

p = 0.04 is not 96% sure.

D8A.ACADEMY/RESOURCES/P-VALUE-NO-STATS

WHAT P SAYS

P(data H0)	if nothing changed
p = 0.04	4% by luck alone
not P(H0 data)	← the flip is the trap.

WHAT IT IS NOT

96% sure	nope. see line 1
effect size	tiny lifts pass too
importance	p ignores money

THE THRESHOLD

alpha = 0.05	a convention, not physics
0.049 vs 0.051	same evidence
1 in 20	← false alarms at 0.05.

READ IT RIGHT

report the CI	a range beats a verdict
lift + CI + n	the honest trio
fix n up front	then stop looking

THE PEEKING TRAP

check daily	p dips under 0.05 by luck
stop at a win	← false positives.
sequential test	if you must peek

GOTCHAS

20 metrics	one wins by chance
p = 0.2	no evidence, not no effect
huge n	everything is significant

WORTH MEMORIZING

01 The false-positive machine, in 9 lines

Two identical variants, one test checked every 500 users. Watch it find a winner.

```
rng = np.random.default_rng(7)
a = rng.binomial(1, 0.10, 20_000) # same rate
b = rng.binomial(1, 0.10, 20_000) # same rate

wins = 0
for n in range(500, 20_000, 500): # peek every 500
    _, p = ztest(a[:n], b[:n])
    wins += p < 0.05
# wins = 6 of 39 peeks flagged "significant"
```

→ A and B are the same coin. Peeking until p dips under 0.05 and stopping there turns a 5% false-alarm rate into a near certainty.

02 Trap: what the sentence actually claims

Three phrasings that go into slide decks every week, and what they really mean.

YOU SAY	IT MEANS	SAY INSTEAD
p = 0.04, so 96% sure	nothing about your idea	unlikely under no effect
p = 0.2, so no effect	no evidence at this n	inconclusive, need more
significant	probably not pure luck	lift of X, CI [a, b]

→ Interview phrasing: a p-value is the probability of the data given no effect, never the probability of the effect given the data.