5.5 Hypothesis testing

Name: _____ Class: _____ Date: _____ **Total: 13 marks**

KEY WORDS null hypothesis • p-value p

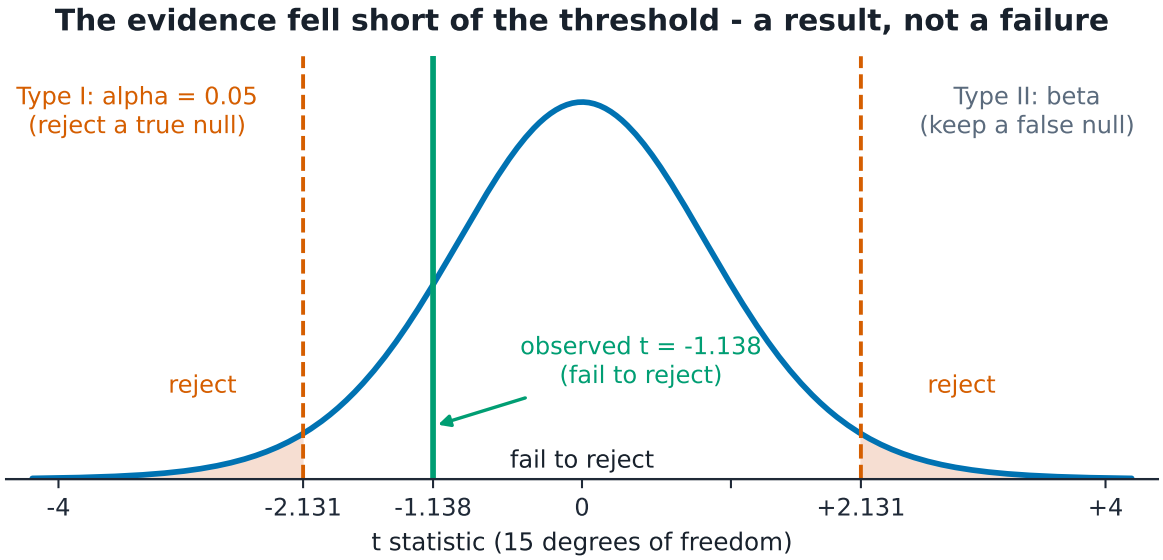
Objective

Set up tests with hypotheses stated correctly; compute t statistics and read thresholds; distinguish statistical from practical significance.

1 Key knowledge and worked examples

■ The machinery and its meaning

$t = (\bar{x} - \mu_0)/SE$; reject past the critical value chosen as your error budget α. The **p-value** p is about the data under the **null hypothesis** , never about the hypothesis's truth. Significance ≠ importance: report the effect size with its interval.



Fail to reject is a result, not a failure

■ Worked example. The 30-minute question

Seed-11 sample: $t = (28.66 - 30)/1.175 = -1.14$ against the 2.131 threshold → fail to reject ✓. The data are consistent with 30 – and with plenty else. NOT "the mean is 30".

exercises.lol exercises.com.cn

2 Practice

A rival claims their mean is under 30. A sample of 36 of their orders: mean 28.2, sd 6.0.

2.1 State H0 and H1 for testing the claim "under 30" (one-sided). [2]

2.2 Compute the t statistic (SE first). [3]

2.3 With critical value -1.69 ($\alpha = 0.05$, one-sided, 35 df), what do you conclude – in the two sentences that are licensed? [2]

2.4 The same shop runs 25 weekly tests on its dashboard; one hits p = 0.04. In two lines, why is that weaker evidence than it looks? [2]

2.5 Big-n vs small-n: match (a) n = 500, difference 0.3, p = 0.01; (b) n = 10, difference 4.0, p = 0.12 – to "statistically significant, practically trivial" and "possibly real, under-evidenced", with one line each on the decision. [4]

exercises.lol exercises.com.cn

Solutions

2.1

- H_0: the mean is 30 (or more).
- H_1: the mean is under 30.

2.2

- $SE = 6.0/\sqrt{36} = 1.0$.
- $t = (28.2 - 30)/1.0$.
- $= \mathbf{-1.8}$.

2.3

- Licensed: $-1.8 < -1.69$, so reject H0 at the 5% level – the data are inconsistent with a 30-minute mean.
- Also licensed: the evidence concerns THIS sample; the effect is about 1.8 minutes (report it with its interval).

2.4

- 25 tests at $\alpha = 0.05$ expect at least one false positive with probability $1 - 0.95^{25} \approx 0.72$.
- This p = 0.04 was selected from a lottery – demand pre-registration and a fresh confirmation.

2.5

- (a) → significant, trivially small: act only if 0.3 matters to the business (it usually does not).
- (b) → a 4-minute gap customers would feel, but the evidence is short (p = 0.12): collect more data before acting.
- The moral: significance answers existence; the effect size answers relevance. (across marks)

exercises.lol exercises.com.cn