INTRODUCTION TO PROBABILITY MODELS

Lecture 37

Qi Wang, Department of Statistics

Nov 26, 2018

CHI-SQUARE(χ^2) HYPOTHESIS TEST

We want to know if there is a relationship between two qualitative (categorical) variables, if there is no relationship, then the two are considered independent.

We will use the crosstab table to test whether there is a relationship. Returning to our example in previous lecture, suppose we wanted to do the following test:

- H_0 : there **is no** relationship between Letter Grades and Class Time
- H_a : There **is** a relationship between Letter Grades and Class Time

Note H_0 is called the null hypothesis, H_a is the alternative hypothesis

We are using the information in the crosstab table to determine whether the data supports the null or alternative hypothesis. We will conduct a hypothesis test!

CHI-SQUARE(χ^2) HYPOTHESIS TEST

1. State the Null and Alternative hypothesis
2. Determine the confidence level and the significance level α
3. Find the test statistic χ^2
4. Determine the degrees of freedom needed to use the χ^2 table
5. Find the χ^2 critical value from the χ^2 table. Compare critical value from the table to the calculated χ^2 value.
6. State the conclusion in terms of the problem

STATE THE NULL AND ALTERNATIVE HYPOTHESIS.

- H_0 : there **is no** relationship between the row and column variables
- H_a : There **is** a relationship between the row and column variables

DETERMINE THE CONFIDENCE LEVEL AND THE SIGNIFICANCE LEVEL

$\alpha = 1 - confidence level$

Typical confidence levels are 90%, 95% and 99% so α is typically 0.10, 0.05, 0.01.

FIND THE TEST STATISTIC

To test the null hypothesis, compare observed cell counts with expected cell counts calculated under the assumption that the null hypothesis is true. The Chi-square statistic, χ^2, is a measure of how far the observed counts in the two-way table are from the expected counts. The formula for the statistic is

$$\chi^2 = \sum \frac{(observed\ cell\ count - expected\ cell\ count)^2}{expected\ cell\ count}$$

$$expected\ cell\ count = \frac{row\ total \times column\ total}{overall\ total}$$

$observed\ cell\ count = actual\ cell\ count$

We sum is over all cells in the table. So to get the overall value of χ^2, calculate each cell's expected count and each cell's partial χ^2. Add all the partial χ^2 values for the overall.

DETERMINE THE DEGREES OF FREEDOM

$$DF = (\#Rows - 1) \times (\#Columns - 1)$$

FIND CRITICAL VALUE

- If the table value $< \chi^2$ (from calculation) then REJECT H_0
- If the table value $> \chi^2$ (from calculation) then DO NOT REJECT H_0

STATE THE CONCLUSION

- If rejecting H_0: Evidence shows that there is a relationship between the row variable and the column variable. (or state that evidence shows the two are not independent)
- If NOT rejecting H_0: Evidence FAILS TO SHOW that there is a relationship between row variable and column variable. (or state that evidence shows that the two are independent.)

IMPORTANT NOTE: We NEVER Accept, and NEVER say Prove. There is always that chance that we are making the incorrect conclusion. Also the decision to reject or not is ALWAYS in terms of the NULL Hypothesis. (e.g. DO NOT say reject H_a)

WHEN IS IT OKAY TO PERFORM A CHI-SQUARE TEST?

This is considered checking the assumptions, very important to make this check!

- If table is larger than 2×2: okay
  - If all cells have an EXPECTED count of at least 1 AND
  - less than 20% of cells have EXPECTED counts under 5
- If table is 2×2 (smallest crosstab table), all cells must have an EXPECTED count of 5 or more

EXAMPLE 1

Psychological factors and social factors can influence the survival of patients with serious diseases. One study examined the relationship between survival of patients with coronary heart disease and pet ownership. Each of 92 patients was classified as having a pet or not by whether they survived for one year. The researchers suspected that having a pet might be connected to the patient status. Here are the data.

Patient status	NO	YES
Alive	28	50
Dead	11	3
Total	39	53

1. Assuming the patient is still alive, what is the probability that he owns a pet? Is this a joint, marginal or conditional probability?
2. What is the probability that a patient owns a pet and is still alive? Is this a joint, marginal or conditional probability?
3. What is the probability that a patient owns a pet? Is this a joint, marginal or conditional probability?
4. State the hypotheses for a χ^2 test for this problem, find the χ^2 test statistic, its degrees of freedom and the χ^2 value from the table. State your conclusions in terms of the original problem.