UNIT 11 · BLOCK 3

Independent-Samples t-Test

Comparing the averages of two separate groups — and deciding whether the gap is real.

Why this matters

Up to now you've asked, "is this *one* group different from a number?" Now comes a question you'll use for the rest of your life: **do these two groups differ?** Treatment vs. placebo. Commuters vs. residents. One menu vs. another. People paid a little vs. people paid a lot. The independent-samples *t*-test is the tool for the most common comparison in all of social science — two separate groups of people, one mean apiece — and Google Sheets does the arithmetic in a single line.

By the end you can…

- Explain what an **independent-samples *t*-test** compares — two *separate* groups of people.
- Tell an **independent** design from a **paired** one, and pick the right test.
- Say where ***t*** comes from (signal ÷ noise) and find $df = n_1 + n_2 - 2$.
- Name the test's three **assumptions** and read a ***t***, a ***p***, and **Cohen's *d***.

Run the test in Google Sheets with one `T.TEST` formula.

1. The big idea

An **independent-samples *t*-test** compares the **means of two separate groups** — different people in each one. Nobody is in both. You already know how to find one mean. Here you find **two** means and ask the key question: *are they far enough apart to be real, or is this just random wiggle?*

Picture the most ordinary study you can: a researcher wants to know whether **commuters** report more stress than **residents** who live on campus. Two clearly separate sets of people. Each person sits in exactly one group, gets measured once, and contributes one stress score. We average each group, look at the gap between the two averages, and ask whether that gap is bigger than the everyday noise you'd expect from grabbing two random handfuls of students. That is the whole job of the independent *t*-test.

Term	Plain English
Independent samples	Two groups, *different* people in each — nobody is in both
t	How far apart the two means are, measured in "noise units"
df	Degrees of freedom — here, $\mathbf{n_1 + n_2 - 2}$; sets which t-curve to use
p	The chance you'd see a gap this big if the groups were really the same
Cohen's *d*	How *big* the difference is, in plain size terms

CHECK **A researcher compares mean stress for `commuters` versus `residents` — two groups made of different students. Which test fits?**

Independent-samples *t*-test. You're comparing the means of *two separate groups* of different people, not one sample against a fixed number. That's exactly what the independent *t*-test is built for.

2. Independent vs. paired — which test?

Before you run anything, ask one question: **are the two sets of scores from different people, or from the same people twice?** Everything hinges on it. Two *separate* groups (commuters vs. residents; $1 group vs. $20 group) call for the **independent** *t*-test. The *same* people measured twice (mood before vs. after a workshop; each student's score in spring vs. fall) call for the **paired** *t*-test — the subject of the next unit. The math differs because paired scores come in linked couples, and the independent test treats every score as its own free-floating observation.

	Independent-samples *t*	Paired-samples *t*
Who's in each column?	Different people in each group	The same people, measured twice
Scores linked?	No — two unrelated piles	Yes — each row is one person's pair
Classic example	Commuters vs. residents; $1 vs. $20	Mood before vs. after a treatment
Equal group sizes?	Not required ($n_1 \neq n_2$ is fine)	Always equal — every score is paired
This unit?	Yes	Unit 12

Quick test

Could a single person appear in *both* columns of your data? If **no** (a person is either a commuter or a resident, never both) **independent**. If the two columns are the *same* people at two times **paired**. That one question settles the choice.

CHECK **You measure each student's anxiety *before* an exam and again *after* it — same students, two scores apiece. Independent or paired?**

Paired *t*-test. The two columns are the *same* people measured twice, so the scores come in linked couples. Independent is only for two groups of *different* people. (We tackle paired in Unit 12.)

3. What *t* actually is

Here's the whole idea in one fraction:

The one formula

***t* = (difference between the two group means) ÷ (standard error of that difference).** The **top** is the gap between the averages — the *signal*. The **bottom** is the noise *within* the groups — the *standard error*, which measures how much two group means would bounce around by chance alone.

So *t* is a **signal-to-noise ratio**. The same one-point gap means very different things depending on the spread inside the groups. If everyone in each group clusters tightly around their average, a one-point gap is huge — clearly bigger than the wiggle. If the scores are wildly scattered, that same one-point gap could easily be random. The standard error in the denominator captures exactly that scatter (and

shrinks as your samples get larger, because big samples give steadier means).

- **Big *t*** (far from 0) the groups are far apart compared to the noise. Strong evidence of a real difference.
- **Small *t*** (near 0) the gap is tiny next to the noise. Probably nothing.

You don't grind through the standard-error formula by hand — Sheets and the calculator below do it. Just hold onto the intuition: **bigger *t* = stronger evidence the two groups truly differ.**

CHECK **Two groups have the very same one-point gap between their means. In which case is *t larger*?**

When the scores are tightly packed. A small standard error (tight groups) sits in the denominator, so the same gap divided by a smaller number yields a bigger *t*. Less noise makes the same signal stand out more.

4. Degrees of freedom: $df = n_1 + n_2 - 2$

To turn your *t* value into a *p* value, the test needs to know *which* t-curve to look at — and that's set by the **degrees of freedom (*df*)**. For an independent-samples *t*-test with two groups, the rule is short:

The *df* rule

***df* = $n_1 + n_2 - 2$.** Add the two group sizes, then subtract 2 — one for each group mean you had to estimate. With 20 people per group: $df = 20 + 20 - 2 = \mathbf{38}$.

Why subtract *two*? Each group's mean "uses up" one degree of freedom (once you know a group's mean and all but one of its scores, the last score is locked in). Two groups, two means, so you lose two. This is the small but important contrast with the one-sample test from

Unit 10, where you estimated just *one* mean and so subtracted only one ($df = n - 1$). Degrees of freedom feed straight into the critical value: more *df* a slightly more forgiving cutoff, because larger samples pin the means down better.

CHECK **An independent *t*-test compares Group 1 ($n = 15$) with Group 2 ($n = 25$). What is *df*?**

38. $df = n_1 + n_2 - 2 = 15 + 25 - 2 = 38$. You subtract 2 — one for each group mean you estimated. (The groups don't need to be the same size.)

5. The three assumptions

The independent *t*-test earns its trustworthy *p* value only if a few background conditions roughly hold. They're not fussy hurdles — they're just the situations where the test's math makes sense.

- **Independence of observations.** Each score comes from a different, unrelated person, and one person's score doesn't influence another's. This is the big one — it's literally what "independent samples" means. Measuring the same people twice and treating them as separate groups breaks it.
- **Roughly normal data.** Within each group, the scores should be approximately bell-shaped — no wild skew or extreme outliers. The test is fairly forgiving here, especially once each group has 30-ish people (a happy consequence of ideas you'll meet with the Central Limit Theorem), but a badly skewed variable is a red flag.
- **Homogeneity of variance.** The two groups should have roughly *equal spread* (similar standard deviations). The basic *t*-test pools the two groups' variability into one estimate, so it assumes that pooling is fair. When the spreads differ a lot, software offers a corrected version (Welch's *t*) that relaxes this assumption.

Why they matter

Break **independence** and the whole *p* value is wrong — that's non-negotiable. Mild violations of **normality** or **equal variance** usually just nudge the result, and good software can correct for unequal spread. Check the assumptions before you trust the verdict, not after.

6. What *p* tells you

The ***p* value** is the chance of seeing a gap this big *if the two groups were actually identical*. A small *p* means "this gap would be a freak accident if nothing were going on," so we believe the difference is real. The line everyone uses is ***p* < .05**. Below it, we call the difference **statistically significant** and we **reject the null hypothesis** (the claim that the two group means are equal).

One precise habit worth building now: when *p* is *not* below your cutoff, you **fail to reject** the null — you do *not* "accept" it or "prove the groups are the same." A non-significant result just means you didn't gather enough evidence to rule out chance; the groups could still differ in a way your study was too small to catch. "Fail to reject" keeps you honest about what the test can and cannot show.

CHECK **Your independent *t*-test returns *p* = .42 — well above .05. The correct conclusion is…**

Fail to reject the null. A large *p* means the gap is easily explained by chance, so we don't have enough evidence to claim a difference. We never "accept" the null or "prove" the groups are equal — absence of evidence isn't evidence of absence.

7. How big? Cohen's *d*

A significant result tells you the difference is *real* — not how *big* it is. For that we use **Cohen's *d***, the gap between the means measured in standard-deviation units. It strips out sample size, so it answers a different question than *p*: not "is there an effect?" but "how much of one?"

Cohen's *d*	Size
around 0.2	small
around 0.5	medium
around 0.8+	large

Quick test

Two questions, two answers. ***p*** asks *is it real?* **Cohen's *d*** asks *how much?* A study can be statistically significant yet tiny (huge samples make even trivial gaps significant), so always report both. *p* says *real*; *d* says *how big*.

8. Worked example — two group means, step by step

Let's run a clean one start to finish. Group A (a campus-life workshop) has mean wellbeing $\bar{x} = 7.0$ with $n = 20$; Group B (no workshop) has mean $\bar{x} = 6.0$ with $n = 20$. Software reports the **standard error of the difference ≈ 0.40**. Predict each step before you reveal it.

Worked example — compare two group means

Group A: $\bar{x} = 7.0$, $n = 20$. Group B: $\bar{x} = 6.0$, $n = 20$. Standard error of the difference ≈ 0.40.

STEP 1 — FIND THE GAP (THE SIGNAL)

Subtract the means: 7.0 − 6.0 = `1.0`. The workshop group averaged one point higher. That gap goes on top of the fraction.

STEP 2 — NOTE THE NOISE (THE STANDARD ERROR)

The standard error of the difference is `0.40`. This is how much two group means would typically bounce around by chance. It goes on the bottom.

STEP 3 — DIVIDE: SIGNAL ÷ NOISE = *T*

t = 1.0 ÷ 0.40 = `2.50`. The gap is two and a half noise-units wide.

STEP 4 — FIND *DF*

$df = n_1 + n_2 - 2 = 20 + 20 - 2 =$ `38`. That tells us which t-curve to compare against.

STEP 5 — COMPARE TO THE CRITICAL VALUE

At $df = 38$, the two-tailed critical t ($p = .05$) is about `2.02`. Our $t = 2.50$ is *bigger*, so it clears the bar.

STEP 6 — DECIDE

Since 2.50 > 2.02, we `reject the null`: the difference is statistically significant. We'd report it as ***t*(38) = 2.50, *p* < .05** — and then add Cohen's d to say how big it is.

Notice what each piece did: the **gap** (1.0) was the signal, the **standard error** (0.40) was the noise, their ratio was ***t*** (2.50), and comparing it to the **critical value** for our ***df*** (38) gave the verdict. Had the standard error been larger — say 0.60 — the same one-point gap would have given only $t \approx 1.67$, which falls *short* of 2.02, and we'd **fail to reject**. Same gap, more noise, different conclusion.

Try it yourself

Here's the live **Independent-Samples t-Test** calculator. Hit **Load class data** to drop in the $1-vs-$20 numbers, then **Run** and read the plain-English result. Paste your own two groups any time — it computes *t*, *df*, *p*, and Cohen's *d* for you.

9. The headline study — $1 vs. $20

One of the most famous experiments in psychology is hiding in our dataset. Festinger and Carlsmith (1959) had people do a deadly-boring task, then paid them to tell the next person it was fun. Half got **$1**; half got **$20**. Afterward everyone rated how much they *actually* enjoyed the boring task. Common sense says the $20 group should like it more — more money, better mood. The opposite happened. (Note the design: two *separate* groups of people, one payment apiece — a textbook independent-samples setup.)

In our data, `task_enjoyment` by payment group came out like this:

Group	Mean enjoyment	*n*
Paid $1	1.42	100
Paid $20	0.04	100

$t(198) = 5.96$, $p < .001$, Cohen's $d = 0.84$ (large).

Notice the *df*: two groups of 100, so $df = 100 + 100 - 2 =$ **198** — exactly the rule from §4. The **$1 group rated the boring task as MORE enjoyable**, and the gap is far too large to be chance. This is

classic **cognitive dissonance**: people paid only a dollar had no good external reason for lying, so to ease the discomfort they convinced *themselves* the task was genuinely fun. The $20 group had a fat excuse ("I did it for the money"), so they felt no need to change their minds. Two separate groups, one mean apiece, and an independent *t*-test tells the whole story.

Report it in APA style

The APA (7th ed.) template for an independent-samples *t* adds one thing to Unit 10's: **both groups' descriptives**, so the reader sees who scored what before the statistics arrive:

```
t(df) = value, p = value, d = value
```

> **The write-up**
>
> "Participants paid $1 rated the task as more enjoyable (*M* = 1.42) than participants paid $20 (*M* = 0.04), *t*(198) = 5.96, *p* < .001, *d* = 0.84."

Same house rules as always: italicized letters, two decimals on statistics, no leading zero on *p*, "*p* < .001" rather than "= .000," and the effect size always rides along. One sentence carries the direction (who scored higher), the means, and the full inferential verdict — a reader who skims only this line still gets the whole finding.

The cow next to the beef

The independent *t*-test isn't just for lab studies. Researchers walked into a real British university cafeteria and made one tiny change to the menu (Murray et al., 2026): next to each meat dish they slipped in a small **photo of the living animal** — a cow beside the beef, a pig beside the pork. Same food, same prices, same line. Two separate sets of diners — one seeing the photos, one not — make this another independent-samples comparison.

Then they counted what people actually ordered. Across **3,674 meal sales**, the odds of choosing a **vegetarian** dish were **22% higher** when the animal photos were on display. Most of us never connect the meat on the plate to the animal it came from — psychologists call that gap the **meat paradox**. The photo breaks the spell: for a moment you *see* the animal, and your willingness to eat it dips.

Our dataset reproduces the study. Each row is a campus dining day; column `menu_condition` tags it as **Animal Photos** or **Standard**, and `plant_based_choice` logs the percent of meals chosen plant-based:

Group	Mean % plant-based	*n*
Animal Photos	45%	100
Standard	37%	100

The result: ***t*(198) = 5.39, *p* < .001, Cohen's *d* = 0.76 (large)**. The animal-photo menu pulled more plant-based orders, the gap is far too big to be chance, and the effect is large — a reminder of where meat comes from can quietly shift behavior, no lecture required. Same test, real-world stakes.

In Google Sheets

You don't compute the standard error by hand — Sheets does it. The two groups usually live in **one** column, tagged by **another** column that says which group each row belongs to. You use `FILTER` to split them apart, then `T.TEST` to compare. In any empty cell:

```
=T.TEST(FILTER(AC2:AC201, AB2:AB201="Animal Photos"),
        FILTER(AC2:AC201, AB2:AB201="Standard"), 2, 2)
```

The `FILTER` pieces pull out each group. The middle **2** means *two-tailed* (we'd accept a difference in either direction). The last **2** is the one that makes it an **independent** test — **type = 2** tells Sheets these are *two independent samples*. (Type 1 would be paired, and type 3 the

unequal-variance version.) The formula returns the ***p* value** directly; if you also want *t* and *df*, the `=TTEST` / `Data ▸ Statistics` add-ons or the calculator above will print them.

Remember the last digit

In `=T.TEST(range1, range2, tails, type)`, the final argument is the test switch: **type = 2 = two independent samples.** That single digit is what separates this unit's test from the paired test next door (type 1).

Stat in the Wild — *t* can be negative

When you run a *t*-test, the number that comes back sometimes has a **minus sign**. Students panic — did I break it? No. A negative *t* just means the *first* group you entered had the lower mean. Swap the two groups and the sign flips, but the **size of *t* and the *p*-value stay exactly the same**.

The lesson: read the *magnitude* of *t* (how far from zero), not its sign. The sign only tells you which group was bigger — which you already knew from the means. Always look at the group averages first, so the sign tells you a story you can already see.

Watch out for

Independent = different people in each group. If the *same* people are measured twice (before/after), that's a different test — the paired *t*-test, next unit. ***df* = $n_1 + n_2 - 2$**, not $n - 1$. ***t* can be negative** (the first group's mean was lower); the size is what matters, not the sign. **Significant ≠ big** — always report Cohen's *d* too. When *p* is high you **fail to reject**, you don't "accept" or "prove" the groups equal. And the groups must be **tagged**: you need a column saying which group each row is in.

Key terms

- **Independent-samples *t*-test**
- **Paired-samples *t*-test**
- **Two-group comparison**
- ***t* value**
- ***p* value**
- **Degrees of freedom ($n_1 + n_2 - 2$)**
- **Standard error of the difference**
- **Independence**
- **Normality**
- **Homogeneity of variance**
- **Statistically significant**
- **Fail to reject**
- **Cohen's *d***
- **Cognitive dissonance**
- **Meat paradox**

PRACTICE — NEVER GRADED

Check yourself

Ten quick true/false items with instant explanations. (Different from your Canvas quiz — real practice.)

Sigma says: Before you pick a test, ask one thing — *are these two groups different people, or the same people twice?* Different people independent *t*-test. The whole choice hinges on that one question.