# Error & confidence cheat sheet — for claims

The maths behind "how sure are we?" — written for the EIS/yeast work, applicable everywhere. **Rule zero: every number in a claim carries three companions — n, an uncertainty, and the conditions it was measured under.**

Bluefruit Software — internal reference

## 1 Vocabulary

- **Accuracy** — closeness to the true value (needs a reference). **Precision** — closeness of repeats to each other. You can be precise and wrong.
- **Repeatability** — same operator, rig, day. **Reproducibility** — different operator/day/rig. Quote which one.
- **Random error** averages away with more runs. **Bias** doesn't — only a reference reveals it. More n never fixes bias.
- **SD** — spread of runs. **SEM = s/√n** — uncertainty of the mean. Label which on every plot.

## 2 The basics

**CV = s/x̄** (report as %) — the workhorse for repeatability rows.

**95% CI of the mean = x̄ ± t·s/√n**, where t depends on n:

| n | 3 | 4 | 5 | 6 | 10 | 20 | 30+ |
|---|---|---|---|---|---|---|---|
| t | 4.30 | 3.18 | 2.78 | 2.57 | 2.26 | 2.09 | ≈2.0 |

At **n=3** the CI is ±4.3 SEM — triplicates give an *indication*, not a quantified claim.

## 3 Comparing two conditions

**Δ = x̄₁ − x̄₂, SE(Δ) = √(s₁²/n₁ + s₂²/n₂)**, CI ≈ Δ ± 2·SE(Δ). If the CI on Δ excludes zero, the separation is real.

**Effect size d = Δ/s** — "separation is N× the noise". d < 1: heavy overlap; d ≈ 2: barely overlapping; d ≥ 5: LED-demo territory.

Eyeball rule: non-overlapping 95% CIs → significant. Overlapping → *not necessarily* insignificant — do the sum.

## 4 How many runs?

Decide *before* the experiment. Per group, 80% power at p<0.05:

$$n \approx 16 / d^2$$

| d = 2 | d = 1 | d = 0.5 |
|---|---|---|
| n ≈ 4 | n ≈ 16 | n ≈ 64 |

Backwards: n=5 per condition only reliably detects d ≳ 1.8. Write the target d into the experiment's acceptance criteria.

## 5 Error budget

Independent sources add **in quadrature**: **σ² = σ_fixture² + σ_prep² + σ_drift² + σ_electronics²**. Measure each separately; improve the dominant term, ignore the rest.

Multiplying/dividing: **relative** errors add in quadrature.

**Ground-truth cap: σ_claim ≥ σ_reference.** If cell counting agrees to ±10%, no viability claim can be tighter than ±10% — however good the instrument.

## 6 Repeatability studies

Report **CV per frequency** across the sweep — repeatability is rarely uniform in frequency.

Separate variance components: **total = between-remount + within-mount**. Recipe: k remounts × m runs each; within = mean of per-mount variances; between = variance of per-mount means − within/m.

The bigger component names the enemy — fixture vs electronics/drift.

## 7 Classification claims

Report **sensitivity** and **specificity** separately — a 90%-accurate classifier on a 90/10 dataset may be useless.

CI on a proportion: **p ± 1.96·√(p(1−p)/n)** (n ≥ ~30; Wilson for smaller).

19/20 correct = 95% accuracy, but the 95% CI is ~**75–99%**. Claiming "95% ± 5%" needs **n ≈ 70+ blind runs**. Small-n accuracies are anecdotes wearing percentages.

## 8 Trend / dose-response

Fit the agreed model: report **slope ± 95% CI**. "Monotonic response" when the slope CI excludes zero.

Stronger claim: means are ordered *and* adjacent CIs on Δ exclude zero.

Plot **every replicate point**, not just means — reviewers and clients' scientists trust scatter.

**Key** **n** number of runs · **x̄** mean · **s / SD** standard deviation · **SEM** standard error of the mean (s/√n) · **CV** coefficient of variation (s/x̄) · **CI** confidence interval · **t** Student's t multiplier · **Δ** difference of two means · **SE(Δ)** standard error of Δ · **d** effect size (Δ/s) · **σ** SD of one error source · **p** probability of a false positive · **EIS** electrochemical impedance spectroscopy · **Rct** charge-transfer resistance · **AC** acceptance criteria

## 9 Honesty rules — the ones that get checked

Significant figures follow uncertainty: **3.2 kΩ ± 0.3 kΩ** — never 3.217 kΩ ± 0.3 kΩ.

State **n on every figure**; label SD vs SEM vs CI on every error bar.

"Inconclusive" with small n means **underpowered**, not "no effect" — say which.

Declare the analysis **before the data**; reporting the best of many frequencies inflates false positives. Post-hoc finds are new hypotheses, not results.

**Blind before you quantify.** Unblinded quantified claims cap at Client-demo rung regardless of the stats.

**Rung mapping** — **Bench anecdote** — none of this · **Client demo** — §2 + §5 (noise floor, effect ≫ noise) · **Quantified claim** — §3–§4 and §7–§8, done blind · **Cited claim** — all of it, versioned, with §9 audited