1

Collapsibility and Suppression for Log-Linear Models

홍 종 선
성균관대학교 통계학과

2005년 4월 29일

.

2

Contents

3

1. Contour Plot & Raindrop Plot

1.1 Odds Ratios in $2\times2\times K$ Table

- $2\times2\times K$ **Table** $\{p_{ijk}\ , i=1,2,\ j=1,2,\ k=1,2,\mathrm{K}\ ,K\}$

p_{111}	p_{121}
p_{211}	p_{221}

p_{112}	p_{122}
p_{212}	p_{222}

$\Lambda \quad \Lambda$

p_{11K}	p_{12K}
p_{21K}	p_{22K}

p_{11+}	p_{12+}
p_{21+}	p_{22+}

- **Define odds** p_k , q_k **and** p_c , q_c **for** $k=1,2,\mathrm{K}\ ,K$

$$p_k=\frac{p_{11k}}{p_{+1k}}\ ,\quad q_k=\frac{p_{12k}}{p_{+2k}} \qquad\qquad p_c=\frac{p_{11+}}{p_{+1+}}\ ,\quad q_c=\frac{p_{12+}}{p_{+2+}}$$

4

- **Odds Ratios for** k **th table**

$$\theta_k=\frac{p_k/(1-p_k)}{q_k/(1-q_k)}=\frac{p_{11k}\ p_{22k}}{p_{12k}\ p_{21k}},\quad k=1,2,\mathrm{K}\ ,K$$

- **Odds Ratio for a collapsed table**

$$\theta_c=\frac{p_c/(1-p_c)}{q_c/(1-q_c)}=\frac{p_{11+}\ p_{22+}}{p_{12+}\ p_{21+}}$$

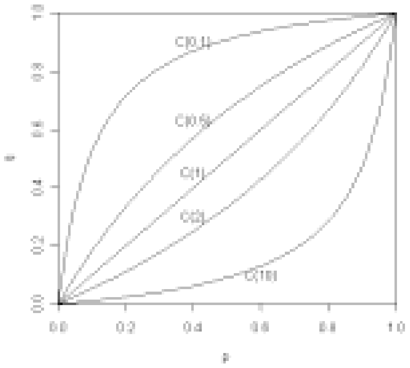
5

1.2 Contour Plot

- Doi, Nakamura, and Yamamoto (2001)

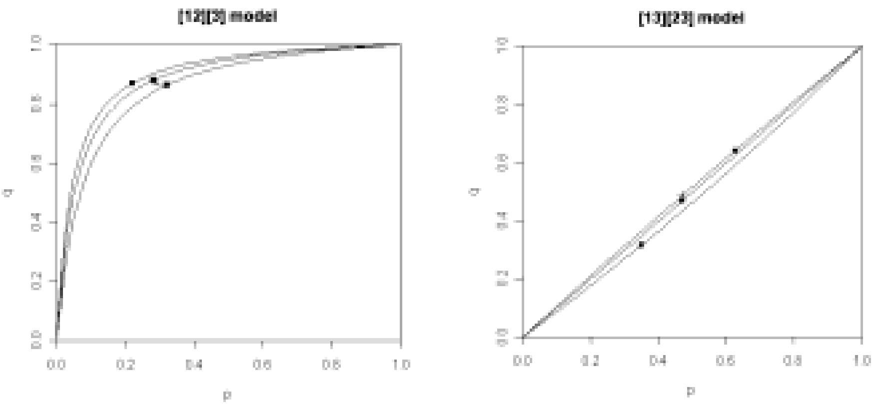
$$C(\theta) = \{(p,q) \in (0,1)^2 : \tfrac{p/(1-p)}{q/(1-q)} = \theta\}, \theta > 0,$$

Shapes of $C(\theta)$ for $\theta = 0.1,\ 0.5,\ 1,\ 2,$ and 10



6

An Example : Contour plots for $2 \times 2 \times 3$ tables



7

1.3 Raindrop Plot

- Barrowman & Myers (2003)

Conditional Likelihood for Odds Ratio(θ)

$$L(\theta) = \binom{x_{1+}}{x_{11}}\binom{x_{2+}}{x_{21}} e^{(\theta)x_{11}} / S(\theta)$$

Approximate $100 \cdot (1-\gamma)\%$ confidence interval for θ

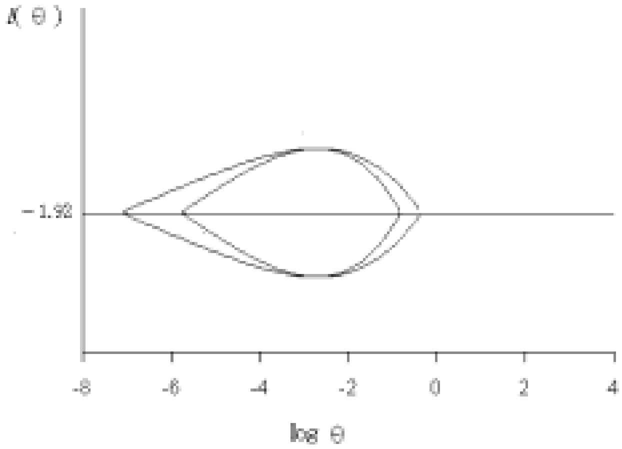
$$\{\theta : 2[\lambda(\theta^{MLE}) - \lambda(\theta)] \le \chi_{1(1-\gamma)}{}^{2}]\}$$

- Set $\lambda(\theta^{MLE}) = 0$
- 95% confidence interval for θ

$$\{\theta : \lambda(\theta) \ge -1.92\}$$

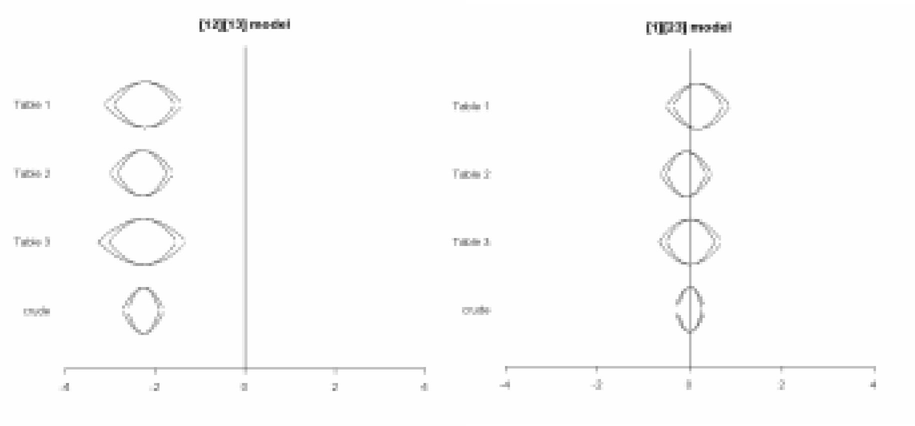
8

95 & 99% confidence interval for $\log(\theta)$



9

An Example : Raindrop plots for a $2 \times 2 \times 3$ table



10

2. Collapsibility

2.1 Three-dimensional log-linear models

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$ **: [123] model**

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$ **: [12][13][23] model**

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)}$ **: [12][13] model**

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)}$ **: [12][23] model**

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)}$ **: [13][23] model**

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}$ **: [12][3] model**

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)}$ **: [13][2] model**

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)}$ **: [1][23] model**

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)}$ **: [1][2][3] model**

11

2.2 Odds Ratio of Log-Linear Model

- **1st group : Models including u_{12} term**

$$p_k = \frac{p_{11k}}{p_{+1k}} = \frac{p_{11+}\,p_{++k}}{p_{+1+}\,p_{++k}} = \frac{p_{11+}}{p_{+1+}} = p_c$$
$$q_k = \frac{p_{12k}}{p_{+2k}} = \frac{p_{12+}\,p_{++k}}{p_{+2+}\,p_{++k}} = \frac{p_{12+}}{p_{+2+}} = q_c$$
for [12][3] model

$$p_k = \frac{p_{11+}\,p_{+1k}/p_{+1+}}{p_{+1+}\,p_{+1k}/p_{+1+}} = \frac{p_{11+}}{p_{+1+}} = p_c$$
$$q_k = \frac{p_{12+}\,p_{+2k}/p_{+2+}}{p_{+2+}\,p_{+2k}/p_{+2+}} = \frac{p_{12+}}{p_{+2+}} = q_c$$
for [12][23] model

$$p_k = \frac{p_{11+}\,p_{+1k}/p_{1++}}{(p_{11+}\,p_{1+k}/p_{1++})+(p_{21+}\,p_{2+k}/p_{2++})}$$
$$q_k = \frac{p_{12+}\,p_{1+k}/p_{1++}}{(p_{12+}\,p_{1+k}/p_{1++})+(p_{22+}\,p_{2+k}/p_{2++})}$$
for [12][13] model

Odds ratios

- [12][3] & [12][23]model : $\theta_k = \theta_c$
- [12][13] model : $\theta_k = \dfrac{p_{11+}/p_{21+}}{p_{12+}/p_{22+}} = \theta_c$

$\therefore \theta_k = \theta_c$ for all k

12

- **2nd group**

$$p_k = \frac{p_{1++}\,p_{+1+}\,p_{++k}}{p_{+1+}\,p_{++k}} = p_{1++}$$
$$q_k = \frac{p_{1++}\,p_{+2+}\,p_{++k}}{p_{+2+}\,p_{++k}} = p_{1++}$$
for [1][2][3] model

$$p_k = \frac{p_{1++}\,p_{+1k}}{p_{+++}\,p_{+1k}} = p_{1++}$$
$$q_k = \frac{p_{1++}\,p_{+2k}}{p_{+++}\,p_{+2k}} = p_{1++}$$
for [1][23] model

$$p_k = \frac{p_{1+k}\,p_{+1+}}{p_{++k}\,p_{+1+}} = \frac{p_{1+k}}{p_{++k}}$$
$$q_k = \frac{p_{1+k}\,p_{+2+}}{p_{++k}\,p_{+2+}} = \frac{p_{1+k}}{p_{++k}}$$
for [13][2] model

$$p_k = \frac{p_{1+k}\,p_{+1k}/p_{++k}}{p_{++k}\,p_{+1k}/p_{++k}} = \frac{p_{1+k}}{p_{++k}}$$
$$q_k = \frac{p_{1+k}\,p_{+2k}/p_{++k}}{p_{++k}\,p_{+2k}/p_{++k}} = \frac{p_{1+k}}{p_{++k}}$$
for [13][23]model

Odds ratios $\theta_k = \theta_c = 1$ for all k

13

- **Other group**

- [12][13][23]model : $\theta_k =$ constant for all k

14

2.3 Definitions of Collapsibility

BFH(1975), Agresti(1984), Christensen(1990)

$$\theta_1 = \Lambda = \theta_k = \theta_c \neq 1$$

<u>Another Def.</u> : Whittemore(1978),
Ducharme & Lepage(1986), Geng(1992)

$\theta_1 = \Lambda = \theta_k = \theta_c \neq 1$ **: Strong Collapsibility**

Note :

$\theta_1 = \Lambda = \theta_k = \theta_c$: Strict Collapsibility

15

2.4 Criteria of Collapsibility via Contour plot & Raindrop plot

1. [12][3], [12][13], [12][23] model :

$$\theta_1 = \Lambda = \theta_k = \theta_c \neq 1$$

=> Collapsible Models or
Strong Collapsible Models

2. [1][2][3], [1][23], [13][2], [13][23] model :

$$\theta_1 = \Lambda = \theta_k = \theta_c = 1$$

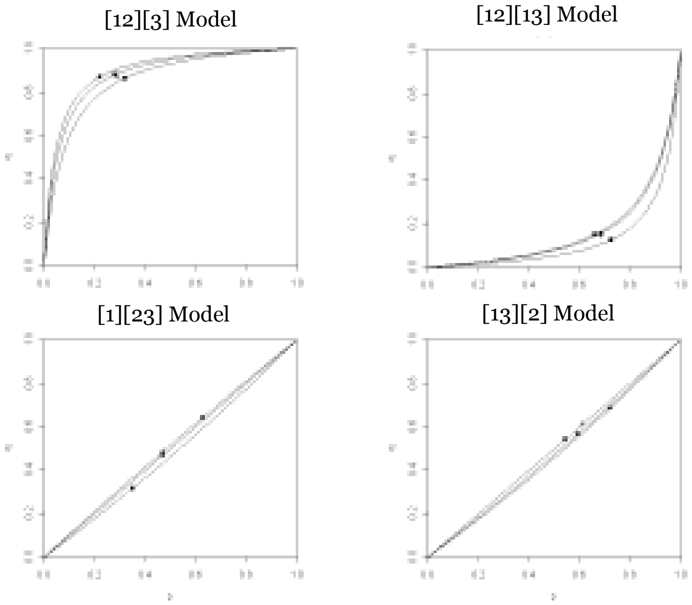
=> Non-Collapsible Models
(might say Strictly Collapsible Models)

3. [12][13][23] model $\theta_1 = \Lambda = \theta_k$

=> Non-Collapsible Model

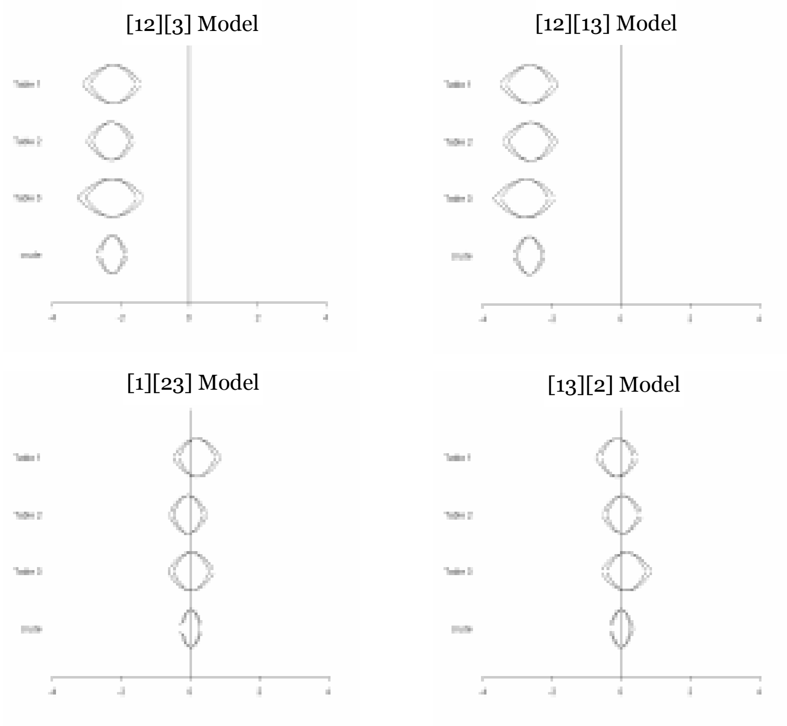
16

- Contour plots for $2 \times 2 \times K$ tables ($K = 3$)



17

- Raindrop plots



18

3. Suppression

3.1 Suppression condition

for linear regression models for X_1 and X_2

$$SSR(X_2|X_1) > SSR(X_2)$$

$SSR(X_2)$: SSR on X_2 alone

$$H_0: \; Y_i = \alpha + \varepsilon_i \qquad (\beta_2 = 0)$$
$$H_1: \; Y_i = \alpha + \beta_2 X_{2i} + \varepsilon_i$$

$SSR(X_2|X_1)$: SSR on addition X_2 already contains X_1

$$H_0^{'}: \; Y_i = \alpha + \beta_1 X_{1i} + \varepsilon_i \qquad (\beta_2 = 0)$$
$$H_1^{'}: \; Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

19

3.2 Suppression in Logit Models

Lynn (2003)

- Suppression condition for logit models

$$L(X_2|X_1\,Y) < L(X_2|\,Y)$$

- $L(X_2|\,Y)$: Log Likelihood ratio statistic to test

$H_0: \ logit(jk) = \omega$ $\qquad (\omega_2 = 0)$

$H_1: \ logit(jk) = \omega + \omega_2$

- $L(X_2|X_1\,Y)$

$H_0^{'}: \ logit(jk) = \omega + \omega_1$ $\qquad (\omega_2 = 0)$

$H_1^{'}: \ logit(jk) = \omega + \omega_1 + \omega_2$

20

3.3 Suppression in Log-Linear Models

- Log-linear model corresponding to logit model

$H_0: \ logit(jk) = \omega$

$H_1: \ logit(jk) = \omega + \omega_2$ $\qquad (\omega_2 = 0)$

$\Rightarrow$ $H_0: [1][23]$

$H_1: [13][23]$ $\qquad (u_{13} = 0)$

$H_0^{'}: \ logit(jk) = \omega + \omega_1$

$H_1^{'}: \ logit(jk) = \omega + \omega_1 + \omega_2$ $\qquad (\omega_2 = 0)$

$\Rightarrow$ $H_0^{'}: [12][23]$

$H_1^{'}: [12][23][13]$ $\qquad (u_{13} = 0)$

•***Note***

$Y \rightarrow V_1$

$X_1 \rightarrow V_2$

$X_2 \rightarrow V_3$

21

Therefore $L(X_2|Y) = G^2([1][23]) - G^2([13][23])$
$\equiv G^2(H_0|H_1)$

$L(X_2|X_1Y) = G^2([12][23]) - G^2([12][23][13])$
$\equiv G^2(H_0^{'}|H_1^{'})$

Suppression condition for log-linear models

$$G^2(H_0^{'}|H_1^{'}) < G^2(H_0|H_1)$$

Note : Suppression condition for logit models

$$L(X_2|X_1Y) < L(X_2|Y)$$

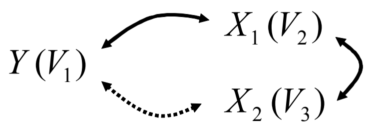
22

$$\left.\begin{array}{ll} H_0 : Y_i = \alpha + \varepsilon_i & \\ H_1 : Y_i = \alpha + \beta_2 X_{2i} + \varepsilon_i & (\beta_2 = 0) \\ H_0^{'} : Y_i = \alpha + \beta_1 X_{1i} + \varepsilon_i & \\ H_1^{'} : Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i & (\beta_2 = 0) \end{array}\right\} SSR(X_2|X_1) > SSR(X_2)$$

$$\left.\begin{array}{ll} H_0 : logit(jk) = \omega & \\ H_1 : logit(jk) = \omega + \omega_2 & (\omega_2 = 0) \\ H_0^{'} : logit(jk) = \omega + \omega_1 & \\ H_1^{'} : logit(jk) = \omega + \omega_1 + \omega_2 & (\omega_2 = 0) \end{array}\right\} L(X_2|X_1Y) < L(X_2|Y)$$

$$\left.\begin{array}{ll} H_0 : [1][23] & \\ H_1 : [13][23] & (u_{13} = 0) \\ H_0^{'} : [12][23] & \\ H_1^{'} : [12][23][13] & (u_{13} = 0) \end{array}\right\} G^2(H_0^{'}|H_1^{'}) < G^2(H_0|H_1)$$

23



- We find that this suppression condition is satisfied for all "collapsible models or strong collapsibility models"

=> [12][3], [12][13], and [12][23] models

24

- **Suppression condition for Collapsible models**

Lynn (2003)

Model	[12][3] model	[12][13] model	[12][23] model
$H_0 : [Y][X_1X_2]$	278.5630	420.0742	290.7003
$H_1 : [Y\,X_2][X_1X_2]$	277.9259	174.4947	241.2220
$G^2(H_0 \mid H_1)$	**0.6371**	**245.5795**	**49.4783**
$H_0' : [Y\,X_1][X_1X_2]$	0.1397	205.1078	0.2050
$H_1' : [Y\,X_1][X_1X_2][Y\,X_2]$	0.0616	0.0184	0.1765
$G^2(H_0' \mid H_1')$	**0.0781**	**205.0894**	**0.0285**

25

• **Suppression condition for Non-collapsible models**

Model	**[13][2] model**	**[1][23] model**	**[13][23] model**
$H_0:[Y][X_1X_2]$	149.4669	0.6904	76.8049
$H_1:[YX_2][X_1X_2]$	0.7189	0.5482	0.4143
$G^2(H_0\|H_1)$	**145.7480**	**0.1422**	**76.3906**
$H_0':[YX_1][X_1X_2]$	146.4669	0.6869	76.8039
$H_1':[YX_1][X_1X_2][YX_2]$	0.6208	0.5257	0.2209
$G^2(H_0'\|H_1')$	**145.8461**	**0.1612**	**76.5830**

26

4. Collapsibility & Suppression via Contour Plot & Raindrop Plot

4.1 Among three-dimensional log-linear models,

- Collapsible models over the third variables :
  - **[12][13] model**
  - **[12][23] model**
  - **[12][3] model**

  (Strong Collapsible Model)

- Non-Collapsible models over the third variables :
  - **[1][2][3] model**
  - **[1][23] model**
  - **[13][2] model**
  - **[13][23] model**

(Strictly Collapsible Model: [12][13], [12][23], [12][3], [1][2][3], [1][23], [13][2], [13][23] models)

[12][13][23] model

27

4.2 Suppression Conditions

- $SSR(X_2|X_1Y) > SSR(X_2\,|\,Y)$ for regression models

- $L(X_2|X_1\,Y) < L(X_2|\,Y)$ for logit models

- $G^2([12][23]\|[12][23][13]) < G^2([1][23]\|[13][23])$

  for log-linear models

28

4.3 If for three-dimensional contingency table
with V_3 being **confounder**,

$$G^2([12][23]\,|\,[12][23][13]) < G^2([1][23]\,|\,[13][23]),$$

then we find that V_3 is a **suppressor.**

=> This condition is satisfied
under the [12][3], [12][13], [12][23] models.

➔ The data including the <u>suppressor variable</u> V_3
might be <u>collapsed over the variable</u> V_3 .

29

4.4 Three dimensional $2 \times 2 \times K$ contingency table can be applied to the contour plots and raindrop plots.

- If most of K contours corresponding to the odds ratios are shown to be <u>convex or concave</u> from origin (0, 0) to (1, 1) coordinate in a contour plot, then the table might be collapsed over third variable.

- If the rain-drops which represents confidence intervals of log-odds ratios <u>do not contain “0” point</u>, then one can say that data is collapsible.

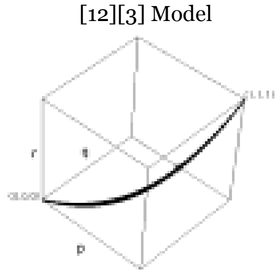
⇒ **Suppression condition is satisfied**.
And the best fitted log-linear model is **collapsible over the third variable.**

30

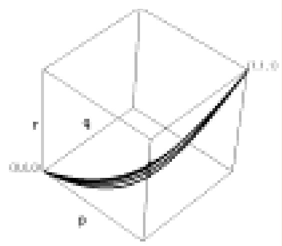
5. Further Studies

- Contour plots for $2 \times 3 \times K$ tables ($K = 3$)

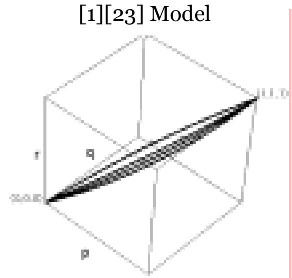
[12][3] Model



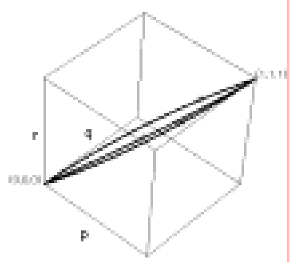
[12][13] Model



[1][23] Model



[13][2] Model



31

- The contour method can be extended $2 \times 3 \times K$ categorical data with the shape of hexahedron with a unit length. In the case that most of K contours are far from the diagonal line from the origin to (1, 1, 1) coordinate, then the table could be regarded as a collapsible data.

$\Rightarrow$ Suppression condition is satisfied.

32

- For $2 \times J \times K$ categorical data, the odds ratios can be generalized as follows, for $j = 2, 3, \ldots, J$, $k = 1, 2, \ldots, K$

$$\theta_{(j-1),jk} = \frac{p_{11k}/p_{21k}}{p_{1jk}/p_{2jk}} \quad \text{and} \quad \theta_{(j-1),jc} = \frac{p_{11+}/p_{21+}}{p_{1j+}/p_{2j+}}$$

- Collapsible model [12][3], [12][13], [12][23] models

$\theta_{jk} = \theta_{jc} \neq 1$ for all j and k

- Non-collapsible model [1][2][3], [1][23], [13][2], [13][23] models

$\theta_{jk} = \theta_{jc} = 1$ for all j and k

Thank you very much.

cshong@skku.ac.kr