

11.2 Inference for a 2 way table

2 way tables can be used to describe relationships between 2 categorical variables.

Ex: check your understanding on p. 699

$$\text{Several/main campus} = \frac{55}{910} = .06$$

$$\text{at least 1^{week}/main campus} = \frac{215}{910} = .236$$

$$\text{at least 1^{day}/main campus} = \frac{640}{910} = .703$$

$$\text{Several/common} = \frac{76}{627} = .12$$

$$\text{week/common} = \frac{157}{627} = .25$$

$$\text{day/common} = \frac{394}{627} = .63$$



A chi square test for homogeneity determines whether the distribution of a categorical variable is the same for each population or treatment.

Ex: For M&M's, we wonder if the color distribution is the same for plain and peanut M&M's

Hypotheses for χ^2 test for homogeneity

H_0 : There is no difference in the distribution of a categorical variable for several populations

H_a : There is a difference in the distribution of categorical variables for several populations.

Find expected counts

When H_0 is true, the expected count of a 2way table is:

$$\text{expected count} = \frac{\text{row total} \cdot \text{column total}}{\text{table total}}$$

Conditions for χ^2 test for homogeneity

- ① Random: data to come from independent random samples or groups of an experiment.
- ② 10% condition: when sampling without replacement, $n < \frac{1}{10} N$ for each group.
- ③ large counts: all expected counts are at least 5.

Calculate χ^2

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

where the sum is over all cells in a 2 way table and the

$$df = (\# \text{ of rows} - 1)(\# \text{ of columns} - 1)$$

Check your understanding p. 711

H_0 : There is no difference in the distribution for quality of life of patients in the US & Canada.

H_a : There is a difference in the distribution for quality of life of patients in the US & Canada.

Significance level is .01

Random: 2 independent random samples.

10%: 311 is less than 10% of HA patients in Canada

2165 is less than 10% of HA patients in US

Large counts: Using calculator expected counts are:

77.37, 71.47, 109.91, 41.70, 10.55, 538.63, 497.53,

765.1, 290.30, 73.45

are all greater than 5.

We will perform a χ^2 test.

Using χ^2 test on calculator,
we entered observed counts

We get $\chi^2 = 11.72$
p-value = .019
df = 4

75	541
71	498
96	779
70	282
19	65

We fail to reject the null hypothesis since .019 > .01. Meaning is not sufficient evidence there is a difference between the distribution of quality of life of HA patients in Canada and the U.S.

The Chi Square test for independence tests whether there is convincing evidence of an association between row and column variables in a 2 way table.

χ^2 test for independence

Suppose conditions are met. We test

H_0 : There is no association between 2 categorical variables in the population of interest.

H_a : There is an association between 2 categorical variables in the population of interest.

Start by finding expected counts. Then calculate

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

where the sum is over all cells in a 2 way table.

If H_0 is true, χ^2 has approximately the χ^2 distribution with $df = (\# \text{ rows} - 1)(\# \text{ of columns} - 1)$

The value is the area to the right of χ^2 under the χ^2 density curve.



check your understanding p. 717

H_0 : There is no association between an exclusive territory clause and business survival for NW franchise firms.

H_a : There is an association between an exclusive territory clause and business survival for new franchise firms

We will use $\alpha = .01$ as a significance level.

Random: random sample

10%: 170 < 10% of all franchises

Large counts: expected counts are 102.74, 20.26, 37.26, 7.74 which are all greater than 5.

We will use the chi square test.

using χ^2 test on calculator or and entering

observed counts: $\begin{bmatrix} 108 & 15 \\ 34 & 13 \end{bmatrix}$ to matrix A, we get

$$\chi^2 = 5.91 \quad p\text{value} = .015 \quad df = 1$$

We fail to reject H_0 since our pvalue of .015 > .01
~~There is~~ not sufficient to say there is an association between an exclusive territory clause & business survival for new franchise firms.

Using χ^2 Square test wisely

1. If data come from 2 or more independent random samples or treatment groups, use χ^2 test for homogeneity. If the data comes from a single sample or treatment group, use χ^2 test for independence.
2. χ^2 tests and 2 sample z tests for difference in proportions are the same.
3. If the expected are not at least 5, we can group categories together by collapsing rows and columns to make them at least 5.

Ex: what sport do you play

Football : 100
Baseball : 46
Basketball : 74
Other : 5