5.7 Decision limitations

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

KEY WORDS p-value p

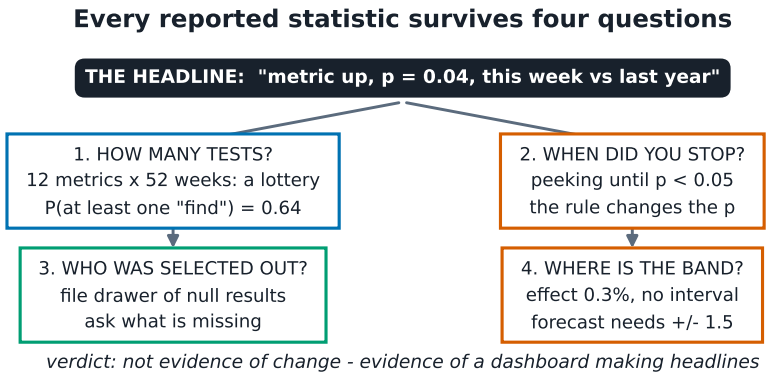
Objective

Spot multiple-testing and stopping-rule traps; audit reported statistics in business claims; demand uncertainty in forecasts.

1 Key knowledge and worked examples

The four gates

How many tests were run? When did you stop? Who was selected out? Where is the band? Twenty tests at $\alpha = 0.05$ expect a "finding" with probability 0.64 – a lottery, not a discovery. A forecast without an interval is a mood.



Every reported statistic survives four questions - or it is a story

Worked example. The dashboard headline

12 metrics scanned weekly; one hits p = 0.04 against last year, effect 0.3%. Question 1 alone retires it: 12 × 52 implicit comparisons. Verdict: not evidence of change – evidence of a dashboard making headlines. The honest response: a pre-registered confirmation on fresh data.

2 Practice

2.1 A chain reports "store 7 improved, p = 0.03" – one of its 14 stores, chosen after looking. Compute the chance that at least one of 14 independent tests clears 0.05, and name TWO gates the claim fails. [4]

2.2 An A/B test was checked daily and stopped on day 9, the first morning $p < 0.05$. In two lines, what does the stopping rule do to the nominal error rate, and what rule would have kept it honest? [2]

2.3 "Next month's revenue: £84,300." With sd £9,000 across months ($n \approx 30$), construct the band the forecast should wear (95%, $t \approx 2$), and say in one line how the owner should use it. [3]

2.4 For each report defect – (a) 12 metrics, one flagged; (b) stopped when p dipped; (c) eleven flat stores omitted; (d) £84,300 with no band – name its gate. [4]

Solutions

2.1

- $P(\geq 1) = 1 - 0.95^{14}$.
- $1 - 0.488 = \mathbf{0.51}$ – a coin flip.
- Gate 1: 14 tests (multiple comparisons).
- Gate 3: the store was selected after looking (selection on the result).

2.2

- Optional stopping inflates the false-positive rate well above the nominal 5% – each peek is another ticket.
- An honest rule fixes n (or a sequential method) in advance and stops only as declared.

2.3

- $SE = 9{,}000/\sqrt{30} = 1{,}643$.
- Band: $84{,}300 \pm 2 \times 1{,}643 = [\mathbf{81{,}000},\ \mathbf{87{,}600}]$.
- Use: plan cash for the band (stress the lower end), and revisit calibration – did past 95% bands hold 95% of the time?

2.4

- (a) How many tests?
- (b) When did you stop?
- (c) Who was selected out?
- (d) Where is the band?

exercises.lol exercises.com.cn