# Algebra 1 · Unit 3 · Assessment Guide

*Two-Variable Statistics*

*For the Parent or Teacher — What to look for*

---

*Unit 3 looks at relationships between TWO variables. The mathematics here (relative frequencies, lines of best fit, residuals, correlation) is mostly mechanical. The deep work is in INTERPRETATION — distinguishing types of frequency, association from causation, and the limits of any single statistic. Watch for those distinctions throughout.*

## Problem 1 — Two-way table (dogs vs. cats; house vs. apartment)

- (a) 84/120 = 70%.
- (b) 84/116 ≈ 72.4%.
- (c) ★ THE diagnostic — these are different CONDITIONAL frequencies. (a) is given-house; (b) is given-dog. The DENOMINATOR is what changes.
- (d) Yes — house dwellers prefer dogs (70%) at a much higher rate than apartment dwellers (32/80 = 40%). The variables are associated.
- Watch for: students who answer (a) and (b) correctly but can't articulate why they're different — they're computing without thinking.

## Problem 2 — Shoe-size vs. height model

- (a) Slope: "For each 1-size increase in shoe size, predicted height increases by 1.4 INCHES." Units matter.
- (b) y-intercept: shoe size = 0 → predicted height = 50 in. NOT sensible — no one has shoe size 0. ★ Strong students raise this as an example of extrapolation outside the data.
- (c) y = 1.4(11) + 50 = 65.4 inches.
- (d) Shoe size 25 is wildly outside any realistic range. Don't trust extrapolations far outside the observed data.

## Problem 3 — Residuals

- (a) Predicted: 2(3) + 5 = 11. Actual: 13. Residual = 13 − 11 = +2.
- (b) Predicted: 2(7) + 5 = 19. Actual: 17. Residual = 17 − 19 = −2.
- (c) Positive: actual > predicted (point above the line). Negative: actual < predicted (point below).
- (d) Most residuals near zero → the linear model fits well; data follows the line closely.
- Watch for sign errors in (b) — common to write |actual − predicted| without sign.

## Problem 4 — High r doesn't equal good linear fit

- ★ The key insight of this section: r summarizes linear association, but a CURVED pattern can still have high |r| if it trends one way overall.
- Strong: "Plot the residuals. If they show a pattern (e.g., U-shape), the linear model is missing something, even if r is high."
- Weak: "Both are equally good" — this misses the whole point about model adequacy.

## Problem 5 — Match scatter plots to r values

- (a) r ≈ 0.97. (b) r ≈ −0.95. (c) r ≈ 0. (d) r ≈ 0.4. (e) r ≈ −0.3.
- Common errors: ignoring the SIGN; not connecting "strong" with magnitude near 1; treating r = 0.4 as "moderate-strong" instead of "weak-to-moderate."

### Problem 6 — Ice cream and drowning

- (a) No, ice cream doesn't cause drowning. They're correlated through a third variable.
- (b) HEAT / SUMMER — hot weather drives both ice cream sales and swimming (and thus drowning risk).
- ★ The single best example of the lurking-variable concept.

### Problem 7 — Causal vs. correlated vs. coincidental

- (a) Likely causal (more study → more learning), though confounders exist (motivation, ability).
- (b) Correlated, NOT causal — bigger fires draw more firefighters AND cause more damage. Common cause: size of fire.
- (c) Coincidental — no plausible mechanism. Classic spurious correlation.
- (d) Likely causal — direct mechanism (energy balance) is well-established.
- Strong: identifies the mechanism (or its absence) rather than just guessing.

### Problem 8 — Lin's shoe-size/math claim

- (a) Problems: (1) r = 0.4 is WEAK; doesn't justify a strong claim. (2) Correlation does not equal causation — there's likely a confounding variable (AGE — older kids have bigger feet AND know more math). (3) Sample ("my friends") may not be representative.
- (b) Better: "Among my friends, there's a weak positive correlation; this might be because older friends have both bigger feet and more math experience."
- ★ A 4-star answer raises BOTH the strength issue AND the confounding issue.

### Problem 9 — r = 0.99, should we trust the model?

- (a) Not automatically — must also check residuals for patterns (a curved relationship can still produce r near 1).
- (b) Look at the scatter plot directly; examine residuals; consider whether the relationship makes physical sense; check for outliers.
- ★ The conceptual peak — r is a SUMMARY, not a verdict.

### Problem 10 — Open design

- Strong response identifies: realistic variables, plausible sign and strength prediction, a real confounder (e.g., "hours of sleep vs. test score — both might be affected by stress"), and explicit awareness that even strong r doesn't prove causation.
- Weak: vague variables ("happiness"), no prediction, no confounder mentioned, or claims causation outright.
- ★ Treat as the modeling/reasoning capstone of the unit.

*— end —*