STAT 226 Lecture 17

Yibi Huang

Example (Mouse Muscle Tension)

- A study to examine relationship between two drugs and muscle tension
- 4-way flat contingency table ($2 \times 2 \times 2 \times 2$) with 4 variables
  - Tension (response): change in muscle tension: High, Low
  - Drug: drug 1, drug 2 .... primary predictor
  - Weight: weight of muscle: High, Low
  - Muscle: muscle type: 1, 2

		Drug 1		Drug 2	
		Muscle 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

The layout of this flat table is bad because ...

A better layout:

	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 4 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

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

Creating Multi-Way Tables in R

In week 4 problem session, we showed how to create multi-way tables in R

```
muscle.tab = array(
  c( 3,21, 3,10,   # Drug-Tension partial given W = High, Type 1
    22,32,45,23,   # Drug-Tension partial given W = Low,  Type 1
    23,11,41,21,   # Drug-Tension partial given W = High, Type 2
     4,12, 6,22),  # Drug-Tension partial given W = Low,  Type 2
  dim = c(2,2,2,2),
  dimnames = list(
    drug = c("1","2"),
    tension = c("High", "Low"),
    weight = c("High", "Low"),
    muscle = c("1","2")
  )
)
muscle.tab = as.table(muscle.tab) # cannot skip this step!
```

See week 4 problem session for how to print a multi-way table as a flat-table.

```
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
```

Various Formats of Multi-Way Table Data

Various Formats of Multi-Way Table Data

There are several formats in R for multi-way table data.

1. Table, created using `array()` or `xtabs()`
   - `ftable()` and CMH test require this format
2. Ungrouped data — data frame
3. Grouped data — long format — data frame
4. Grouped data — wide format — data frame

Data must be either either **ungrouped** or in **wide format** if grouped to fit a `glm()` model.

Ungrouped Data to Tables — `xtabs()`

`xtabs()` can convert a data frame of ungrouped data into to multi-way tables.

```
muscle.ug = read.table(
  "http://www.stat.uchicago.edu/~yibi/s226/mousemuscle_ungrouped.txt",
  header=TRUE)
muscle.ug[1:8,]
  drug tension weight muscle
1    1    High   High      1
2    1    High   High      1
3    1    High   High      1
4    2    High   High      1
5    2    High   High      1
6    2    High   High      1
7    2    High   High      1
8    2    High   High      1
# ...(291 more rows omitted)...
```

Ungrouped Data to Tables (2)

```
muscle.tab = xtabs(~ weight + muscle + drug + tension, data=muscle.ug)
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
```

Long Format of Grouped Data

- One column for each variable of the multi-way table
- One column (Freq) for the cell counts of the multi-way table

weight	muscle	drug	tension	Freq
High	1	1	High	3
Low	1	1	High	22
High	2	1	High	23
Low	2	1	High	4
High	1	2	High	21
Low	1	2	High	32
High	2	2	High	11
Low	2	2	High	12
High	1	1	Low	3
Low	1	1	Low	45
High	2	1	Low	41
Low	2	1	Low	6
High	1	2	Low	10
Low	1	2	Low	23
High	2	2	Low	21
Low	2	2	Low	22

Wide Format of Grouped Data

- One column for each explanatory variable
- k columns for the response if it has k levels

```
drug weight muscle tension.High tension.Low
   1   High      1            3           3
   1   High      2           23          41
   1    Low      1           22          45
   1    Low      2            4           6
   2   High      1           21          10
   2   High      2           11          21
   2    Low      1           32          23
   2    Low      2           12          22
```

Table to Long Format — `as.data.frame()`

`as.data.frame()` can convert a multi-way table to a data frame in long format.

```
muscle.long = as.data.frame(muscle.tab)
muscle.long
   weight muscle drug tension Freq
1    High      1    1    High    3
2     Low      1    1    High   22
3    High      2    1    High   23
4     Low      2    1    High    4
5    High      1    2    High   21
6     Low      1    2    High   32
7    High      2    2    High   11
8     Low      2    2    High   12
9    High      1    1     Low    3
10    Low      1    1     Low   45
11   High      2    1     Low   41
12    Low      2    1     Low    6
13   High      1    2     Low   10
```

Long Format to Wide Format — dcast()

`dcast()` in the `reshape2` library can convert data frames from long to wide format.

```
# install.packages("reshape2")  # only install ONCE!
library(reshape2)
muscle.wide = dcast(muscle.long,
                    drug+weight+muscle ~ tension,
                    value.var="Freq")
muscle.wide
  drug weight muscle High Low
1    1   High      1    3   3
2    1   High      2   23  41
3    1    Low      1   22  45
4    1    Low      2    4   6
5    2   High      1   21  10
6    2   High      2   11  21
7    2    Low      1   32  23
8    2    Low      2   12  22
```

Wide Format to Long Format

Use the `melt()` function in the `reshape2` library to convert data from wide format to long format.

```
melt(muscle.wide, id.vars=c("weight","muscle","drug"))
   weight muscle drug variable value
1    High      1    1     High     3
2    High      2    1     High    23
3     Low      1    1     High    22
4     Low      2    1     High     4
5    High      1    2     High    21
6    High      2    2     High    11
7     Low      1    2     High    32
8     Low      2    2     High    12
9    High      1    1      Low     3
10   High      2    1      Low    41
11    Low      1    1      Low    45
12    Low      2    1      Low     6
13   High      1    2      Low    10
14   High      2    2      Low    21
```

Wide Format to Long Format (2)

```
muscle.long = melt(muscle.wide, id.vars=c("weight","muscle","drug"))
names(muscle.long)
[1] "weight"   "muscle"   "drug"     "variable" "value"
names(muscle.long)[4] = "tension"
names(muscle.long)[5] = "Freq"
muscle.long
   weight muscle drug tension Freq
1    High      1    1    High    3
2    High      2    1    High   23
3     Low      1    1    High   22
4     Low      2    1    High    4
5    High      1    2    High   21
6    High      2    2    High   11
7     Low      1    2    High   32
8     Low      2    2    High   12
9    High      1    1     Low    3
10   High      2    1     Low   41
11    Low      1    1     Low   45
12    Low      2    1     Low    6
```

Long Format to Table

`xtabs()` can also convert grouped data in long format to tables.

```
muscle.tab = xtabs(Freq ~ weight + muscle + drug + tension, data= muscl
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
```

Summary of Conversions Between Data Formats

Ungrouped

`xtabs()` ↓

Table

`as.data.frame()` ↓↑ `xtabs()`

Long Format

`dcast()` ↓↑ `melt()`

Wide Format

Logistic Models for Multi-way Tables

Logistic Models for Multi-way 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

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

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

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

can be written in terms of the dummy variables as

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

- 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 to drop one of the dummy variables A_1, B_1, C_1 for each categorical predictor from the model, which is equivalent to setting the coefficients for those dummy variables to 0.

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

As A_1, B_1, and C_1 are dropped, the interaction terms that involve those levels are also dropped. So the coefficients for those interaction terms are set to 0

$$\beta_{1j}^{AB} = \beta_{i1}^{AB} = \beta_{1k}^{BC} = \beta_{j1}^{BC} = \beta_{1k}^{AC} = \beta_{i1}^{AC} = 0$$

$$\beta_{1jk}^{ABC} = \beta_{i1k}^{ABC} = \beta_{ij1}^{ABC} = 0$$

- the effective number of parameters for a main effect is

$$\text{number of levels} - 1$$

- the effective number 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.

- 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 the Model $A + B + C$:

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \log(\text{odds of } \{Y = 1\}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C$$

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

- $1 + (a-1) + (b-1) + (c-1)$ effective parameters in total since $\beta_1^A = \beta_1^B = \beta_1^C = 0$
- Conditional OR between $\{Y = 0, 1\}$ and $\{A = 1, i\}$ given $B = j$ and $C = k$ is

$$\frac{\text{odds of } (Y = 1 \text{ given } A = i, B = j, C = k)}{\text{odds of } (Y = 1 \text{ given } A = 1, B = j, C = k)}$$
$$= \frac{\exp(\cancel{\alpha} + \beta_i^A + \cancel{\beta_j^B} + \cancel{\beta_k^C})}{\exp(\cancel{\alpha} + \beta_1^A + \cancel{\beta_j^B} + \cancel{\beta_k^C})} = \exp(\beta_i^A - \beta_1^A) = \exp(\beta_i^A)$$

  which doesn't change with the levels of B and C.
- Interpretation for $\exp(\beta_j^B)$ and $\exp(\beta_k^C)$: Likewise
- Homogeneous YA, YB, and YC association

Example (Mouse Muscle Tension)

```
# wide-format data
names(muscle.wide)
[1] "drug"   "weight" "muscle" "High"   "Low"
names(muscle.wide)[c(1,2,3)] = c("D","W","M")
muscle.wide$M = as.factor(muscle.wide$M)
muscle.wide$D = as.factor(muscle.wide$D)
```

Let's first fit a model with W, M, and D main effects only.

```
glm1 = glm(cbind(High,Low) ~ W + M + D, family=binomial,
           data=muscle.wide)
```

Fitted model coefficients:

```
glm1$coef
(Intercept)        WLow          M2          D2
   -0.03424    -0.41866    -0.70899     0.58657
```

For the model $W + M + D$: $\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \alpha + \beta_i^W + \beta_j^M + \beta_k^D$ where $\pi_{ijk} = P(\text{High Tension} \mid W = i, M = j, D = k)$,

```
glm1$coef
(Intercept)        WLow          M2          D2
   -0.03424    -0.41866    -0.70899     0.58657
```

R gives the estimated coefficients:

$$\widehat{\alpha} \approx -0.0342, \quad \widehat{\beta}_L^W \approx -0.419, \quad \widehat{\beta}_2^M \approx -0.709, \quad \widehat{\beta}_2^D \approx 0.587.$$

- What are the values for $\widehat{\beta}_H^W$, $\widehat{\beta}_1^M$ and $\widehat{\beta}_1^D$?
- What is the estimated value for $\pi = P(\text{tension} = \text{High})$ for low muscle weight, Type 1 muscle, when Drug 1 is applied?

For the model $W + M + D$: $\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \alpha + \beta_i^W + \beta_j^M + \beta_k^D$ where $\pi_{ijk} = P(\text{High Tension} \,|\, W = i, M = j, D = k)$,

```
glm1$coef
(Intercept)        WLow          M2          D2
   -0.03424    -0.41866    -0.70899     0.58657
```

R gives the estimated coefficients:

$$\widehat{\alpha} \approx -0.0342, \quad \widehat{\beta_L^W} \approx -0.419, \quad \widehat{\beta_2^M} \approx -0.709, \quad \widehat{\beta_2^D} \approx 0.587.$$

- What are the values for $\widehat{\beta_H^W}$, $\widehat{\beta_1^M}$ and $\widehat{\beta_1^D}$? All zero!
- What is the estimated value for $\pi = P(\text{tension} = \text{High})$ for low muscle weight, Type 1 muscle, when Drug 1 is applied?

For the model $W + M + D$: $\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \alpha + \beta_i^W + \beta_j^M + \beta_k^D$ where $\pi_{ijk} = P(\text{High Tension} \mid W = i, M = j, D = k)$,

```
glm1$coef
(Intercept)        WLow          M2          D2
   -0.03424    -0.41866    -0.70899     0.58657
```

R gives the estimated coefficients:

$$\widehat{\alpha} \approx -0.0342, \quad \widehat{\beta}_L^W \approx -0.419, \quad \widehat{\beta}_2^M \approx -0.709, \quad \widehat{\beta}_2^D \approx 0.587.$$

- What are the values for $\widehat{\beta}_H^W$, $\widehat{\beta}_1^M$ and $\widehat{\beta}_1^D$? All zero!
- What is the estimated value for $\pi = P(\text{tension} = \text{High})$ for low muscle weight, Type 1 muscle, when Drug 1 is applied?
  When W = L, M = 1, D = 1,

$$\widehat{\pi} = \frac{\exp(\widehat{\alpha} + \widehat{\beta}_L^W + \widehat{\beta}_1^M + \widehat{\beta}_1^D)}{1 + \exp(\widehat{\alpha} + \widehat{\beta}_L^W + \widehat{\beta}_1^M + \widehat{\beta}_1^D)}$$
$$= \frac{\exp(-0.0342 + (-0.419) + 0 + 0)}{1 + \exp(-0.0342 + (-0.419) + 0 + 0)} \approx 0.39$$

For the main effect model

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \alpha + \beta_i^W + \beta_j^M + \beta_k^D$$

how to interpret the parameter estimates below?

$$\begin{aligned}\widehat{\alpha} \approx -0.0342, \quad & \widehat{\beta}_H^W = 0, & \widehat{\beta}_1^M = 0, \quad & \widehat{\beta}_1^D = 0, \\ & \widehat{\beta}_L^W \approx -0.419, & \widehat{\beta}_2^M \approx -0.709, \quad & \widehat{\beta}_2^D \approx 0.587\end{aligned}$$

- The odds of High tension if Drug 2 is applied are $e^{\widehat{\beta}_2^D} = e^{0.587} \approx 1.8$ times the odds if Drug 1 is applied on the same type of muscle of the same weight.
- The odds of High tension for Type 2 muscle are $e^{\widehat{\beta}_2^M} = e^{-0.709} \approx 0.49$ times the odds for Type 1 muscle of the same weight with the same drug applied.
- The odds of High tension for high-weight muscle are $e^{-\widehat{\beta}_L^W} = e^{0.419} \approx 1.52$ times the odds for low-weight muscle of the same type with the same drug applied.

Model $A * B + C$

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \log(\text{odds of } \{Y = 1\}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB}$$

- By model hierarchy, must include A and B if $A * B$ is included in the model.
  So the model $A * B + C$ is equivalent to $A + B + C + A * B$

Model $A * B + C$

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \log(\text{odds of } \{Y = 1\}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB}$$

- By model hierarchy, must include A and B if $A * B$ is included in the model.
  So the model $A * B + C$ is equivalent to $A + B + C + A * B$
- Number of parameters
  $= 1 + (a - 1) + (b - 1) + (c - 1) + (a - 1)(b - 1)$ because of the constraints

$$\beta_1^A = \beta_1^B = \beta_1^C = 0$$
$$\beta_{i1}^{AB} = 0 \quad \text{for } i = 1, \ldots, a$$
$$\beta_{1j}^{AB} = 0 \quad \text{for } j = 1, \ldots, b$$

Model $A * B + C$: Interpretation

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \log(\text{odds of } \{Y=1\}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB}$$

The conditional OR between $\{Y = 0, 1\}$ and $\{A = 1, i\}$ given $B = j$ and $C = k$ is

$$\frac{\text{odds of } (Y = 1 \text{ given } A = i, B = j, C = k)}{\text{odds of } (Y = 1 \text{ given } A = 1, B = j, C = k)}$$

$$= \frac{\exp(\cancel{\alpha} + \beta_i^A + \cancel{\beta_j^B} + \cancel{\beta_k^C} + \beta_{ij}^{AB})}{\exp(\cancel{\alpha} + \beta_1^A + \cancel{\beta_j^B} + \cancel{\beta_k^C} + \beta_{1j}^{AB})}$$

$$= \exp(\beta_i^A - \underbrace{\beta_1^A}_{=0} + \beta_{ij}^{AB} - \underbrace{\beta_{1j}^{AB}}_{=0}) = e^{\beta_i^A + \beta_{ij}^{AB}} = \begin{cases} e^{\beta_i^A} & \text{if } B = 1 \\ e^{\beta_i^A + \beta_{ij}^{AB}} & \text{if } B = j \end{cases}$$

which changes with the levels of B (but not C)

- YA association is NOT homogeneous
- can show likewise that the conditional ORs of YB change with A (but not C). ⇒ No homogeneous YB association

Model $A * B + C$: Interpretation

Under the $A * B + C$ model

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \log(\text{odds of } \{Y=1\}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB},$$

the conditional OR between $\{Y = 0, 1\}$ and $\{C = 1, k\}$ given $A = i$ and $B = j$ is

$$\frac{\text{odds of } (Y = 1 \text{ given } A = i, B = j, C = k)}{\text{odds of } (Y = 1 \text{ given } A = i, B = j, C = 1)}$$

$$= \frac{\exp(\cancel{\alpha} + \cancel{\beta_i^A} + \cancel{\beta_j^B} + \beta_k^C + \cancel{\beta_{ij}^{AB}})}{\exp(\cancel{\alpha} + \cancel{\beta_i^A} + \cancel{\beta_j^B} + \beta_1^C + \cancel{\beta_{ij}^{AB}})} = \exp(\beta_k^C - \beta_1^C) = e^{\beta_k^C}$$

which doesn't change with the levels of A and B.

- homogeneous YC association given A and B.
- If one further assumes that $\beta_k^C = 0$ for all k, then Y and C would be conditionally independent given A, B

Example (Mouse Muscle Tension) — $W + M * D$

```
glm2 = glm(cbind(High,Low) ~ W + M * D, family=binomial,
           data=muscle.wide)
glm2$coef
(Intercept)        WLow          M2          D2       M2:D2
    -0.4682     -0.2011     -0.0596      1.0717     -1.0676
```

$$
\begin{array}{llll}
\widehat{\alpha} \approx -0.4682, & \widehat{\beta}_H^W = 0, & \widehat{\beta}_1^M = 0, & \widehat{\beta}_1^D = 0, \\
 & \widehat{\beta}_L^W \approx -0.2011, & \widehat{\beta}_2^M \approx -0.0596, & \widehat{\beta}_2^D \approx 1.0717 \\
\widehat{\beta}_{11}^{MD} = 0, & \widehat{\beta}_{12}^{MD} = 0, & \widehat{\beta}_{21}^{MD} = 0, & \widehat{\beta}_{22}^{MD} = -1.0676
\end{array}
$$

The estimated π when $W = L$, $M = 1$, $D = 1$ is

$$
\begin{aligned}
\widehat{\pi} &= \frac{\exp(\widehat{\alpha} + \widehat{\beta}_L^W + \widehat{\beta}_1^M + \widehat{\beta}_1^D + \widehat{\beta}_{11}^{MD})}{1 + \exp(\widehat{\alpha} + \widehat{\beta}_L^W + \widehat{\beta}_1^M + \widehat{\beta}_1^D + \widehat{\beta}_{11}^{MD})} \\
&= \frac{\exp(-0.4676 + (-0.2011) + 0 + 0 + 0)}{1 + \exp(-0.4676 + (-0.2011) + 0 + 0 + 0)} \approx 0.34
\end{aligned}
$$

Example (Mouse Muscle Tension) — Drug Effect Under $W+M*D$

Under the model $W + M * D$

$$\log(\text{odds of high tension}) = \alpha + \beta_i^W + \beta_j^M + \beta_k^D + \beta_{jk}^{MD}$$

The conditional OR between {Tension = H, L} and {$D = 1, 2$} given $W = j$ and $M = k$ is

$$\frac{\text{odds of high tension given } W = i, M = j, D = 2}{\text{odds of high tension given } W = i, M = j, D = 1}$$

$$= \frac{\exp(\cancel{\widehat{\alpha}} + \cancel{\widehat{\beta}_i^W} + \cancel{\widehat{\beta}_j^M} + \widehat{\beta}_2^D + \widehat{\beta}_{j2}^{MD})}{\exp(\cancel{\widehat{\alpha}} + \cancel{\widehat{\beta}_i^W} + \cancel{\widehat{\beta}_j^M} + \widehat{\beta}_1^D + \widehat{\beta}_{j1}^{MD})} = \exp(\widehat{\beta}_2^D - \underbrace{\widehat{\beta}_1^D}_{=0} + \widehat{\beta}_{j2}^{MD} - \underbrace{\widehat{\beta}_{j1}^{MD}}_{=0}) = e^{\widehat{\beta}_2^D - \widehat{\beta}_{j2}^{MD}}$$

$$= \begin{cases} e^{1.071+0} \approx 2.9 & \text{for Type 1 muscle} \\ e^{1.071+(-1.068)} \approx 1.004 & \text{for Type 2 muscle} \end{cases}$$

Conclusion: Drug 1 and 2 have nearly identical effects on Type 2 muscle, while Drug 1 is significantly more effective in reducing muscle tension than drug 2.

Example (Mouse Muscle Tension) — Weight Effects

Under the model $W + M * D$

$$\log(\text{odds of high tension}) = \alpha + \beta_i^W + \beta_j^M + \beta_k^D + \beta_{jk}^{MD}$$

the estimated conditional OR between {Tension = H, L} and $\{W = H, L\}$ given $M = j$, $D = k$ is $e^{\widehat{\beta}_H^W - \widehat{\beta}_L^W} \approx e^{0-(-0.2011)} \approx 1.22$, which doesn't change with the levels of M or D.

Interpretation: If muscle weight is high, the odds of high tension were 1.22 times the odds when the muscle weight is low, given the same drug and same muscle type.

Wald CI for Conditional OR

```
summary(glm2)$coef
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.4682     0.3648 -1.2836 0.199284
WLow         -0.2011     0.2941 -0.6837 0.494173
M2           -0.0596     0.4155 -0.1434 0.885937
D2            1.0717     0.3408  3.1450 0.001661
M2:D2        -1.0676     0.5198 -2.0536 0.040014
```

95% Wald CI for the conditional OR for W & Tension given D & M:

$$\exp(-\widehat{\beta}_L^W \pm 1.96\text{SE}) = \exp(0.2011 \pm 1.96 \times 0.2941) \approx (0.687, 2.176).$$

As the CI contains 1, we see that Tension and W could be *conditionally independent* given M & D.

Test of Conditional Independence

Under Model $W + M * D$,

conditional indep. of Tension & W given M & D $\iff \beta_L^W = 0.$

Both Wald and LR tests of $\beta_L^W = 0$ give P-values ≈ 0.49 (next page)

- Tension and W could be conditionally indep. given M and D.

```
> summary(glm2)
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.4682     0.3648  -1.284  0.19928
WL           -0.2011     0.2941  -0.684  0.49417 <-- Wald test p-value
M2           -0.0596     0.4155  -0.143  0.88594
D2            1.0717     0.3408   3.145  0.00166 **
M2:D2        -1.0675     0.5198  -2.054  0.04001 *
> drop1(glm2, test="Chisq")
Model:
T ~ W + M * D
       Df Deviance    AIC    LRT Pr(>Chi)
<none>      1.0596 41.176
W       1   1.5289 39.646 0.4693  0.49332  <-- LR test p-value
M:D     1   5.3106 43.427 4.2510  0.03923 *
```

Model $A * B + B * C$

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \log(\text{odds of } \{Y = 1\}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{jk}^{BC}$$

- By model hierarchy, it's equivalent to $A + B + C + A * B + B * C$ as A, B, and C must be included if $A * B$ and $B * C$ have been included in the model

Model $A * B + B * C$

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \log(\text{odds of } \{Y = 1\}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{jk}^{BC}$$

- By model hierarchy, it's equivalent to $A + B + C + A * B + B * C$ as A, B, and C must be included if $A * B$ and $B * C$ have been included in the model
- Number of parameters $= 1 + (a-1) + (b-1) + (c-1) + (a-1)(b-1) + (b-1)(c-1)$ because of the constraints

$$\begin{aligned}
\beta_1^A = \beta_1^B = \beta_1^C &= 0 \\
\beta_{1j}^{AB} &= 0 \quad \text{for } j = 1, \ldots, b \\
\beta_{i1}^{AB} &= 0 \quad \text{for } i = 1, \ldots, a \\
\beta_{1k}^{BC} &= 0 \quad \text{for } k = 1, \ldots, c \\
\beta_{j1}^{BC} &= 0 \quad \text{for } j = 1, \ldots, b
\end{aligned}$$

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

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \log(\text{odds of } \{Y=1\}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{jk}^{BC}$$

The conditional OR between $\{Y = 0, 1\}$ and $\{A = 1, i\}$ given $B = j$ and $C = k$ is

$$\frac{\text{odds of } (Y=1 \text{ given } A=i, B=j, C=k)}{\text{odds of } (Y=1 \text{ given } A=1, B=j, C=k)}$$

$$= \frac{\exp(\cancel{\alpha} + \beta_i^A + \cancel{\beta_j^B} + \cancel{\beta_k^C} + \beta_{ij}^{AB} + \cancel{\beta_{jk}^{BC}})}{\exp(\cancel{\alpha} + \beta_1^A + \cancel{\beta_j^B} + \cancel{\beta_k^C} + \beta_{1j}^{AB} + \cancel{\beta_{jk}^{BC}})}$$

$$= \exp(\beta_i^A - \underbrace{\beta_1^A}_{=0} + \beta_{ij}^{AB} - \underbrace{\beta_{1j}^{AB}}_{=0}) = e^{\beta_i^A + \beta_{ij}^{AB}} = \begin{cases} e^{\beta_i^A} & \text{if } B = 1 \\ e^{\beta_i^A + \beta_{ij}^{AB}} & \text{if } B = j \end{cases}$$

which changes with the levels of B (but not C)

- YA association is NOT homogeneous
- Likewise, can show the conditional OR of YC changes with B (but not A). $\Rightarrow$ No homogeneous YC association

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

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \log(\text{odds of } \{Y = 1\}) = \alpha + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{jk}^{BC}$$

The conditional OR between $\{Y = 0, 1\}$ and $\{B = 1, j\}$ given $A = i$ and $C = k$ is

$$\frac{\text{odds of } (Y = 1 \text{ given } A = i, B = j, C = k)}{\text{odds of } (Y = 1 \text{ given } A = i, B = 1, C = k)}$$

$$= \frac{\exp(\cancel{\alpha} + \cancel{\beta_i^A} + \beta_j^B + \cancel{\beta_k^C} + \beta_{ij}^{AB} + \beta_{jk}^{BC})}{\exp(\cancel{\alpha} + \cancel{\beta_i^A} + \underbrace{\beta_1^B}_{=0} + \cancel{\beta_k^C} + \underbrace{\beta_{i1}^{AB}}_{=0} + \underbrace{\beta_{1k}^{BC}}_{=0})} = e^{\beta_j^B + \beta_{ij}^{AB} + \beta_{jk}^{BC}}$$

$$= \begin{cases} e^{\beta_j^B} & \text{if } i = k = 1 \\ e^{\beta_j^B + \beta_{ij}^{AB}} & \text{if } i \neq 1, k = 1 \\ e^{\beta_j^B + \beta_{jk}^{BC}} & \text{if } i = 1, k \neq 1 \\ e^{\beta_j^B + \beta_{ij}^{AB} + \beta_{jk}^{BC}} & \text{if } i \neq 1, k \neq 1 \end{cases}$$

which changes with the levels of both A and C.

If No 3-way Interaction...

Under the Model $A * B + B * C$ or $A * B + B * C + A * C$

$$\text{YB odds ratio given } A = i \text{ and } C = k \text{ is} \begin{cases} e^{\beta_j^B} & \text{if } i = k = 1 \\ e^{\beta_j^B + \beta_{ij}^{AB}} & \text{if } i \neq 1, k = 1 \\ e^{\beta_j^B + \beta_{jk}^{BC}} & \text{if } i = 1, k \neq 1 \\ e^{\beta_j^B + \beta_{ij}^{AB} + \beta_{jk}^{BC}} & \text{if } i \neq 1, k \neq 1 \end{cases}$$

So

$$\frac{YB \text{ odds ratio when } A = i \text{ and } C = k}{YB \text{ odds ratio when } A = 1 \text{ and } C = k} = \begin{cases} \frac{e^{\beta_j^B + \beta_{ij}^{AB}}}{e^{\beta_j^B}} = e^{\beta_{ij}^{AB}} & \text{when } k = 1; \\ \frac{e^{\beta_j^B + \beta_{ij}^{AB} + \beta_{jk}^{BC}}}{e^{\beta_j^B + \beta_{jk}^{BC}}} = e^{\beta_{ij}^{AB}} & \text{when } k \neq 1. \end{cases}$$

which doesn't change w/ the level of C.

Model $A * B * C$

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

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

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