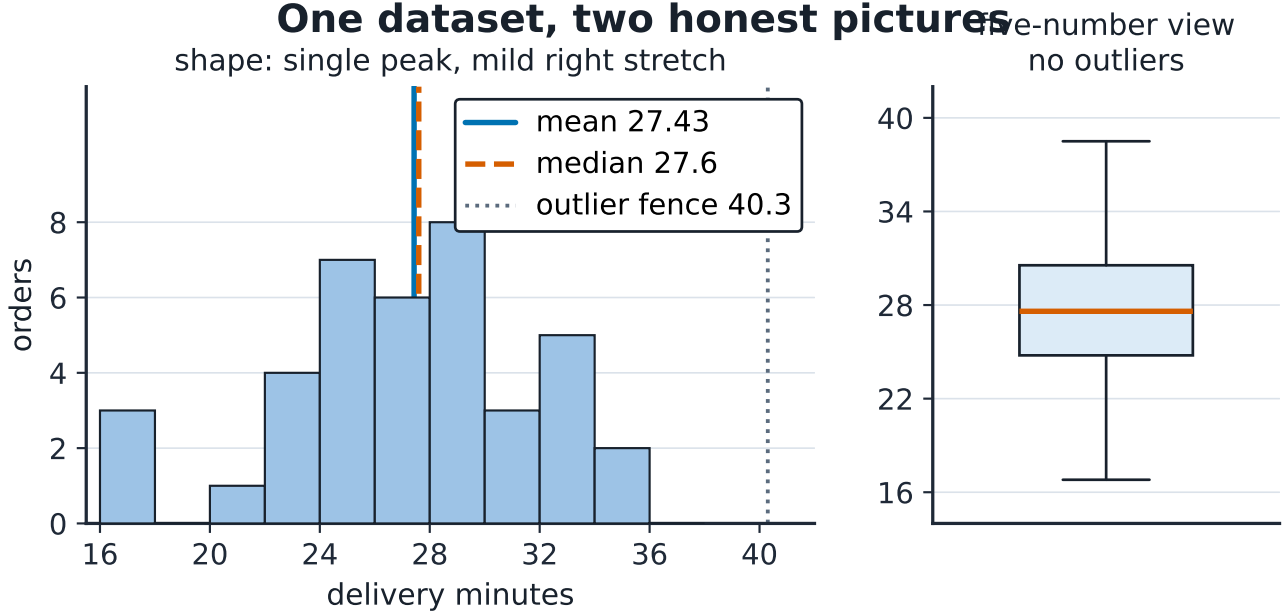
Exercise walk-through

5.6
Regression and correlation

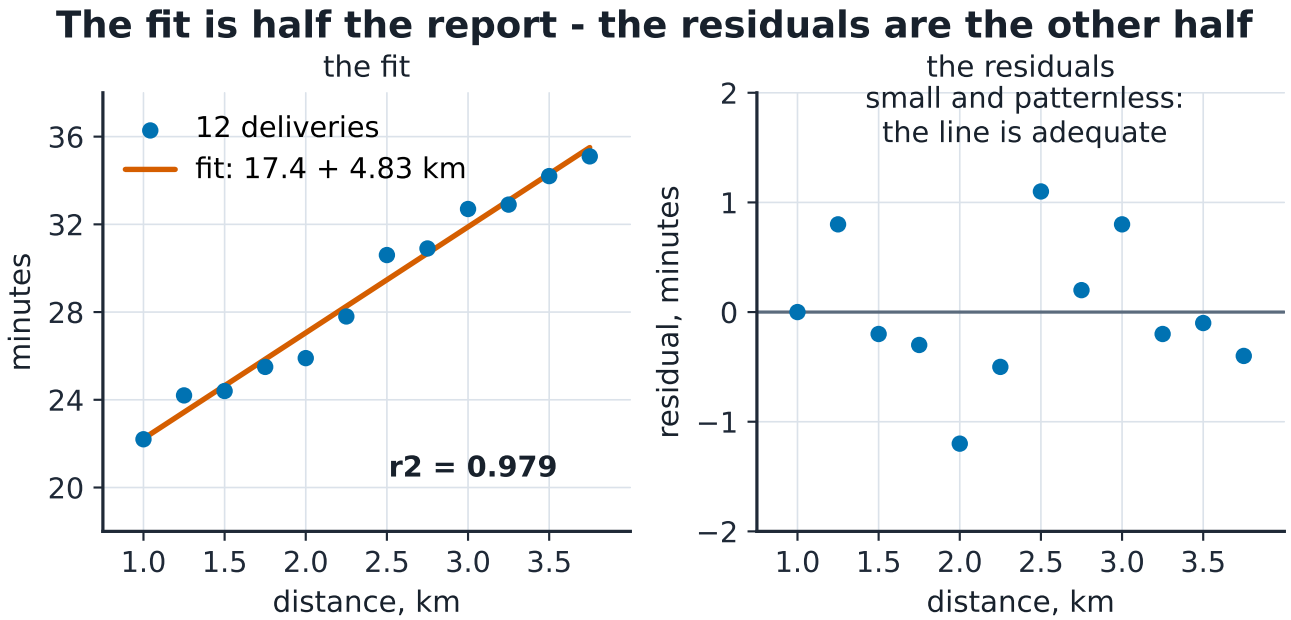
University Business

Descriptive measures and visualisation: the four questions



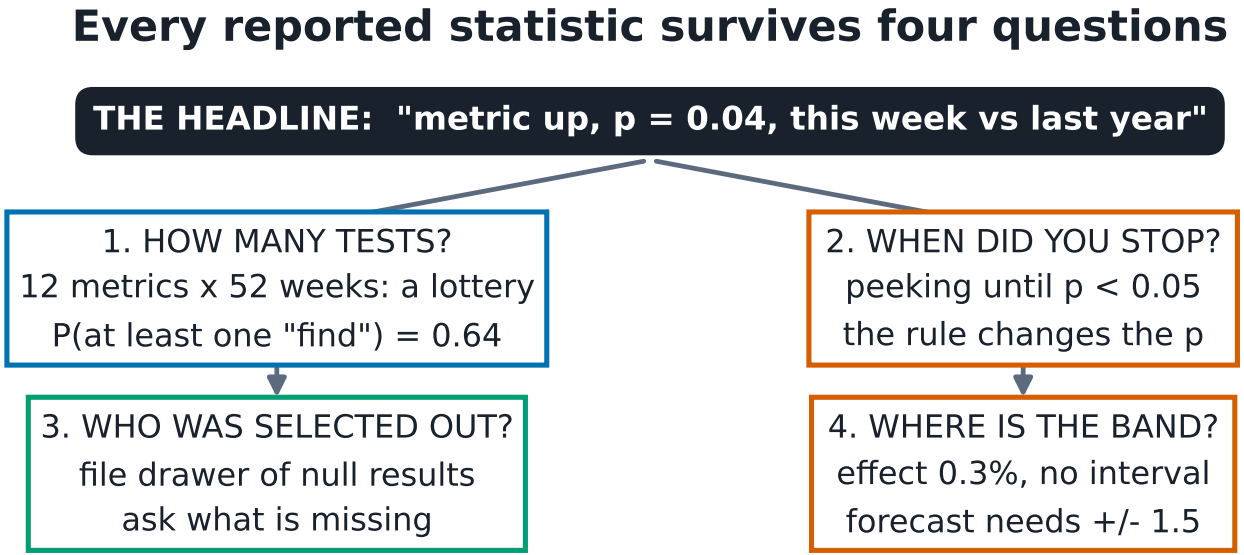
Mean 27.43, **median** 27.6, fence at 40.3: centre, spread and outliers.

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 Interpret the slope and predict the score at setting 4. [3]

2.2 Compute r^2 and say what it does – and does not – mean. [3]

Exercises 2.1·2.2

2.1 Interpret the slope and predict the score at setting 4. [3]

2.1

- Each one-step setting increase travels with +0.62 score on average in this sample.
- At 4: $2.1 + 0.62 \times 4 = 2.1 + 2.48$.
- $=$ **4.6** (4.58).

2.2 Compute r^2 and say what it does – and does not – mean. [3]

2.2

- $r^2 = 0.71^2 =$ **0.50**.
- It means half the score variance moves with the setting in this sample.
- It does NOT mean 50%-accurate predictions or any causal share – variance shared, nothing more.

Exercises 2.3·2.4

2.3 The U-shaped residuals: what do they tell you about the straight line, and what would you try next? [2]

2.4 "Each setting step CAUSES +0.62 score." Rewrite the sentence honestly, and name the study that could earn the causal version. [3]

Exercises 2.3·2.4

2.3 The U-shaped residuals: what do they tell you about the straight line, and what would you try next? [2]

2.3

- The U-shape says the true relation curves; a straight line systematically misses at the ends.
- Try a quadratic term (or bin the settings) – and say so before quoting any slope.

2.4 "Each setting step CAUSES +0.62 score." Rewrite the sentence honestly, and name the study that could earn the causal version. [3]

2.4

- Honest: "setting and score move together ($r^2 = 0.50$); higher settings travel with higher scores in these 10 batches."
- A causal rate needs temperature randomly assigned across batches (holding recipe and time fixed).
- Then the slope becomes an effect of the intervention, not an association.

Exercise 2.5

2.5 True or false, one line each: (a) $r = 0.71$ means the fit is 71% accurate; (b) the intercept 2.1 is the score at setting zero, a setting that does not exist; (c) predicting at setting 9 is extrapolation. [3]

Exercise 2.5

2.5 True or false, one line each: (a) $r = 0.71$ means the fit is 71% accurate; (b) the intercept 2.1 is the score at setting zero, a setting that does not exist; (c) predicting at setting 9 is extrapolation. [3]

2.5

- (a) False – r is strength and direction of association; accuracy claims need prediction errors.
- (b) True – and that is exactly why the intercept is not interpreted here.
- (c) True – no data beyond setting 5 vouches for the line.