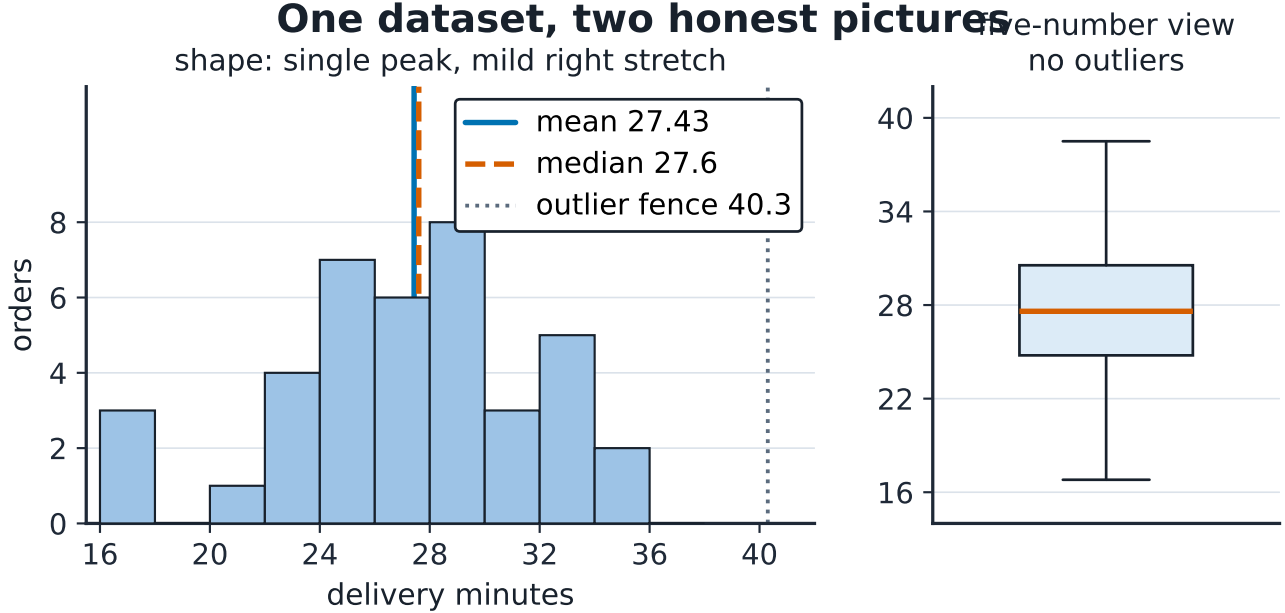
Exercise walk-through

5.1
Data and sampling

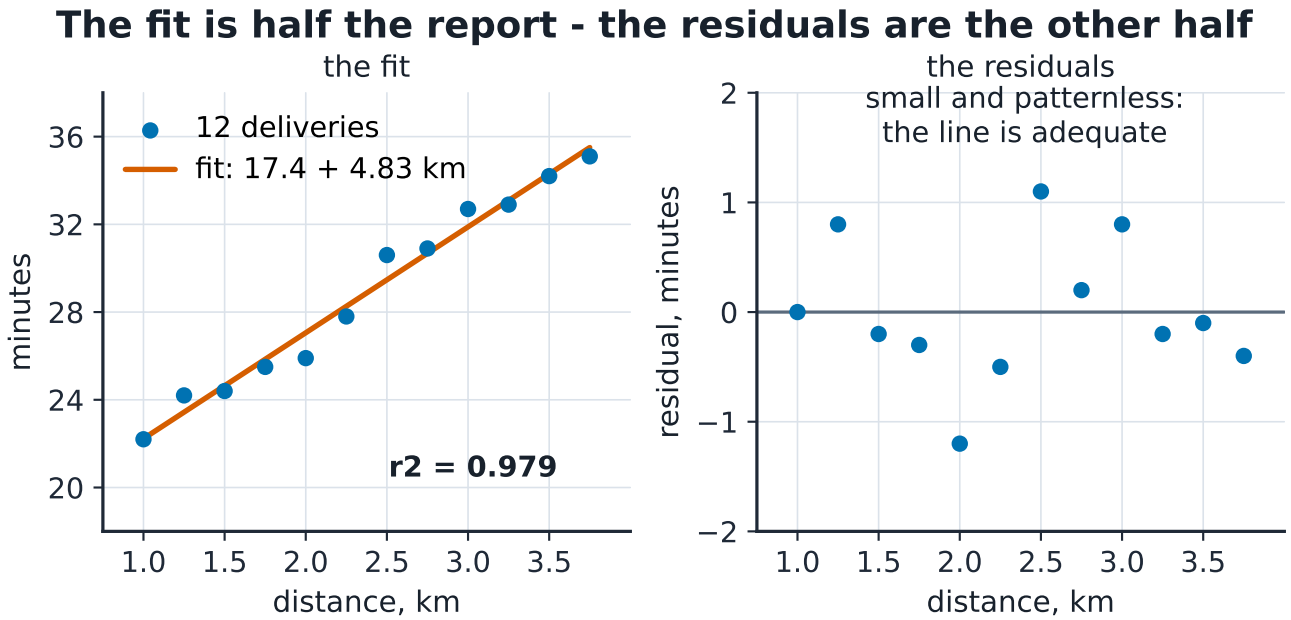
University Business

Descriptive measures and visualisation: the four questions



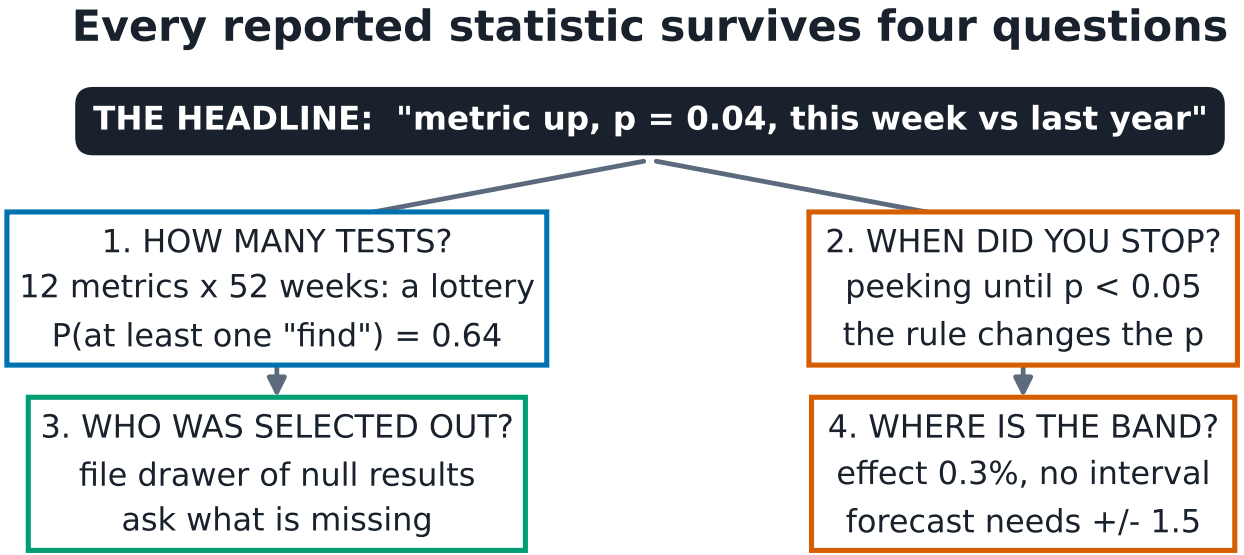
Mean 27.43, **median** 27.6, fence at 40.3: centre, spread and outliers.

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 For each field – delivery minutes; hour of day (11–21); rider code; satisfaction 1–5 – state quantitative or categorical, and the measurement scale. [4]

2.2 Nimbus wants to estimate average weekday delivery time to within ±1 minute. Rank these designs and justify: (a) every 10th order of the week, (b) a random sample of 200 weekday orders, (c) the 200 orders nearest the shop. [3]

Exercises 2.1·2.2

2.1 For each field – delivery minutes; hour of day (11–21); rider code; satisfaction 1–5 – state quantitative or categorical, and the measurement scale. [4]

2.1

- Minutes: quantitative, ratio (zero = no time).
- Hour: categorical, ordinal.
- Rider code: categorical, nominal.
- Satisfaction: categorical, ordinal (ordered ratings).

2.2 Nimbus wants to estimate average weekday delivery time to within ±1 minute. Rank these designs and justify: (a) every 10th order of the week, (b) a random sample of 200 weekday orders, (c) the 200 orders nearest the shop. [3]

2.2

- (b) random 200: unbiased, and n gives the ±1.
- (a) systematic is near-random unless orders cycle in tens.
- (c) convenience: nearest orders are systematically faster – biased at any size.

Exercises 2.3·2.4

2.3 The survey of 500 customers has a 22% response rate, and responders order 30% more often than average. Name the bias and state in one line what it does to the estimated satisfaction. [3]

2.4 A dataset has 40 rows and 3 missing delivery times. Two analysts report means of 27.43 (37 cases) and 25.52 (zeros filled). Which is defensible, and what must accompany it? [3]

Exercises 2.3·2.4

2.3 The survey of 500 customers has a 22% response rate, and responders order 30% more often than average. Name the bias and state in one line what it does to the estimated satisfaction. [3]

2.3

- Non-response bias (with selection inside the response).
- Frequent orderers are over-represented.
- Satisfaction is biased toward loyal customers' view – state the direction, report the rate.

2.4 A dataset has 40 rows and 3 missing delivery times. Two analysts report means of 27.43 (37 cases) and 25.52 (zeros filled). Which is defensible, and what must accompany it? [3]

2.4

- The 37-case mean is defensible.
- Zero-filling invents three instant deliveries and drags the mean down by 1.9.
- The report must state the missingness and its likely direction (busy-runs forgotten → true mean likely a little higher).

Exercise 2.5

2.5 True or false, one line each: (a) a bigger convenience sample becomes representative; (b) stratified sampling needs the strata sizes known; (c) the population is every order the service could have taken that week. [3]

Exercise 2.5

2.5 True or false, one line each: (a) a bigger convenience sample becomes representative; (b) stratified sampling needs the strata sizes known; (c) the population is every order the service could have taken that week. [3]

2.5

- (a) False – size multiplies bias.
- (b) True – proportional allocation needs strata sizes (or estimates).
- (c) True – the population is defined by the claim you want to make.