Lecture 19: Categorical Data Analysis — Multinomial Probabilities and Two-Way Tables

STAT 515 — Statistical Methods I

Dr. Yen-Yi Ho

Department of Statistics, University of South Carolina

Outline

1. Counts in categories: the multinomial experiment
2. Testing multinomial probabilities: goodness of fit
3. Two-way (contingency) tables
4. Expected counts under independence
5. The χ^2 test for independence
6. Test of homogeneity: fixed row totals
7. Summary: which χ^2 test, when

Outline

1. Counts in categories: the multinomial experiment
2. Testing multinomial probabilities: goodness of fit
3. Two-way (contingency) tables
4. Expected counts under independence
5. The χ^2 test for independence
6. Test of homogeneity: fixed row totals
7. Summary: which χ^2 test, when

Where We Are

Every procedure so far needed a **measurement**: a mean, a variance, a difference of means, a slope.

This chapter

The data are **counts in categories** — which candidate, which opinion, which TV show, congestive or not. There is nothing to average. The only information is **how many fell where**.

- First: **one** categorical variable, compared against a set of hypothesized probabilities — a **goodness-of-fit** test.
- Then: **two** categorical variables measured on the same subjects, cross-classified in a **two-way table**. The question becomes whether the two variables are **related**.

Example 1: Three Candidates

Voter preference

Three political candidates are running for the same elective position. A random sample of **150** eligible voters is asked which candidate they prefer. Test at $\alpha = .05$ whether voters, in total, have a preference for any one of the candidates.

	Candidate 1	**Candidate 2**	**Candidate 3**	**Total**
Observed count	$n_1 = 61$	$n_2 = 53$	$n_3 = 36$	$n = 150$

p_i = the probability that a voter chooses candidate i.

Example 2: Heart Failure and Drinking Habits

Journal of the American Medical Association, 2001

A sample of **1,913** patients suffering from acute myocardial infarction. Each patient was classified according to **drinking habits** (drinks per week) and **type of heart failure**.

	Abstain	**7 or fewer**	**7 or more**
Congestive	146	106	29
Not congestive	750	590	292

Is there a relationship between heart failure and drinking habits?

A Multinomial Experiment

The four properties

- The experiment consists of n **identical** and **independent** trials.
- There are k possible outcomes to each trial. These outcomes are sometimes called **classes**, **categories**, or **cells**.
- The probabilities of the k outcomes, denoted $p_1, p_2, \ldots, p_k$ with $p_1 + p_2 + \cdots + p_k = 1$, **remain the same from trial to trial**.
- The random variables of interest are the **cell counts** $N_1, N_2, \ldots, N_k$ — the number of observations falling in each of the k categories — with $N_1 + N_2 + \cdots + N_k = n$.

Example: ask 500 randomly chosen people which of four opinions on a policy they hold. Each person is a trial; the four opinions are the categories; the four counts are the data.

Probability distribution of the cell counts

$$p(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1!\, n_2! \cdots n_k!}\, p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}, \qquad n = n_1 + n_2 + \cdots + n_k$$

- The leading factor counts the **orderings**; the product of powers is the probability of any *one* such ordering.
- Each individual count is binomial: $N_i \sim \text{Bin}(n, p_i)$, so

$$E(N_i) = n\, p_i \qquad \text{and} \qquad \text{Var}(N_i) = n\, p_i q_i.$$

- $E(N_i) = np_i$ **is the only piece we actually need today** — it is where the expected counts come from.

The Binomial Is the Case $k = 2$

With only two categories — success and failure — write $n_2 = n - n_1$ and $p_2 = 1 - p_1$:

$$\frac{n!}{n_1!\,(n-n_1)!}\,p_1^{n_1}(1-p_1)^{n-n_1} \;=\; \binom{n}{n_1}p_1^{n_1}(1-p_1)^{n-n_1}$$

The pattern

The multinomial is to the binomial what ANOVA is to the two-sample t test: the same idea, released from the restriction to two groups. Chapter 13 is the categorical counterpart of the work we did in Chapters 9 and 10.

So a question about a *single* proportion is still a one-sample z test. The χ^2 machinery below is what you reach for when there are **three or more** categories.

Outline

1. Counts in categories: the multinomial experiment
2. Testing multinomial probabilities: goodness of fit
3. Two-way (contingency) tables
4. Expected counts under independence
5. The χ^2 test for independence
6. Test of homogeneity: fixed row totals
7. Summary: which χ^2 test, when

Goodness-of-fit test

$$H_0: \; p_1 = p_{1,0}, \; p_2 = p_{2,0}, \; \ldots, \; p_k = p_{k,0}$$

H_a : at least one of the p_i does not equal its hypothesized value

$$E_i = n\,p_{i,0}, \qquad \chi^2 = \sum_{i=1}^{k} \frac{(n_i - E_i)^2}{E_i} \;\dot{\sim}\; \chi^2_{\mathrm{df}\,=\,k-1} \text{ under } H_0$$

- n_i is the **observed** count in category i; $E_i = np_{i,0}$ is the **expected** count *if* H_0 *were true*.
- Reject H_0 when $\chi^2 > \chi^2_{\alpha,\,k-1}$.
- The hypothesized $p_{i,0}$ need **not** be equal — they only have to sum to 1.

Why This Statistic?

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - E_i)^2}{E_i}$$

- $(n_i - E_i)$ is how far that category missed. **Squaring** it stops overshoots and undershoots from cancelling.
- Dividing by E_i puts every category on the **same scale**: being off by 10 matters a great deal when $E_i = 15$ and hardly at all when $E_i = 1000$.
- Summing collects the evidence from all k categories into **one** number.

The shape of the test

$\chi^2 = 0$ means a perfect match. **Any** disagreement, in either direction, makes χ^2 larger — so the rejection region is **always the upper tail**, with all of α on the right.

Degrees of Freedom: Why $k - 1$?

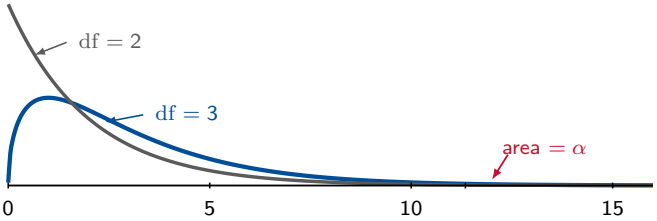
- There are k cell counts, but they are not free to vary independently: they must add up to n.
- Once you know $k - 1$ of them, the last one is **determined**. So there are **k-1** free pieces of information.

3 candidates $\Rightarrow \mathrm{df} = 2$ 4 opinions $\Rightarrow \mathrm{df} = 3$

Conditions required for a valid χ^2 test

- A **multinomial experiment** has been conducted. This is generally satisfied by taking a **random sample** from the population of interest.
- The sample size n is large enough that, **for every cell**, the expected count $E_i = np_{i,0}$ is at least 5.

The χ^2 Distribution



- The χ^2 distribution lives on $[0, \infty)$ — **never negative** — and is **right-skewed**.
- One parameter, the degrees of freedom. More df shifts the curve right and flattens it, so the critical value grows with k.
- Large χ^2 means observed counts far from expected, which is evidence against H_0. **Only the right tail rejects.**

Example 1: Three Candidates

Voter preference

Three political candidates are running for the same elective position. A random sample of **150** eligible voters is asked which candidate they prefer. Test at $\alpha = .05$ whether voters, in total, have a preference for any one of the candidates.

	Candidate 1	**Candidate 2**	**Candidate 3**	**Total**
Observed count	$n_1 = 61$	$n_2 = 53$	$n_3 = 36$	$n = 150$

p_i = the probability that a voter chooses candidate i.

Example 1: The Test

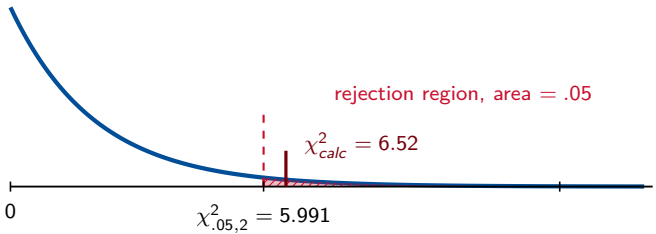
1. $H_0: p_1 = p_2 = p_3 = \frac{1}{3}$ (**no preference**); H_a: at least one $p_i \neq \frac{1}{3}$ (**some preference exists**).
2. Expected counts: $E_i = n\,p_{i,0} = 150 \cdot \frac{1}{3} = \mathbf{50}$ for each candidate — all ≥ 5 ✓
3. Test statistic:

$$\chi^2 = \frac{(61-50)^2}{50} + \frac{(53-50)^2}{50} + \frac{(36-50)^2}{50} = \frac{121+9+196}{50} = \mathbf{6.52}$$

4. $\mathrm{df} = k - 1 = 2$; rejection region $\{\chi^2 > \chi^2_{.05,2}\} = \{\chi^2 > 5.991\}$.
5. $6.52 > 5.991$, so we **reject** H_0 at $\alpha = .05$ (p-value $= .038$).

We conclude there is some preference for at least one of the candidates. **The test does not say which one** — only that "all equal" is implausible.

Example 1: The Picture



The observed statistic sits just inside the rejection region — this is a **borderline** result, which the p-value of .038 says more precisely than the reject/fail-to-reject verdict does.

Example 2: Opinions on Marijuana Laws

Example 2

An educational television station has broadcast a series of programs on the physiological and psychological effects of smoking marijuana. Now that the series is finished, the station wants to see whether citizens within the viewing area have **changed their minds** about how the possession of marijuana should be considered legally.

Before the series was shown it was determined that 7% of the citizens favored legalization, 18% favored decriminalization (an offender could be fined or imprisoned), 65% favored the existing law, and 10% had no opinion. A summary of the opinions of a random sample of 500 people *after* the series is given below. Test at $\alpha = .01$.

	Legalization	**Decriminalization**	**Existing laws**	**No opinion**	**Total**
Observed	$n_1 = 39$	$n_2 = 99$	$n_3 = 336$	$n_4 = 26$	$n = 500$

Example 2: Hypotheses and Expected Counts

p_i = the proportion holding opinion i **after** the series was aired.

H_0 : $p_1 = .07,\ \ p_2 = .18,\ \ p_3 = .65,\ \ p_4 = .10$ (the distribution is unchanged)

H_a : at least one of the probabilities differs (the distribution changed)

	Legalization	**Decrim.**	**Existing**	**No opinion**
$p_{i,0}$	.07	.18	.65	.10
$E_i = n\,p_{i,0}$	$500(.07) = 35$	$500(.18) = 90$	$500(.65) = 325$	$500(.10) = 50$
Observed n_i	39	99	336	26

All four expected counts are well above 5, and the 500 people are a random sample ✓ Note the E_i sum to 500, as they must.

Example 2: The Test Statistic

$$\chi^2 = \sum_{i=1}^{4} \frac{(n_i - E_i)^2}{E_i} = \frac{(39-35)^2}{35} + \frac{(99-90)^2}{90} + \frac{(336-325)^2}{325} + \frac{(26-50)^2}{50}$$

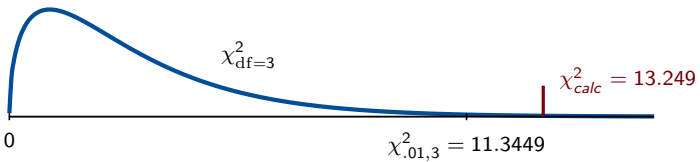
$$= 0.457 + 0.900 + 0.372 + \mathbf{11.520} = \mathbf{13.249}, \qquad \mathrm{df} = k - 1 = 3$$

Read the cell contributions

Almost all of the statistic comes from one cell. The "no opinion" group contributes 11.52 of the 13.25: 50 people were expected and only 26 turned up. The other three categories are close to what H_0 predicted.

Example 2: Decision and Conclusion

$$RR = \{\, \chi^2 > \chi^2_{.01,\,3} \,\} = \{\, \chi^2 > 11.3449 \,\}$$



Since $13.249 > 11.3449$ the statistic **falls in the rejection region**;
p-value $= P(\chi^2_3 > 13.249) =$ **0.0041** $< .01$.

Conclusion

At the .01 significance level we **reject** H_0 and conclude that the distribution of opinions has **changed** since the series was aired — driven mainly by the drop in the "no opinion" group.

Goodness of Fit in R

```
obs <- c(39, 99, 336, 26)
p0  <- c(.07, .18, .65, .10)
chisq.test(obs, p = p0)
```

```
    Chi-squared test for given probabilities

data:  obs
X-squared = 13.249, df = 3, p-value = 0.004127
```

Feed `chisq.test` the **raw counts** and the hypothesized probabilities — never percentages, and never counts you have already converted to proportions. `chisq.test(obs)$expected` prints the E_i, and `$residuals` gives $(n_i - E_i)/\sqrt{E_i}$, whose squares are the cell contributions from the previous slide.

Outline

1. Counts in categories: the multinomial experiment
2. Testing multinomial probabilities: goodness of fit
3. Two-way (contingency) tables
4. Expected counts under independence
5. The χ^2 test for independence
6. Test of homogeneity: fixed row totals
7. Summary: which χ^2 test, when

Example 1: Heart Failure and Drinking Habits

Journal of the American Medical Association, 2001

A sample of **1,913** patients suffering from acute myocardial infarction. Each patient was classified according to **drinking habits** (drinks per week) and **type of heart failure**.

	Abstain	7 or fewer	7 or more
Congestive	146	106	29
Not congestive	750	590	292

Is there a relationship between heart failure and drinking habits?

A Two-Way (Contingency) Table

Contingency table

A table of counts in which each subject is classified by **two** categorical variables at once: one defines the r **rows**, the other the c **columns**, and n_{ij} counts the subjects in row i, column j.

	Abstain	**7 or fewer**	**7 or more**	**Row total**
Congestive (C)	$146 = n_{11}$	$106 = n_{12}$	$29 = n_{13}$	$R_1 = 281$
Not congestive (C^c)	$750 = n_{21}$	$590 = n_{22}$	$292 = n_{23}$	$R_2 = 1632$
Column total	$C_1 = 896$	$C_2 = 696$	$C_3 = 321$	$n = 1913$

$R_i = i$th **row marginal**, $C_j = j$th **column marginal**, $n = \sum_i R_i = \sum_j C_j$ the grand total. Here $r = 2$, $c = 3$ — a 2×3 **table**.

What Would "No Relationship" Look Like?

Write A = abstain, F = 7 or fewer, M = 7 or more, and C = congestive heart failure.

- Within each drinking group, the proportion congestive is $P(C \mid A) = n_{11}/C_1$, and so on:

$$\frac{146}{896} = .163, \qquad \frac{106}{696} = .152, \qquad \frac{29}{321} = .090.$$

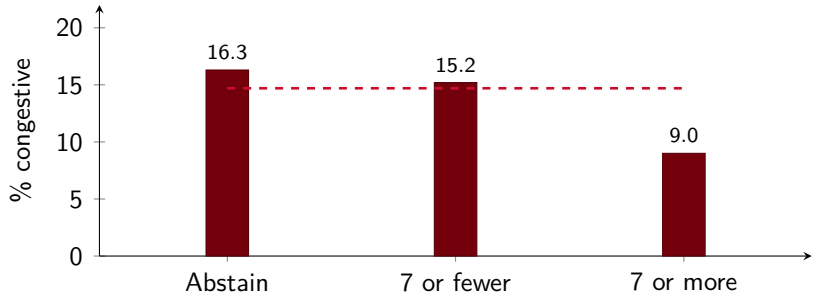
- Overall, $P(C) = R_1/n = 281/1913 = .147$.

The idea

If the two variables were **independent**, knowing the drinking group would tell you nothing about the chance of congestive failure: $P(C \mid A) = P(C \mid F) = P(C \mid M) = P(C)$. **The three percentages are not equal** — but is that more than sampling variation?

The Same Picture, as Percentages

Percent congestive within each drinking group, against the overall 14.7%:



- The dashed line is the overall rate — what every bar would sit near if the two variables were independent.
- The **heavy-drinking** group looks low. The χ^2 test decides whether that gap is real.

Outline

1. Counts in categories: the multinomial experiment
2. Testing multinomial probabilities: goodness of fit
3. Two-way (contingency) tables
4. Expected counts under independence
5. The χ^2 test for independence
6. Test of homogeneity: fixed row totals
7. Summary: which χ^2 test, when

From Independence to an Expected Count

Suppose the row variable and the column variable are **independent**. Then the conditional and unconditional probabilities agree:

$$P(C \mid A) = P(C) \quad \Longrightarrow \quad \frac{n_{11}}{C_1} = \frac{R_1}{n} \quad \Longrightarrow \quad n_{11} = \frac{R_1 C_1}{n}.$$

Repeating the argument for every column,

$$\frac{n_{12}}{C_2} = \frac{R_1}{n} \Rightarrow n_{12} = \frac{R_1 C_2}{n}, \qquad \frac{n_{13}}{C_3} = \frac{R_1}{n} \Rightarrow n_{13} = \frac{R_1 C_3}{n}.$$

Expected cell count under independence

$$E_{ij} = \frac{R_i \, C_j}{n} = \frac{(\text{row total}) \times (\text{column total})}{\text{grand total}}$$

Why That Formula Makes Sense

- Independence means

$$P(\text{row } i \text{ and column } j) = P(\text{row } i)\,P(\text{column } j) = \frac{R_i}{n} \cdot \frac{C_j}{n}.$$

- Multiply that probability by the n subjects:

$$E_{ij} = n \cdot \frac{R_i}{n} \cdot \frac{C_j}{n} = \frac{R_i C_j}{n}.$$

Two things to notice

- The E_{ij} use **only the margins** — the row and column totals, which we are not questioning. The interior counts are what we test.
- Expected counts **need not be whole numbers**. $E_{11} = 131.61$ is perfectly fine; it is an average, not a count.

Example 1: The Expected Counts

With $R_1 = 281$, $R_2 = 1632$, $C_1 = 896$, $C_2 = 696$, $C_3 = 321$, $n = 1913$:

$$E_{11} = \frac{R_1 C_1}{n} = \frac{281(896)}{1913} = 131.613, \qquad E_{12} = \frac{R_1 C_2}{n} = \frac{281(696)}{1913} = 102.235, \ \ldots$$

Observed (expected)	**Abstain**	**7 or fewer**	**7 or more**	**Total**
Congestive	146 (131.61)	106 (102.24)	29 (**47.15**)	281
Not congestive	750 (764.39)	590 (593.76)	292 (**273.85**)	1632
Total	896	696	321	1913

- **Check:** the expected counts reproduce the same margins ($131.61 + 102.24 + 47.15 = 281$). Always a good arithmetic check.
- The biggest gap is in the "7 or more / congestive" cell: **29 observed** against **47.15 expected**.

Outline

1. Counts in categories: the multinomial experiment
2. Testing multinomial probabilities: goodness of fit
3. Two-way (contingency) tables
4. Expected counts under independence
5. The χ^2 test for independence
6. Test of homogeneity: fixed row totals
7. Summary: which χ^2 test, when

The χ^2 Test for Independence

Test for independence in an $r \times c$ table

H_0: the two variables are **independent**

H_a: the two variables are **dependent** (related)

$$\chi^2 = \sum_{\text{all cells}} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}, \qquad E_{ij} = \frac{R_i C_j}{n}$$

Reject H_0 if $\chi^2 > \chi^2_{\alpha,(r-1)(c-1)}$.

- Degrees of freedom $= (r-1)(c-1)$: **(# rows -1)(# columns -1)**; for Example 1, $(2-1)(3-1) = 2$.
- Same engine as goodness of fit — compare observed with expected, cell by cell — only the recipe for E and the df have changed.

Where the Degrees of Freedom Come From

In a 2×3 table the margins are fixed. Once you choose **2** interior counts in the first row, **everything else is forced**:

	Abstain	**7 or fewer**	**7 or more**	**Total**
Congestive	**free**	**free**	forced	281
Not congestive	forced	forced	forced	1632
Total	896	696	321	1913

- $(r-1)(c-1) = (2-1)(3-1) = 2$ cells are free to vary.
- A 3×4 table would give $(3-1)(4-1) = 6$; a 2×2 table gives $(2-1)(2-1) = 1$.

When Does This Test Work?

Conditions

1. The n subjects are a **random sample**, and each is counted in exactly **one** cell.
2. The χ^2 approximation works well if **at least 80% of the expected counts are ≥ 5** and **all expected counts are ≥ 1**.

- The condition is on the **expected** counts E_{ij}, not on the observed counts.
- Example 1: the smallest expected count is 47.15 — comfortably satisfied.
- If the condition fails, combine categories that make sense together, or use an exact test.

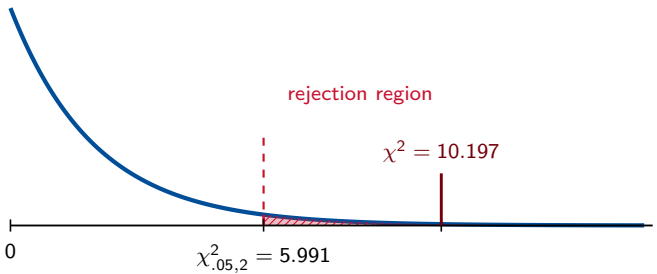
Example 1: Carrying Out the Test

1. H_0: heart-failure type and drinking habits are **independent**; H_a: they are **dependent**.
2. Expected counts $E_{ij} = R_i C_j / n$ (previous slide); all ≥ 5.
3. Test statistic:

$$\chi^2 = \frac{(146 - 131.613)^2}{131.613} + \frac{(106 - 102.235)^2}{102.235} + \frac{(29 - 47.152)^2}{47.152}$$
$$+ \frac{(750 - 764.387)^2}{764.387} + \frac{(590 - 593.765)^2}{593.765} + \frac{(292 - 273.848)^2}{273.848} = \mathbf{10.197}$$

4. $\mathrm{df} = (2-1)(3-1) = 2$; rejection region $\{\chi^2 > \chi^2_{.05,2}\} = \{\chi^2 > 5.991\}$.

Example 1: The Picture and the Conclusion



Conclusion

Since 10.197 falls in the rejection region, we **reject** H_0 at $\alpha = .05$. There is sufficient evidence that type of heart failure and drinking habits are **related** in this population.

Example 1 in R

```
> dat <- matrix(c(146,106,29,750,590,292), byrow=T, ncol=3)
> chisq.test(dat)$expected
          [,1]       [,2]       [,3]
[1,] 131.6132 102.2352  47.15159
[2,] 764.3868 593.7648 273.84841
```

- `byrow=T` fills the matrix one **row** at a time — match the order to the way you read the table.
- `$expected` returns the E_{ij}; use it to check the "80% at least 5" condition before you trust the test.

Example 1 in R: The Test

```
> chisq.test(dat)

        Pearson's Chi-squared test

X-squared = 10.197, df = 2, p-value = 0.006107
```

- The p-value is an **upper-tail** area: $p = P_{H_0}(\chi^2 \geq 10.197) = .006107$.
- $.006107 < \alpha = .05$, so **reject** H_0 — the same decision as the rejection-region approach, as always.
- R reports `df = 2`, confirming $(r-1)(c-1) = 2$.

Which Cell Drove the Result?

The individual terms $(n_{ij} - E_{ij})^2/E_{ij}$ add to 10.197:

	Abstain	7 or fewer	7 or more	Row sum
Congestive	1.573	0.139	**6.988**	8.700
Not congestive	0.271	0.024	**1.203**	1.498
Total		$\chi^2 = 10.197$		

- Over 80% of the statistic comes from the **"7 or more"** column — fewer congestive cases there than independence predicts.
- **This is description, not a second test.** The χ^2 test gives one verdict for the whole table; cell contributions only suggest *where* to look.

Outline

1. Counts in categories: the multinomial experiment
2. Testing multinomial probabilities: goodness of fit
3. Two-way (contingency) tables
4. Expected counts under independence
5. The χ^2 test for independence
6. Test of homogeneity: fixed row totals
7. Summary: which χ^2 test, when

Example 2: Favorite TV Program

A developmental psychologist

Takes a random sample of **100 first-grade boys** *and* a random sample of **100 first-grade girls**, and asks each child which of three television programs they prefer. Is there a difference between the preferences of boys and girls at $\alpha = .05$?

	Lone Ranger	**Sesame Street**	**The Simpsons**	**Total**
Boy	50	30	20	**100**
Girl	25	40	35	**100**
Total	75	70	55	200

Notice: the row totals are 100 and 100 **by design**, not by accident.

Fixed Margins: A Different Sampling Plan

- This contingency table has **fixed row marginals**, since the row totals are **known in advance** — the researcher chose to sample 100 of each.
- The goal of the analysis does not change: determine whether the two qualitative variables are dependent. Sometimes we refer to it as testing **homogeneity**.
- The **procedure for conducting a χ^2 analysis for a contingency table with fixed marginals is identical** to the one outlined above.
- One reason to choose this alternative sampling plan is to obtain **sufficient observations in each cell**.

Independence vs. Homogeneity — The Hypotheses

Independence

One sample of n subjects, each classified two ways.

H_0: the row and column variables are independent.

Homogeneity

Several samples of fixed size, one per row.

H_0: the column distribution is the *same* in every row population.

For Example 2, H_0 says $(p_{\text{LR}}, p_{\text{SS}}, p_{\text{Simp}})_{\text{boys}} = (p_{\text{LR}}, p_{\text{SS}}, p_{\text{Simp}})_{\text{girls}}$, and H_a says the two preference distributions **differ**.

Different question, different sampling plan — identical arithmetic: $E_{ij} = R_i C_j / n$, $\text{df} = (r-1)(c-1)$, upper tail.

Example 2: Expected Counts

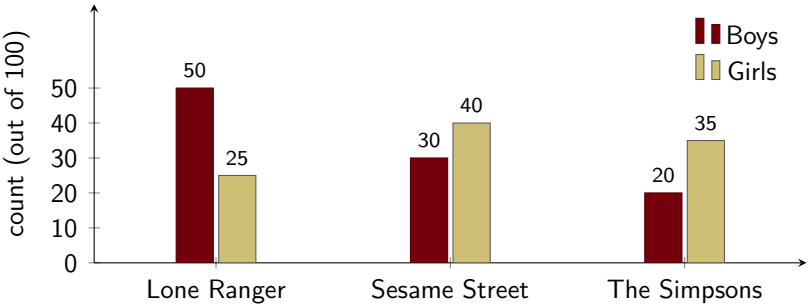
$R_1 = R_2 = 100;\ C_1 = 75,\ C_2 = 70,\ C_3 = 55;\ n = 200$:

$$E_{11} = \frac{100(75)}{200} = 37.5, \qquad E_{12} = \frac{100(70)}{200} = 35, \qquad E_{13} = \frac{100(55)}{200} = 27.5.$$

Observed (expected)	**Lone Ranger**	**Sesame Street**	**The Simpsons**	**Total**
Boy	50 (37.5)	30 (35)	20 (27.5)	100
Girl	25 (37.5)	40 (35)	35 (27.5)	100
Total	75	70	55	200

- With equal row totals the two rows get the **same** expected counts — exactly what "same distribution in both groups" means.
- All expected counts are well above 5.

Example 2: Observed Preferences



If the two distributions were homogeneous, each pair of bars would sit at the same height — **37.5**, **35**, **27.5**.

Example 2: Carrying Out the Test

1. H_0: boys and girls have the **same** distribution of program preference; H_a: the distributions **differ**.
2. Test statistic:

$$\chi^2 = \frac{(50-37.5)^2}{37.5} + \frac{(30-35)^2}{35} + \frac{(20-27.5)^2}{27.5} + \frac{(25-37.5)^2}{37.5} + \frac{(40-35)^2}{35} + \frac{(35-27.5)^2}{27.5} = \mathbf{13.853}$$

3. $\mathrm{df} = (2-1)(3-1) = 2$; rejection region $\{\chi^2 > 5.991\}$.
4. $13.853 > 5.991 \Rightarrow$ **reject** H_0 at $\alpha = .05$.

Example 2 in R

```
> dat <- matrix(c(50,30,20,25,40,35), byrow=T, ncol=3)
> chisq.test(dat)$expected
     [,1] [,2] [,3]
[1,] 37.5   35 27.5
[2,] 37.5   35 27.5

> chisq.test(dat)

        Pearson's Chi-squared test

X-squared = 13.8528, df = 2, p-value = 0.0009815
```

$p = .00098 < .05$: there is strong evidence that **boys and girls differ** in their program preferences — boys favor the Lone Ranger, girls the Simpsons.

A Note on the Data You Feed In

Common mistakes

- Enter the data **one row at a time, not counting the totals**. Including a margin row or column inflates n and invalidates the test.
- Feed in **raw counts**, never percentages or averages. The χ^2 statistic depends on sample size; percentages throw that away.
- Each subject must appear in **exactly one** cell — no "check all that apply" tables.

- `chisq.test(obs)` gives the test; `chisq.test(obs)$expected` gives the expected counts it used.
- For a 2×2 table R applies a continuity correction by default; `correct=F` turns it off to match the hand computation.

Outline

1. Counts in categories: the multinomial experiment
2. Testing multinomial probabilities: goodness of fit
3. Two-way (contingency) tables
4. Expected counts under independence
5. The χ^2 test for independence
6. Test of homogeneity: fixed row totals
7. Summary: which χ^2 test, when

Summary: Which χ^2 Test?

Test	Data	Expected count	df
Goodness of fit	one variable, k categories	$E_i = np_{i0}$	$k-1$
Independence	one sample, cross-classified	$E_{ij} = R_i C_j / n$	$(r-1)(c-1)$
Homogeneity	r samples of fixed size	$E_{ij} = R_i C_j / n$	$(r-1)(c-1)$

- All three use the **same statistic** $\chi^2 = \sum \dfrac{(\text{obs} - \text{exp})^2}{\text{exp}}$ and the **same upper-tail rejection region**.
- What changes is where the expected counts come from and how many degrees of freedom are left.
- **Independence** vs. **homogeneity** is decided by the **sampling plan**, not by the arithmetic.

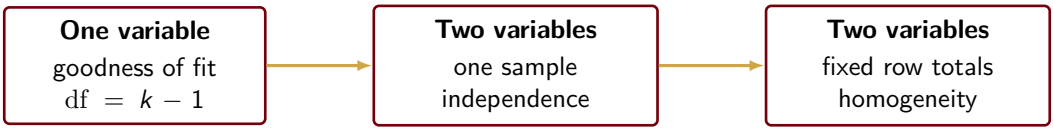
What the Test Does Not Tell You

- **Not direction.** Rejecting H_0 says the variables are related, not which way. Look at the percentages and the cell contributions to describe the relationship.
- **Not strength.** χ^2 grows with n: a tiny association in a huge sample can be highly significant.
- **Not causation.** Example 1 is observational — drinking habits and heart failure are associated; nothing here says one causes the other.

Checklist before you report a χ^2 test

Random sample(s) · each subject in exactly one cell · raw counts · all $E_{ij} \geq 1$ and at least 80% of them ≥ 5 · df $= (r-1)(c-1)$.

Looking Back at Chapter 13



- One idea the whole way through: **compare what you observed with what H_0 predicts**, standardized cell by cell. The statistic $\sum(\text{obs} - \text{exp})^2/\text{exp}$ never changes; only the source of the expected counts and the degrees of freedom do.
- Read Sections 13.1–13.3; Examples 13.2 (pp. 776–777) and 13.3 (p. 787) in the text are two of the worked problems from today.