Topics in statistics III/IV (MATH3361/4071) Michaelmas term, 2018-2019

Handouts: Contingency Tables [a, b, c]

Lecturer: Georgios Karagiannis georgios.karagiannis@durham.ac.uk

[a] Basic reading list: [1, 4, 2, 3, in the Section References]
[b] Author: Georgios P. Karagiannis.
[c] Acknowledgments to P. J. Craig for proofreading, and students in 2018 for spotting typos in the handouts.

1 $I \times J$ way contingency tables

The set-up

- Let X and Y denote two categorical response variables (also called classifiers)
- X has I categories and Y with J categories.
- Classifications of subjects on both variables have IJ possible combinations.

The objective

- We are interested in performing inference on the probabilities

$$\pi_{ij} = \Pr(X = i, Y = j),$$

  $i = 1, ..., I$, $j = 1, ..., J$ of responses (X, Y) of a subject chosen randomly from some population.
- E.g., the distribution $\pi_{ij} = \Pr(X = i, Y = j)$ of responses (X, Y) of a subject chosen randomly from some population.

We define the theoretical probabilities:

- $\pi_{ij} = \Pr(X = i, Y = j)$ is the probability getting an outcome $(X = i, Y = j)$
- $\pi_{i+} = \sum_j \pi_{ij} = \Pr(X = i)$ is the marginal probability of getting an outcome $X = i$ regardless the outcome of Y
- $\pi_{+j} = \sum_i \pi_{ij} = \Pr(Y = j)$ is the marginal probability of getting an outcome $Y = j$ regardless the outcome of X
- $\pi_{++} = \sum_i \sum_j \pi_{ij}$

One can tabulate the distribution $\pi_{ij} = \Pr(X = i, Y = j)$ of responses (X, Y) as in the following classification table (Table 1).

		Y					total
		1	$\cdots$	j	$\cdots$	J	
X	1	π_{11}	$\cdots$	π_{1j}	$\cdots$	π_{1J}	π_{1+}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	i	π_{i1}	$\cdots$	π_{ij}	$\cdots$	π_{iJ}	π_{i+}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	π_{I1}	$\cdots$	π_{Ij}	$\cdots$	π_{IJ}	π_{I+}
total		π_{+1}	$\cdots$	π_{+j}	$\cdots$	π_{+J}	π_{++}

Table 1: The distribution $\pi_{ij} = \Pr(X = i, Y = j)$ of responses (X, Y) in a IJ-table.

To learn the theoretical probabilities in Table 1, we collect a sample, (by using a sampling scheme from Section 1.1), and compute $(n_{i,j};\ \forall i = 1, ..., I,\ \forall j = 1, ..., J)$ where n_{ij} is the observed number of outcomes $(X = i, Y = j)$. We can specify related quantities, as:

- n_{ij} the observed number (or counts, frequency) of outcomes $(X = i, Y = j)$
- $n_{i+} = \sum_j n_{ij}$ is the observed marginal counts of outcomes $X = i$ regardless the outcome of Y
- $n_{+j} = \sum_i n_{ij}$ is the observed marginal counts of outcomes $Y = j$ regardless the outcome of X
- $n_{++} = \sum_{i,j} n_{ij}$ is the observed total counts (aka) the number of observations

Here, the $I \times J$ contingency table of X and Y [1] of observable frequencies $(n_{i,j})$, associated to the joint probabilities $(\pi_{i,j})$, is the rectangular table

- which has I rows for categories of X and J columns for categories of Y
- whose cells represent the IJ possible outcomes of the responses (X, Y)
- which displays $(n_{i,j})$ the observed number of outcomes $(X = i, Y = j)$ for each case

A schematic of the $(n_{i,j})$ contingency table displaying counts is presented in Table 2.

		Y					total
		1	$\cdots$	j	$\cdots$	J	
X	1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	n_{1+}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	n_{i+}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	n_{I1}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	n_{I+}
total		n_{+1}	$\cdots$	n_{+j}	$\cdots$	n_{+J}	n_{++}

Table 2: A $I \times J$ Contingency table of X and Y displaying the the observed number of outcomes $(X = i, Y = j)$

[1] also called (n_{ij})-contingency table, or 2-way contingency table, or classification table

In a similar manner, a contingency table / classification table can display the following quantities

- $p_{i,j} = \frac{n_{i,j}}{n_{+,+}}$ the proportion of the outcomes $(X = i, Y = j)$,
- $\mu_{i,j}$ the expected number of outcomes $(X = i, Y = j)$,
- $N_{i,j}$ the number of outcomes $(X = i, Y = j)$ (before the observations are collected)

When variable Y is treated as a response and variable X is treated as an explanatory variable, contingency tables can display the following conditional quantities

- $\pi_{i|j} = \frac{\pi_{ij}}{\pi_{+j}} = \Pr(X = i|Y = j)$: the conditional probability getting an outcome $X = i$ given that the outcome $Y = j$
- $\pi_{j|i} = \frac{\pi_{ij}}{\pi_{i+}} = \Pr(Y = j|X = i)$: the conditional probability getting an outcome $Y = j$ given that the outcome $X = i$
- $p_{i|j} = \frac{p_{ij}}{p_{i+}} = \frac{n_{ij}}{n_{i+}}$: the proportion of the outcomes $X = i$, given that $Y = j$
- $p_{j|i} = \frac{p_{ij}}{p_{+j}} = \frac{n_{ij}}{n_{+j}}$: the proportion of the outcomes $Y = j$, given that $X = i$

Two schematics of tables representing the conditional probabilities $\pi_{j|i}$, and conditional proportions $p_{j|i}$ of the outcomes $Y = j$, given that $X = i$ (aka conditioning on the rows) are in Tables 3 and 4.

$\pi_{j\|i} = \frac{\pi_{ij}}{\pi_{i+}}$		Y					total
		1	$\cdots$	j	$\cdots$	J	
X	1	$\pi_{1\|1}$	$\cdots$	$\pi_{j\|1}$	$\cdots$	$\pi_{J\|1}$	$\pi_{+\|1} = 1$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	i	$\pi_{1\|i}$	$\cdots$	$\pi_{j\|i} = \frac{\pi_{ij}}{\pi_{i+}}$	$\cdots$	$\pi_{J\|i}$	$\pi_{+\|i} = 1$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	$\pi_{1\|I}$	$\cdots$	$\pi_{j\|I}$	$\cdots$	$\pi_{J\|I}$	$\pi_{+\|I} = 1$

Table 3: A table displaying the conditional proportions $\pi_{j|i}$ of the outcomes $Y = j$, given that $X = i$

$p_{j\|i} = \frac{n_{ij}}{n_{i+}}$		Y					total
		1	$\cdots$	j	$\cdots$	J	
X	1	$p_{1\|1}$	$\cdots$	$p_{j\|1}$	$\cdots$	$p_{J\|1}$	$\pi_{+\|1} = 1$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	i	$p_{1\|i}$	$\cdots$	$p_{j\|i} = \frac{n_{ij}}{n_{i+}}$	$\cdots$	$p_{J\|i}$	$\pi_{+\|i} = 1$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	$p_{1\|I}$	$\cdots$	$p_{j\|I}$	$\cdots$	$p_{J\|I}$	$\pi_{+\|I} = 1$

Table 4: A table displaying the conditional proportions $p_{j|i}$ of the outcomes $Y = j$, given that $X = i$

Example 1. (Smoking vs. depression)

Consider a sample of size 3213 collected in 1980-1983 in St. Luis Epidemic Survey.
Assume Classification variables

X: Ever smoked (Yes, No)

Y: Major depression (Yes, No)

The observed counts are tabulated below.

	n_{ij}	Y: yes	Y: no	total
X	yes	$n_{11} = 144$	$n_{12} = 1729$	$n_{1+} = 1873$
	no	$n_{21} = 50$	$n_{22} = 1290$	$n_{2+} = 1340$
total		$n_{+1} = 194$	$n_{+2} = 3019$	$n_{++} = 3213$

The joint proportions are

	p_{ij}	Y: yes	Y: no	total
X	yes	$\frac{144}{3213} = 0.04$	0.53	0.58
	no	0.01	0.4	0.41
total		0.06	0.93	1

The conditional proportions are

	$p_{j\|i} = \frac{n_{ij}}{n_{i+}}$	Y: yes	Y: no	total
X	yes	$\frac{n_{11}}{n_{1+}} = \frac{144}{1873} = 0.076$	0.92	1
	no	0.037	0.96	1

1.1 Sampling schemes

The observations (data set) are generated from sampling distributions specified by the sampling scheme (aka experiment) performed.

1.1.1 Poisson sampling scheme

Poisson sampling describes the scenario where the data-set (n_{ij}) is collected (sampled) independently in a given spatial, temporal, or other interval.
Then, n_{ij} is a realization of a Poisson random variable

$$N_{ij} \overset{\text{ind}}{\sim} \text{Poi}(\mu_{ij})$$

where

$$\Pr(N_{ij} = n_{ij}|\mu_{ij}) = \frac{\exp(-\mu_{ij})\mu_{ij}^{n_{ij}}}{n_{ij}!}$$

with

$$\text{E}(N_{ij}) = \mu_{ij} \qquad \text{and} \qquad \text{Var}(N_{ij}) = \mu_{ij}$$

for $i = 1, ..., I$, and $j = 1, ..., J$.
The likelihood is

$$L(\boldsymbol{\mu}) \quad = \quad \prod_{\forall i,j} \frac{\exp(-\mu_{ij})\mu_{ij}^{n_{ij}}}{n_{ij}!} \quad \propto \quad \exp(-\sum_{\forall i,j} \mu_{ij}) \prod_{\forall i,j} \mu_{ij}^{n_{ij}} \tag{1}$$

where $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, ..., \mu_{IJ})$.

1.1.2 Multinomial sampling scheme

Multinomial sampling scheme describes the scenario where the data-set (n_{ij}) is collected (sampled) idependently given that the total sample size n_{++} is fixed (aka predetermined). Then $\boldsymbol{n} = (n_{11}, n_{12}, ..., n_{IJ})$ is a realization of a Multinomial random variable

$$\underbrace{(N_{11}, N_{12}, ...N_{IJ})}_{=\boldsymbol{N}} \overset{\text{ind}}{\sim} \text{Mult}(n_{++}, \underbrace{(\pi_{11}, ..., \pi_{IJ})}_{=\boldsymbol{\pi}}) \tag{2}$$

where

$$\Pr(\boldsymbol{N} = \boldsymbol{n} | n_{++}, \boldsymbol{\pi}) = \frac{n_{++}!}{\prod_{\forall i,j} n_{ij}!} \prod_{\forall i,j} \pi_{ij}^{n_{ij}} \tag{3}$$

$\boldsymbol{N} = (N_{11}, N_{12}, ...N_{IJ})$ and $\boldsymbol{\pi} = (\pi_{11}, ..., \pi_{IJ})$ with

$$\text{E}(\boldsymbol{N}) = n_{++}\boldsymbol{\pi} \qquad \text{and} \qquad \text{Var}(\boldsymbol{N}) = n_{++}(\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T)$$

The likelihood is

$$L(\boldsymbol{\pi}) \quad = \Pr(\boldsymbol{N} = \boldsymbol{n} | n_{++}, \boldsymbol{\pi}) \quad = \frac{n_{++}!}{\prod_{\forall i,j} n_{ij}!} \prod_{\forall i,j} \pi_{ij}^{n_{ij}} \quad \propto \prod_{\forall i,j} \pi_{ij}^{n_{ij}} \tag{4}$$

as n_{++} is predetermined and fixed. Notice that if $\mu_{ij} = n_{++}\pi_{ij}$

$$L(\boldsymbol{\pi}) \quad \propto \prod_{\forall i,j} (\frac{\mu_{ij}}{n_{++}})^{n_{ij}} \quad \propto \prod_{\forall i,j} \mu_{ij}^{n_{ij}} \tag{5}$$

Relation between Multinomial and Poisson sampling.

We can produce the Multinomial distribution (describing the above experiment) based on the Poisson distr.

Theorem 2. *If r.v. $Y_i \sim Poi(\mu_i)$ for $i = 1, ..., k$ and they are independent then*

$$\Pr(Y_1 = y_1, ..., Y_k = y_k | \sum_{i=1}^{k} Y_i = n) = \frac{n!}{\prod_{i=1}^{k} y_i!} \prod_{i=1}^{k} \pi_i^{y_i} 1(y \in \mathcal{Y})$$

where $\mathcal{Y} = \{y \in \{0, ..., n\}^k | \sum_{i=1}^{k} \pi_i\}$, and $\pi_i = \mu_i / n$.

Proof. It is easy to show that $\sum_{i=1}^{k} Y_i \sim \text{Poi}(\sum_{i=1}^{k} \mu_i)$ —(see Probability I). Then

$$\Pr(Y_1, ..., Y_k | \sum_{i=1}^{k} Y_i = n) = \frac{\Pr(Y_1 = y_1, ..., Y_k = y_k) 1(y \in \mathcal{Y})}{\Pr(\sum_{i=1}^{k} Y_i = n_{++})} = \frac{\prod_{\forall i} \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!} 1(y \in \mathcal{Y})}{\prod_{\forall i,j} \frac{\exp(-\sum_{i=1}^{k}\mu_i)(\sum_{i=1}^{k}\mu_i)^n}{n_{++}!}}$$

$$= \frac{n!}{\prod_{i=1}^{k} y_i!} \prod_{i=1}^{k} \underbrace{(\mu_i/n)}_{=\pi_i}{}_i^{y_i} 1(y \in \mathcal{Y})$$

□

1.1.3 Product Multinomial sampling scheme

Product Multinomial sampling scheme describe the scenario where the sample (n_{ij}) is collected randomly and given that the marginal row size n_{i+} is fixed (aka predetermined) for $i = 1, ..., I$. Then $(n_{i1}, n_{i2}, ...n_{iJ})$ is a realization of the Multinomial random variable

$$(N_{i1}, N_{i2}, ...N_{iJ}) \stackrel{\text{ind}}{\sim} \text{Mult}(n_{i+}, \boldsymbol{\pi}_i^*) \tag{6}$$

where $\boldsymbol{\pi}_i^* = (\pi_{1|i}, ..., \pi_{J|i})$ for $i = 1, ..., I$. In (6), 'ind' means that there is a Multinomial distribution for each row i and that rows are independent each other.

The likelihood is

$$L(\boldsymbol{\pi}) \quad = \prod_{\forall i} \left[\frac{n_{i+}!}{\prod_{\forall j} n_{ij}!} \prod_{\forall j} (\pi_i^*)^{n_{ij}} \right] \quad \propto \prod_{\forall i,j} \pi_{ij}^{n_{ij}}$$

as n_{i+} is predetermined and fixed. Notice that if $\mu_{ij} = n_{++}\pi_{ij}$

$$L(\boldsymbol{\pi}) \quad \propto \prod_{\forall i,j} (\frac{\mu_{ij}}{n_{++}})^{n_{ij}} \quad \propto (\frac{1}{n_{++}})^{\sum_{\forall i,j} n_{i,j}} \prod_{\forall i,j} \mu_{ij}^{n_{ij}} \quad \propto \prod_{\forall i,j} \mu_{ij}^{n_{ij}} \tag{7}$$

as the sum $\sum_{\forall i,j} n_{i,j} = \sum_{\forall i} \sum_{\forall i} n_{i,j} = \sum_{\forall j} n_{i,+}$ is fixed as a sum of predetermined marginal row sizes n_{i+}.

Remark 3. The three sampling schemes lead to the same MLE for $\boldsymbol{\mu}$'s or $\boldsymbol{\pi}$'s given that $\sum_{\forall i,j} \mu_{ij} = n_{++}$ is fixed. This is obvious in Multinomial sampling scheme and product Multinomial sampling scheme from 5 and 7. Regarding the Poisson, it is $\sum_{\forall i,j} \mu_{ij} = n_{++}$ and hence (1) becomes $L(\underset{\sim}{\mu}) \propto \prod_{\forall i,j} \mu_{ij}^{n_{ij}}$.

1.2 Maximum likelihood estimates

Lagrange Multipliers[a]

Problem: Let functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g_k : \mathbb{R}^n \to \mathbb{R}$, for $k = 1, ..., K$.

$$\begin{cases} \text{maximize} & f(x) \\ \text{subject to} & g_k(x) = 0,\ \forall k = 1, ..., K \end{cases} \tag{8}$$

Assume ∇f, ∇g_1, ..., ∇g_K are continuous

Define Lagrange function is

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{k=1}^{K} \lambda_k g_k(x)$$

where $\lambda = (\lambda_1, ..., \lambda_K)$.

Solution The solution x^* of (8) is the solution of

$$0 = \nabla_{x,\lambda} \mathcal{L}(x, \lambda)|_{(x,\lambda)=(x^*,\lambda^*)}$$

[a] **Please revise your calculus.**

Consider a Multinomial sampling scheme, and that X and Y are independent. We need to find the MLE of $\boldsymbol{\pi}$.

Because X and Y are independent, I have

$$\pi_{ij} = \pi_{i+}\pi_{+j}$$

The log likelihood is

$$\ell(\pi) = \sum_{ij} n_{ij} \log(\pi_{ij}) \ = \sum_{i} n_{i+} \log(\pi_{i+}) + \sum_{j} n_{+j} \log(\pi_{+j}) + \text{const}$$

Then Lagrange function is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda}) &= \ell(\boldsymbol{\pi}) - \lambda_1(\sum_i \pi_{i+} - 1) - \lambda_2(\sum_j \pi_{+j} - 1) \\ &= \sum_i n_{i+} \log(\pi_{i+}) + \sum_j n_{+j} \log(\pi_{+j}) - \lambda_1(\sum_i \pi_{i+} - 1) - \lambda_2(\sum_j \pi_{+j} - 1) \end{aligned}$$

It is

$$0 = \nabla_{\pi,\lambda} \mathcal{L}(\pi, \lambda)|_{(\pi,\lambda)=(\pi^*,\lambda^*)}$$

namely

$$\begin{cases} 0 = & \frac{\mathrm{d}}{\mathrm{d}\pi_{i+}} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \\ 0 = & \frac{\mathrm{d}}{\mathrm{d}\pi_{+j}} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \\ 0 = & \frac{\mathrm{d}}{\mathrm{d}\lambda_1} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \\ 0 = & \frac{\mathrm{d}}{\mathrm{d}\lambda_2} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \end{cases} \xrightarrow{\text{calc.}} \begin{cases} \hat{\pi}_{i+} = \frac{n_{i+}}{n_{++}} \\ \hat{\pi}_{+j} = \frac{n_{+j}}{n_{++}} \\ \hat{\lambda}_1 = n_{++} \\ \hat{\lambda}_2 = n_{++} \end{cases}$$

So

$$\hat{\pi}_{ij} = \hat{\pi}_{i+}\hat{\pi}_{+j} = \frac{n_{i+}}{n_{++}} \frac{n_{+j}}{n_{++}}$$

Example 4. Consider a Multinomial sampling scheme. We need to find the MLE of π.

The log likelihood is

$$\ell(\boldsymbol{\pi}) \propto \sum_{ij} n_{ij} \log(\pi_{ij})$$

Then Lagrange function is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda}) &= \ell(\boldsymbol{\pi}) - \lambda(\sum_{ij} \pi_{ij} - 1) \\ &= \sum_{ij} n_{ij} \log(\pi_{ij}) - \lambda(\sum_{ij} \pi_{ij} - 1)\end{aligned}$$

It is

$$0 = \nabla_{\boldsymbol{\pi},\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)}$$

namely

$$\ldots \overset{\text{calc.}}{\Longrightarrow} \begin{cases} 0 = & \frac{\mathrm{d}}{\mathrm{d}\pi_{ij}} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \\ 0 = & \frac{\mathrm{d}}{\mathrm{d}\lambda_1} \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\lambda})|_{(\boldsymbol{\pi},\boldsymbol{\lambda})=(\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*)} \end{cases} \overset{\text{calc.}}{\Longrightarrow} \begin{cases} \hat{\pi}_{ij} = \frac{n_{ij}}{n_{++}} \\ \hat{\lambda} = n_{++} \end{cases}$$

So

$$\hat{\pi}_{ij} = \frac{n_{ij}}{n_{++}}$$

1.3 Goodness of fit test (Model selection)

The pair of hypothesis test is as follows

$$\begin{cases} \mathrm{H}_0 : & X, Y \text{ are independent} \\ \mathrm{H}_1 : & X, Y \text{ are not independent} \end{cases} \Longrightarrow \begin{cases} \mathrm{H}_0 : & \pi_{ij} = \pi_{i+}\pi_{+j} \\ \mathrm{H}_1 : & X, Y \text{ are not independent} \end{cases}$$

The test can be based on the following two statistics:

- Pearson's statistic

$$X^2 \overset{\mathrm{H}_0}{=} \sum_{\forall i,j} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \xrightarrow[D]{} \chi^2_{\mathrm{df}}$$

- Likelihood ratio statistic

$$G^2 \overset{\mathrm{H}_0}{=} 2 \sum_{\forall i,j} n_{ij} \log(\frac{n_{ij}}{\hat{\mu}_{ij}}) \xrightarrow[D]{} \chi^2_{\mathrm{df}}$$

- The degrees of freedom are

$$\mathrm{df} = (I - 1)(J - 1)$$

  and

$$\hat{\mu}_{ij} = \frac{n_{i+} n_{+j}}{n_{++}} : \text{the MLE of } \mu_{ij} \text{ under the model in } \mathrm{H}_0$$

The rejection areas at sig. level a are:

- For Pearson's statistic
$$R(\{n_{ij}\}) = \{X^2_{\text{obs}} \geq \chi^2_{\text{df},1-a}\}$$
- For Likelihood ratio statistic
$$R(\{n_{ij}\}) = \{G^2_{\text{obs}} \geq \chi^2_{\text{df},1-a}\}$$

The p-values are:

$$\text{pvalue} = \begin{cases} 1 - F_{\chi^2_{\text{df}}}(X^2_{\text{obs}}) & \text{, Pearson's statistic} \\ 1 - F_{\chi^2_{\text{df}}}(G^2_{\text{obs}}) & \text{, Likelihood ratio statistic} \end{cases}$$

1.4 Analysis of residuals

If H_0 is rejected, we need to investigate which cell caused this rejection.
This can be done by analyzing the residuals and detecting large deviations.

- Pearson's residuals
$$e^P_{ij} = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}$$
If Poisson sampling is performed then
$$e^P_{ij} \xrightarrow[D]{\text{H}_0} \text{N}(0, 1)$$
- Standardized residuals
$$e^S_{ij} = e^P_{ij} \frac{1}{\sqrt{v_{ij}}} \tag{9}$$
$$\approx e^P_{ij} \frac{1}{\sqrt{\hat{v}_{ij}}} = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}\sqrt{\hat{v}_{ij}}} \tag{10}$$
where
$$v_{ij} = (1 - \pi_{i+})(1 - \pi_{+j})$$
$$\hat{v}_{ij} = (1 - \frac{n_{i+}}{n_{++}})(1 - \frac{n_{+j}}{n_{++}})$$
Regardless the sampling scheme
$$e^S_{ij} \xrightarrow[D]{\text{H}_0} \text{N}(0, 1)$$
Why to standardize? The Pearson's residuals e^S_{ij} are not homogeneous if the sampling scheme is not Poisson, but $e^P_{ij} \xrightarrow[D]{\text{H}_0} \text{N}(0, v_{ij})$. Hence the adjustment in (9) standardizes them. Because v_{ij} are actually unknown parameters, they are approximated by their sampling analog in (10).

- Deviance residuals

$$e_{ij}^{D} = \text{sign}(n_{ij} - \hat{\mu}_{ij})\sqrt{2n_{ij}\log(\frac{n_{ij}}{\hat{\mu}_{ij}})} \tag{11}$$

  where

$$e_{ij}^{D} \xrightarrow[D]{\text{H}_0} \text{N}(0,1)$$

Criterion If $|e_{ij}^{S}| > z_{1-\frac{a}{2}}$ then, at sig. level. a, the (i,j)th cell may be characterized as the influential causing the rejection of H_0.

- So you actually need to plot $\{e_{ij}^{S}\}$ against (i,j).
- Note that $|e_{ij}^{S}| > |e_{ij}^{P}|$ because _______________

1.5 Odds ratio in contingency tables

Before we go into the odds ratio in $I \times J$ tables recall concepts

Relative risk: Assume 2 events X and Y with probability of success π_X and π_Y respectively. The relative risk is

$$r = \frac{\pi_X}{\pi_Y}$$

Interpretation

$$\begin{aligned} r = 1 \Longrightarrow &\pi_X = \pi_Y \Longrightarrow && X \text{ is equaly likely to happen with } Y \\ r > 1 \Longrightarrow &\pi_X > \pi_Y \Longrightarrow && X \text{ is more likely to happen than } Y \\ r < 1 \Longrightarrow &\pi_X < \pi_Y \Longrightarrow && X \text{ is less likely to happen than } Y \end{aligned}$$

Odds: Assume 1 event X with probability of success π_X. The odds of X are

$$\Omega_X = \frac{\pi_X}{1 - \pi_X}$$

Interpretation

$$\begin{aligned} \Omega_X = 1 \Longrightarrow &\pi_X = 1 - \pi_X \Longrightarrow && \text{success is equally likely to happen} \\ \Omega_X > 1 \Longrightarrow &\pi_X > 1 - \pi_X \Longrightarrow && \text{success is more likely to happen} \\ \Omega_X < 1 \Longrightarrow &\pi_X < 1 - \pi_X \Longrightarrow && \text{success is less likely to happen} \end{aligned}$$

Odds ratio: Assume 2 events X and Y with probability of success π_X and π_Y respectively. The odds ratio of X vs Y are

$$\theta = \frac{\Omega_X}{\Omega_Y} = \frac{\pi_X/(1 - \pi_X)}{\pi_Y/(1 - \pi_Y)}$$

Interpretation

$$\theta = 1 \Longrightarrow \Omega_X = \Omega_Y$$
$$\theta > 1 \Longrightarrow \Omega_X > \Omega_Y$$
$$\theta < 1 \Longrightarrow \Omega_X < \Omega_Y$$

...more details on the interpretation will be given with respect to the contingency tables in what follows.

§ **Odds ratio for 2×2 tables**

Assume 2×2 tables

		Y		total
	$\pi_{i,j}$	1	2	
X	1	π_{11}	π_{12}	π_{1+}
	2	π_{21}	π_{22}	π_{2+}
total		π_{+1}	π_{+2}	π_{++}

$\Longrightarrow$

		Y		total
	$\pi_{j\mid i}$	1	2	
X	1	$\pi_{j=1\mid i=1}$	$\pi_{j=2\mid i=1}$	1
	2	$\pi_{j=1\mid i=2}$	$\pi_{j=2\mid i=2}$	1

		Y	
	$\pi_{i\mid j}$	1	2
X	1	$\pi_{i=1\mid j=1}$	$\pi_{i=1\mid j=2}$
	2	$\pi_{i=2\mid j=1}$	$\pi_{i=2\mid j=2}$
total		1	1

The odds ratio of events $A = \{Y = 1|X = 1\}$, and $B = \{Y = 1|X = 2\}$ is

$$\begin{aligned}\theta &= \frac{\Omega_A}{\Omega_B} = \frac{\pi_{j=1|i=1}/(1-\pi_{j=1|i=1})}{\pi_{j=1|i=2}/(1-\pi_{j=1|i=2})} \\ &= \frac{\pi_{j=1|i=1}/(1-\pi_{j=1|i=1})}{\pi_{j=1|i=2}/(1-\pi_{j=1|i=2})} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \frac{\pi_{i=1|j=1}/(1-\pi_{i=1|j=1})}{\pi_{i=1|j=2}/(1-\pi_{i=1|j=2})} \end{aligned} \tag{12}$$

Interpretation

- $\theta = 1 \Longleftrightarrow \Omega_A = \Omega_B \Longleftrightarrow \pi_{j=1|i=1} = \pi_{j=1|i=2}$
  $\Longleftrightarrow$ Y is equally likely to happen regardless whether $X = 1$ or $X = 2$
  $\Longleftrightarrow$ X and Y are independent
- $\theta > 1 \Longleftrightarrow \Omega_A > \Omega_B \Longleftrightarrow \pi_{j=1|i=1} > \pi_{j=1|i=2}$
  $\Longleftrightarrow$ Y is more likely to happen if $X = 1$
  $\Longleftrightarrow$ positive dependence
- $\theta < 1 \Longleftrightarrow \Omega_A < \Omega_B \Longleftrightarrow \pi_{j=1|i=1} < \pi_{j=1|i=2}$
  $\Longleftrightarrow$ Y is less likely to happen if $X = 1$
  $\Longleftrightarrow$ negative dependence

The dependence becomes stronger as odds ratio θ moves away from 1.

Due to the in-variance property in (12), odds ratio is a suitable/consistent metric of dependence regardless which total n_{++} , n_{i+}, n_{+j} was fixed during the data collection (experimental design).

1.5.1 Odds ratio for $I \times J$ tables

- Decompose the $I \times J$ table into a minimal set of $(I-1)(J-1)$ 2×2 tables able to fully describe the problem in terms of odds ratio.
- However, this decomposition is not unique. Below, we present popular some choices.

§ **Nominal Odds ratios**

- Suitable for nominal classification variables
- They are defined in terms of a reference level, e.g. I-th and J-th
- The Nominal Odds ratios with reference levels I and J are

$$\theta_{ij}^{IJ} = \frac{\pi_{ij}/\pi_{iJ}}{\pi_{Ij}/\pi_{IJ}} = \frac{\pi_{ij}\pi_{IJ}}{\pi_{iJ}\pi_{Ij}}, \qquad \forall i = 1, ..., I-1; \qquad \forall j = 1, ...J-1$$

§ **Local odds ratios**

- Suitable for ordinal classification variables
- Compute each level of the ordinal classification tot he immediate next
- The local Odds ratios are

$$\theta_{ij}^{L} = \frac{\pi_{i,j}/\pi_{i,j+1}}{\pi_{i+1,j}/\pi_{i+1,j+1}} = \frac{\pi_{i,j}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i,j+1}}, \qquad \forall i = 1, ..., I-1; \qquad \forall j = 1, ...J-1$$

$\Longrightarrow$

	j	$j+1$
i	$*$	$*$
$i+1$	$*$	$*$

- They can show us a taste of the trend
- They can recover all the odd rations, as:

$$\theta_{i,j}^{i+k,j+\ell} = \frac{\pi_{i,j}/\pi_{i,j+\ell}}{\pi_{i+k,j}/\pi_{i+k,j+\ell}} = \prod_{\rho=0}^{k-1}\prod_{\xi=0}^{\ell-1}\theta_{i+\rho,j+\xi}^{L} \qquad \forall k = 1, ..., I-i; \qquad \forall \ell = 1, ...J-j$$

$\Longrightarrow$

	j	$j+\ell$
i	$*$	$*$
$i+k$	$*$	$*$

- Note that $\theta_{ij}^{L} = \theta_{ij}^{i+1,j+1}$
- Note that for $k = I-i$ and $\ell = J-j$ we get the nominal odds ratios with reference levels I and J.

§ § Maximum Likelihood Estimator

The MLE estimate of $\theta^L_{i,j}$ is

$$\hat{\theta}^L_{i,j} = \frac{\hat{\pi}_{i,j}\hat{\pi}_{i+1,j+1}}{\hat{\pi}_{i,j+1}\hat{\pi}_{i+1,j}} = \frac{n_{i,j}, n_{i+1,j+1}}{n_{i,j+1}n_{i+1,j}}$$

§ § Asymptotic distribution

The asymptotic distribution of $\hat{\theta}^L_{i,j}$ (given that n_{++} is large) is such that Using

$$\frac{\log(\hat{\theta}^L_{i,j}) - \log(\theta^L_{i,j})}{\sqrt{\frac{1}{n_{i,j}} + \frac{1}{n_{i,j+1}} + \frac{1}{n_{i+1,j}} + \frac{1}{n_{i+1,j+1}}}} \xrightarrow[D]{} \mathrm{N}(0,1)$$

§ § Confidence interval

The $(1-a)$ CI of $\log(\theta^L_{i,j})$ is

$$(\log(\hat{\theta}^L_{i,j}) \pm z_{1-\frac{a}{2}}\sqrt{\frac{1}{n_{i,j}} + \frac{1}{n_{i,j+1}} + \frac{1}{n_{i+1,j}} + \frac{1}{n_{i+1,j+1}}})$$

§ § Hypothesis test

- pair of hypothesis: $\mathrm{H}_0 : \theta^L_{i,j} = \theta_0$ v.s. $\mathrm{H}_0 : \theta^L_{i,j} \neq \theta_0$
- rejection area: $\{\left|\frac{\log(\hat{\theta}^L_{i,j}) - \log(\theta_0)}{\sqrt{\frac{1}{n_{i,j}} + \frac{1}{n_{i,j+1}} + \frac{1}{n_{i+1,j}} + \frac{1}{n_{i+1,j+1}})}}\right| \geq z_{1-\frac{a}{2}}\}$
- p-value$= 2(1 - P_{N(0,1)}(Z \leq \frac{\log(\hat{\theta}^L_{i,j}) - \log(\theta_0)}{\sqrt{\frac{1}{n_{i,j}} + \frac{1}{n_{i,j+1}} + \frac{1}{n_{i+1,j}} + \frac{1}{n_{i+1,j+1}})}}))$

Likewise for other alternative hypothesis

1.6 Independence test for small samples (Fisher's exact test)

The Hypergeometric experiment[a]

- Let, r.v. X be the number of successes in n draws without replacement from a finite population of size N, that contains exactly K objects with the feature defining success and $N-K$ objects with a feature defining failure.
- Then X is distributed according to the Hypergeometric distribution , as

  $$X \sim \mathrm{Hg}(N, K, n)$$

  with probability

  $$\Pr(X = x|N, K, n) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} 1(X \in \mathcal{X})$$

  where $\mathcal{X} = \{x \in \mathbb{N} | x \in \ (\ \max(0, n+K-N)\ ,\ \min(K, n)\ )\ \}$

[a]Notation is restricted in the scope of this box.

Assume a 2×2 table whose marginal counts n_{i+}, n_{+j} are fixed and predetermined when the experiment was perfumed. Then the random counts of the $(1,1)$th cell are distributed as

$$N_{11} \sim \mathrm{Hg}(n_{++}, n_{1+}, n_{+1})$$

with

$$\Pr(N_{11} = t | n_{++}, n_{1+}, n_{+1}) = \frac{\binom{n_{1+}}{t}\binom{n_{++}-n_{1+}}{n_{+1}-t}}{\binom{n_{++}}{n_{+1}}} 1(X \in \mathcal{X})$$

where $\mathcal{X} = \{x \in \mathbb{N} | x \in \ (\ \max(0, n_{1+} + n_{+1} - n_{++})\ ,\ \min(n_{1+}, n_{+1})\ )\ \}$.
Let

$$\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{22}}$$

Fisher's exact test:

- For the pair of hypothesis

$$\mathrm{H}_0 : \theta = 1 \text{ vs. } \mathrm{H}_1 : \theta > 1$$

  the pvalue is

$$\text{p-value} = \Pr(N_{11} \geq n_{11}) = \sum_{t=n_{11}}^{n} \Pr_{\mathrm{Hg}}(N_{11} = t)$$

- For the pair of hypothesis

$$\mathrm{H}_0 : \theta = 1 \text{ vs. } \mathrm{H}_1 : \theta < 1$$

  the pvalue is

$$\text{p-value} = \mathrm{P}(N_{11} \leq n_{11}) = \sum_{t=0}^{n_{11}} \Pr_{\mathrm{Hg}}(N_{11} = t)$$

Example 5. (Tea-Milk) Dr Bristol was claiming that she could tell if the milk was poured before the tea or if the tea was poured before the milk in the cup. Fisher did not believed her.
Fisher set 8 cups of tea. In 4 cups he poured first the tea and then the milk, while in the rest 4 cups he poured first the milk and hen the tea. Dr Bristol drank all of them and then she said which 4 has the milk poured first.
This experiment guaranteed that both the column and row marginal counts n_{i+}, n_{+j} where fixed. The 2 way observed count table is presented below

		poured first (guess)		total
	$n_{i,j}$	Milk	Tea	
poured first (reality)	Milk	3	1	4
	Tea	1	3	4
total		4	4	8

Solution. Test

$$\mathrm{H}_0: \quad \theta = 1 \,(\text{independence}); \quad \mathrm{H}_1 : \theta > 1 \,(\text{positive dependence})$$

I have $N_{11} \sim \mathrm{Hg}(n_{++} = 8, n_{1+} = 4, n_{+1} = 4)$ where $N_{11} \in \{0, 4\}$. Also I have $n_{11}^{obs} = 3$. So the pvalue is

$$\begin{aligned}
\text{p-value} =& \mathrm{P}(N_{11} \geq 3) = \sum_{t=3}^{8} \mathrm{P}_{\mathrm{Hg}}(N_{11} = t) \\
=& \mathrm{P}_{\mathrm{Hg}}(N_{11} = 3) + \mathrm{P}_{\mathrm{Hg}}(N_{11} = 4) \\
& + \cancelto{0}{\mathrm{P}_{\mathrm{Hg}}(N_{11} = 5)} + \cancelto{0}{\mathrm{P}_{\mathrm{Hg}}(N_{11} = 6)} + \cancelto{0}{\mathrm{P}_{\mathrm{Hg}}(N_{11} = 7)} + \cancelto{0}{\mathrm{P}_{\mathrm{Hg}}(N_{11} = 8)} \\
& = 0.229 + 0.014 + 0 + 0 + 0 + 0 = 0.243 > 0.05
\end{aligned}$$

So Dr. Bristol cannot tell if the Milk or the Tea was poured first in the cup, at sig. level 5%

2 $I \times J \times K$ contingency tables

Consider a $I \times J \times K$ contingency table $(n_{i,j,k})$ for $i = 1, ..., I$, $j = 1, ..., J$, and $k = 1, ..., K$, with classification variables X (the rows), Y(the columns), Z (the layers). A schematic of a $2 \times 2 \times 2$ contingency table is given in Table 5 below:

Z	X	Y		Total
		1	2	
1	1	n_{111}	n_{121}	n_{1+1}
	2	n_{211}	n_{221}	n_{2+1}
2	1	n_{112}	n_{122}	n_{1+2}
	2	n_{212}	n_{222}	n_{2+2}
Total		n_{+1+}	n_{+2+}	n_{+++}

Table 5: A schematic of a $2 \times 2 \times 2$ table

- We can define the joint probability distr. of (X, Y, Z) as

$$\pi_{ijk} = \Pr(X = i, Y = j, Z = k)$$

- Proportions, observed and random counts are defined similar to the $I \times J$ contingency table cases...

Definition 6. The XY- partial table (denoted as $(n_{ij(k)})$) results by controlling over Z and keeping it at a fixed level.
...Consider associated quantities:

- Partial probabilities

$$\pi_{ij|k} = \Pr(X = i, Y = j|Z = k) = \frac{\pi_{ijk}}{\pi_{++k}}, \forall k = 1, ..., K$$

- Partial proportions

$$p_{ij|k} = \frac{n_{ijk}}{n_{++k}}, \forall k = 1, ..., K$$

- Partial Local odd ratios conditional on Z

$$\theta^{XY}_{ij(k)} = \frac{\pi_{i,j,k}/\pi_{i,j+1,k}}{\pi_{i+1,j,k}/\pi_{i+1,j+1,k}} = \frac{\pi_{i,j,k}\pi_{i+1,j+1,k}}{\pi_{i+1,j,k}\pi_{i,j+1,k}},$$

  $i = 1, ..., I - 1$, $j = 1, ..., J - 1$, and $k = 1, ..., K$. It expresses the partial association of X and Y over level k of classification variable Z.

- MLE of Partial Local odd ratios conditional on Z

$$\hat{\theta}^{XY}_{ij(k)} = \frac{n_{i,j,k}n_{i+1,j+1,k}}{n_{i+1,j,k}n_{i+1,j,k}},$$

  $i = 1, ..., I - 1$, $j = 1, ..., J - 1$, and $k = 1, ..., K$

Definition 7. The XY- marginal table (denoted as (n_{ij})) results by collapsing XYZ -table at Z.
...Consider associated quantities:

- Marginal probabilities

$$\pi_{ij} = \Pr(X = i, Y = j) = \sum_{k=1}^{K} \pi_{ijk}$$

- Marginal proportions

$$p_{ij} = \sum_{k=1}^{K} p_{ijk}, \forall k = 1, ..., K$$

- Marginal Local odd ratios

$$\theta^{XY}_{ij} = \frac{\pi_{i,j,+}/\pi_{i,j+1,+}}{\pi_{i+1,j,+}/\pi_{i+1,j+1,+}} = \frac{\pi_{i,j,+}\pi_{i+1,j+1,+}}{\pi_{i+1,j,+}\pi_{i+1,j,+}},$$

  $i = 1, ..., I - 1$, and $j = 1, ..., J - 1$. It expresses the marginal association of X and Y by ignoring classification variable Z.

- MLE of Marginal Local odd ratios conditional on Z

$$\hat{\theta}^{XY}_{ij} = \frac{n_{i,j,+}n_{i+1,j+1,+}}{n_{i+1,j,+}n_{i+1,j,+}},$$

  $i = 1, ..., I - 1$, and $j = 1, ..., J - 1$

2.1 Types of independency in $I \times J \times K$ tables

We specify a number of important (in)dependencies modeled by $I \times J \times K$ contingency tables:

1. X, Y, Z are jointly independent iff

$$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}, \quad \forall i, j, k$$

   This type of independency is symbolized as $[X, Y, Z]$

2. Y jointly independent from X, Z iff

$$\pi_{ijk} = \pi_{+j+}\pi_{i+k}, \quad \forall i, j, k$$

   This type of independency (in 3 way contingency tables with classification variables X,Y, and Z) is symbolized as $[Y, XZ]$.

3. X and Y independent conditionally on Z iff

$$\pi_{ijk} = \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}}, \quad \forall i, j, k$$

   because,

$$\begin{aligned}
\Pr(X = i, Y = j, Z = k) &= \Pr(X = i, Y = j|Z = k)\Pr(Z = k) \\
&= \Pr(X = i, |Z = k)\Pr(Y = j|Z = k)\Pr(Z = k) \\
&= \frac{\Pr(X = i, Z = k)}{\Pr(Z = k)}\Pr(Y = j, Z = k)
\end{aligned}$$

   - If X and Y independent conditionally on Z then

$$\begin{aligned}
\theta^{XZ}_{i(j)k} =& \frac{\pi_{i,j,k}\pi_{i+1,j,k+1}}{\pi_{i+1,j,k}\pi_{i,j,k+1}} = \frac{\frac{\pi_{i,+,k}\cancel{\pi_{+,j,k}}}{\cancel{\pi_{+,+,k}}}\frac{\pi_{i+1,+,k+1}\cancel{\pi_{+,j,k+1}}}{\cancel{\pi_{+,+,k+1}}}}{\frac{\pi_{i+1,+,k}\cancel{\pi_{+,j,k}}}{\cancel{\pi_{+,+,k}}}\frac{\pi_{i,+,k+1}\cancel{\pi_{+,j,k+1}}}{\cancel{\pi_{+,+,k+1}}}} \\
=& \frac{\pi_{i,+,k}\pi_{i+1,+,k+1}}{\pi_{i+1,+,k}\pi_{i,+,k+1}} \\
=& \theta^{XZ}_{ik}, \begin{cases} \forall i = & 1, ..., I-1 \\ \forall j = & 1, ..., J \\ \forall k = & 1, ..., K-1 \end{cases}
\end{aligned}$$

     and

$$\theta^{YZ}_{(i)jk} = \theta^{YZ}_{jk}, \begin{cases} \forall i = & 1, ..., I \\ \forall j = & 1, ..., J-1 \\ \forall k = & 1, ..., K-1 \end{cases}$$

     where $\theta^{XZ}_{i(j)k}$ are the conditions XZ local odds ratios given Y, θ^{XZ}_{ik} are the marginal XZ local odds ratios, $\theta^{YZ}_{(i)jk}$ are the conditions YZ local odds ratios given X, and θ^{YZ}_{jk} are the marginal YZ local odds ratios.

It is symbolized as $[XZ, YZ]$.

4. X and Y are marginally independent iff

$$\pi_{ij+} = \pi_{i++}\pi_{+j+}, \quad \forall i, j$$

Here, we actually ignore Z.

Remark 8. Conditional independence of X and Y given Z does NOT necessarily imply marginal independence of X and Y (the proof is left for homework)

$$\pi_{ij+} = \sum_k \pi_{ij} = \sum_k \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}} \neq \pi_{i++}\pi_{+j+}$$

in general.

2.2 Mantel-Haenszel test for $2 \times 2 \times K$ tables

Motivation: Is the association of X, Y homogeneous (aka the same) across the levels of Z ? This question can be investigated by testing $\mathrm{H}_0 : \forall k\, \theta_{(k)}^{XY} = \theta^*$, vs. $\mathrm{H}_0 : \exists k, k'\, \theta_{(k)}^{XY} \neq \theta_{(k')}^{XY}$.

Following, we will discuss the particular case where $\theta^* = 1$ which tests if X, Y are independent at each level of Z.

Mantel-Haenszel test The hypothesis pair is:

$$\begin{cases} \mathrm{H}_0 : & X, Y \text{are independent accross the partial tables at each level of } Z \\ \mathrm{H}_1 : & X, Y \text{are not independent accross the partial tables at each level of } Z \end{cases} \iff \begin{cases} \mathrm{H}_0 : & \theta_{(k)}^{XY} = 1,\ \forall k \\ \mathrm{H}_1 : & \theta_{(k)}^{XY} \neq 1,\ \exists k \end{cases}$$

The statistic is

$$T_{MH} = \frac{[\sum_k (n_{11k} - \mu_{11k})]^2}{\sum_k \sigma_{11k}^2} \xrightarrow{D} \chi_1^2$$

where

$$\mu_{11k} = \frac{n_{1+k}n_{+1k}}{n_{++k}}$$

and

$$\sigma_{11k}^2 = \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n_{++k}^2(n_{++k} - 1)}$$

Reasonably, we reject the Null hypothesis for large values.

- The Rejection area at a sig. level is

$$R(\{n_{11k}\}) = \{T_{MH}^{obs} \geq \chi_{1,1-a}^2\}$$

- The p-value is

$$\text{p-value} = 1 - \mathrm{Pr}_{\chi_1^2}(T_{MH} \leq T_{MH}^{obs})$$

Practice

Exercise sheet

Exercise #1
Exercise #2
Exercise #3
Exercise #4
Exercise #5
Examples in [4]

References

[1] Agresti, A. (2013). *Categorical data analysis.* Wiley.

[2] Agresti, A. (2018). *An introduction to categorical data analysis.* Wiley.

[3] Bishop, Y. M., S. E. Fienberg, and P. W. Holland (2007). *Discrete multivariate analysis: theory and practice.* Springer Science & Business Media.

[4] Kateri, M. (2014). Contingency table analysis. *Statistics for Industry and Technology 525.*