1$)**.

**b.** Interpret the interval in context.

- df = 23, $t^* \approx 2.069$
- ME ≈ 0.621
- CI = -0.8 ± 0.621 → (-1.42, -0.18)
- **Interpretation:** Each extra hour of exercise **lowers resting heart rate by 0.18–1.42 bpm**

## Q6. Harder Calculation (Inference Insight)

A dataset produces:

- $b = -1.8$
- $SE_b = 0.5$
- $n = 20$

Without using a table, determine:

- Whether 0 is likely in the 95% CI
- Whether the slope is significantly different from 0

*(Use $t^* \approx 2$ for estimation)*

---

## Q7. Stretch (Deep Understanding)

Two models are built:

| Model | Slope | SE of slope |
|---|---|---|
| A | 3.0 | 0.2 |
| B | 3.0 | 1.0 |

Which model gives a **narrower confidence interval** and why?

What does this imply about the data?

**Q6:**

Approx CI:

$$-1.8 \pm 2(0.5) = -1.8 \pm 1.0$$

$$(-2.8,\ -0.8)$$

- 0 is **not in interval**
- Slope is **significantly negative**

---

**Q7:**

- Model A has **narrower CI** (smaller SE)
- Implies:
  - Less variability
  - Stronger linear relationship
  - More precise estimate

# Inference for Linear Regression

1. **Slope (β) — 斜率**
2. **Least Squares Regression Line — 最小二乘回归直线**
3. **Residual — 残差**
4. **Standard Error of the Slope — 斜率的标准误**
5. **Confidence Interval (for slope) — 斜率的置信区间**

1. **Slope $\beta$** – True population slope; measures linear relationship.
2. **Sampling variability** – Sample slope $b$ varies; measured by $SE_b$.
3. **t-inference** – Use t-test ($t = b/SE_b, df = n - 2$) since $\sigma$ unknown.
4. **CI & hypothesis test link** – 0 in CI → not significant; 0 outside CI → significant.
5. **Conditions (LINER)** – Linearity, Independence, Normality, Equal variance.

A researcher studies the relationship between the number of hours students study per week ($x$) and their exam scores ($y$). A sample of 15 students produces the least-squares regression line:

$$\hat{y} = 50 + 5x$$

The standard error of the slope is $SE_b = 1.8$.

**(a) Construct a 95% confidence interval for the slope of the population regression line.**

**(b) Interpret the confidence interval in context.**

**(c) Using the confidence interval, test at the 5% significance level whether there is a linear relationship between study hours and exam scores. State your conclusion.**

**(d) Explain one condition that must be satisfied for this inference to be valid.**

**(a) 95% CI for slope:**

$$5 \pm 2.160(1.8) \approx (1.11, 8.89)$$

**(b) Interpretation:**

Each extra study hour increases exam score by **1.11–8.89 points** on average.

**(c) Hypothesis test:**

0 not in CI → **reject** $H_0$ → significant positive linear relationship.

**(d) Condition:**

Linearity (relationship between x and y is approximately linear).

### Derivation

The standard error of the slope can be derived from the variability of the residuals and the variability in the explanatory variable $x$.

1. The slope of the regression line is

$$b_1 = r\frac{s_y}{s_x}$$

2. The standard error of the regression slope is related to the residual standard deviation:

$$SE_{b_1} = \frac{s}{\sqrt{\square}}$$

3. Recall that

$$SS_x = \sum (x_i - \bar{x})^2$$

Therefore,

$$\sqrt{SS_x} = \sqrt{\boxed{\sum (x_i - \bar{x})^2}}$$

4. Hence,

$$\boxed{SE_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}}$$