Week 29

AP Statistics

Homework bundle for this week
本周作业合集

exercises.lol

Contents 目录

Topic 9 quiz 2
Exercise sheet 9.1 5
Exercise sheet 9.2 8
Past-paper questions 9.2 11
Exercise sheet 9.3 15
Exercise sheet 9.4 18
Exercise sheet 9.5 21
Past-paper questions 9.5 25
Exercise sheet 9.6 29

AP Statistics Name: ____________ Score: ______

1. The slope b from a sample is...
   - ☐ an estimate of the true slope β
   - ☐ always exactly β
   - ☐ a fixed constant
   - ☐ the correlation
2. For slope inference with n = 20, what are the degrees of freedom (n − 2)?

   ____
3. A 95% interval for a slope is (2.1, 6.3). You can conclude...
   - ☐ convincing evidence of a positive linear relationship
   - ☐ no relationship
   - ☐ the slope is 0
   - ☐ nothing
4. The usual null hypothesis for a slope test is...
   - ☐ H0: $\beta = 0$
   - ☐ H0: b = 0
   - ☐ Ha: $\beta \neq 0$
   - ☐ H0: r = 1
5. b = 4.2, SE_b = 1.0. Compute the slope test statistic t = b/SE_b.

   ____
6. A question about a linear relationship between two quantitative variables calls for...
   - ☐ slope t-inference
   - ☐ a two-proportion z-test
   - ☐ a chi-square test
   - ☐ a one-mean t-test
7. The sampling distribution of the sample slope b is centered at the true slope β.

   ☐ True ☐ False

8. $b = 4.2$, SE_b = 1.0, $t^* = 2.10$. Find the margin of error $t^* \cdot$ SE_b.

____

9. A 95% interval for a slope is $(-1.0, 3.4)$. You can conclude...

- ☐ no convincing evidence of a linear relationship
- ☐ a strong positive relationship
- ☐ a strong negative relationship
- ☐ the slope is exactly 0

10. A two-sided output p-value is 0.001. For a one-sided Ha (with b in the claimed direction), what is the p-value?

____

9. Inference for Quantitative Data: Slopes

AP Statistics

1. **an estimate of the true slope β**

   b estimates the unknown population slope β, with uncertainty.

2. **18**

   df = n − 2 = 18 (slope and intercept estimated).

3. **convincing evidence of a positive linear relationship**

   The interval is entirely positive (above 0).

4. **H0: $\beta = 0$**

   H0 is about the population slope β being 0.

5. **4.2**

   4.2 / 1.0 = 4.2.

6. **slope t-inference**

   Two quantitative variables + a trend → slope inference.

7. **True**

   b is unbiased for β.

8. **2.1**

   $2.10 \times 1.0 = 2.10$.

9. **no convincing evidence of a linear relationship**

   The interval contains 0, so a flat slope is plausible.

10. **0.0005**

    Halve the two-sided value: 0.001 / 2 = 0.0005.

AP STATISTICS • EXERCISE SHEET

9.1 Introducing Statistics: Do Those Points Align?

Name: ______________ Class: __________ Date: ____________ **Total: 8 marks**

KEY WORDS linear 线性 • sample slope 样本斜率 • inference 推断 • regression 回归 • true (population) slope 总体斜率

Objective

Build the skills to answer exam questions on **inference for the slope of a regression model**.

You must be able to:

- explain that inference for the slope tests whether a **linear relationship** exists
- state that a slope of 0 means **no linear relationship**
- identify the sample slope b as an estimate of the population slope β

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ The question

Inference for regression asks whether the **true (population) slope** β is different from 0 — i.e. whether there is a real **linear relationship** between x and y.

■ Slope of zero

If $\beta = 0$, changes in x predict **no** change in y — no linear relationship. The **sample slope** b (from the least-squares line) is our estimate of β.

2 Practice

2.1 State what inference for the slope tests. [1]

2.2 State what a slope of 0 means for the relationship. [1]

2.3 State the parameter estimated by the sample slope b. [1]

3 Exam-style questions

3.1 Inference for regression tests whether the true slope is [1]

- **A** 1
- **B** 0 (no linear relationship)
- **C** the mean
- **D** negative

3.2 The sample slope b estimates the [1]

- **A** intercept
- **B** population slope β
- **C** correlation
- **D** residual

3.3 A regression of weight on height is analysed.

(a) Name the parameter of interest. [1]

(b) State what a slope of 0 would mean. [1]

(c) Name the value that estimates it. [1]

Solutions

2.1 whether there is a linear relationship (whether the true slope $\beta \neq 0$).

2.2 no linear relationship between x and y.

2.3 the population slope β.

3.1 **B**.

3.2 **B**.

3.3 (a) the population slope β. (b) no linear relationship between height and weight. (c) the sample slope b.

AP STATISTICS • EXERCISE SHEET

9.2 Confidence Intervals for the Slope of a Regression Model

Name: ____________ Class: ________ Date: __________ **Total: 9 marks**

KEY WORDS inference 推断 • sample slope 样本斜率

Objective

Build the skills to answer exam questions on **a confidence interval for the slope**.

You must be able to:

- construct the interval $b \pm t^* \cdot SE(b)$
- use degrees of freedom $n - 2$
- compute the margin of error

1 Worked examples

Study these first. Each one shows the method for a question type used later.

> **The interval for the slope**
>
> $$b \pm t^* \cdot SE(b),$$
>
> where b is the sample slope, $SE(b)$ is its standard error (from computer output), and the degrees of freedom are $n - 2$.
>
> **Example**
>
> $b = 2.5$, $SE(b) = 0.5$, $t^* = 2.05$: ME $= 2.05 \times 0.5 = 1.025$.

2 Practice

2.1 Write the confidence-interval formula for the slope. [1]

2.2 State the degrees of freedom for slope inference with $n = 30$. [1]

2.3 $b = 2.5$, $SE(b) = 0.5$, $t^* = 2.05$. Find the margin of error. [2]

3 Exam-style questions

3.1 A confidence interval for the slope uses [1]

- **A** z^*
- **B** t^* with $\text{df} = n - 2$
- **C** χ^2
- **D** no critical value

3.2 The degrees of freedom for inference about a slope is [1]

- **A** n
- **B** $n - 1$
- **C** $n - 2$
- **D** $2n$

3.3 $b = 1.8$, $SE(b) = 0.4$, $n = 25$, $t^* = 2.07$.

(a) Find the margin of error. [2]

(b) Give the confidence interval. [1]

Solutions

2.1 $b \pm t^* \cdot SE(b)$.

2.2 $n - 2 = 28$.

2.3 $\text{ME} = 2.05 \times 0.5 = 1.025$.

3.1 **B**.

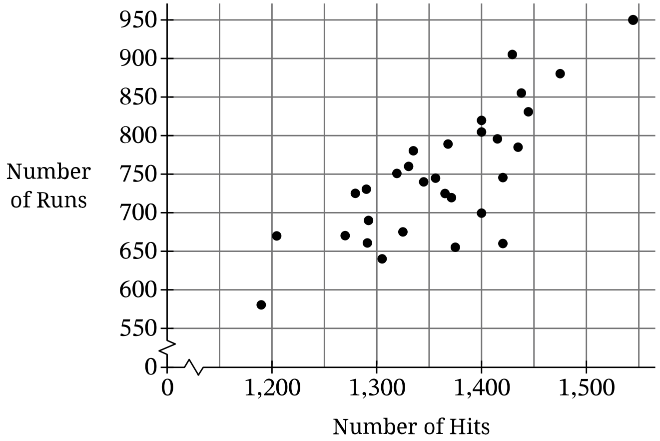
3.2 **C**.

3.3 (a) $\text{ME} = 2.07 \times 0.4 = 0.828$. (b) $(0.97,\ 2.63)$.

The following information applies to parts A and B.

6. To win a game, a baseball team needs to score runs. To score runs, players need to get on base. The primary way players get on base is by hitting the ball. Figure 1 shows the relationship between the number of hits and the number of runs for 30 randomly selected professional baseball teams.

Figure 1. Scatterplot of Number of Runs by Number of Hits



Part A

Consider the scatterplot in Figure 1.

i. Describe the relationship between the number of hits and the number of runs for these baseball teams, in context.

ii. Based on the data from the sample of 30 teams, an equation of the least-squares regression line model for predicting the number of runs from the number of hits is as follows.

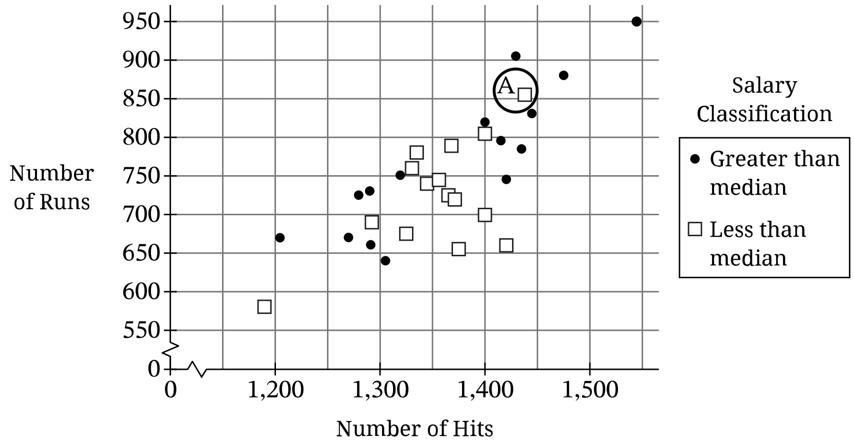
$$\text{Predicted number of runs} = -372.2 + 0.823(\text{number of hits})$$

Using the given regression equation, calculate the point estimate for the predicted number of runs a team would score if they were to achieve their goal of 1,250 hits. Show your work.

Part B

Each team has a total salary equal to the sum of the salaries of all players on the team. The median total salary for the sample of 30 teams is $160 million. In Figure 2, points with dots represent teams with a total salary greater than the median, and points with squares represent teams with a total salary less than the median.

Figure 2. Scatterplot of Number of Runs by Number of Hits and Salary Classification



Consider the scatterplot in Figure 2.

i. Compare the team represented by the point that is circled and labeled A with the other teams that have the same total salary classification.

ii. For each salary classification, consider the linear relationship between the number of hits and the number of runs. For teams with a total salary greater than the median, is the strength of the linear relationship stronger than, weaker than, or similar to the strength of the linear relationship for teams with a total salary less than the median? Explain.

Part C

Different types of intervals can be used when working with linear regression models. One type of interval is a confidence interval for the slope of the least-squares regression line. Two other types of intervals in the context of this problem are as follows:

- A confidence interval that estimates the mean number of runs for all teams with a specific number of hits.
- A prediction interval that predicts the number of runs for a single team with a specific number of hits.

All three intervals use the same basic formula:

$$\text{Point Estimate} \pm t^{*}(\text{standard error}),$$

where critical value t^* comes from a t-distribution that has $n - 2$ degrees of freedom. The same critical value is used when finding each of these intervals with the same level of confidence.

Consider the point estimate for the predicted number of runs found in part A (ii) for the team whose goal is to achieve 1,250 hits next year. Recall that this value was calculated using the linear regression model based on the data from the sample of 30 teams.

i. What is the critical value for a 95% confidence level that would be used for the confidence interval for the mean number of runs as well as for the prediction interval for the number of runs? Indicate the answer to two decimal places.

ii. The standard error for the confidence interval for the mean number of runs is 17.48. Assuming the conditions for inference are met, calculate the 95% confidence interval for the mean number of runs for all teams with 1,250 hits. Show your work.

iii. The standard error for the prediction interval for the number of runs is 56.78. Assuming the conditions for inference are met, calculate the 95% prediction interval for the number of runs for a single team with 1,250 hits. Show your work.

Part D

i. Would a distribution of sample means be expected to have more variability or less variability than a distribution of individual observations? Explain.

ii. The standard error used in the confidence interval that estimates the mean number of runs for all teams with x hits can be found using the following formula.

$$\sqrt{s^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)}$$

The standard error used in the prediction interval that predicts the number of runs for a single team with x hits can be found using the following formula.

$$\sqrt{s^2 + s^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)}$$

In both standard error formulas, s is the standard deviation of the residuals, n is the sample size, and $\bar{x}$ is the sample mean number of hits. Based on the answer from part D (i) and the standard error formulas for the confidence interval and the prediction interval, explain why the prediction interval calculated in part C (iii) is wider than the confidence interval calculated in part C (ii).

STOP

END OF EXAM

AP STATISTICS • EXERCISE SHEET

9.3 Justifying a Claim About the Slope of a Regression Model Based on a Confidence Interval

Name: ____________ Class: ________ Date: __________ **Total: 8 marks**

KEY WORDS linear 线性 • regression 回归

Objective

Build the skills to answer exam questions on **justifying a claim about the slope from a confidence interval**.

You must be able to:

- use a CI for the slope to judge whether a **linear relationship** exists
- recognise that an interval containing 0 means **no significant linear relationship**

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Using the slope interval

The confidence interval gives the plausible values of the true slope β:

- if it **contains** 0, there is **no significant linear relationship** (β could be 0);
- if it is entirely **positive** or **negative**, there **is** a significant linear relationship.

■ Reading a worked interval

Regression output gives the slope's 95% interval as $(0.4,\ 1.6)$. It is entirely **positive** and excludes 0, so there **is** a significant positive linear relationship. An interval of $(-0.3,\ 1.1)$ would contain 0 – no significant linear relationship.

2 Practice

2.1 State how a CI for the slope judges "no linear relationship". [1]

2.2 A 95% CI for the slope is $(0.5, 1.5)$. State whether there is evidence of a linear relationship. [1]

2.3 State what a slope interval containing 0 implies. [1]

3 Exam-style questions

3.1 If a confidence interval for the slope contains 0, there is [1]

- **A** a strong relationship
- **B** no significant linear relationship
- **C** an error
- **D** a curved pattern

3.2 A 95% CI for the slope is $(1.2, 2.8)$. This suggests [1]

- **A** no relationship
- **B** a positive linear relationship
- **C** a negative linear relationship
- **D** a curve

3.3 A 95% CI for the slope is $(-0.2, 0.6)$.

(a) State whether 0 is in the interval. [1]

(b) State whether there is a significant linear relationship. [1]

(c) State the plausible conclusion. [1]

Solutions

2.1 if the interval contains 0, there is no significant linear relationship.

2.2 yes — the interval is entirely positive, so there is a positive linear relationship.

2.3 the true slope could be 0, so no significant linear relationship.

3.1 **B**.

3.2 **B**.

3.3 (a) yes. (b) no. (c) there is no significant linear relationship.

9.4 Setting Up a Test for the Slope of a Regression Model

Name: ____________ Class: ________ Date: __________ **Total: 10 marks**

KEY WORDS sample slope 样本斜率 • linear 线性 • regression 回归

Objective

Build the skills to answer exam questions on **setting up a test for the slope**.

You must be able to:

- state the hypotheses $H_0 : \beta = 0$ and H_a
- write the test statistic $t = \dfrac{b}{SE(b)}$
- state the degrees of freedom $n - 2$

1 Worked examples

Study these first. Each one shows the method for a question type used later.

Hypotheses

- **Null** $H_0 : \beta = 0$ (no linear relationship).
- **Alternative** $H_a : \beta \neq 0$ (or one-sided).

Test statistic

$$t = \frac{b}{SE(b)}, \qquad \mathrm{df} = n - 2,$$

where b is the sample slope and $SE(b)$ its standard error.

Example

$b = 3.0$, $SE(b) = 0.6$: $t = \dfrac{3.0}{0.6} = 5.0$.

2 Practice

2.1 State the null hypothesis for a slope test. [1]

2.2 Write the test-statistic formula for a slope. [1]

2.3 For $b = 3.0$ and $SE(b) = 0.6$, find the test statistic t. [2]

3 Exam-style questions

3.1 The null hypothesis in a slope test is usually [1]

- **A** $\beta = 1$
- **B** $\beta = 0$
- **C** $b = 0$
- **D** $\beta \neq 0$

3.2 The test statistic for a slope is [1]

- **A** $\frac{b}{SE(b)}$
- **B** $b \cdot SE(b)$
- **C** $\frac{SE(b)}{b}$
- **D** $b - SE(b)$

3.3 A regression has $b = 2.4$, $SE(b) = 0.8$, $n = 22$.

(a) Write H_0. [1]

(b) Compute t. [2]

(c) State the degrees of freedom. [1]

Solutions

2.1 $H_0 : \beta = 0$.

2.2 $t = \frac{b}{SE(b)}$.

2.3 $t = \frac{3.0}{0.6} = 5.0$.

3.1 B.

3.2 A.

3.3 (a) $H_0 : \beta = 0$. (b) $t = \frac{2.4}{0.8} = 3.0$. (c) $n - 2 = 20$.

AP STATISTICS • EXERCISE SHEET

9.5 Carrying Out a Test for the Slope of a Regression Model

Name: __________ Class: ________ Date: ________ **Total: 9 marks**

KEY WORDS linear 线性 • regression 回归

Objective

Build the skills to answer exam questions on **carrying out a test for the slope**.

You must be able to:

- compute the slope t statistic and find the **p-value**
- compare it to α and reach a decision
- state a conclusion about the linear relationship in context

1 Worked examples

Study these first. Each one shows the method for a question type used later.

The procedure

1. State $H_0 : \beta = 0$ and H_a; check conditions.
2. Compute $t = \dfrac{b}{SE(b)}$ with $\text{df} = n - 2$.
3. Find the **p-value** from the t-distribution.
4. Compare to α: a **small** p-value $\rightarrow$ **reject** $H_0 \rightarrow$ there **is** a linear relationship.

A worked test

A regression gives slope $b = 2.5$ with $SE(b) = 1.0$ from $n = 12$ points. Then

$$t = \frac{2.5}{1.0} = 2.5, \quad \text{df} = 12 - 2 = 10,$$

a p-value $\approx 0.03 < 0.05$ – reject $H_0 : \beta = 0$, so a linear relationship exists.

2 Practice

2.1 State the steps to carry out a slope test. [2]

2.2 State what a small p-value indicates about the slope. [1]

2.3 A slope test gives a p-value of 0.001 at $\alpha = 0.05$. State the decision. [1]

3 Exam-style questions

3.1 A slope test with a small p-value suggests [1]

- **A** no relationship
- **B** a real linear relationship
- **C** bias
- **D** a curved pattern

3.2 A slope test rejects H_0 when the p-value is [1]

- **A** above α
- **B** below α
- **C** exactly zero
- **D** exactly one

3.3 A slope test gives $t = 3.0$, df $= 20$, p-value 0.007 at $\alpha = 0.05$.

(a) State the decision. [1]

(b) State whether there is a linear relationship. [1]

(c) State the conclusion in context. [1]

Solutions

2.1 state hypotheses and check conditions; compute t; find the p-value; compare to α and conclude.

2.2 the slope is significantly different from 0 — a real linear relationship.

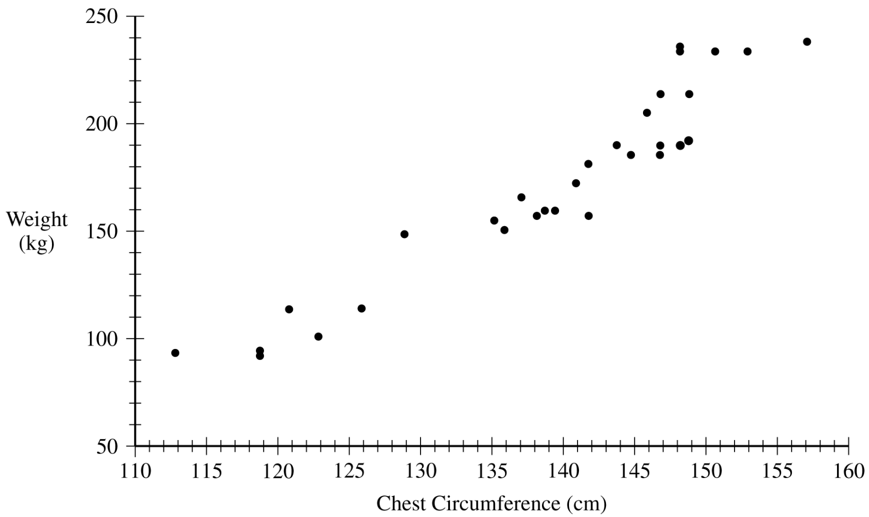
2.3 reject H_0 ($0.001 < 0.05$).

3.1 **B**.

3.2 **B**.

3.3 (a) reject H_0. (b) yes. (c) there is convincing evidence of a linear relationship between the variables.

. Wildlife biologists are interested in the health of tule elk, a species of deer found in California. An important measurement of tule elk health is their weight. The weight of a tule elk is difficult to measure in the wild. However, chest circumference, which is believed to be related to the weight of a tule elk, can easily be measured from a safe distance using a harmless laser. A study was done to investigate whether chest circumference, in centimeters (cm), could be used to accurately estimate the weight, in kilograms (kg), of male tule elk. For the study, wildlife biologists captured 30 male tule elk, measured their chest circumference and weight, and then released the elk. The data for the 30 male tule elk are shown in the scatterplot.



(a) Describe the relationship between chest circumference and weight of male tule elk in context.

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 5** on this page.

Following is the equation of the least-squares regression line relating chest circumference and weight for male tule elk.

$$\text{predicted weight} = -350.3 + 3.7455(\text{chest circumference})$$

(b) The weight of one male tule elk with a chest circumference of 145.9 cm is 204.3 kg.

(i) Using the equation of the least-squares regression line, calculate the predicted weight for this male tule elk. Show your work.

(ii) Calculate the residual for this male tule elk. Show your work.

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 5** on this page.

The equation of the least-squares regression line relating chest circumference and weight for male tule elk is repeated here.

$$\text{predicted weight} = -350.3 + 3.7455(\text{chest circumference})$$

(c) Interpret the slope of the least-squares regression line in context.

(d) The sambar, another species of deer, is similar in size to the tule elk. The slope of the population regression line relating chest circumference and weight for all male sambars is 4.5 kilograms per centimeter. A wildlife biologist wants to determine whether the slope of the population regression line for male tule elk is different than that for male sambars. Let β represent the slope of the population regression line for male tule elk. The wildlife biologist conducted a test of the following hypotheses using the sample of 30 tule elk.

$$H_0: \beta = 4.5$$
$$H_a: \beta \neq 4.5$$

The test statistic was calculated to be 3.408. Assume all conditions for inference were met.

(i) Determine the p-value of the test.

GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 5** on this page.

(ii) At a significance level of $\alpha = 0.05$, what conclusion should the wildlife biologist make regarding the slope of the population regression line for male tule elk? Justify your response.

GO ON TO THE NEXT PAGE.

AP STATISTICS • EXERCISE SHEET

9.6 Skills Focus: Selecting an Appropriate Inference Procedure

Name: ____________ Class: ________ Date: __________ **Total: 8 marks**

KEY WORDS inference 推断 • regression 回归

Objective

Build the skills to answer exam questions on **selecting an appropriate inference procedure**.

You must be able to:

- choose the right procedure across proportions, means, chi-square, and slopes
- match the procedure to the data type and the question asked

1 Worked examples

Study these first. Each one shows the method for a question type used later.

Matching procedure to question

- **One/two proportions** — a categorical variable's proportion, or comparing two.
- **One/two means (t)** — a quantitative variable's mean, or comparing two.
- **Chi-square** — goodness of fit, homogeneity, or independence for categorical counts.
- **Slope (t)** — the relationship between **two quantitative** variables.

Then decide: **estimate** (confidence interval) or **test a claim** (hypothesis test).

Worked choices

"Confidence interval for the mean height" – quantitative, one sample, estimate → **one-sample t interval**. "Test whether exam score depends on hours studied (both numeric)" – a relationship between two quantitative variables → **slope t test**.

2 Practice

2.1 State which procedure to use for a single population mean. [1]

2.2 State which procedure to use for the relationship between two quantitative variables. [1]

2.3 State how to decide between a confidence interval and a hypothesis test. [1]

3 Exam-style questions

3.1 Inference about the relationship between two quantitative variables uses a test for the [1]

- **A** proportion
- **B** mean
- **C** slope
- **D** chi-square statistic

3.2 To test whether one categorical variable matches a claimed distribution, use a [1]

- **A** t-test
- **B** chi-square goodness-of-fit test
- **C** slope test
- **D** z-interval

3.3 A study asks whether hours studied predicts exam score.

(a) Name the type of variables. [1]

(b) Name the inference procedure. [1]

(c) State the null hypothesis. [1]

Solutions

2.1 a one-sample t procedure for a mean.

2.2 a t procedure for the slope of the regression line.

2.3 use a confidence interval to estimate a parameter; use a hypothesis test to evaluate a claim.

3.1 **C**.

3.2 **B**.

3.3 (a) two quantitative variables. (b) a test for the slope of a regression model. (c) $H_0 : \beta = 0$.