UNIT 15 · BLOCK 4 · BONUS

Chi-Square Tests

For when your data are categories, not numbers.

Why this matters

Every test you've learned — *t*, *r*, *F* — needs a **mean**. But some of the most interesting variables have no mean at all: diet, major, yes-or-no, on-campus vs. off. You can't average "vegan." What you *can* do is **count** — and the **chi-square (χ²) test** is the tool built for counts. This bonus unit adds the last piece to your toolkit: one test for "do these counts match a claim?" and one for "are these two category variables related?" Same hypothesis-testing rhythm you already know; brand-new kind of data.

By the end you can…

- Recognize when data are **categorical counts** — and why that rules out every mean-based test you know.
- Run a **goodness-of-fit** test: build **expected counts** from a claimed distribution and compare them to what you observed.
- Compute **$\chi^2 = \Sigma\,(O - E)^2 \div E$** and decide with the right ***df***.
- Run a **test of independence** on a 2×2 table and say, in plain English, whether two category variables are related.

Check the **rule of 5**, and **report a χ^2 result in APA style**.

1. A new kind of data — counts in categories

Look back at what every test in this course has eaten for breakfast: **means**. The *t*-tests compared means. ANOVA compared means. Even correlation needed two columns of numbers you could average. That works because sleep hours, stress scores, and compassion ratings are **quantitative** — averaging them makes sense.

But plenty of variables are **nominal** — pure categories, straight from Unit 1: diet (omnivore / vegetarian / vegan), gender, class year, commuter vs. resident, "did they choose the plant-based option — yes or no?" There is no such thing as the *mean diet*. All you can do with a nominal variable is **count how many people land in each box** — and it turns out that's enough. The **chi-square test** asks one question, in two flavors:

- **Goodness of fit** (one variable): do my observed counts match the counts some claim *expected*? *(Is our class's diet mix like the nation's?)*
- **Test of independence** (two variables): are two categorical variables *related*, or independent? *(Does diet depend on gender?)*

The fix in one line

No means? Count the boxes. Chi-square compares the counts you *got* to the counts you'd *expect* if nothing interesting were going on — and asks whether the mismatch is bigger than chance.

CHECK **A researcher wants to know whether preferred social-media platform (TikTok / Instagram / YouTube) differs across her sample. Why can't she run an ANOVA?**

Platform is nominal. ANOVA compares group *means* on a quantitative outcome. "Preferred platform" has no mean — you can only count how many people fall in each category. Counts in categories are exactly chi-square's job.

2. Goodness of fit — do my counts match the claim?

Here's a real question about our own class. National surveys put the U.S. adult diet mix at roughly **90% omnivore, 7% vegetarian, 3% vegan**. Our **SJNY Student Life** dataset has 200 students. If our class were just a miniature copy of the nation, the 200 students should split about the same way. Do they?

Diet	Observed (our class)	National claim
Omnivore	136	90%
Vegetarian	46	7%
Vegan	18	3%

That's the **goodness-of-fit** setup: one categorical variable, a set of **observed counts**, and a claimed distribution to test them against. The null hypothesis is that the claim is true — our class *fits* the national mix, and any mismatch is sampling noise. The alternative: the class genuinely differs from the claim.

3. Expected counts — what the claim predicts

To compare counts to a claim, first turn the claim into counts. Each **expected count** is just the claimed percentage applied to your *N*:

```
E = (claimed proportion) × N

Omnivore:    .90 × 200 = 180
Vegetarian:  .07 × 200 =  14
Vegan:       .03 × 200 =   6
```

Line them up and the drama is already visible — we observed **46** vegetarians where the nation predicts **14**, and **18** vegans where it predicts **6**:

Diet	Observed (O)	Expected (E)
Omnivore	136	180
Vegetarian	46	14
Vegan	18	6

Two things to notice

First, **expected counts don't have to be whole numbers** — an E of 14.5 people is fine; it's a long-run average, not a head count. Second, **the E's must add to the same *N* as the O's** (180 + 14 + 6 = 200). If they don't, your percentages were off before the test even started.

4. The χ^2 statistic — score the mismatch

Now one number measures how badly the observed counts miss the expected ones. For each category: take the miss (O − E), square it (so misses in either direction count, and big misses count a lot), and divide by E (a miss of 10 matters more when you expected 6 than when you expected 180). Then add up the categories:

$$\chi^2 = \Sigma\ (O - E)^2 \div E$$

Worked example — is our class a mini-America?

Observed: 136 / 46 / 18. Expected: 180 / 14 / 6. One $(O - E)^2/E$ per category, then sum. Predict each step before revealing it.

STEP 1 — OMNIVORES

$(136 - 180)^2 \div 180 = (-44)^2 \div 180 = 1936 \div 180 \approx$ `10.8`.

Fewer omnivores than the nation predicts.

STEP 2 — VEGETARIANS

$(46 - 14)^2 \div 14 = 32^2 \div 14 = 1024 \div 14 \approx$ `73.1`. A huge contribution — more than three times the expected count showed up.

STEP 3 — VEGANS

$(18 - 6)^2 \div 6 = 12^2 \div 6 = 144 \div 6 =$ `24.0`. Triple the expected count here too.

STEP 4 — ADD THEM UP

$\chi^2 = 10.8 + 73.1 + 24.0 \approx$ `107.9`. One number scoring the total mismatch across all three categories.

STEP 5 — COMPARE AND DECIDE

With $df = k - 1 = 2$, the critical χ^2 at $\alpha = .05$ is `5.99`. Our 107.9 obliterates it ($p < .001$) — `reject` the null. Our class is *not* a mini-America: it has far more vegetarians and vegans than the national mix predicts.

Worth pausing on *why* that's plausible rather than alarming: a psychology statistics course at a New York college is not a random

draw of American adults. Selection tells the story — and the test just told us, precisely, that selection is real here. To the extent that your sample isn't drawn like the claim's population, goodness of fit will notice.

5. Degrees of freedom & the decision

Chi-square's *df* comes from **categories, not people** — a common trip-up. With the totals fixed, only so many cells are free to vary:

Test	Formula	Our examples
Goodness of fit	$df = k - 1$	3 diets $df = 2$
Independence	$df = (\text{rows} - 1)(\text{cols} - 1)$	2×2 table $df = 1$

The decision works exactly like *F*: χ^2 can't go below zero (every term is squared), so all the action is in the **upper tail**. Small χ^2 = counts near expectations = nothing to report. Big χ^2 = counts far from expectations = reject. The usual critical values you'll bump into: **3.84** for $df = 1$ and **5.99** for $df = 2$, at $\alpha = .05$. And say it the way you've said it all semester: "**fail to reject**," never "accept."

The fine print — the rule of 5

Chi-square is trustworthy when every **expected** count is at least about **5**. (Observed counts can be anything — it's the E's that matter.) Small expected counts make the χ^2 approximation shaky — the fix is bigger samples or merged categories. Our vegan E of exactly 6 squeaks by; the calculators below flag any violation for you. And one more habit: chi-square needs raw **counts**, never percentages — feeding it "68%" instead of "136 of 200" breaks the arithmetic.

CHECK **A goodness-of-fit test compares observed counts across 5 majors for 300 students. What is the *df*?**

***df* = 4. Chi-square's df comes from the number of *categories*: k − 1 = 5 − 1 = 4. The 300 students set the expected counts, but they don't set the df — that's the classic trip-up.**

6. The test of independence — are two category variables related?

The second flavor crosses **two** categorical variables. Here's a question with a stereotype riding on it: *is diet related to gender?* Pop culture says yes — meat is coded masculine (remember Unit 6's `masculinity` × `meat_commitment` correlation of +.35), so you might expect the women in our class to be likelier than the men to skip meat. Count everyone into a **2×2 contingency table** (diet collapsed to omnivore vs. "veg" — vegetarian or vegan; the 20 non-binary students form too small a group for stable expected counts here, so we compare the 180 men and women):

Observed	Omnivore	Veg (any)	Total
Man	54	26	**80**
Woman	68	32	**100**
Total	**122**	**58**	**180**

The null hypothesis: the two variables are **independent** — knowing someone's gender tells you nothing about their diet. Independence makes a precise prediction for every cell, built from the margins: **E = (row total × column total) ÷ N**. Men-omnivore: 80 × 122 ÷ 180 ≈ **54.2**. Compare that to the observed 54… and you can already feel where this is going. Every observed count sits almost exactly on its expected count:

```
χ²(1, N = 180) = 0.01, p = .94
```

The verdict — and the lesson

Fail to reject. In our class, diet and gender are as close to perfectly independent as real data ever get — men and women skip meat at essentially the same rate (32.5% vs. 32.0%). The stereotype made a testable prediction; the counts shrugged. That's the discipline working exactly as designed: a non-significant result is *informative* when it contradicts a confident cultural claim. (One class of 200 isn't the last word — but notice that "everyone knows" wasn't evidence.)

CHECK

A test of independence on a 2×2 table gives $\chi^2(1, N = 180) = 0.01$, $p = .94$. The right conclusion is…

Fail to reject. The observed counts sit almost exactly where independence predicts, so there's no evidence of a relationship. But as always, we never *prove* the null — we simply found nothing that contradicts it.

Try it yourself — the two chi-square calculators

Both flavors, live on the page. Hit **Load class data** in each to reproduce this unit's two examples — the diet mix that crushed the national claim, and the gender × diet table that didn't budge.

Report it in APA style

Chi-square's APA (7th ed.) sentence has one twist: because the *df* alone doesn't reveal the sample size (it comes from categories, remember), the ***N* rides inside the parentheses**:

```
χ²(df, N = total) = value, p = value
```

The write-ups — one significant, one not

"The class's diet distribution differed significantly from the national distribution, $\chi^2(2, N = 200) = 107.90$, $p < .001$, with vegetarians and vegans overrepresented relative to national rates."

"Diet was not significantly related to gender, $\chi^2(1, N = 180) = 0.01$, $p = .943$."

Same house rules as Units 10–14: italicized letters, two decimals on the statistic, no leading zero on *p*, "< .001" instead of "= .000," and a plain-English clause saying *what the numbers mean* — overrepresented where, related to what. A statistics sentence with no story attached is only half a result.

In Google Sheets

Sheets speaks fluent chi-square. Get your observed counts with `COUNTIF`, build the expected counts with plain arithmetic, and one function returns the *p*-value:

```
=COUNTIF(X2:X201, "Vegan")              observed count per category
=0.03 * 200                              expected count per category (claim × N)
=CHISQ.TEST(obs_range, exp_range)        the p-value, directly
```

For the test of independence, lay out the 2×2 **observed** table, build the matching **expected** table with `= row_total * col_total / N` in each cell, and point `CHISQ.TEST` at both ranges. If you have the χ^2 statistic and want the *p* (or vice versa):

```
=CHISQ.DIST.RT(107.9, 2)    p for a given χ² and df
=CHISQ.INV.RT(0.05, 2)      the critical χ² (5.99 for df = 2)
```

(SPSS note for your future research-methods self: **Analyze ▸ Descriptive Statistics ▸ Crosstabs**, tick *Chi-square* under Statistics — the whole independence test, expected counts included, in one dialog.)

Watch out for

Counts, never percentages or means — chi-square eats whole-number tallies of people in categories. ***df* comes from categories**, not people: k − 1 for goodness of fit, (r − 1)(c − 1) for independence — and the *N* goes inside the APA parentheses precisely because df won't reveal it. **Check the rule of 5** on the *expected* counts. χ^2 is never negative, so it's always an upper-tail test. A significant independence test says the variables are **related** — it doesn't say which direction or how strongly, and (chorus, everyone) it doesn't say *why*: relatedness isn't causation. And when χ^2 comes up small, say "**fail to reject**" — the counts matching the claim is not proof the claim is true.

Key terms

Chi-square (χ^2) · **Nominal / categorical data** · **Observed count (O)** · **Expected count (E)** · **Goodness of fit** · **Test of independence** · **Contingency table** · **Rule of 5** · **df = k − 1** · **df = (r − 1)(c − 1)** · **Upper-tail test** · **Independence**

Check yourself

Ten quick true/false items with instant explanations. Real practice — this bonus unit won't appear on Exam 4, but these ideas turn up constantly in published research.

Sigma says: No mean? No problem. **Count the boxes, predict the boxes, score the mismatch:** $\chi^2 = \Sigma(O - E)^2/E$. One flavor checks your counts against a claim; the other checks whether two category variables are related. Everything else — null, *df*, critical value, reject or fail to reject — you already knew before you got here.