Research & Analysis / Data Interpretation

Turn non-significant findings into actionable insights — was the study underpowered, or is there really nothing there?
Difficulty: Advanced
Model: GPT-4 / Claude / Gemini
Use Case: A/B Test Analysis, Research Reports, Decision Making
Updated: May 2026
Why This Prompt Exists
“Not statistically significant” is not the same as “no effect” — but most people treat it that way.

You get:

- abandoning good ideas because your test was underpowered
- implementing bad ideas because you mistook noise for signal
- frustrated stakeholders saying “the data shows nothing” when it shows uncertainty
- stopping tests too early and calling null results conclusive
- missing moderate effects that matter for business but not for p-values

But null results have explanations:

- true null: there really is no effect (effect size = 0)
- underpowered: effect exists but sample too small to detect it
- measurement failure: effect exists but you measured it poorly
- heterogeneity: effect exists but only for some subgroups (averaged out)
- wrong metric: effect exists but on a different outcome than you measured

Without proper interpretation, you throw away learning.

This prompt analyzes null results and tells you what they actually mean.

The Prompt

```
Assume the role of a statistical consultant who interprets null findings.

Your task is to analyze a non-significant result and recommend next steps.

Generate:

1. RESULT SUMMARY
   - What was tested
   - P-value and effect size (with confidence interval)
   - Sample size

2. POWER ANALYSIS
   - What effect size could this study detect? (given sample size)
   - Is the study underpowered for practically meaningful effects?
   - Post-hoc power (interpret with caution)

3. POSSIBLE EXPLANATIONS (ranked)
   - True null (effect is zero or trivial)
   - Underpowered (effect exists but study too small)
   - Measurement issue (poor reliability, wrong construct)
   - Heterogeneity (effects cancel out across subgroups)

4. CONFIDENCE INTERVAL INTERPRETATION
```

```
   - Range of plausible effect sizes
   - Does this interval exclude practically meaningful effects?
   - What effects are still possible?

5. RECOMMENDATION
   - Stop, effect is truly negligible (upper bound of CI is trivial)
   - Run larger study (effect might exist, CI includes meaningful
values)
   - Improve measurement (CI is wide but plausible effect size is
meaningful)
   - Analyze subgroups (heterogeneity suspected)

INPUTS:

Test description:
[E.G., "A/B test of new checkout button"]

Null result:
[E.G., "p = 0.23, effect size = +0.5% conversion, 95% CI [-0.8%,
+1.8%], N=5,000 per group"]

Practically meaningful effect size (minimum detectable that matters):
[E.G., "2% conversion lift would be worth implementing"]

Business context:
[E.G., "High-traffic e-commerce site"]

RULES:
- Never say "no effect" when you mean "not statistically significant"
```

```
- Interpret the confidence interval, not just the p-value
- Distinguish between "statistically significant" and "practically meaningful"
- Note that failure to reject the null is not acceptance of the null
```

How To Use It

- Run this on every A/B test that comes back non-significant before killing the variant.
- Use the confidence interval to guide decisions — the point estimate + uncertainty.
- Calculate the minimum detectable effect for your sample size before running tests.
- Distinguish between “stop” (effect is definitely tiny) and “needs more data” (effect might be meaningful).
- Present null results as confidence intervals, not p-values, to stakeholders.

Example Input

Test description:
“A/B test of personalized email subject lines vs. generic”

Null result:
“p = 0.45, effect size = +0.3% open rate, 95% CI [-0.5%, +1.1%], N=10,000 per group”

Practically meaningful effect size:
“1% increase in open rate would be worth implementing”

Business context:
“Email marketing to 2M subscribers monthly”

Why It Works
Most organizations treat null results as failures — “the test didn’t work.”

This framework improves outcomes by forcing:

- power analysis (could you have detected an effect if it existed?)
- explanation ranking (not all nulls are the same)
- confidence interval interpretation (range of possible truths)
- practical significance check (does the effect need to be large to matter?)
- clear recommendation (stop, run larger study, or improve measurement)

Great null interpretation doesn't call a test "failed" — it extracts the maximum learning from uncertainty.

Build Better AI Systems

Subscribe for advanced prompt engineering, AI coding tools, debugging frameworks, and practical strategies for developers and engineers.

Carefully engineered prompts for people doing real work.

Share this:

- Share on Facebook (Opens in new window) Facebook
- Share on X (Opens in new window) X

See also Statistical Output Translator