A researcher wants to determine whether there is a significant linear relationship between the number of hours studied (x) and exam score (y).

Which of the following is the correct set of hypotheses for testing the slope of the population regression line?

A.

$$H_0 : \beta_0 = 0$$

$$H_a : \beta_0 \neq 0$$

B.

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

C.

$$H_0 : b_1 = 0$$

$$H_a : b_1 \neq 0$$

D.

$$H_0 : \mu = 0$$

$$H_a : \mu \neq 0$$

Regression Line	回归线
Slope (β_1)	总体斜率
Sample Slope (b_1)	样本斜率
Null Hypothesis (H_0)	原假设
Alternative Hypothesis (H_a)	备择假设
Test Statistic	检验统计量
p-value	p值
Significance Level (α)	显著性水平
Linear Relationship	线性关系
Statistical Significance	统计显著性

1. **State hypotheses:**
   - $H_0 : \beta_1 = 0$ (no linear relationship between x and y)
   - $H_a : \beta_1 \neq 0$ (or one-sided if direction is specified)
2. **Check conditions:** Linearity, nearly normal residuals, constant variability (homoscedasticity), independent observations.
3. **Calculate test statistic:**

$$t = \frac{b_1 - 0}{SE_{b_1}}$$

   where b_1 is the sample slope and SE_{b_1} is its standard error.
4. **Determine degrees of freedom:** $df = n - 2$ and find the p-value from the t-distribution.
5. **Make a conclusion:** Compare p-value to α and interpret whether there is significant evidence of a linear relationship between x and y.

A researcher wants to determine whether there is a significant linear relationship between the number of hours studied (x) and exam score (y).

Which of the following is the correct set of hypotheses for testing the slope of the population regression line?

A.

$$H_0 : \beta_0 = 0$$

$$H_a : \beta_0 \neq 0$$

B.

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

C.

$$H_0 : b_1 = 0$$

$$H_a : b_1 \neq 0$$

D.

$$H_0 : \mu = 0$$

$$H_a : \mu \neq 0$$

A teacher wants to investigate whether there is a linear relationship between the number of hours students study (x) and their exam scores (y).

A random sample of 20 students is selected. A least-squares regression analysis produces the following results:

- Sample slope: $b_1 = 3.2$
- Standard error of the slope: $SE_{b_1} = 1.1$

The teacher wishes to test whether the population slope differs from zero at the 5% significance level.

Questions

(a) State the null and alternative hypotheses. (1 mark)

(b) Calculate the test statistic. (2 marks)

(c) The p-value is 0.009. Interpret this p-value in context. (1 mark)

(d) State your conclusion in the context of the problem. (2 marks)

(a)

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

(b)

$$t = \frac{b_1 - 0}{SE_{b_1}}$$

$$t = \frac{3.2}{1.1} = 2.91$$

(c)

If the true slope is 0, there is a 0.9% chance of obtaining a sample slope this extreme or more extreme.

(d)

Since

$$p = 0.009 < 0.05,$$

reject H_0.

There is convincing evidence of a linear relationship between study hours and exam scores in the population.

Stretch and Challenge – Hypothesis Test for the Slope of a Regression Model

A sports scientist investigates whether weekly training hours (x) can predict race times (y) for runners.

A random sample of 15 runners produces the following regression output:

- Sample slope: $b_1 = -0.80$
- Standard error of the slope: $SE_{b_1} = 0.25$

The scientist wishes to test:

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 < 0$$

at the $\alpha = 0.05$ significance level.

Questions

(a) Explain why a one-sided alternative hypothesis is appropriate. (1 mark)

(b) Calculate the test statistic. (2 marks)

(c) The p-value is 0.003. State the conclusion in context. (2 marks)

(d) Suppose the sample size were increased from 15 to 60 while the observed slope remained approximately the same. Explain how this would affect the power of the test. (1 mark)

(e) A student concludes that increasing training hours causes race times to decrease. Explain why this conclusion may not be justified. (2 marks)

(a)

More training is expected to reduce race times, so a negative slope is anticipated.

(b)

$$t = \frac{-0.80}{0.25} = -3.20$$

(c)

Since

$$0.003 < 0.05,$$

reject H_0.

There is convincing evidence that increased training hours are associated with lower race times.

(d)

Increasing the sample size decreases the standard error and increases the power of the test.

(e)

The data are observational, not from a randomized experiment. Confounding variables (fitness, age, diet, genetics, etc.) may explain the relationship, so causation cannot be established.

Confidence Intervals for Slopes

Population Slope β_1	总体斜率
Sample Slope b_1	样本斜率
Confidence Interval	置信区间
Standard Error	标准误
Critical Value	临界值
t-Distribution	t分布
Degrees of Freedom	自由度
Margin of Error	误差范围
Confidence Level	置信水平
Linear Association	线性关联

1. **Parameter:** The population slope is β_1.
2. **Statistic:** The sample slope is b_1.
3. **Purpose:** Estimate the true population slope using sample data.
4. **Formula:**

$$b_1 \pm t^* SE_{b_1}$$

5. **Degrees of freedom:**

$$df = n - 2$$

6. **Critical value:** Use the **t-distribution**, not the normal distribution.
7. **Interpretation:** We are confident that the interval contains the **true population slope** β_1.
8. **Zero:** If the interval **contains 0**, a linear relationship may not be statistically significant.
9. **Significance:** If the interval **does not contain 0**, there is evidence of a linear association.
10. **Context:** Interpret the slope in terms of the change in the response variable for each **1-unit increase in the explanatory variable**.

A researcher investigates whether the number of hours students sleep each night (x) is related to their reaction time (y).

A random sample of **25 students** is selected. A least-squares regression model gives:

- Sample slope: $b_1 = -4.2$
- Standard error of the slope: $SE_{b_1} = 1.5$

The researcher wants a **95% confidence interval** for the population slope β_1. The appropriate critical value is $t^* = 2.069$.

(a) Calculate the 95% confidence interval for β_1. **(2 marks)**

(b) Interpret the confidence interval in context. **(2 marks)**

(c) Does the interval provide convincing evidence of a linear association between sleep and reaction time? Justify. **(1 mark)**

(d) Interpret the slope $b_1 = -4.2$ in context. **(1 mark)**

(a)

$$-4.2 \pm 2.069(1.5)$$

$$\boxed{(-7.30, -1.10)}$$

(b) We are **95% confident** that the true mean change in reaction time is between **–7.30 and –1.10 units** for each additional hour of sleep.

(c) Yes. The interval does not contain 0, providing evidence of a linear association.

(d) For each additional hour of sleep, predicted reaction time decreases by about **4.2 units**, on average.

TRUE or FALSE

1. T/F

The parameter being estimated is the population slope β_1.

TRUE or FALSE

2. T/F

The sample slope b_1 is a statistic used to estimate β_1.

TRUE or FALSE

---

3. T/F

A confidence interval for a regression slope uses a t-distribution.