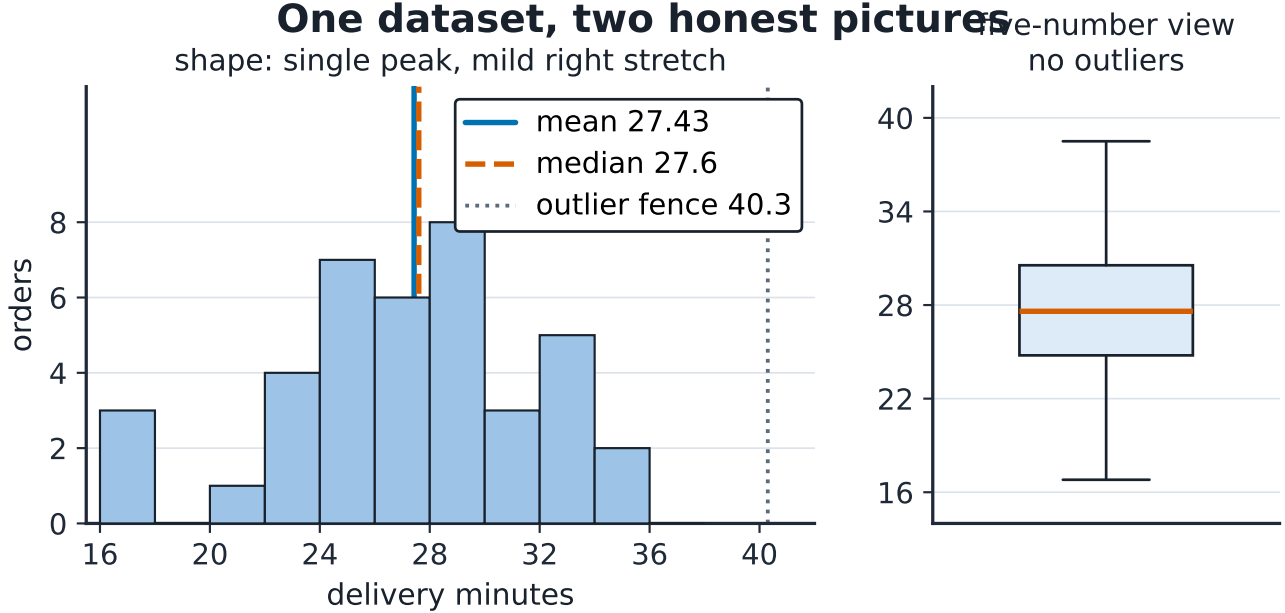
Exercise walk-through

5.5
Hypothesis testing

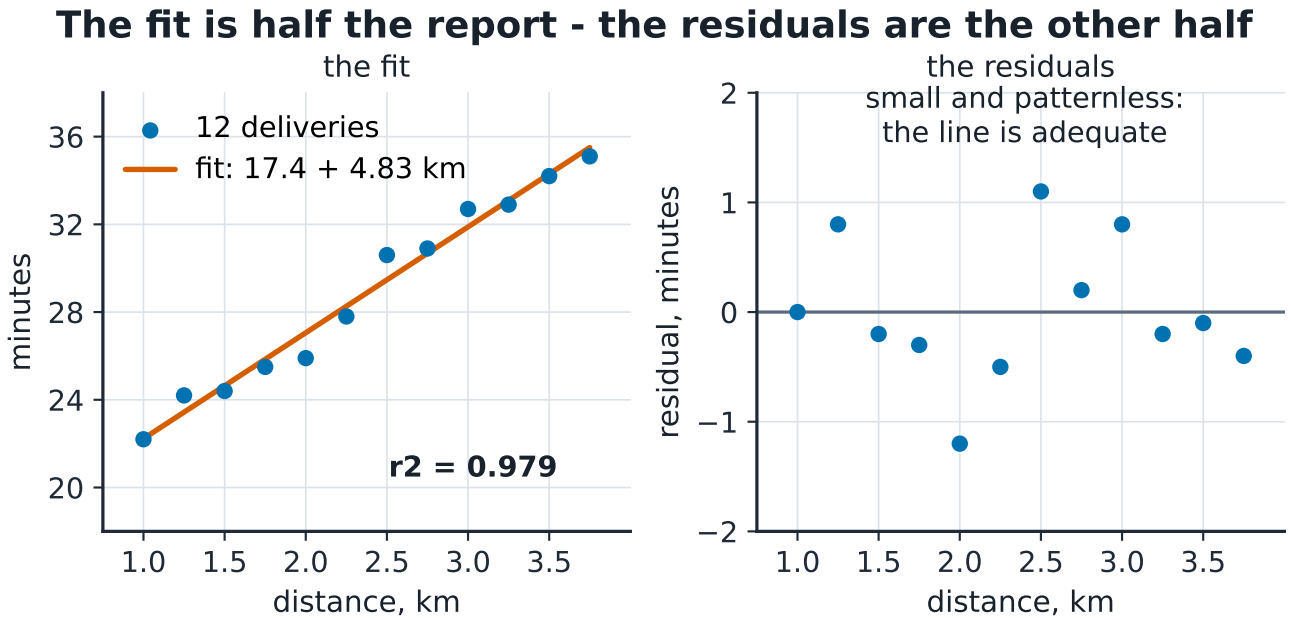
University Business

Descriptive measures and visualisation: the four questions



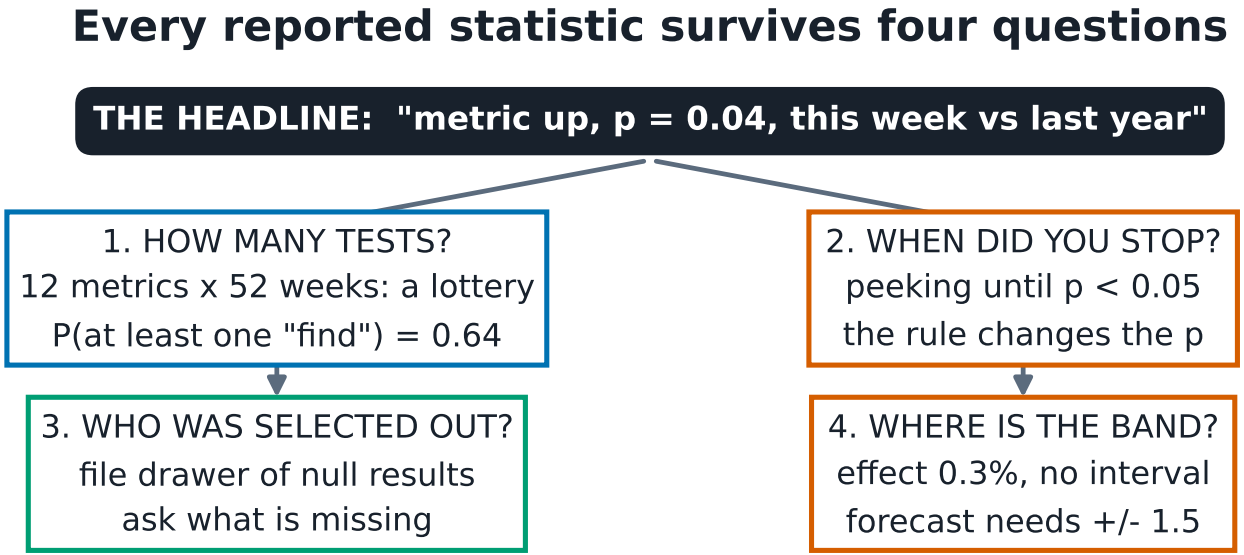
Mean 27.43, **median** 27.6, fence at 40.3: centre, spread and outliers.

Regression: the line and its police



Time = 17.4 + 4.83 km, r2 = 0.979, with patternless residuals.

Limitations: the four gates



verdict: not evidence of change - evidence of a dashboard making headlines

Challenge a headline through four gates: tests, stopping, selection, error.

Exercises 2.1·2.2

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

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

Exercises 2.1–2.2

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

2.1

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

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

2.2

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

Exercises 2.3·2.4

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]

Exercises 2.3·2.4

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]

2.3

- 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 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]

2.4

- 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.

Exercise 2.5

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]

Exercise 2.5

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]

2.5

- (a) Statistically significant but trivially small.
- Act only if a 0.3 difference matters to the business (it usually does not).
- (b) A 4-minute gap customers would feel, but the evidence is weak (p = 0.12).
- Collect more data before acting, rather than dismissing the effect.
- The moral: significance answers existence; the effect size answers relevance. (across marks)