Statistics and Data Analysis

Unit 03 – Lecture 04: Chi-square Tests (Goodness-of-Fit and Independence)

Tofik Ali

School of Computer Science, UPES Dehradun

February 17, 2026

https://github.com/tali7c/Statistics-and-Data-Analysis

Quick Links

Overview Chi-square Tests Expected Counts Exercises Demo Summary

Agenda

Learning Outcomes

- Explain when chi-square tests are used (counts/frequencies)

Learning Outcomes

- Explain when chi-square tests are used (counts/frequencies)
- Compute expected counts for a contingency table

Learning Outcomes

- Explain when chi-square tests are used (counts/frequencies)
- Compute expected counts for a contingency table
- Compute chi-square statistic (basic)

Learning Outcomes

- Explain when chi-square tests are used (counts/frequencies)
- Compute expected counts for a contingency table
- Compute chi-square statistic (basic)
- State assumptions (expected counts not too small)

Learning Outcomes

- Explain when chi-square tests are used (counts/frequencies)
- Compute expected counts for a contingency table
- Compute chi-square statistic (basic)
- State assumptions (expected counts not too small)
- Interpret independence vs association

Chi-square Tests: Key Points

- GOF: one categorical variable

Chi-square Tests: Key Points

- GOF: one categorical variable
- Independence: two categorical variables

Chi-square Tests: Key Points

- GOF: one categorical variable
- Independence: two categorical variables
- Compare observed O to expected E under H0

Chi-square Tests: Key Formula

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

Expected Counts: Key Points

- E_rc = (row total)(col total)/N

Expected Counts: Key Points

- E_rc = (row total)(col total)/N
- df = (R-1)(C-1) for independence

Exercise 1: Expected counts

Row totals 60/40, column totals 70/30, N=100. Compute E11.

Solution 1

- E11 = 60*70/100 = 42

Exercise 2: Interpret reject

If you reject H0 in independence test, what do you conclude?

Solution 2

- Evidence of association between variables.
- Not direction; not causation.

Exercise 3: Assumption

Why do we worry about small expected counts?

Solution 3

- Chi-square approximation can break when E is very small.

Mini Demo (Python)

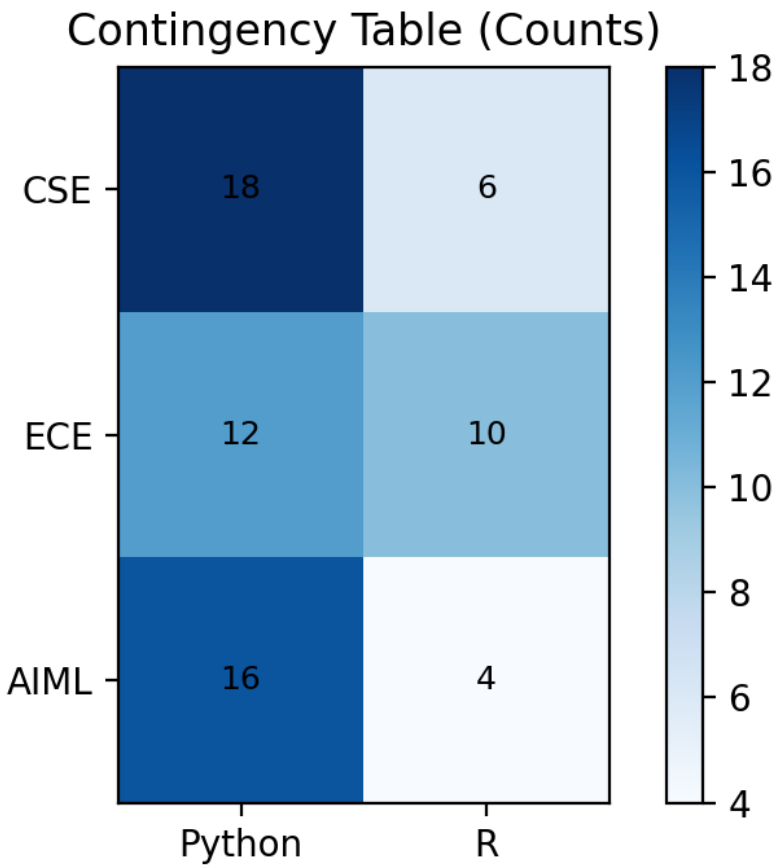
Run from the lecture folder:

```
python demo/demo.py
```

Outputs:

- `images/demo.png`
- `data/results.txt`

Demo Output (Example)



Summary

- Key definitions and the main formula.

Summary

- Key definitions and the main formula.
- How to interpret results in context.

Summary

- Key definitions and the main formula.
- How to interpret results in context.
- How the demo connects to the theory.

Exit Question

Why do chi-square tests use expected counts instead of only raw percentages?