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 *Muscle*		*Drug 2* *Type*	
Tension	*Weight*	1	2	1	2
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 ...

A better table:

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

Conditional odds ratios between Drug and Tension:

	Muscle	
Wt.	Type 1	Type 2
High	$\frac{3\times10}{3\times21}\approx0.48$	$\frac{23\times21}{41\times11}\approx1.07$
Low	$\frac{22\times23}{45\times32}\approx0.35$	$\frac{4\times22}{6\times12}\approx1.22$

- 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 inner most layer of the table

Example (Mouse Muscle Tension)

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

		Muscle			
		Type 1		Type 2	
		Tension			
Weight	*Drug*	High	Low	High	Low
High	1	50%	50%	36%	64%
	2	68%	32%	34%	66%
Low	1	33%	67%	40%	60%
	2	58%	42%	35%	65%

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

Sex	Adult				Child			
Age	Female		Male		Female		Male	
Survived	No	Yes	No	Yes	No	Yes	No	Yes
Class								
1st	4	140	118	57	0	1	0	5
2nd	13	80	154	14	0	13	0	11
3rd	89	76	387	75	17	14	35	13
Crew	3	20	670	192	0	0	0	0

Manipulating Flat Contingency Tables in 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))
```



```
> data.frame(weight, muscle, drug, tension, Freq)
   weight muscle drug tension Freq
1    High      1    1    High    3
2    High      1    2    High   21
3    High      2    1    High   23
4    High      2    2    High   11
5     Low      1    1    High   22
6     Low      1    2    High   32
7     Low      2    1    High    4
8     Low      2    2    High   12
9    High      1    1     Low    3
10   High      1    2     Low   10
11   High      2    1     Low   41
12   High      2    2     Low   21
13    Low      1    1     Low   45
14    Low      1    2     Low   23
15    Low      2    1     Low    6
16    Low      2    2     Low   22
```


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

```
> muscle.tab = xtabs(Freq ~ weight + muscle + drug + tension)
> muscle.tab
, , drug = 1, tension = High
      muscle
weight  1  2
  High  3 23
  Low  22  4

, , drug = 2, tension = High
      muscle
weight  1  2
  High 21 11
  Low  32 12

, , drug = 1, tension = Low
      muscle
weight  1  2
  High  3 41
  Low  45  6

, , drug = 2, tension = Low
      muscle
weight  1  2
  High 10 21
  Low  23 22
```

`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.

```
> 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

> ftable(muscle.tab, row.vars=c("tension","weight"),
                     col.vars=c("drug","muscle"))
               drug    1     2
               muscle  1  2  1  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
```


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`:

```
> 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`:

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


Flat Table for Conditional Distributions

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

```
> 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

> round(100*ftable(muscle.p.tab,row.vars=c("weight","drug")))
            muscle      1         2
            tension High Low High Low
weight drug
High   1              50  50   36  64
       2              68  32   34  66
Low    1              33  67   40  60
       2              58  42   35  65
```


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..

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

2. Find the conditional distribution using `prop.table`.

```
> 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
```

Conditional Distributions in Marginal Tables

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

```
> 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
```


Logistic Models for Multi-way Contingency Tables

Let's start w/ 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

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

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

$$A + B + C + A * B + B * C + A * C + A * B * C$$

The model formula is

$$\mathrm{logit}(\pi_{ijk}) = \alpha + \underbrace{\beta_i^A + \beta_j^B + \beta_k^C}_{\text{main effects}} + \underbrace{\beta_{ij}^{AB} + \beta_{jk}^{BC} + \beta_{ik}^{AC}}_{\text{two-way interactions}} + \beta_{ijk}^{ABC}$$

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

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

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

can be written in terms of the dummy variables as

$$\begin{aligned}\mathrm{logit}(\pi_{ijk}) = \alpha &+ \sum_{\ell=1}^{a} \beta_\ell^A A_\ell + \sum_{m=1}^{b} \beta_m^B B_m + \sum_{n=1}^{c} \beta_n^C C_n \\ &+ \sum_{\ell=1}^{a}\sum_{m=1}^{b} \beta_{\ell m}^{AB} A_\ell B_m + \sum_{m=1}^{b}\sum_{n=1}^{c} \beta_{mn}^{BC} B_m C_n + \sum_{\ell=1}^{a}\sum_{n=1}^{c} \beta_{\ell n}^{AC} A_\ell C_n \\ &+ \sum_{\ell=1}^{a}\sum_{m=1}^{b}\sum_{n=1}^{c} \beta_{\ell mn}^{ABC} A_\ell B_m C_n\end{aligned}$$

- How many parameters are there?

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, need constraints on the parameters.

Some commonly used constraints are

- Baseline constraints:

$$\beta_1^A = \beta_1^B = \beta_1^C = 0$$
$$\beta_{1j}^{AB} = \beta_{i1}^{AB} = \beta_{1k}^{BC} = \beta_{j1}^{BC} = \beta_{1k}^{AC} = \beta_{i1}^{AC} = 0$$
$$\beta_{1jk}^{ABC} = \beta_{i1k}^{ABC} = \beta_{ij1}^{ABC} = 0$$

- Sum-to-zero constraints:

$$\sum_{i=1}^{a} \beta_1^A = \sum_{j=1}^{b} \beta_1^B = \sum_{k=1}^{c} \beta_1^C = 0$$
$$\sum_{i=1}^{a} \beta_{ij}^{AB} = \sum_{j=1}^{b} \beta_{ij}^{AB} = \sum_{j=1}^{b} \beta_{jk}^{BC} = \sum_{k=1}^{c} \beta_{jk}^{BC} = \sum_{i=1}^{a} \beta_{ik}^{AC} = \sum_{k=1}^{c} \beta_{ik}^{AC} = 0$$
$$\sum_{i=1}^{a} \beta_{1jk}^{ABC} = \sum_{j=1}^{b} \beta_{ijk}^{ABC} = \sum_{k=1}^{c} \beta_{ijk}^{ABC} = 0$$

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

$$
\begin{aligned}
&1+\underbrace{(a-1)}_{A\text{ main effects}}+\underbrace{(b-1)}_{B\text{ main effects}}+\underbrace{(c-1)}_{C\text{ main effects}}\\
&\quad+\underbrace{(a-1)(b-1)}_{AB\text{ interactions}}+\underbrace{(b-1)(c-1)}_{BC\text{ interactions}}+\underbrace{(a-1)(c-1)}_{AC\text{ interactions}}\\
&\quad+\underbrace{(a-1)(b-1)(c-1)}_{ABC\text{ interactions}}\\
&=abc
\end{aligned}
$$

There are several simplifications of the 3-way interaction model, such as

- Model $A*B+B*C+A*C$:

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

- Model $A*B+A*C$

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

- Model $A+B*C$:

$$\text{logit}(\pi_{ijk})=\alpha+\beta_i^A+\beta_j^B+\beta_k^C+\beta_{jk}^{BC}$$

- Model $A+B+C$:

$$\text{logit}(\pi_{ijk})=\alpha+\beta_i^A+\beta_j^B+\beta_k^C$$

- Model $A*B$:

$$\text{logit}(\pi_{ijk})=\alpha+\beta_i^A+\beta_j^B+\beta_{ij}^{AB}$$

- In all the models, constraints on the main effects and interactions are the same as those on the corresponding parameters in the 3-way interaction models.

- Generally, models must maintain *hierarchy* — cannot include an interaction terms without including the relevant main effects and lower order interactions

Interpretation of Model $A+B+C$ and its Coefficients

In Model $A+B+C$:

$$\text{logit}(\pi_{ijk})=\alpha+\beta_i^A+\beta_j^B+\beta_k^C$$

- There are $1+(a-1)+(b-1)+(c-1)$ effective parameters.
- Under the baseline constraint $\beta_1^A=\beta_1^B=\beta_1^C=0$,
  - The odds of $Y=1$ when $A=i$ is the odds of $Y=1$ when $A=1$ multiplied by a factor of $e^{\beta_i^A}$, regardless of B and C
  - Interpretation for $e^{\beta_j^B}$ and $e^{\beta_k^C}$: 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.

	$Z=1$		$Z=2$	
	$X=1$	$X=2$	$X=1$	$X=2$
$Y=1$	a	b	A	B
$Y=2$	c	d	C	D

Homogeneous XY association given Z means

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

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

which means homogeneous YZ association given X.

	$X=1$		$X=2$	
	$Z=1$	$Z=2$	$Z=1$	$Z=2$
$Y=1$	a	A	b	B
$Y=2$	c	C	d	D

Interpretation of Model $A * B$ and its Coefficients

$$\text{logit}(\pi_{ijk}) = \alpha + \beta_i^A + \beta_j^B + \beta_{ij}^{AB}$$

- Number of parameters: $1 + (a-1) + (b-1) + (a-1)(b-1)$
- C is not in the model
  $\Rightarrow$ Y and C are conditionally independent given A, B
- Under the baseline constraint $\beta_1^A = \beta_1^B = \beta_{1j}^{AB} = \beta_{i1}^{AB} = 0$, the conditional odds ratio for

	$Y=1$	$Y=0$
$A=i$		
$A=1$		

  equals

$$\begin{cases} e^{\beta_i^A} & \text{when } B=1; \\ e^{\beta_i^A + \beta_{ij}^{AB}} & \text{when } B=j \end{cases}$$

  So $e^{\beta_{ij}^{AB}}$ is a ratio of two odds ratios.
- Odds ratios of YA change with B (but not C),
  $\Rightarrow$ no homogeneous YA association
- Likewise, can show that odds ratios of YB changes with A (but not C). $\Rightarrow$ No homogeneous YB association

Interpretation of Model $A * B * C$ and its Coefficients

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

Under the baseline constraint, the conditional odds ratio for

	$Y=1$	$Y=0$
$A=i$		
$A=1$		

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

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

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

Model $A * B + B * C + A * C$:

- YA odds ratios change with both B and C
- YB odds ratios change with both A and C
- YC odds ratios change with both A and B
- no 3-way interactions means that

$$\frac{YA \text{ odds ratio when } B=j_1}{YA \text{ odds ratio when } B=j_2}$$

  do not change with C

Model $A * B + B * C$:

- YA odds ratios change with B but not C
- YC odds ratios change with A but not B
- YB odds ratios change with both A and C