Data Analysis and Machine-Learning

Chapter 12.

*Chi-Square Test using JAMOVI*



1. Introduction

Data Source for exercise: ‘Telco Customer Churn’ from Kaggle.

Recall that chi-square test is applicable for variables (both dependent and independent) with categorical dataset (nominal, ordinal).

2. One-way Chi-Square

Jamovi => Frequencies => 2 Outcomes => Input Gender

**Proportion Test (2 Outcomes)**

| Binomial Test | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | | **Level** | | **Count** | | **Total** | | **Proportion** | | **p** | |
| gender |  | Female |  | 3488 |  | 7043 |  | 0.495 |  | 0.432 |  |
|  |  | Male |  | 3555 |  | 7043 |  | 0.505 |  | 0.432 |  |
| Note. Hₐ is proportion ≠ 0.5 | | | | | | | | | | | |
|  | | | | | | | | | | | |

No difference in ratio between Male and Female (p=.432)

3. Chi-Square with N Outcomes (One-Way)

Jamovi => Frequencies => N Outcomes

Variable: PaymentMethod => Check Expected Outcomes

When Expected Proportions are set as 1:1:2:1 (accordingly to the hypothesis for the experiment), p<.001, which means that the proportion of 1:1:2:1 is not statistically true.

**Proportion Test (N Outcomes)**

| Proportions - PaymentMethod | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |  |  |  |  |  |  |  |
| **Level** | | | | | | | |  | | **Count** | | **Proportion** | |
| Bank transfer (automatic) | | | | | | |  | Observed |  | 1544 |  | 0.219 |  |
|  | | | | | | | | Expected |  | 1409 |  | 0.200 |  |
| Credit card (automatic) | | | | | | |  | Observed |  | 1522 |  | 0.216 |  |
|  | | | | | | | | Expected |  | 1409 |  | 0.200 |  |
| Electronic check | | | | | | |  | Observed |  | 2365 |  | 0.336 |  |
|  | | | | | | | | Expected |  | 2817 |  | 0.400 |  |
| Mailed check | | | | | | |  | Observed |  | 1612 |  | 0.229 |  |
|  | | | | | | | | Expected |  | 1409 |  | 0.200 |  |
|  | | | | | | | | | | | | | |
| χ² Goodness of Fit | | | | | |
|  |  |  |  |  |  |
| **χ²** | | **df** | | **p** | |
| 124 |  | 3 |  | < .001 |  |
|  | | | | | |

4. Chi-Square with N Outcomes (Two-Way)

Frequencies => Contingency Tables (Independent Samples – Chi-square test of association)

Rows: Phone Service

Columns: Gender

Independence Test:

Null hypothesis: Respectively independent

Alternative hypothesis: ‘Phone Service’ and ‘Gender’ are not respectively independent. (e.g., if p<.05, there is a difference in gender accordingly to the phone service)

**Contingency Tables**

| Contingency Tables | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | **gender** | | | | | | | | | | | |  | | | | | |
| **PhoneService** | | | **Female** | | | | | | **Male** | | | | | | **Total** | | | | | |
| No | |  | 331 | | | |  | | 351 | | |  | | | 682 | | | |  | |
| Yes | |  | 3157 | | | |  | | 3204 | | |  | | | 6361 | | | |  | |
| Total | |  | 3488 | | | |  | | 3555 | | |  | | | 7043 | | | |  | |
|  | | | | | | | | | | | | | | | | | | | | |
| χ² Tests | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | |  | | |  | |  | | | |  | |  | |  | | | |  |
|  | | | | | | | | **Value** | | | | | | **df** | | | | **p** | | | | |
| χ² | | | | |  | | | 0.296 | |  | | | | 1 | |  | | 0.586 | | | |  |
| χ² continuity correction | | | | |  | | | 0.254 | |  | | | | 1 | |  | | 0.614 | | | |  |
| Likelihood ratio | | | | |  | | | 0.296 | |  | | | | 1 | |  | | 0.586 | | | |  |
| Fisher's exact test | | | | |  | | |  | |  | | | |  | | | | 0.600 | | | |  |
| N | | | | |  | | | 7043 | |  | | | |  | | | |  | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | |
| * No difference (chi-squared and chi-squared continuity correction <.05)    Comparative Measures | | | | | | | | | | | | | | | | | | | |
|  | | | | | | **95% Confidence Intervals** | | | | | | | | | | | | | |
|  | | **Value** | | | | **Lower** | | | | | | | **Upper** | | | | | | |
| Odds ratio |  | 0.957 | |  | | 0.817 | | | | |  | | 1.12 | | | |  | | |
| Relative risk |  | 0.978 | | ᵃ | | 0.902 | | | | |  | | 1.06 | | | |  | | |
| ᵃ rows compared | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | |

The following describes coefficients for nominal variables, i.e. association measurements. Coefficients imply degree of associations. For 2X2 cross-tabulation, phi-coefficients are used. For other types of cross-tabulations, cramer’s V is appropriate.

Contingency coefficient, phi-coefficient, and Cramer’s V all display values between 0~1, with highest relevance at 1.

| Nominal | | | |
| --- | --- | --- | --- |
|  |  |  |  |
|  | | **Value** | |
| Contingency coefficient |  | 0.00649 |  |
| Phi-coefficient |  | 0.00649 |  |
| Cramer's V |  | 0.00649 |  |
|  | | | |