Multi-way Contingency Tables

Outline

- Multiway Contingency Tables
  - Flat Contingency Tables
  - How to effectively convey information in multiway contingency tables?
  - Manipulation of ‘Flat” Contingency Tables in R
- Logistic Models for Multi-way Contingency Tables

Example (Mouse Muscle Tension)

A study to examine relationship between two drugs and muscle tension

Response: Tension — change in muscle tension: High, Low

Explanatory variables

- Drug: drug 1, drug (primary)
- Weight: weight of muscle: High, Low
- Muscle: muscle type: 1, 2

A four-way “flat” contingency table ($2 \times 2 \times 2 \times 2$):

```
Drug 1  Drug 2
Muscle Type 1 Type 2
Tension Weight
High High 3 23 21 11
Low 22 4 32 12
Low High 3 41 10 21
Low 45 6 23 22
```

This flat table is bad because ...

Multiway - 1

A better table:

```
<table>
<thead>
<tr>
<th>Muscle Type</th>
<th>Type 1</th>
<th>Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Conditional odds ratios between Drug and Tension:

```
<table>
<thead>
<tr>
<th>Drug</th>
<th>Muscle Type 1</th>
<th>Muscle Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>1 3×10 (\frac{5\times21}{23} \approx 0.48)</td>
<td>23 (\frac{4\times22}{11}) \approx 1.07</td>
</tr>
<tr>
<td>Low</td>
<td>22 (\frac{23\times32}{6\times12} \approx 1.22)</td>
<td></td>
</tr>
</tbody>
</table>
```

The table splits in to four partial tables for the primary predictor (Drug) and the response (Tension), controlling for the other two variables.

Conditional odds ratios between Drug and Tension can be easily computed from this table but not from the table on the previous slide.

Tip: response and the primary predictor (if any) should be placed in the innermost layer of the table.

Multiway - 2

Example (Mouse Muscle Tension)

Conditional distributions of Tension given Drug, Weight, and Muscle type:

```
<table>
<thead>
<tr>
<th>Muscle Type</th>
<th>Type 1</th>
<th>Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tension 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tension 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Observation:

- For Type 1 muscle, Drug 1 looks more effective in lowering muscle tension than Drug 2 does
- For Type 2 muscle, the effect of the two drugs looks similar
Another Example of “Flat” Contingency Tables

Example (Titanic)

Four-way table: \((2 \times 2 \times 4 \times 2)\)

Breakup of people on Titanic by Sex, Age, Class, and Survival

<table>
<thead>
<tr>
<th></th>
<th>Adult</th>
<th>Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Survived</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>1st</td>
<td>4</td>
<td>140</td>
</tr>
<tr>
<td>2nd</td>
<td>13</td>
<td>80</td>
</tr>
<tr>
<td>3rd</td>
<td>89</td>
<td>76</td>
</tr>
<tr>
<td>Crew</td>
<td>3</td>
<td>20</td>
</tr>
</tbody>
</table>

Manipulating Flat Contingency Tables in R

```r
> mouse.muscle = read.table("mousemuscle.dat",header=T)
> mouse.muscle
W M D tension.high tension.low
1  High 1 1 3  3
2  High 1 2 21 10
3  High 2 1 23 41
4  High 2 2 11 21
5  Low 1 1 22 45
6  Low 1 2 32 23
7  Low 2 1 4 6
8  Low 2 2 12 22

> attach(mouse.muscle)
> Freq = c(tension.high,tension.low)
> weight = rep(W,2)
> muscle = rep(M,2)
> drug = rep(D,2)
> tension = c(rep("High",8),rep("Low",8))

xtabs creates multiway tables in R, but the output is awkward.

```
ftable() can print flat multi-way tables. Row and column variables in a flat table can be specified using row.vars and col.vars, and they are ordered from outer to inner layers.

```r
> ftable(muscle.tab, row.vars=c("weight","drug"),
          col.vars=c("muscle","tension"))
  muscle 1 2
tension High Low High Low
weight drug
High 1 3 3 23 41
2 21 10 11 21
Low 1 22 45 4 6
2 32 23 12 22
```

Multiway - 9

### Flat Marginal Tables

If a variable in the contingency table is neither specified in row.vars nor in col.vars, then it is ignored. The output is a marginal table of the specified variables, e.g., the following is a marginal table for drug, muscle, and tension, ignoring weight:

```r
> ftable(muscle.tab, row.vars="drug", col.vars=c("muscle","tension"))
  muscle 1 2
tension High Low High Low
drug
1 25 48 27 47
2 53 33 23 43
```

Marginal table for drug and tension, ignoring weight and muscle:

```r
> ftable(muscle.tab,row.vars="drug", col.vars="tension")
tension High Low
drug
1 52 95
2 76 76
```

Multiway - 10

### Flat Table for Conditional Distributions

First compute the conditional distributions of tension given the rest using prop.table, and then print it using ftable.

```r
> muscle.p.tab = prop.table(muscle.tab,1:3)
> ftable(muscle.p.tab,row.vars=c("weight","drug"),
          col.vars=c("muscle","tension"))
  muscle 1 2
tension High Low High Low
weight drug
High 1 0.5000000 0.5000000 0.3593750 0.6406250
2 0.6774194 0.3225806 0.3437500 0.6562500
Low 1 0.3283582 0.6716418 0.4000000 0.6000000
2 0.5818182 0.4181818 0.3529412 0.6470588
```

Multiway - 11

### Conditional Distributions in Marginal Tables

Say, we want the condition distribution of tension given drug and muscle, ignoring weight.

1. Create the marginal table, with “response” as the only column variable, and other var. in the marginal table as the row var..

```r
> temp1 = ftable(muscle.tab, row.vars=c("muscle","drug"),
                col.vars="tension")
> temp1
tension High Low
muscle drug
1 1 25 48
2 2 53 33
2 1 27 47
2 2 23 43
```

2. Find the conditional distribution using prop.table.

```r
> temp2 = prop.table(temp1, 1)
> temp2
tension High Low
muscle drug
1 1 0.3424658 0.6575342
2 0.6162791 0.3837209
2 1 0.3648649 0.6351351
2 0.3484848 0.6515152
```

Multiway - 12
Conditional Distributions in Marginal Tables

3 If necessary, reshape the condition distribution table using `ftable`

```r
> ftable(temp2, row.vars="drug", col.vars=c("muscle","tension"))
muscle  1  2
tension High Low High Low
drug 1 0.3424658 0.6575342 0.3648649 0.6351351 2 0.6162791 0.3837209 0.3484848 0.6515152

# converted into percentages
> round(100*ftable(temp2, row.vars="drug", col.vars=c("muscle","tension")))
muscle  1  2
tension High Low High Low
drug 1 34 66 36 64 2 62 38 35 65
```

Multiway - 13

Let $A_i, i = 1, \ldots, a$ be the dummy variables for levels of $A$

$B_j, j = 1, \ldots, b$ be the dummy variables for levels of $B$

$C_k, k = 1, \ldots, c$ be the dummy variables for levels of $C$

The model formula

$$\logit(\pi_{ijk}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{jk}^{BC} + \beta_{ik}^{AC} + \beta_{ijk}^{ABC}$$

can be written in terms of the dummy variables as

$$\logit(\pi_{ijk}) = \alpha + \sum_{l=1}^a \beta_l^A A_l + \sum_{m=1}^b \beta_m^B B_m + \sum_{n=1}^c \beta_n^C C_n + \sum_{l=1}^a \sum_{m=1}^b \beta_{lm}^{AB} A_l B_m + \sum_{m=1}^b \sum_{n=1}^c \beta_{mn}^{BC} B_m C_n + \sum_{l=1}^a \sum_{n=1}^c \beta_{ln}^{AC} A_l C_n + \sum_{l=1}^a \sum_{m=1}^b \sum_{n=1}^c \beta_{lmn}^{ABC} A_l B_m C_n$$

Multiway - 14

Logistic Models for Multi-way Contingency Tables

Let’s start with models for 4-way tables (1 response + 3 predictors)

- categorical predictors: $A, B, C$, with $a, b, c$ levels respectively
- response: $Y = 0$ or $1$

Let

$$\pi_{ijk} = P(Y = 1|A = i, B = j, C = k)$$

The most complex model for a 4-way table is the three way interaction model, denoted as $A \times B \times C$, including all main effects and 2-way, 3-way interactions

$$A + B + C + A \times B + B \times C + A \times C + A \times B \times C$$

The model formula is

$$\logit(\pi_{ijk}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{jk}^{BC} + \beta_{ik}^{AC} + \beta_{ijk}^{ABC}$$

for $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$.

Multiway - 15

In the 3-way on the previous page, many parameters are redundant because

$$A_1 + \cdots + A_a = 1, \quad B_1 + \cdots + B_b = 1, \quad C_1 + \cdots + C_c = 1.$$ 

So, we need constraints on the parameters.

Some commonly used constraints are

- Baseline constraints:

  $$\beta_1^A = \beta_1^B = \beta_1^C = 0$$

  $$\beta_{1j}^A = \beta_{1j}^B = \beta_{1j}^C = \beta_{1k}^A = \beta_{1k}^B = \beta_{1k}^C = 0$$

  $$\beta_{1jk}^A = \beta_{1jk}^B = \beta_{1jk}^C = 0$$

- Sum-to-zero constraints:

  $$\sum_{i=1}^a \beta_i^A = \sum_{j=1}^b \beta_j^B = \sum_{k=1}^c \beta_k^C = 0$$

  $$\sum_{i=1}^a \beta_{ij}^A = \sum_{j=1}^b \beta_{ij}^B = \sum_{k=1}^c \beta_{jk}^C = 0$$

  $$\sum_{i=1}^a \beta_{ijk}^A = \sum_{j=1}^b \beta_{ijk}^B = \sum_{k=1}^c \beta_{ijk}^C = 0$$

Multiway - 16
Regardless of the constraints used,
- the effective # of parameters for a main effect is number of levels - 1
- the effective # of parameters for an interaction is the product of (number of levels - 1) for each factor involved in the interaction.

The total number of effective parameters is

\[ 1 + (a-1) + (b-1) + (c-1) \]
\[ + (a-1)(b-1) + (b-1)(c-1) + (a-1)(c-1) \]
\[ + (a-1)(b-1)(c-1) \]
\[ = abc \]

There are several simplifications of the 3-way interaction model, such as
- Model \( A * B + B * C + A * C \):
  \[ \logit(\pi_{ijk}) = \alpha + \beta_A^i + \beta_B^j + \beta_C^k + \beta_{AB}^{ij} + \beta_{BC}^{jk} + \beta_{AC}^{ik} \]
- Model \( A * B + A * C \):
  \[ \logit(\pi_{ijk}) = \alpha + \beta_A^i + \beta_B^j + \beta_C^k + \beta_{AB}^{ij} + \beta_{AC}^{ik} \]
- Model \( A + B * C \):
  \[ \logit(\pi_{ijk}) = \alpha + \beta_A^i + \beta_B^j + \beta_C^k + \beta_{BC}^{jk} \]
- Model \( A + B + C \):
  \[ \logit(\pi_{ijk}) = \alpha + \beta_A^i + \beta_B^j + \beta_C^k \]
- Model \( A * B \):
  \[ \logit(\pi_{ijk}) = \alpha + \beta_A^i + \beta_B^j + \beta_{AB}^{ij} \]

Multiway - 17

Multiway - 18

Interpretation of Model \( A + B + C \) and its Coefficients

In Model \( A + B + C \):
\[ \logit(\pi_{ijk}) = \alpha + \beta_A^i + \beta_B^j + \beta_C^k \]
- There are 1 + (a - 1) + (b - 1) + (c - 1) effective parameters.
- Under the baseline constraint \( \beta_A^1 = \beta_B^1 = \beta_C^1 = 0 \).
  - The odds of \( Y = 1 \) when \( A = i \) is the odds of \( Y = 1 \) when \( A = 1 \) multiplied by a factor of \( e^{\beta_A^i} \), regardless of \( B \) and \( C \).
  - Interpretation for \( e^{\beta_B^j} \) and \( e^{\beta_C^k} \): Ditto
  - Homogeneous \( YA, YB, \) and \( YC \) association,
Homogeneous Association Revisit

In a 3-way table, if $XY$ has homogeneous association given $Z$, then so do $YZ$ given $X$ and $XZ$ given $Y$.

$$
\begin{array}{ccc}
Z = 1 & Z = 2 \\
X = 1 & X = 2 \\
Y = 1 & a & b & A & B \\
Y = 2 & c & d & C & D \\
\end{array}
$$

Homogeneous $XY$ association given $Z$ means

$$
\theta_{XY(1)} = \frac{ad}{cb} = \frac{AD}{CB} = \theta_{XY(2)}
$$

$$
\iff \quad \theta_{YZ(1)} = \frac{aC}{cA} = \frac{bD}{dB} = \theta_{YZ(2)}
$$

which means homogeneous $YZ$ association given $X$.

$$
\begin{array}{ccc}
X = 1 & X = 2 \\
Z = 1 & Z = 2 \\
Y = 1 & a & A & b & B \\
Y = 2 & c & C & d & D \\
\end{array}
$$

Multiway - 21

Interpretation of Model $A \ast B \ast C$ and its Coefficients

$$
\text{logit}(\pi_{ijk}) = \alpha + \beta_{i}^{A} + \beta_{j}^{B} + \beta_{k}^{C} + \beta_{ij}^{AB} + \beta_{ik}^{AC} + \beta_{jk}^{BC} + \beta_{ijk}^{ABC}
$$

Under the baseline constraint, the conditional odds ratio for

$$
\begin{array}{|c|c|c|}
\hline
A & Y = 1 & Y = 0 \\
\hline
1 & & \\
\hline
\end{array}
$$

equals

$$
\begin{cases}
\frac{e^{\beta_{i}^{A}}}{e^{\beta_{i}^{A}+\beta_{ij}^{AB}}} & \text{when } B = 1, C = 1; \\
\frac{e^{\beta_{i}^{A}+\beta_{ij}^{AB}}}{e^{\beta_{i}^{A}+\beta_{ik}^{AC}+\beta_{jk}^{BC}+\beta_{ijk}^{ABC}}} & \text{when } B = j, C = k.
\end{cases}
$$

So

$$
\begin{cases}
\frac{\text{YA odds ratio when } B = j}{\text{YA odds ratio when } B = 1} = \frac{e^{\beta_{ij}^{AB}}}{e^{\beta_{ij}^{AB}+\beta_{jk}^{BC}+\beta_{ijk}^{ABC}}} & \text{when } C = 1; \\
\frac{e^{\beta_{ij}^{AB}+\beta_{jk}^{BC}+\beta_{ijk}^{ABC}}}{e^{\beta_{ij}^{AB}+\beta_{jk}^{BC}+\beta_{ijk}^{ABC}}} & \text{when } C = k.
\end{cases}
$$

The 3-way interaction $e^{\beta_{ijk}^{ABC}}$ is the ratio of the ratios of odds ratios.