Statistics and Data Analysis
Unit 03 – Lecture 04 Notes

Tofik Ali

February 14, 2026

Topic

Chi-square tests for counts: GOF and independence; expected counts; assumptions.

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

Detailed Notes

These notes are designed to be read alongside the slides. They expand each slide bullet into plain-language explanations, small worked examples, and common pitfalls. When a formula appears, emphasize (1) what each symbol means, (2) the assumptions needed to use it, and (3) how to interpret the final number in the problem context.

Chi-square Tests

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

Expected Counts

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

Exercises (with Solutions)

Exercise 1: Expected counts

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

Solution

- E11 = 60*70/100 = 42

Exercise 2: Interpret reject

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

Solution

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

Exercise 3: Assumption

Why do we worry about small expected counts?

Solution

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

Exit Question

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

Demo (Python)

Run from the lecture folder:

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

Output files:

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

References

- Montgomery, D. C., & Runger, G. C. *Applied Statistics and Probability for Engineers*, Wiley.
- Devore, J. L. *Probability and Statistics for Engineering and the Sciences*, Cengage.
- McKinney, W. *Python for Data Analysis*, O'Reilly.