Topics in Statistics III/IV (MATH3361/4071) Michaelmas term, 2020

Handout 2: 3-way contingency tables

Lecturer & author: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Aim

To get an introduction in the analysis of 3way contingency tables
Notation, conditional independences

Reading list:

- Kateri, M. (2014; Chapters 1, 2). Contingency table analysis. Methods and implementation using R. Birkhauser
- Agresti, A. (2003; Chapters 1, 2, and 3). Categorical data analysis (Vol. 482). John Wiley & Sons.
- Lauritzen, S. L. (1996; Chapter 2). Graphical models (Vol. 17). Clarendon Press. [Maybe useful for Level 4]

1 $I \times J \times K$ contingency tables: Notation & tools

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 1 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 1: A schematic of a $2 \times 2 \times 2$ table

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

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

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

Example 1. The 674 subjects classified in Table 2 were the defendants in indictments involving cases with multiple murders in Florida between 1976 and 1987. The variables in Table 2 are

Y: death penalty verdict, with categories ($j =$ 1: Yes, 2: No)

X: race of Defendant, with categories ($i =$ 1: White, 2: Black)

Z: race of Victim, with categories ($k =$ 1: White, 2: Black)

		Death Penalty (Y)	
Victim's Race (Z)	Defendant's Race (X)	Yes	No
White	White	53	414
	Black	11	37
Black	White	0	16
	Black	4	139

Table 2: Death Penalty Verdict by Defendant's Race and Victims' Race

Definition 2. 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
$$\pi_{ij|k} = \mathsf{P}(X = i, Y = j|Z = k) = \frac{\pi_{ijk}}{\pi_{++k}}, \forall k = 1, ..., K$$
- Partial proportions
$$p_{ij|k} = \frac{n_{ijk}}{n_{++k}}, \forall k = 1, ..., K$$
- Partial Local odd ratios conditional on Z
$$\theta_{ij(k)}^{XY} = \frac{\pi_{i,j,k}/\pi_{i,j+1,k}}{\pi_{i+1,j,k}/\pi_{i+1,j+1,k}} = \frac{\pi_{i,j,k}\pi_{i+1,j+1,k}}{\pi_{i+1,j,k}\pi_{i,j+1,k}},$$
$i = 1, ..., I-1$, $j = 1, ..., J-1$, 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
$$\hat{\theta}_{ij(k)}^{XY} = \frac{n_{i,j,k}n_{i+1,j+1,k}}{n_{i+1,j,k}n_{i+1,j,k}},$$
$i = 1, ..., I-1$, $j = 1, ..., J-1$, and $k = 1, ..., K$

Definition 3. The XY- marginal table (denoted as (n_{ij})) results by collapsing XYZ -table at Z.

...Consider associated quantities:

- Marginal probabilities
$$\pi_{ij} = \mathsf{P}(X = i, Y = j) = \sum_{k=1}^{K} \pi_{ijk}$$
- Marginal proportions
$$p_{ij} = \sum_{k=1}^{K} p_{ijk}, \forall k = 1, ..., K$$
- Marginal Local odd ratios
$$\theta_{ij}^{XY} = \frac{\pi_{i,j,+}/\pi_{i,j+1,+}}{\pi_{i+1,j,+}/\pi_{i+1,j+1,+}} = \frac{\pi_{i,j,+}\pi_{i+1,j+1,+}}{\pi_{i+1,j,+}\pi_{i+1,j,+}},$$
$i = 1, ..., I-1$, 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
$$\hat{\theta}_{ij}^{XY} = \frac{n_{i,j,+}n_{i+1,j+1,+}}{n_{i+1,j,+}n_{i+1,j,+}},$$

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

2 Types of independency

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 $[X, Y, Z]$

Example The race of defendant, race of victim, and penalty decision are all independent; aka knowledge of the values of one variable don't affect the possibilities of the levels of the other.

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 $[Y, XZ]$.

Example The race of defendant and victim do not affect the penalty decision.

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,

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

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

Example The penalty decision and the race of defendant are independent (no extra info is provided) given that the race of the victim is known.

4. The conditional relationship between any pair of variables given the third one is the same at each level of the third variable; but not necessarily independent.

The join probability $\{\pi_{ijk}\}$ is not available in a separable closed form, but the relation implies that if we know all two-way tables, we have sufficient information to compute $\{\pi_{ijk}\}$.

It can be repressed as with respect to the odds ratios.

- the XY partial odds ratios at each level of Z to be identical: $\theta_{ij(k)}^{XY} = \theta_{ij}^{*}$
- the ZY partial odds ratios at each level of X to be identical: $\theta_{(i)jk}^{ZY} = \theta_{jk}^{*}$
- the ZX partial odds ratios at each level of Y to be identical: $\theta_{i(j)k}^{ZX} = \theta_{ik}^{*}$

Example The penalty decision and the race of defendant have the same association across the levels of the victim's race.

$$\theta_{1,1,(k)}^{XY} = \theta_{1,1}^{*} \iff \frac{\pi_{1,1,(k)}}{1 - \pi_{1,1,(k)}} = \underbrace{\theta^{*}}_{\text{const.}}\frac{\pi_{1,2,(k)}}{1 - \pi_{1,2,(k)}}, \;\; \forall k = 1, 2$$

Defendants of different race have different probabilities to get a death penalty given we know the race of the victim; in particular for different defendant races the odds of getting a death penalty change by a constant θ^* across different races of victims.

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

5. X and Y are marginally independent iff

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

Here, we actually ignore Z.

Example The penalty decision and the race of defendant are independent.

Example 4. If X and Y independent conditionally on Z show that $\theta^{YZ}_{(i)jk} = \theta^{YZ}_{jk}$, where $\theta^{XZ}_{i(j)k}$ are the conditions XZ local odds ratios given Y and .and θ^{YZ}_{jk} are the marginal YZ local odds ratios.

Solution. It is

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

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.

Example 5. Assume a matrix $2 \times 2 \times 2$. Assume that the XY partial odds ratios at each level of Z to be identical. Then:

1. the ZY partial odds ratios at each level of X to be identical

2. the ZX partial odds ratios at each level of Y to be identical

Solution.

1. It is

$$\theta^* = \theta^{XY}_{(k)} = \frac{\pi_{11k}\pi_{22k}}{\pi_{21k}\pi_{12k}}$$

so

$$\frac{\theta^{ZY}_{(1)}}{\theta^{ZY}_{(2)}} = \frac{\frac{\pi_{111}\pi_{212}}{\pi_{211}\pi_{112}}}{\frac{\pi_{121}\pi_{222}}{\pi_{221}\pi_{122}}} = \frac{\frac{\pi_{111}\pi_{221}}{\pi_{211}\pi_{121}}}{\frac{\pi_{112}\pi_{222}}{\pi_{212}\pi_{122}}} = \frac{\theta^{XY}_{(1)}}{\theta^{XY}_{(2)}} = 1 \implies \theta^{ZY}_{(1)} = \theta^{ZY}_{(2)}$$

2. yours...

Remark 6. 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.

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

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:

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

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

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

Extensions: A more general versions of the Mantel-Haenszel test exist testing e.g. (1.) tables $I \times J \times K$ or the more general hypothesis: if the association of X, Y homogeneous (aka the same) across the levels of Z ($\text{H}_0: \forall k \, \theta_{(k)}^{XY} = \theta^*$, vs. $\text{H}_0: \exists k, k' \, \theta_{(k)}^{XY} \neq \theta_{(k')}^{XY}$) –not discussed here.