UNIT 14 · BLOCK 4

Two-Way ANOVA

Two factors at once — and the famous picture that reads itself.

Why this matters

This is the finale — you've earned it. All year you've asked, "does **one** thing matter?" Now you can ask, "do **two** things matter — and do they matter *together*?" That last word is the whole game. The real world almost never lets one cause act alone: a drug works differently by age, a teaching method helps beginners but not experts, a stressor lands hard on one group and bounces off another. A two-way ANOVA is built to catch exactly that. By the end of today you'll look at a single **line graph** and say, out loud, what **three** different results are hiding inside it. One picture, three answers. Let's finish strong.

By the end you can…

- Explain that a **two-way ANOVA** crosses **two factors** and returns **three** tests — two main effects and one interaction.
- Tell a **cell mean** from a **marginal mean**, and define a **main effect** and an **interaction** in plain English.
- **Read an interaction graph** — the single most important skill in this unit.

Lay out the four **cell means** in a 2×2 table, add the **marginal means**, and judge all three results yourself.

1. The big idea — two factors, three tests

Back in Unit 13, one-way ANOVA let you compare three or more groups on **one** grouping variable — that variable is called a **factor**. A **two-way ANOVA** takes the natural next step: it studies **two factors at the same time**. The word "two-way" literally means "two factors." A **2×2 design** (say it "two-by-two") is the simplest case — **two factors, each with two levels**. Two levels times two levels gives **four groups**, one for every combination.

Here's the payoff and the twist. Because you're crossing two factors, you don't get one answer — you get **three**, and each one gets its own *F* test:

Result (its own *F*-test)	Plain English	The question it answers
Main effect of Factor A (the rows)	The overall effect of Factor A, averaging over Factor B	Does Factor A move scores *overall*?
Main effect of Factor B (the columns)	The overall effect of Factor B, averaging over Factor A	Does Factor B move scores *overall*?
A × B interaction	The two factors *team up*	Does the effect of one **depend on** the level of the other?

Two main effects plus one interaction — **three separate *F* ratios**, each tested for significance on its own. That third one, the interaction, is new this unit, and it's the star of the show. Crucially, the interaction is *not* "the two main effects added together" — it's a genuinely separate question that the first two tests can't answer.

Why one test isn't enough

You could run two separate one-way ANOVAs — one per factor — but you'd *never see the interaction*. The interaction is the part that says "the effect of one factor changes depending on the other," and it only appears when both factors live in the same analysis. That's the whole reason to run a two-way ANOVA instead of two one-ways.

CHECK **A study crosses two factors (each with two levels) in one analysis. How many *F*-tests does the two-way ANOVA produce?**

Three. Two factors give two main effects (one per factor), plus the interaction between them — three separate *F* tests, each judged on its own. That third test, the interaction, is the reason you run a two-way ANOVA instead of two one-ways.

2. Cell means and marginal means — the raw material

Everything in a two-way ANOVA is built from averages of a 2×2 grid. Two kinds of averages matter, and keeping them straight is most of the battle.

- A **cell mean** is the average of **one combination** of the two factors — one inner box of the grid. A 2×2 has **four cell means**.
- A **marginal mean** sits in the *margin* — the edge — of the table. A **row marginal** averages across a whole row (collapsing the column factor); a **column marginal** averages down a whole column (collapsing the row factor). The marginals are exactly what the **main effects** compare.

Here's the layout. The four inner numbers are **cell means**; the bold edge numbers are **marginal means**:

	B_1 (Col 1)	B_2 (Col 2)	Row marginal
A_1 (Row 1)	cell A_1B_1	cell A_1B_2	**mean of row A_1**
A_2 (Row 2)	cell A_2B_1	cell A_2B_2	**mean of row A_2**
Col marginal	**mean of col B_1**	**mean of col B_2**	**grand mean**

Read it like a map. The **main effect of Factor A** compares the two **row marginals** (far-right column). The **main effect of Factor B** compares the two **column marginals** (bottom row). The **interaction** lives *inside* the four cells — it asks whether the gap between cells changes from one row to the next. Same four numbers, three different questions.

3. Main effects — the "overall" effects

A **main effect** is the effect of **one** factor, **averaging over** the other. It's the marginal-means comparison, in words:

- **Main effect of Factor A (the rows):** ignore Factor B for a second. Lump everyone together by Factor A only — that's the two row marginals. Do they differ? If yes, Factor A moved scores overall.
- **Main effect of Factor B (the columns):** ignore Factor A. Lump everyone together by Factor B only — the two column marginals. Do *those* differ? If yes, Factor B moved scores overall.

Main effects are the *big picture* — one factor at a time, the other one blurred out. Each gets its own *F* ratio (the same logic as Unit 13: between-groups variance over within-groups variance), and you decide each one with the familiar rule — **if $p < .05$, reject the null** for that effect; otherwise **fail to reject** it. "Fail to reject" — never "accept" — because absence of evidence isn't evidence of absence.

4. The interaction — the "it depends" effect

An **interaction** means the effect of one factor **depends on the level of the other**. That's the whole definition: **it depends**. The factors aren't acting alone — they're teaming up. In the table, you spot it by asking whether the gap between the two cells in one row is *different* from the gap in the other row. If Factor B helps a lot in Row 1 but barely at all in Row 2, B's effect **depends on** A — that's an interaction.

> **The one definition to keep**
>
> **An interaction means the effect of one factor DEPENDS on the other.** A study tip might help *beginners* a lot but do nothing for *experts* — the tip's effect **depends on** who you are. That dependence *is* the interaction.

And here's the subtle part that trips people up: **an interaction can hide a main effect.** If one group's scores rise while another's fall by about the same amount, the two changes cancel when you average them — so the main effect can look flat (non-significant) even though something dramatic is happening. The story isn't "nothing happened." The story is "it depends," and only the interaction term tells you that.

5. Reading the interaction graph — the key skill

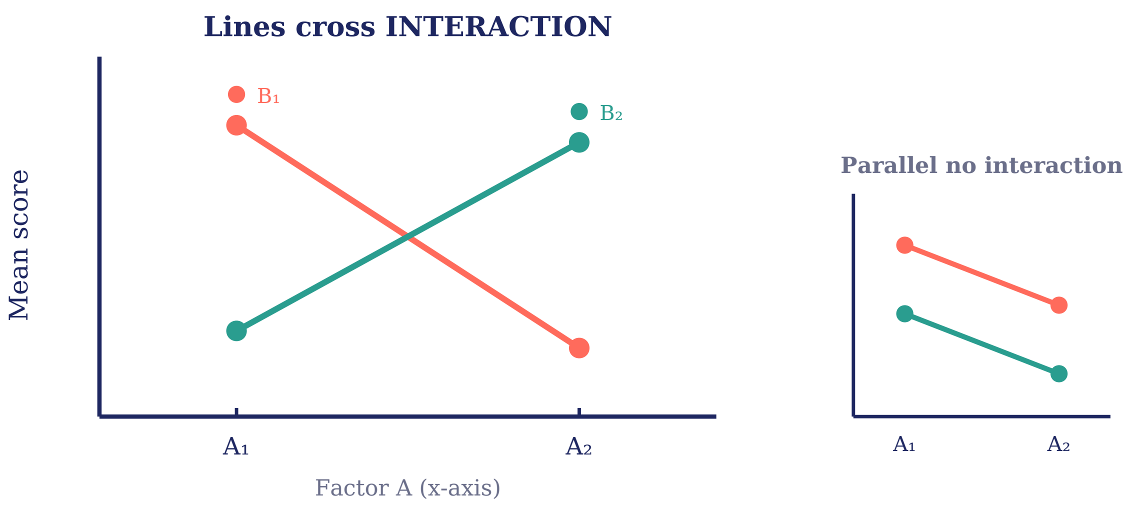
Here's the trick that makes all three results pop out of **one picture**. Put one factor on the **x-axis** and draw **one line per group** of the other factor. Each plotted point is a **cell mean**. Now read it in three quick looks:

1. **Main effect of the x-axis factor?** Do the lines, *on average*, **slope up or down** as you move across the x-axis? If they slope, that factor moved scores overall.

2. **Main effect of the "lines" factor?** Is **one line generally higher** than the other? If yes, that group scored higher overall.
3. **Interaction?** Are the lines **non-parallel**? **Parallel lines = no interaction.** Lines that **cross or fan apart = an interaction.**

> **Burn this in**
>
> The interaction is the *only* thing you read from whether the lines are **parallel**. **Parallel no interaction. Not parallel (crossing or fanning) interaction.** That one sentence is most of this unit. Slopes tell you about main effects; *parallelness* tells you about the interaction.

Here's the picture itself. On the left, the two lines **cross** — the effect of Factor B flips direction depending on Factor A. That non-parallelness *is* the interaction. On the right, for contrast, two perfectly **parallel** lines: there's a gap (a main effect) but the lines stay in step, so **no interaction**.



Each dot is a cell mean. Crossing or fanning lines signal an interaction; parallel lines do not.

CHECK **On an interaction line graph, non-parallel lines (crossing or fanning apart) indicate…**

An interaction. Non-parallel lines — whether they cross or just fan apart — mean the effect of one factor depends on the level of the other. *Parallel* lines are the ones that signal no interaction. Parallelness is the only thing that tells you about the interaction; slopes and heights tell you about the main effects.

6. The centerpiece — the Two-Way ANOVA calculator

Here's the grand-finale calculator, embedded right here. Hit **"Load class data (gender × threat)"** to drop in our stereotype-threat study, then **Run**. You'll get all three *F* tests at once — the two main effects and, the one to watch, the interaction. The worked example below walks you through exactly what you'll see.

Worked example — read a 2×2 table cold

Before the famous study, let's build the skill with clean numbers. Imagine a memory experiment crossing **Factor A = study method** (Massed vs. Spaced) with **Factor B = test delay** (Immediate vs. Delayed). The outcome is the number of words recalled. Here are the four **cell means** — predict each step before you reveal it.

	Immediate (B_1)	Delayed (B_2)
Massed (A_1)	30	20
Spaced (A_2)	30	40

Worked example — judge all three results from the four cell means

Two factors, two levels each. Compute the marginals, then call the two main effects and the interaction.

STEP 1 — ROW MARGINALS (FACTOR A)

Average across each row. Massed: (30 + 20) ÷ 2 = `25`. Spaced: (30 + 40) ÷ 2 = `35`. These two numbers are the main effect of **study method**.

STEP 2 — COLUMN MARGINALS (FACTOR B)

Average down each column. Immediate: (30 + 30) ÷ 2 = `30`. Delayed: (20 + 40) ÷ 2 = `30`. These two are the main effect of **test delay**.

STEP 3 — MAIN EFFECT OF FACTOR A?

Row marginals are `25 vs 35` — they differ by 10. **Yes**, a main effect of study method: spaced practice recalls more *overall*.

STEP 4 — MAIN EFFECT OF FACTOR B?

Column marginals are `30 vs 30` — identical. **No** main effect of delay overall… but hold that thought, because the cells tell a different story.

STEP 5 — INTERACTION?

Is A's effect the same across B's levels? For *Immediate*, MassedSpaced goes 3030 (gap = 0). For *Delayed*, 2040 (gap = +20). The gaps differ, so **yes — an interaction.** Spacing only helps on the delayed test. The flat column marginal in Step 4 *hid* this: an interaction can mask a main effect.

CHECK

In that memory study, the two column marginals for test delay were both 30 — yet delay clearly mattered for the spaced group. What does that tell you?

An interaction hid the effect. The delay helped one method and hurt the other by about the same amount, so the two changes canceled in the marginal average. The flat main effect doesn't mean "nothing happened" — it means "it depends," which is exactly what the interaction term catches.

The textbook interaction — stereotype threat

Now the example to remember for the exam. In a classic line of research (Spencer, Steele, & Aronson-style), people take a hard math test. Half are quietly **reminded** beforehand of a negative stereotype about their group's math ability — that's the **threat** condition — and half are not — the **control** condition. The two factors are **gender** (Woman, Man) — our row factor, shown as the two lines — and **threat condition** (Control, Threat) — our column factor, on the x-axis; the outcome is the **math score**.

Here are the four **cell means** — the average math score for each combination — from our N = 200, with the **marginal means** added on the edges:

	Control	Threat	Row marginal
Woman	76.4	68.9	**72.7**
Man	79.2	80.6	**79.9**
Col marginal	**77.8**	**74.8**	**76.3**

Now picture the line graph — put **condition on the x-axis** (Control Threat) and draw **one line per gender**:

- The **Women's line drops** sharply (76.4 68.9). Under threat, women's scores fall about 7.5 points.
- The **Men's line stays flat** — even ticks up a hair (79.2 80.6). The reminder does nothing to men.

The two lines are **not parallel** — they fan apart. **That non-parallelness *is* the interaction.** The effect of the threat reminder **depends on** gender: it hurts women's scores but leaves men's untouched. The marginals tell the "overall" story (men higher overall, 79.9 vs 72.7; a slight dip under threat overall, 77.8 vs 74.8) — but if you'd *only* looked at those marginals and ignored the cells, you'd badly miss the real headline.

The three results, read off the picture

Interaction: F(1, 196) = 11.1, p = .001 — the lines really do fan apart; threat's effect depends on gender, so we **reject** the null of no interaction. **Main effect of gender (rows): F = 29.3, p < .001** — the Men's row marginal sits higher (79.9 vs 72.7). **Main effect of condition (columns): F = 5.4, p = .02** — averaging the columns, scores dip a little under threat (77.8 74.8). The interaction is the headline: *this is THE graph for reading an interaction.*

How big? Partial eta-squared

A two-way ANOVA reports **three** effect sizes — one per test. The standard choice is **partial η^2** (partial eta-squared): each effect's SS divided by that SS plus the error SS. Read it just like Unit 13's η^2 — the share of variance an effect accounts for, with the other effects set aside:

```
partial η² = SS_effect ÷ (SS_effect + SS_within)
```

For our stereotype-threat data: gender partial $\eta^2 \approx$ **.13**, condition $\approx$ **.03**, interaction $\approx$ **.05**. Common benchmarks: ~.01 small, ~.06 medium, ~.14 large. So the gender gap is a large effect, the overall threat dip is small, and the interaction — the headline — lands at the small-to-medium boundary: real and meaningful, but subtler than the marginal gender gap. The calculator above prints partial η^2 for all three effects.

Report it in APA style

A two-way ANOVA gets **three APA sentences** — one per test — and the interaction, when significant, leads the story. Same house rules as Units 10–13 (italicized letters, two decimals, no leading zero on *p*, effect sizes included):

The write-up

"A 2 × 2 ANOVA revealed a significant gender × condition interaction, $F(1, 196) = 11.10$, $p = .001$, partial $\eta^2 = .05$: women scored lower under threat ($M = 68.9$) than in the control condition ($M = 76.4$), whereas men's scores did not differ (*M*s = 80.6 and 79.2). There were also main effects of gender, $F(1, 196) = 29.30$, $p < .001$, partial $\eta^2 = .13$, and condition, $F(1, 196) = 5.40$, $p = .021$, partial $\eta^2 = .03$."

The anatomy: interaction first (it's the finding), immediately unpacked with the cell means that show *what depends on what*, then the main effects in a supporting sentence. When an interaction is significant, the main effects are footnotes to the "it depends" story — the write-up's order should say so.

7. Why the graph beats the table

You could stare at the four numbers and eventually work it out. But the line graph hands you all three results in a single glance. Train your eye on the three questions, in order, and they become automatic:

What you look at	What it tells you
Do the lines slope, on average?	Main effect of the **x-axis factor**
Is one line higher than the other?	Main effect of the **"lines" factor**
Are the lines parallel, or not?	The **interaction** (parallel = none; crossing/fanning = yes)

An interaction in one sentence

An interaction means the effect of one factor DEPENDS on the other. In the picture, that dependence shows up as lines that aren't parallel — women feel the threat, men don't, so the two lines refuse to stay in step.

CHECK

On an interaction graph, you notice that the red line sits well above the blue line at both x-positions, and the two lines are perfectly parallel. What can you conclude?

A main effect of the "lines" factor, but no interaction. **One line consistently higher = a main effect of the factor drawn as lines. Parallel = no interaction. The two readings are independent: height gives a main effect, parallelness gives the interaction.**

In Google Sheets

Sheets has **no one-click two-way ANOVA**, so we do the part that matters: the four **cell means** and the **interaction graph**. Use `=AVERAGEIFS`, which says "average the score column, but *only* the rows where gender matches **and** condition matches":

```
=AVERAGEIFS(score, gender, "Woman", cond, "Control")   76.4
=AVERAGEIFS(score, gender, "Woman", cond, "Threat")    68.9
=AVERAGEIFS(score, gender, "Man",   cond, "Control")   79.2
=AVERAGEIFS(score, gender, "Man",   cond, "Threat")    80.6
```

For the **marginal means**, drop one condition and average by gender alone (and vice versa):

```
=AVERAGEIF(gender, "Woman", score)    72.7   (Woman row marginal)
=AVERAGEIF(cond,   "Threat", score)   74.8   (Threat column marginal)
```

Lay the four cell means in a little 2×2 table, then **Insert ▸ Chart ▸ Line chart** with **condition on the x-axis** and **one line per gender**. That picture is your interaction graph — you'll watch the Women's line dive while the Men's line holds flat. Read the two main effects and the interaction right off it. The three *F*-values come from the provided template (or this page's calculator above).

Watch out for

Parallel = no interaction. Don't overthink it — the interaction lives in the *parallelness* of the lines, nothing else. **An interaction can hide a main effect, and that's fine** — if two opposite slopes cancel, the overall average looks flat even though "it depends" is the real story. **Lines crossing OR fanning apart both count** as an interaction. Decide each of the three tests with $p < .05$, and say "**fail to reject**" — never "accept" — when a test comes up non-significant. And the headline of an interaction is always "*it depends*," never a flat one-size-fits-all claim.

Key terms

- Two-way ANOVA
- Factor
- 2×2 design
- Level
- Main effect
- Interaction
- Cell mean
- Marginal mean
- Grand mean
- Crossover interaction
- Parallel lines
- Line graph
- Stereotype threat

You made it

That's the course. You started by squeezing 200 numbers into one average. You're finishing by reading two factors and their teamwork off a single line. The whole point of a two-way ANOVA is that one skill — spotting the interaction — so keep practicing reading those graphs until it's automatic. Well done.

Still hungry? There's one tool this course never needed but published research uses constantly: the test for *categories and counts*. Bonus Unit 15 — Chi-Square Tests is short, won't appear on Exam 4, and rounds out your toolkit.

INTERACTIVE · THE WHOLE SKILL

Train your eye — read the graph

A line graph appears; you call the main effect of the **ROW variable** (the lines), the **COLUMN variable** (the x-axis), and the interaction. It cycles through **all 8 possible patterns** — and "Show all 8 patterns" lays out the entire cheat-sheet at once.

PRACTICE — NEVER GRADED

Check yourself

Ten quick true/false items with instant explanations. The finale — real practice, never graded.

> **Sigma says:** When you see a two-way line graph, ask the three questions in order — do the lines slope, is one higher, are they parallel? The first two are main effects; the last one is the interaction. Parallel means no interaction. That's the whole unit in your pocket.