Statistical Analysis & Error Control

Companion to the Week 8 episode. Most statistical damage is misinterpreting the right test — not using the wrong one.

1. What a p-value IS — and 4 things it is NOT

IS: the probability of data at least as extreme as observed, *IF the null (usually "no effect") were true.* A statement about your data, under an assumption.

Common misreading	Why it's wrong (ASA, 2016)
✗ "p = the probability the null is true"	It's computed *assuming* the null — it can't tell you the probability *of* that assumption.
✗ "p = probability the result is due to chance"	No.
✗ "1 – p = probability the effect is real / will replicate"	No — p-values are highly variable study to study.
✗ "small p = big / important effect"	Says **nothing** about size — a trivial effect is highly significant with large N.

> **"p < .05" does NOT mean "95% likely to be real."** It means: if there were truly no effect, data like ours would occur < 5% of the time. That's all.

2. The dichotomy trap

p = 0.049 and p = 0.051 are **essentially identical evidence** — yet the bright line calls one a "discovery" and one "nothing." The ASA: don't base conclusions only on crossing a threshold. .05 is an arbitrary **convention**, not a law of nature.

The dichotomy tempts two opposite errors
"Significant" ⇒ true & important — ignoring a trivial effect size or low-power inflation (Week 7).
"Non-significant" ⇒ no effect — when the study may just be too small (absence of evidence ≠ evidence of absence).

Treat p as **continuous** evidence (.01 > .04 > .2), not a switch that flips at a line.

3. Think in estimates — the "New Statistics" (Cumming, 2014)

Stop asking "is there an effect?" Ask **"how big is it, and how precisely do I know it?"** Report the **effect size** (meaningful units) + its **confidence interval** (best estimate AND uncertainty in one picture). Interpret the whole interval — especially its ends — against your SESOI.

Old (dichotomy)	Estimation (richer, honest)
"The improvement was significant (p = .03)."	"The improvement was ~4 s, 95% CI 1 to 7 s."
Stops there.	Direction + likely size + uncertainty. Could be as small as 1 s (trivial?) or 7 s (a real edge) — *and that honest "we don't yet know which" is the finding*.

Bonus: a "non-significant" result with a **tight** CI hugging zero (e.g. −0.5 to +0.5 s) is *positively informative* — it shows any real effect is small. The ritual "we found nothing" can never say that.

4. Multiplicity & doing the analysis honestly

One test → ~5% false positive. **Twenty tests → expect ~1 "significant" by chance** even if nothing is real (the engine under researcher degrees of freedom, Week 6). The more outcomes/subgroups/time-points, the more your real error rate inflates.

Protection is procedural: run **preregistered confirmatory analyses first** (error rate intact) → report as primary; then explore freely but **label everything else exploratory** (a hypothesis for next time). The honesty *is* the error control — you get rigour *and* discovery.

5. Do this — this week

Write the analysis honestly

1. State your **one** preregistered primary analysis.
2. Commit to reporting its **effect size + confidence interval vs your SESOI** — not a bare p-value, never a bare "significant/not."
3. List exploratory analyses **separately**, flagged as such.
4. Practise on a paper: cover the p-value; look only at the effect size + CI; ask **"how big, how sure?"** (You'll often find the CI so wide it rules little out.)

References

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician, 70*(2), 129–133.
Cumming, G. (2014). The new statistics: why and how. *Psychological Science, 25*(1), 7–29.
Gelman, A., & Carlin, J. (2014). Beyond power calculations. *Perspectives on Psychological Science, 9*(6), 641–651.