## Chi-Square Tests

Nathaniel E. Helwig

Associate Professor of Psychology and Statistics
University of Minnesota


October 17, 2020

Copyright © 2020 by Nathaniel E. Helwig

## Table of Contents

1. Goodness of Fit
2. Tests of Association (for 2-way Tables)
3. Conditional Association Tests (for 3-way Tables)

## Table of Contents

1. Goodness of Fit

## 2. Tests of Association (for 2-way Tables)

## 3. Conditional Association Tests (for 3-way Tables)

## A Primer on Categorical Data Analysis

In the previous chapter, we looked at inferential methods for a single proportion or for the difference between two proportions.

In this chapter, we will extend these ideas to look more generally at contingency table analysis.

All of these methods are a form of "categorical data analysis", which involves statistical inference for nominal (or categorial) variables.

## Categorical Data with $J>2$ Levels

Suppose that $X$ is a categorical (i.e., nominal) variable that has $J$ possible realizations: $X \in\{0, \ldots, J-1\}$. Furthermore, suppose that

$$
P(X=j)=\pi_{j}
$$

where $\pi_{j}$ is the probability that $X$ is equal to $j$ for $j=0, \ldots, J-1$.

Assume that the probabilities satisfy $\sum_{j=0}^{J-1} \pi_{j}=1$, so that $\left\{\pi_{j}\right\}_{j=0}^{J-1}$ defines a valid probability mass function for the random variable $X$.

- If $J=2$, this is a Bernoulli distribution
- If $J>2$, this is a multinomial distribution


## Testing Hypotheses about Multiple Probabilities

Suppose we have a sample of data $x_{1}, \ldots, x_{n}$ where $x_{i} \stackrel{\text { iid }}{\sim} F$ with $F$ denoting the probability mass function defined by $\left\{\pi_{j}\right\}_{j=0}^{J-1}$.

We want to test a null hypothesis of the form $H_{0}: \pi_{j}=\pi_{j 0} \forall j$ versus the alternative hypothesis $H_{1}:(\exists j)\left(\pi_{j} \neq \pi_{j 0}\right)$.

- The symbol $\forall$ should be read as "for all"
- The symbol $\exists$ should be read as "there exists"
$H_{0}$ states that the probability for the $j$-th category is equal to $\pi_{j 0}$ for all $j=0, \ldots, J-1$, and $H_{1}$ states that there exists at least one category where the probability for the $j$-th category is not equal to $\pi_{j 0}$.


## Observed and Expected Frequencies

Given an iid sample of $n$ observations of the random variable $X$, the observed frequency for the $j$-th category is given by

$$
f_{j}=\sum_{i=1}^{n} I\left(x_{i}=j\right)
$$

for $j=0, \ldots, J-1$, which is the number of $x_{i}$ belonging to category $j$.

Given an iid sample of $n$ observations of the random variable $X$, the expected frequency for the $j$-th category is given by

$$
m_{j}=n \pi_{j}
$$

for $j=0, \ldots, J-1$, which is the sample size $n$ multiplied by $\pi_{j}$.

## Test Statistic

To test $H_{0}: \pi_{j}=\pi_{j 0} \forall j$ versus $H_{1}:(\exists j)\left(\pi_{j} \neq \pi_{j 0}\right)$, define

$$
X^{2}=\sum_{j=0}^{J-1} \frac{\left(f_{j}-m_{j 0}\right)^{2}}{m_{j 0}}
$$

where $m_{j 0}=n \pi_{j 0}$ is the expected frequency assuming that $H_{0}$ is true.

If the observed frequencies are far from what would be expected assuming $H_{0}$ is true, then the value of $X^{2}$ would be relatively large.

As the sample size $n \rightarrow \infty$, the $X^{2}$ test statistic approaches a $\chi^{2}$ distribution with $J-1$ degrees of freedom (Pearson, 1900).

## Example 1: Vaping Prevalence

Suppose a researcher is interested in studying the prevalence of vaping among college students in the United States.

The researcher asked a random sample of $n=1000$ college students the typical number of pods they vape per day, and finds the following:

| \# Pods | Observed Frequency | Expected Frequency |
| :---: | :---: | :---: |
| 0 | 780 | 830 |
| $0-1$ | 140 | 110 |
| $1-2$ | 60 | 50 |
| $>2$ | 20 | 10 |

Expected frequencies are based on last year's data, which found: $\pi_{00}=0.83, \pi_{10}=0.11, \pi_{20}=0.05$, and $\pi_{30}=0.01$.

## Example 1: Vaping Prevalence (continued)

To test the null hypothesis $H_{0}: \pi_{j}=\pi_{j 0} \forall j$ versus the alternative hypothesis $H_{1}:(\exists j)\left(\pi_{j} \neq \pi_{j 0}\right)$, the $X^{2}$ test statistic is
$X^{2}=\frac{(780-830)^{2}}{830}+\frac{(140-110)^{2}}{110}+\frac{(60-50)^{2}}{50}+\frac{(20-10)^{2}}{10}=23.19387$

Comparing this to a $\chi^{2}$ distribution with three degrees of freedom

$$
p=P\left(\chi_{3}^{2}>23.19387\right)=0.00003679=3.679 \times 10^{-5}
$$

Thus, we have reason to suspect that the prevalence of vaping this year is significantly different from last year.

## Example 1: Vaping Prevalence (in R)

> f <- c(780, 140, 60, 20)
$>\mathrm{piO}<-\mathrm{c}(0.83,0.11,0.05,0.01)$
> chisq.test(f, p = piO)

Chi-squared test for given probabilities
data: f
X-squared $=23.194, \mathrm{df}=3, \mathrm{p}$-value $=3.679 \mathrm{e}-05$

## Table of Contents

## 1. Goodness of Fit

2. Tests of Association (for 2-way Tables)

## 3. Conditional Association Tests (for 3-way Tables)

## Joint Distribution for Two Categorical Variables

Suppose that $A$ and $B$ are both categorical (i.e., nominal) variables with $A \in\{1, \ldots, J\}$ and $B \in\{1, \ldots, K\}$.

Furthermore, suppose that...

- $P(A=j)=\pi_{j}$. for $j=1, \ldots, J$ (with $\left.\sum_{j=1}^{J} \pi_{j} .=1\right)$
- $P(B=k)=\pi_{\cdot k}$ for $k=1, \ldots, K\left(\right.$ with $\left.\sum_{k=1}^{K} \pi_{\cdot k}=1\right)$

And assume that the probability of observing the joint event is given by

$$
P(A=j \cap B=k)=\pi_{j k}
$$

where the joint probabilities satisfy $\sum_{j=1}^{J} \sum_{k=1}^{K} \pi_{j k}=1$.

## $2 \times 2$ Contingency Table

Given an iid sample of $n$ observations $\left(a_{i}, b_{i}\right) \stackrel{\text { iid }}{\sim} F$ from the joint probability distribution $F$, the $J \times K$ table that contains the observed frequency for each combination of $A$ and $B$ is referred to as a contingency table, which is also known as a cross tabulation.

|  | $B=1$ | $B=2$ | $\cdots$ | $B=k$ | $\cdots$ | $B=K$ | Row Totals |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $A=1$ | $f_{11}$ | $f_{12}$ | $\cdots$ | $f_{1 k}$ | $\cdots$ | $f_{1 K}$ | $f_{1 \cdot}$ |
| $A=2$ | $f_{21}$ | $f_{22}$ | $\cdots$ | $f_{2 k}$ | $\cdots$ | $f_{2 K}$ | $f_{2}$. |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $A=j$ | $f_{j 1}$ | $f_{j 2}$ | $\cdots$ | $f_{j k}$ | $\cdots$ | $f_{j K}$ | $f_{j}$. |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $A=J$ | $f_{J 1}$ | $f_{J 2}$ | $\cdots$ | $f_{J k}$ | $\cdots$ | $f_{J K}$ | $f_{J .}$ |
| Column Totals | $f_{\cdot 1}$ | $f_{\cdot 2}$ | $\cdots$ | $f_{\cdot k}$ | $\cdots$ | $f_{\cdot K}$ | $f . .=n$ |

## $2 \times 2$ Contingency Table (continued)

In the previous table, $f_{j k}$ denotes the number of observations that are cross-classified in category $j$ of variable $A$ and category $k$ of variable $B$

$$
f_{j k}=\sum_{i=1}^{n} I\left(a_{i}=j\right) I\left(b_{i}=k\right)
$$

for all $j, k \in\{1, \ldots, J\} \times\{1, \ldots, K\}$.

The row and column totals are the marginal observed frequencies for variables $A$ and $B$, which are defined as

$$
\begin{aligned}
& f_{j \cdot}=\sum_{i=1}^{n} I\left(a_{i}=j\right)=\sum_{k=1}^{K} f_{j k} \\
& f_{\cdot k}=\sum_{i=1}^{n} I\left(b_{i}=k\right)=\sum_{j=1}^{J} f_{j k}
\end{aligned}
$$

for all $j \in\{1, \ldots, J\}$ and all $k \in\{1, \ldots, K\}$.

## Test of Statistical Independence

To test if $A$ and $B$ are independent of one another, we want to test $H_{0}: \pi_{j k}=\pi_{j .} \pi_{\cdot k} \forall j, k$ versus $H_{1}:(\exists j, k)\left(\pi_{j k} \neq \pi_{j} . \pi_{\cdot k}\right)$.

Assuming that $H_{0}$ is true, the expected number of observations cross-classified in cell $(j, k)$ of the contingency table would be

$$
m_{j k}=n \pi_{j .} \pi_{\cdot k}
$$

which is the sample size $n$ multiplied by the marginal probabilities.

Use $\hat{m}_{j k}=n \hat{\pi}_{j} \cdot \hat{\pi}_{\cdot k}$ in practice (because $\pi_{j}$. and $\pi_{\cdot k}$ are unknown)

## Test Statistic

Given the estimates of the expected cell counts (assuming $H_{0}$ is true), we can use the chi-square test statistic

$$
X^{2}=\sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left(f_{j k}-\hat{m}_{j k}\right)^{2}}{\hat{m}_{j k}}
$$

to test the null hypothesis of independence between $A$ and $B$.

Assuming that $H_{0}$ is true, we have that $X^{2} \dot{\sim} \chi_{(J-1)(K-1)}^{2}$ as $n \rightarrow \infty$, which is Pearson's chi-square test for association (Pearson, 1900)

The chi-square approximation is because, assuming $H_{0}$ is true, we have

$$
z_{j k}=\frac{f_{j k}-\hat{m}_{j k}}{\sqrt{\hat{m}_{j k}}} \xrightarrow{d} N(0,1)
$$

as the sample size $n \rightarrow \infty$.

## Example 2: Race and The Death Penalty

This contingency table is from Table 4 of Radelet and Pierce (1991), which cross-classifies individuals by race and death penalty sentence:

|  | Death Penalty |  |  |
| :---: | :---: | :---: | :---: |
| Defendant | Yes | No | Total |
| White | 53 | 430 | 483 |
| Black | 15 | 176 | 191 |
| Total | 68 | 606 | 674 |

Suppose that we want to test the null hypothesis that race and death penalty sentence are independent.

## Example 2: Race and The Death Penalty (continued)

The marginal probability estimates for the rows are

$$
\hat{\pi}_{1} .=483 / 674=0.7166172 \quad \text { and } \quad \hat{\pi}_{2 .}=191 / 674=0.2833828
$$

and the marginal probability estimates for the columns are

$$
\hat{\pi}_{\cdot 1}=68 / 674=0.1008902 \quad \text { and } \quad \hat{\pi}_{\cdot 2}=606 / 674=0.8991098
$$

This implies that the null hypothesized (estimates of the) expected frequency for each cell is

|  | Death Penalty |  |
| :---: | :---: | :---: |
| Defendant | Yes | No |
| White | 48.72997 | 434.27 |
| Black | 19.27003 | 171.73 |

## Example 2: Race and The Death Penalty (continued)

The Pearson's chi-square test statistic is given by

$$
\begin{aligned}
X^{2} & =\frac{(53-48.72997)^{2}}{48.72997}+\frac{(430-434.27)^{2}}{434.27}+\frac{(15-19.27003)^{2}}{19.27003}+\frac{(176-171.73)^{2}}{171.73} \\
& =1.468519
\end{aligned}
$$

Comparing this to a $\chi_{1}^{2}$ distribution, the p-value for testing $H_{0}$ is

$$
p=P\left(\chi_{1}^{2}>1.468519\right)=0.2255796
$$

Don't have sufficient evidence to reject the null hypothesis that race and death penalty sentence are independent.

## Example 2: Race and The Death Penalty (in R)

We can confirm this result using the chisq.test function in $R$ :
> xtab <- matrix $(c(53,15,430,176), 2,2)$
> colnames(xtab) <- c("Yes", "No")
> rownames(xtab) <- c("White", "Black")
> xtab

| Yes |  |  |  | No |
| :--- | ---: | ---: | :---: | :---: |
| White | 53 | 430 |  |  |
| Black | 15 | 176 |  |  |
| $>$ | chisq. test (xtab, correct $=$ FALSE) |  |  |  |

Pearson's Chi-squared test
data: xtab
X-squared $=1.4685, \mathrm{df}=1, \mathrm{p}$-value $=0.2256$

## Relation to Testing Proportion Difference

The result of Pearson's chi-squared test is the exact same as the asymptotic test for the difference in proportions that was conducted in the previous chapter (via the prop.test function).

For a $2 \times 2$ contingency table, it can be shown that the $X^{2}$ test statistic is identical to the $Z^{2}$ test statistic that was used for the asymptotic test of the difference between two proportions.

Testing the null hypothesis $H_{0}: \pi_{j k}=\pi_{j} \cdot \pi_{\cdot k} \forall j, k$ is equivalent to testing the null hypothesis $H_{0}: \pi_{1} .=\pi_{2}$.

- $A$ and $B$ being independent implies equal probabilities of success


## Table of Contents

## 1. Goodness of Fit

2. Tests of Association (for 2-way Tables)
3. Conditional Association Tests (for 3-way Tables)

## Three-Way Contingency Tables

A three-way contingency table cross-classifies observations on three different categorical variables, such as the below table.

Previously, we looked at a 2-way table (Defendant's Race by Death Penalty), however these data could actually be arranged into a 3-way table where the third variable is the Victim's Race.

Note that all of these defendants were on trial for committing multiple homicides, so we can look at how both the defendant's and victim's race affects the probability of receiving a death penalty sentence.

## Example of Three-Way Table

|  | Death Penalty |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Victim | Defendant | Yes | No | Total |
| White | White | 53 | 414 | 467 |
|  | Black | 11 | 37 | 48 |
| Black | White | 0 | 16 | 16 |
|  | Black | 4 | 139 | 143 |
| Total | White | 53 | 430 | 483 |
|  | Black | 15 | 176 | 191 |

$A=$ Defendant's Race
$B=$ Death Penalty Verdict
$C=$ Victim's Race

## Marginal and Partial Tables

The previous two-way table that we looked at was a "marginal table", which is formed by aggregating the three-way table across the levels of one variable (in this case Victim's Race).

We could also look at the two relevant "partial tables", which are two-way slices of the three-way table.

| White Victim <br> Defendant | Death Penalty <br> Yes |  |  |
| :---: | :---: | :---: | :---: |
| No | Total |  |  |
| White | 53 | 414 | 467 |
| Black | 11 | 37 | 48 |
| Total | 64 | 451 | 515 |


| Black Victim | Death Penalty |  |  |
| :---: | :---: | :---: | :---: |
| Defendant | Yes | No | Total |
| White | 0 | 16 | 16 |
| Black | 4 | 139 | 143 |
| Total | 4 | 155 | 159 |

## Marginal and Conditional Independence

Variables $A$ and $B$ are said to be marginally independent if they are independent after aggregating across the levels of a third variable $C$.

Variables $A$ and $B$ are said to be conditionally independent given $C$ if they are independent at all levels of the third variable $C$.

Marginal independence does not imply conditional independence. Conditional independence does not imply marginal independence.

- Simpson's paradox: when marginal and conditional effects differ


## Example 3: Conditional Independence Test

Previously, we tested the null hypothesis that $A$ (Defendant's Race) and $B$ (Death Sentence Verdict) are marginally independent after collapsing across both of the levels of $C$ (Victim's Race).

- No evidence to reject the null hypothesis of independence

We will test the conditional independence between $A$ and $B$ given $C$.

- $H_{0}: \pi_{j k(\ell)}=\pi_{j \cdot(\ell)} \pi_{\cdot k(\ell)} \forall j, k, \ell$
- $H_{1}:(\exists j, k, \ell)\left(\pi_{j k(\ell)} \neq \pi_{j \cdot(\ell)} \pi_{\cdot k(\ell)}\right)$
- $\ell \in\{1,2\}$ denotes the Victim's Race


## Example 3: White Victim

When the victim is white, the estimated probabilities of receiving the death penalty are

$$
\hat{\pi}_{1 \cdot(1)}=53 / 467=0.1134904 \quad \text { and } \quad \hat{\pi}_{2 \cdot(1)}=11 / 48=0.2291667
$$

Note that the probability of receiving the death penalty is twice as high for black defendants when the victim is white.

```
> white.victim <- matrix(c(53, 11, 414, 37), 2, 2)
> colnames(white.victim) <- c("Yes", "No")
> rownames(white.victim) <- c("White", "Black")
> chisq.test(white.victim, correct = FALSE)
Pearson's Chi-squared test
data: white.victim
X-squared = 5.3518, df = 1, p-value = 0.0207
```


## Example 3: Black Victim

When the victim is black, the estimated probabilities of receiving the death penalty are

$$
\hat{\pi}_{1 \cdot(2)}=0 / 16=0 \quad \text { and } \quad \hat{\pi}_{2 \cdot(2)}=4 / 143=0.02797203
$$

```
> black.victim <- matrix(c(0, 4, 16, 139), 2, 2)
> colnames(black.victim) <- c("Yes", "No")
> rownames(black.victim) <- c("White", "Black")
> chisq.test(black.victim, correct = FALSE)
Pearson's Chi-squared test
data: black.victim
X-squared = 0.4591, df = 1, p-value = 0.498
Warning message:
In chisq.test(black.victim, correct = FALSE) :
    Chi-squared approximation may be incorrect
```


## Example 3: Conclusions

Regardless of whether we can trust the results for the black victims...

It seems apparent that the Defendant's Race and the Death Penalty Sentence are not conditionally independent given the Victim's Race.

Evidence suggests that black defendants are more likely than white defendants to receive the death penalty if the homicide victim is white.

## References

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine 50(302), 157-175.
Radelet, M. L. and G. L. Pierce (1991). Choosing those who will die: Race and the death penalty in Florida. Florida Law Review $43(1), 1-34$.

