Practical Class

The Chi-Square Test of Independence and Cramér's V

chief assist. prof. Kostadin Kostadinov, MD, PhD, MPH, MEcon

Academic Year 2025/2026

Department of "Social Medicine and Public Health"

Outline

1. When and Why
2. Example 1 — Smoking and Low Birth Weight
3. Cramér's V — Effect Size
4. Example 2 — Treatment Regimen and *H. pylori* Eradication
5. Practical Guidance

When and Why

The Problem

In medicine we frequently ask: **is there an association between two categorical variables?**

- Does smoking status relate to low birth weight?
- Does treatment regimen relate to eradication success?
- Does blood group relate to infection risk?

When **both** variables are categorical (nominal or ordinal), we cannot use correlation or t-tests. We need a test designed for **frequency data** arranged in a contingency table.

The **Pearson chi-square test of independence** (χ^2) answers exactly this question.

The Contingency Table

A contingency table cross-classifies observations by two categorical variables.

	Category B_1	Category B_2	Row total
Category A_1	O_{11}	O_{12}	R_1
Category A_2	O_{21}	O_{22}	R_2
Column total	C_1	C_2	N

Each cell contains the observed frequency (O) — the count of individuals who fall into that combination of categories. Row totals (R_i), column totals (C_j), and the grand total (N) are the marginal totals.

The Core Logic

The chi-square test compares what we actually observe with what we would expect if the two variables were independent.

Step 1 — Calculate expected frequencies under the null hypothesis of independence:

$$E_{ij} = \frac{R_i \times C_j}{N}$$

The expected count for any cell is the product of its row total and column total, divided by the grand total.

Intuition: if smoking and birth weight are truly unrelated, the proportion of low-birth-weight babies should be the **same** among smokers and non-smokers.

The Core Logic (continued)

Step 2 — Measure the discrepancy between observed (O) and expected (E) in every cell:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

Each cell contributes $\frac{(O-E)^2}{E}$ — a large positive value when the observed count departs substantially from the expected count.

The Core Logic (continued)

Step 3 — Degrees of freedom:

$$\mathrm{df} = (r - 1) \times (c - 1)$$

where r = number of rows and c = number of columns.

Step 4 — Compare the computed χ^2 to the critical value from the χ^2 distribution at the chosen significance level (α).

Decision Rule

- H_0: The two variables are **independent** (no association).
- H_1: The two variables are **not independent** (association exists).

If $\chi^2_{\text{calc}} \geq \chi^2_{\text{crit}} \Rightarrow$ Reject H_0 — evidence of an association.

If $\chi^2_{\text{calc}} < \chi^2_{\text{crit}} \Rightarrow$ Fail to reject H_0 — no evidence of an association.

df	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
1	3.841	6.635	10.828
2	5.991	9.210	13.816
3	7.815	11.345	16.266
4	9.488	13.277	18.467
5	11.070	15.086	20.515

Assumptions

The chi-square test requires several conditions to yield valid results:

- Independence of observations — each individual contributes to only one cell of the table.
- Adequate expected frequencies — all expected cell counts should be ≥ 5. When this condition is violated, consider Fisher's exact test (for 2×2 tables) or collapsing categories.
- Random sampling — data should come from a random or representative sample.
- The test uses categorical (nominal or ordinal) variables — not continuous measurements.

Caution: the χ^2 test detects **association**, not **causation**. A significant result does not establish a causal link.

Example 1 — Smoking and Low Birth Weight

Example 1 — Setting

A maternity hospital records data on **200 singleton pregnancies**. For each mother, two variables are recorded:

- Smoking status during pregnancy — smoker vs non-smoker
- Birth weight — low birth weight (< 2 500 g) vs normal birth weight (≥ 2 500 g)

Research question: Is there a statistically significant association between maternal smoking during pregnancy and low birth weight?

H_0: Smoking status and birth weight are independent.

H_1: Smoking status and birth weight are not independent.

Significance level: $\alpha = 0.05$

Step 1 — Observed Frequencies

	Low BW	Normal BW	Total
Smoker	30	50	**80**
Non-smoker	15	105	**120**
Total	**45**	**155**	200

Among the 80 smokers, 30 delivered a low-birth-weight infant (37.5%).

Among the 120 non-smokers, 15 delivered a low-birth-weight infant (12.5%).

These proportions look different — but is the difference large enough to rule out chance?

Step 2 — Expected Frequencies

Under H_0 (independence), each expected cell count is:

$$E_{ij} = \frac{R_i \times C_j}{N}$$

$$E_{\text{smoker, low BW}} = \frac{80\times45}{200} = 18.0$$

$$E_{\text{smoker, normal BW}} = \frac{80\times155}{200} = 62.0$$

$$E_{\text{non-smoker, low BW}} = \frac{120\times45}{200} = 27.0$$

$$E_{\text{non-smoker, normal BW}} = \frac{120\times155}{200} = 93.0$$

✓ All expected frequencies ≥ 5 — the χ^2 test is valid.

Step 2 — Expected Frequencies (table)

	Low BW	Normal BW	Total
Smoker	30 (18.0)	50 (62.0)	80
Non-smoker	15 (27.0)	105 (93.0)	120
Total	45	155	200

Blue values = expected frequencies under H_0.

Notice: we observe 30 low-BW infants among smokers, but only 18 were expected. Among non-smokers we observe 15, but 27 were expected. The discrepancies are substantial.

Step 3 — Compute χ^2

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Cell	O	E	$(O - E)^2$	$\frac{(O-E)^2}{E}$
Smoker, Low BW	30	18.0	144.0	8.000
Smoker, Normal BW	50	62.0	144.0	2.323
Non-smoker, Low BW	15	27.0	144.0	5.333
Non-smoker, Normal BW	105	93.0	144.0	1.548

$$\chi^2 = 8.000 + 2.323 + 5.333 + 1.548 = 17.204$$

Step 4 — Degrees of Freedom and Decision

$$\text{df} = (r-1)(c-1) = (2-1)(2-1) = 1$$

Critical value at $\alpha = 0.05$, $\text{df} = 1$: $\quad \chi^2_{\text{crit}} = 3.841$

$$\chi^2_{\text{calc}} = 17.204 \quad \gg \quad \chi^2_{\text{crit}} = 3.841$$

Decision: Reject H_0.

There is a statistically significant association between maternal smoking during pregnancy and low birth weight ($\chi^2 = 17.20$, $\text{df} = 1$, $p < 0.001$).

The result is significant even at $\alpha = 0.001$ (critical value $= 10.828$).

Cramér's V — Effect Size

Why Effect Size Matters

A significant χ^2 tells us **that** an association exists — but not **how strong** it is.

- The χ^2 statistic is influenced by sample size: with a large enough sample, even trivial associations become statistically significant. A study of 100 000 patients might find $p < 0.001$ for a clinically negligible association.
- We need a measure that quantifies the strength of association independently of sample size.

Cramér's V provides exactly this — a standardised effect size for contingency tables of **any** dimension.

Cramér's V — Formula

$$V = \sqrt{\frac{\chi^2}{N \times (k-1)}}$$

where:

- χ^2 = the computed chi-square statistic
- N = total sample size
- $k = \min(r, c)$ — the smaller of the number of rows or columns

Cramér's V ranges from 0 (no association) to 1 (perfect association).

For a **2×2 table**, $k - 1 = 1$, so $V = \sqrt{\frac{\chi^2}{N}}$, which equals the absolute value of the phi coefficient ($|\varphi|$).

Interpreting Cramér's V

Conventional benchmarks depend on the degrees of freedom ($k - 1$):

Effect size	$k - 1 = 1$	$k - 1 = 2$	$k - 1 = 3$
Small	0.10	0.07	0.06
Medium	0.30	0.21	0.17
Large	0.50	0.35	0.29

These benchmarks (adapted from Cohen) are guidelines, not rigid thresholds. Clinical and epidemiological context should always inform interpretation of effect magnitude.

Cramér's V for Example 1

From the smoking–birth weight analysis:

$$\chi^2 = 17.204, \quad N = 200, \quad k = \min(2, 2) = 2$$

$$V = \sqrt{\frac{17.204}{200 \times (2-1)}} = \sqrt{\frac{17.204}{200}} = \sqrt{0.086} = 0.293$$

Interpretation: a Cramér's V of 0.293 represents a small-to-medium effect size ($k - 1 = 1$).

Maternal smoking is significantly associated with low birth weight, and the strength of this association is clinically meaningful — roughly 30% of the theoretical maximum association.

Example 2 — Treatment Regimen and *H. pylori* Eradication

Example 2 — Setting

A gastroenterology department compares three antibiotic regimens for **Helicobacter pylori** eradication in **180 patients** with confirmed peptic ulcer disease:

- Triple therapy (PPI + amoxicillin + clarithromycin) — 60 patients
- Sequential therapy (PPI + amoxicillin → PPI + clarithromycin + metronidazole) — 60 patients
- Bismuth quadruple therapy (PPI + bismuth + metronidazole + tetracycline) — 60 patients

Outcome: eradication success (confirmed by urea breath test at 4 weeks) vs failure.

This is a **3 × 2** contingency table — Cramér's V becomes more informative here because $k - 1 = 2$.

Step 1 — Observed Frequencies

Regimen	Success	Failure	Total
Triple therapy	35	25	**60**
Sequential therapy	42	18	**60**
Bismuth quadruple	50	10	**60**
Total	**127**	**53**	180

Eradication rates: triple 58.3%, sequential 70.0%, bismuth quadruple 83.3%.

The rates appear to differ — but is this variation beyond chance?

H_0: Treatment regimen and eradication outcome are independent.

H_1: They are not independent. $\alpha = 0.05$

Step 2 — Expected Frequencies

Since all three groups have the same total ($R_i = 60$), the expected values are identical across rows:

$$E_{\text{any regimen, success}} = \frac{60 \times 127}{180} = 42.33$$

$$E_{\text{any regimen, failure}} = \frac{60 \times 53}{180} = 17.67$$

Step 2 — Expected Frequencies (table)

Regimen	Success	Failure	Total
Triple therapy	35 (42.33)	25 (17.67)	60
Sequential therapy	42 (42.33)	18 (17.67)	60
Bismuth quadruple	50 (42.33)	10 (17.67)	60
Total	127	53	180

✓ All expected frequencies ≥ 5.

Step 3 — Compute χ^2

Cell	O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Triple, Success	35	42.33	53.73	1.269
Triple, Failure	25	17.67	53.73	3.040
Sequential, Success	42	42.33	0.11	0.003
Sequential, Failure	18	17.67	0.11	0.006
Bismuth, Success	50	42.33	58.83	1.390
Bismuth, Failure	10	17.67	58.83	3.329

$$\chi^2 = 1.269 + 3.040 + 0.003 + 0.006 + 1.390 + 3.329 = 9.037$$

Step 4 — Degrees of Freedom and Decision

$$\mathrm{df} = (r-1)(c-1) = (3-1)(2-1) = 2$$

Critical value at $\alpha = 0.05$, df $= 2$: $\quad \chi^2_{\mathrm{crit}} = 5.991$

$$\chi^2_{\mathrm{calc}} = 9.037 \quad \gg \quad \chi^2_{\mathrm{crit}} = 5.991$$

Decision: Reject H_0.

There is a statistically significant association between antibiotic regimen and *H. pylori* eradication outcome ($\chi^2 = 9.04$, df $= 2$, $p = 0.011$).

The three regimens do not produce equivalent eradication rates. Bismuth quadruple therapy shows the highest success, while triple therapy shows the lowest.

Step 5 — Cramér's V for Example 2

$$V = \sqrt{\frac{\chi^2}{N \times (k-1)}}$$

- Here $k = \min(3, 2) = 2$, so $k - 1 = 1$.

$$V = \sqrt{\frac{9.037}{180 \times 1}} = \sqrt{0.0502} = 0.224$$

- With $k - 1 = 1$, this corresponds to a small-to-medium effect (Cohen's benchmark for medium at $k - 1 = 1$ is 0.30).
- The choice of antibiotic regimen is significantly associated with eradication success, though additional factors (antibiotic resistance patterns, patient compliance) likely also contribute to outcome variability.

Which Cells Drive the Result?

Examining the cell contributions to χ^2 reveals **where** the association lies:

- Triple therapy — failure: 3.040 and bismuth — failure: 3.329 are the largest contributors.
- Sequential therapy contributes almost nothing (0.003 + 0.006) — its observed values are very close to expected.

The association is driven by triple therapy underperforming (more failures than expected) and bismuth quadruple therapy outperforming (fewer failures than expected).

Sequential therapy sits close to the overall average eradication rate.

Practical Guidance

The Complete Workflow

1	State H_0 (independence) and H_1 (association); choose α
2	Construct the contingency table with **observed** frequencies
3	Compute **expected** frequencies: $E_{ij} = \frac{R_i \times C_j}{N}$
4	Verify assumptions — all $E_{ij} \geq 5$
5	Calculate $\chi^2 = \sum \frac{(O-E)^2}{E}$
6	Determine df $= (r-1)(c-1)$ and find χ^2_{crit}
7	Compare χ^2_{calc} to χ^2_{crit} and make a decision
8	Compute Cramér's $V = \sqrt{\chi^2/(N(k-1))}$ for effect size

Common Pitfalls

- Using percentages instead of counts — the χ^2 formula requires raw frequencies, not proportions or percentages.
- Small expected frequencies — when any $E < 5$, the χ^2 approximation becomes unreliable. Use Fisher's exact test (2×2) or merge categories.
- Confusing significance with importance — a large sample can make χ^2 significant even for trivially small associations. Always report Cramér's V.
- Non-independent observations — repeated measurements on the same patient violate the independence assumption (use McNemar's test for paired 2×2 data).

Comparison of the Two Examples

	Example 1	Example 2
Design	2 × 2	3 × 2
Variables	Smoking × BW	Regimen × Eradication
N	200	180
χ^2	17.20	9.04
df	1	2
p	< 0.001	0.011
Decision	Reject H_0	Reject H_0
Cramér's V	0.293	0.224
Effect size	Small–medium	Small–medium

Both studies find significant associations, but with moderate rather than strong effect sizes — typical in epidemiological research where outcomes are determined by multiple factors.

Thank you for your attention!