Accelerated 6 · Unit 8 · Assessment Guide

Data Sets, Distributions, and Sampling

For the Parent or Teacher — What to look for

Unit 8 introduces statistical thinking. The big ideas: (1) center is summarized differently by mean and median; (2) variability is as important as center; (3) random samples allow valid inference, biased samples don't. Many problems test whether the student understands the DIFFERENCE between describing data and making claims about a population.

Part A — Quick Stats

- Mean: (4+7+7+8+10+12+14)/7 = 62/7 ≈ 8.86. Median: 8. Mode: 7. Range: 14−4 = 10.
- Watch for confusion between median and mean for a small set.

Problem 1 — Summary statistics

- Sum = 65. Mean = 6.5. Median (5th-6th avg) = (6+7)/2 = 6.5. Mode = 4 (appears twice) and 8 (appears twice) — bimodal. Range = 10.
- (e) The distribution is slightly RIGHT-SKEWED (12 pulls the tail right; mean > median… but here mean = median ≈ 6.5. So fairly symmetric).
- Strong: notices mean ≈ median → roughly symmetric. Catches the bimodal nature.
- Weak: identifies one mode (misses one of 4, 8).

Problem 2 — Class scores

- Sum = 1115. Mean ≈ 74.3. Median = 75 (8th value). Mode = 70 (appears 3 times).
- (b) Median is more 'typical' here because of low outliers (the 50 and 60s pull mean down).
- (c) Adding 30: mean changes a lot, median barely (slips a slot).
- ★ Strong: explains how mean is SENSITIVE to outliers, median is RESISTANT.
- Weak: 'mean is always best' or 'they're the same' — no reasoning about robustness.

Problem 3 — Consistency: Alex vs Sam

- Alex mean = 13. Sam mean = 13 (same!).
- But Alex's range is 2; Sam's is 20.
- ★ KEY: same mean, very different variability. This is the heart of why variability matters.
- Strong: 'I'd pick Alex — same mean but much less variability, so 13 next game is more likely.'
- Weak: picks Sam because of one big game (25); doesn't think about probability of repeating it.

Problem 4 — IQR of 3,5,7,8,10,12,15,18,20

- Median = 10 (5th value).
- Lower half: 3,5,7,8. Q1 = (5+7)/2 = 6.
- Upper half: 12,15,18,20. Q3 = (15+18)/2 = 16.5.
- IQR = 10.5.
- Strong: correct quartile calculations.
- Common error: includes median in both halves (yields different Q1/Q3). Either convention is OK if used consistently, but mixing them is a mistake.

Problem 5 — Sampling methods

- (a) Mai's sample is BIASED — her friends likely have similar habits.
- (b) Han's is RANDOM (or close to it).
- ★ Strong: clearly articulates that bias compromises generalizability.
- Weak: declares Mai's sample 'random because she didn't pick on purpose' — misses the population/sample distinction.

Problem 6 — Sample-to-population estimate

- 30/50 = 60% prefer apple. Estimated population apple-lovers: 2400.
- Assumption: the sample is REPRESENTATIVE of the population.
- Second sample (26/50 = 52%): estimate becomes 2080.
- Difference: 320 (significant!). This shows SAMPLING VARIABILITY.
- ★ Strong: addresses that different samples give different estimates; both could be true 'enough.'
- Weak: claims the second sample is 'wrong' — doesn't understand sample variation.

Problem 7 — Probability with marbles

- Total = 12. P(red) = 4/12 = 1/3. P(NOT blue) = 9/12 = 3/4. P(red or green) = 9/12 = 3/4.
- Expected reds in 120: 120 · 1/3 = 40.
- Strong: gives clean reduced fractions.
- ★ Strong+: notes P(NOT blue) = P(red or green) (since red and green are 'not blue').

Problem 8 — Add a value to keep median, raise mean

- Original sum = 49. To get mean = 10 with 7 numbers, total must be 70. So Diego added 21.
- Median stays 8: the added value (21) goes to the END of the sorted list, so the new median is just the original middle (8 or thereabouts; specifically, average of 3rd and 4th in 7 numbers).
- Original sorted: 5,7,8,8,9,12. Adding 21 → 5,7,8,8,9,12,21. Median = 8. ✓
- ★ Strong: ANY value ≥ 12 that makes the mean 10 works, but the mean condition uniquely fixes the value at 21. So there's only ONE answer.
- Weak: gives a non-21 answer; or claims many possibilities without checking the mean.

Problem 9 — Defect rate sampling

- 3/200 = 1.5% defect rate → 50000 · 0.015 = 750 estimated defects.
- Second sample: 8/200 = 4%. New estimate (combining): 11/400 = 2.75% → 1375 defects.
- ★ Strong: combines samples for a BETTER estimate; doesn't just take the second alone.
- ★ Strong+: addresses that with small samples, the proportion can vary significantly; large samples are more reliable.
- Weak: replaces the first estimate entirely with the second; or averages percents incorrectly.

Problem 10 — Class average pulled down by one

- Sum of 10 scores = 780. Sum of 9 students at 80 = 720. So 10th score = 60.
- Why does the average drop? The one low score of 60 pulls the mean DOWN by 2 from 80 → 78. This is a 20-point gap divided by 10 students.
- ★ Strong: explains that 'each student's score contributes equally to the average, but one outlier can shift the average noticeably.'
- Weak: gets 60 but can't articulate why averages are sensitive to single values.

— end —