Chi-square Test

Tom Bruning

2018-02-16

Learning Objectives

In this chapter you will learn:

- How and when to use chi-square test for contingency tables
- How to use Marascuilo procedure for determining pair-wise differences when evaluationg more than two proproportions
- How and when to use non-parametric tests

Contingency Tables

- Useful in situations comparing multiple population proportions
- Used to classify sample observations according to two or more characteristics
- Also called cross-classification table (cross-tab)

Contingency Table Example

Suppose you think that there might be a relationship (dependency) between right/left handedness and gender.

Dominant Hand: Right vs. Left Gender: Male vs. Female

Since there are two categories for each variable, this is called a 2x2 table.

We sample 300 Children and note their gender and dominant hand.

Table 1: Sample Results organized in contingency table

Gender	Left	Right	Total
Female	12	108	120
Male	24	156	180
Total	36	264	300

χ^2 *Test For Difference Between 2 Proportions*

$H_0 : \pi_1 = \pi_2$ (Proportion of females whare are left handed is equal to proportion of males who are left handed)

$H_0 : \pi_1 \neq \pi_2$ (The two proportions are not the same)

If H_0 is true, then the proportion of left handed females should be the same as the proportion of left handed males. The conclusion is that gender has no effect on handedness (i.e. handedness is independent of gender).

The Chi-square Test Statistic

The statistic is:

$$\chi^2_{STAT} = \sum_{all\ cells} \frac{(f_o - f_e)^2}{f_e} \tag{1}$$

Where:
f_o is the observed frequency of each cell
f_e is the expected frequency of each cell ir H_0 is true
(Assume: each cell has an expected frequency of at least 5)

Decision Rule

The χ^2_{STAT} test statistic approximately follows a chi-squared distribution with (rows -1) time (columns -1) degrees of freedom.

Decision Rule:
If $\chi^2_{STAT} > \chi^2_\alpha$ reject H_0.

Computing the Overall Proportion

This overall proportion is:

$$\bar{p} = \frac{x_1 + ... + x_k}{n_1 + ... + n_k} \tag{2}$$

Where:
k is the number of columns
x_1 is number of items of interest from the first group
x_k is number of items of interest from the k^{th} group
n_1 is number of items in the first group
n_k is number of items in the k^{th} group
For this example:
$\bar{p} = \frac{12+24}{180+120} = \frac{36}{300} = .12$
Finding Expected Frequencies

To obtain the expected frequency for left handed females, multiply the number of females by the overall proportion $\bar{p}$ of left handers; To obtain the expected frequency for left handed males, multiply the number of males by the overall proportion $\bar{p}$ of left handers.

If the two proportions are equal, then:

$P(Left\ Handers|Female) = P(Left\ Handers|Male) = .12$ (i.e. we would expect $(.12)(120) = 14.4$ females to be left handed and $(.12)(180) = 21.6$ males to be left handed)

Table 2: Observed vs. Expected Table

Gender	Left	Right	Total
Female	O=12, E=14.4	O=108, E=105.6	120
Male	O=24, E=21.6	O=156, E=158.4	180
Total	36	264	300

Following the 6 step hypothesis process we have:

1. $H_0 : \pi_1 = \pi_2$
   $H_1 : \pi_1 \neq \pi_2$ where pi_1 is right handed females, and
   pi_2 is right handed males
2. $\alpha = .05$ $n = 300$
3. χ^2 Test and and sampling distribution
4. Critical value is 3.841 (from χ_2 table with d.f=1 (rows-1)(columns-1))
5. Determine test statistic:
   $\chi^2_{STAT} = \sum_{all\ cells} \frac{(f_o - f_e)^2}{f_e}$
   $\chi^2_{STAT} = \frac{(12-14.4)_2}{14.4} + \frac{(108-105.6)_2}{105.6} + \frac{(24-21.6)_2}{21.6} + \frac{(156-158.4)_2}{158.4} = 0.7576$
6. Decision Rule: If $\chi^2_{STAT} >$ Critical Value, reject H_0, thus
   since 0.7576 is not greater than 3.841 **do not** reject H_0.
   Since we do not reject H_0 there is sufficient evidence that the two proportions are different at$\alpha = .05$.

The Marascuilo Procedure for χ^2

- Used when the null hypothesis of equal proportions is rejected
- Enables you to make comparisons between all pairs
- Start with the observed differences, p_j–$p_{j'}$, for all pairs (for $j \neq j'$) then compare the absolute difference to a calculated critical range Critical Range for the Marascuilo Procedure:
- Critical Range $= \sqrt{\chi^2}\sqrt{\frac{p_j(1-p_j)}{n_j} + \frac{p_{j'}(1-p_{j'})}{n_{j'}}}$

  (Note: the critical range is different for each pairwise comparison)

A particular pair of proportions is significantly different if:
$|p_j – p_{j'}| >$ critical range for j and j'

Marascuilo Procedure Example

A University is thinking of switching to a trimester academic calendar. A random sample of 100 administrators, 50 students, and 50 faculty members were surveyed:

Opinion	Administrators	Students	Faculty	Total
Favor	63	20	37	120
Opposed	37	30	13	80
Total	100	50	50	200

Using a 1% level of significance, which groups have a different attitude?

Chi-Square Test Results

$H_0 : \pi_1 = \pi_2 = \pi_3$ H_1: Not all of the j are equal (j = 1, 2, 3)

Opinion	Administrators	Students	Faculty	Total
Favor	0=63,e=60	o=20,e=30	o=37, e=30	120
Opposed	o=37,e=40	o=30, e=20	o=13, e=20	80
Total	100	50	50	200

$\chi^2_{STAT} = \sum_{All\ cells} \frac{(f_o - f_e)^2}{f_e} = 12.792$

Marascuilo Procedure: Solution

Marascuilo Procedure

Group	Sample Proportion	Sample Size	Comparison	Absolute Difference	Std. Error of Difference	Critical Range	Results
1	0.63	100	1 to 2	0.23	0.084445249	0.2563	Means are not different
2	0.4	50	1 to 3	0.11	0.078606615	0.2386	Means are not different
3	0.74	50	2 to 3	0.34	0.092994624	0.2822	Means are different

Other Data

Level of significance	0.01	Chi-sq Critical Value	9.2103
d.f	2		
Q Statistic	3.034854		