

---

# Lecture 22: Introduction to Log-linear Models

Dipankar Bandyopadhyay, Ph.D.

BMTRY 711: Analysis of Categorical Data Spring 2011

Division of Biostatistics and Epidemiology

Medical University of South Carolina

# Log-linear Models

---

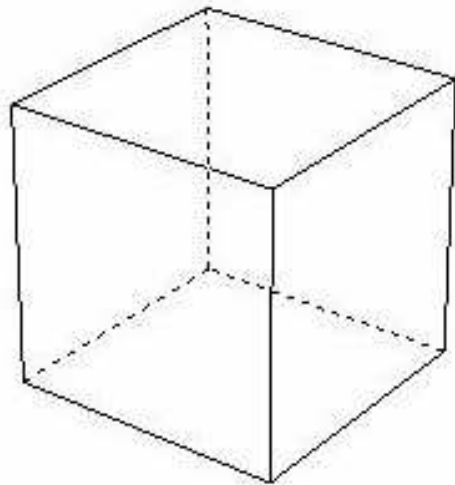
- Log-linear models are a Generalized Linear Model
- A common use of a log-linear model is to model the cell counts of a contingency table
- The systematic component of the model describe how the expected cell counts vary as a result of the explanatory variables
- Since the response of a log linear model is the cell count, no measured variables are considered the response

# Recap from Previous Lectures

---

- Lets suppose that we have an  $I \times J \times Z$  contingency table.
- That is, There are  $I$  rows,  $J$  columns and  $Z$  layers.

(picture of cube)



# Conditional Independence

---

We want to explore the concepts of independence using a log-linear model.

But first, let's review some probability theory.

Recall, two variables  $A$  and  $B$  are independent if and only if

$$P(AB) = P(A) \times P(B)$$

Also recall that Bayes Law states for any two random variables

$$P(A|B) = \frac{P(AB)}{P(B)}$$

and thus, when  $X$  and  $Y$  are independent,

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$$

# Conditional Independence

---

Definitions:

In layer  $k$  where  $k \in \{1, 2, \dots, Z\}$ ,  $X$  and  $Y$  are conditionally independent at level  $k$  of  $Z$  when

$$P(Y = j|X = i, Z = k) = P(Y = j|Z = k), \quad \forall i, j$$

If  $X$  and  $Y$  are conditionally independent at ALL levels of  $Z$ , then  $X$  and  $Y$  are **CONDITIONALLY INDEPENDENT**.

# Application of the Multinomial

---

Suppose that a single multinomial applies to the entire three-way table with cell probabilities equal to

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

Let

$$\begin{aligned}\pi_{.jk} &= \sum_X P(X = i, Y = j, Z = k) \\ &= P(Y = j, Z = k)\end{aligned}$$

Then,

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

by application of Bayes law. (The event  $(Y = j) = A$  and  $(X = i, Z = k) = B$ ).

---

Then if  $X$  and  $Y$  are conditionally independent at level  $z$  of  $Z$ ,

$$\begin{aligned}\pi_{ijk} &= P(X = i, Z = k)P(Y = j|X = i, Z = k) \\ &= \pi_{i \cdot k}P(Y = j|Z = k) \\ &= \pi_{i \cdot k}P(Y = j, Z = k)/P(Z = k) \\ &= \pi_{i \cdot k}\pi_{\cdot jk}/\pi_{\cdot \cdot k}\end{aligned}$$

for all  $i, j$ , and  $k$ .

## $(2 \times 2)$ table

---

- Let's suppose we are interested in a  $(2 \times 2)$  table for the moment
- Let  $X$  describe the row effect and  $Y$  describe the column effect
- If  $X$  and  $Y$  are independent, then

$$\pi_{ij} = \pi_{i\cdot} \pi_{\cdot j}$$

- Then the expected cell count for the  $ij^{th}$  cell would be

$$n\pi_{ij} = \mu_{ij} = n\pi_{i\cdot} \pi_{\cdot j}$$

Or,

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y$$

- This model is called the log-linear model of independence



# Interaction term

---

- In terms of a regression model, a significant interaction term indicates that the response varies as a function of the combination of  $X$  and  $Y$
- That is, changes in the response as a function of  $X$  require the specification of  $Y$  to explain the change
- This implies that  $X$  and  $Y$  are **NOT INDEPENDENT**
- Let  $\lambda_{ij}^{XY}$  denote the interaction term
- Testing  $\lambda_{ij}^{XY} = 0$  is a test of independence

# Log-linear Models for $(2 \times 2)$ tables

---

- Unifies all probability models discussed.
- We will use log-linear models to describe designs in which
  1. Nothing is fixed (Poisson)
  2. The total is fixed (multinomial sampling or double dichotomy)
  3. One margin is fixed (prospective or case-control)
- Represents expected cell counts as functions of row and column effects and interactions
- Makes no distinction between response and explanatory variables.
- Can be generalized to larger dimensions ( $R \times C$ ,  $2 \times 2 \times 2$ ,  $2 \times 2 \times K$ , etc.)

As before, for random counts, double dichotomy, prospective, and case-control designs

Variable (Y)

1

2

Variable (X)

1

2

$Y_{11}$	$Y_{12}$	$Y_{1+}$
$Y_{21}$	$Y_{22}$	$Y_{2+}$
$Y_{+1}$	$Y_{+2}$	$Y_{++}$

The expected counts are  $\mu_{jk} = E(Y_{jk})$

		Variable (Y)		
		1	2	
Variable (X)	1	$\mu_{11}$	$\mu_{12}$	$\mu_{1+}$
	2	$\mu_{21}$	$\mu_{22}$	$\mu_{2+}$
		$\mu_{+1}$	$\mu_{+2}$	$\mu_{++}$

# Example

An example of such a  $(2 \times 2)$  table is

Cold incidence among French Skiers (Pauling, Proceedings of the national Academy of Sciences, 1971).

		OUTCOME		
		COLD	NO COLD	Total
T R E A T M E N T	VITAMIN C	17	122	139
	NO VITAMIN C	31	109	140
Total		48	231	279

Regardless of how these data were actually collected, we have shown that the estimate of the odds ratio is the same for all designs, as is the likelihood ratio test and Pearson's chi-square for independence.

# Using SAS Proc Freq

---

```
data one;
input vitc cold count;
cards;
1 1 17
1 2 122
2 1 31
2 2 109
;

proc freq;
  table vitc*cold / chisq measures;
  weight count;
run;
```

```
/* SELECTED OUTPUT */
```

```
Statistics for Table of vitc by cold
```

Statistic	DF	Value	Prob
Chi-Square	1	4.8114	0.0283 (Pearsons')
Likelihood Ratio Chi-Square	1	4.8717	0.0273

```
Estimates of the Relative Risk (Row1/Row2)
```

Type of Study	Value	95% Confidence Limits
Case-Control (Odds Ratio)	0.4900	0.2569 0.9343

Instead of just doing this analysis for a  $(2 \times 2)$  table, we will now discuss a 'log-linear' model for a  $(2 \times 2)$  table

# Expected Counts

Expected cell counts  $\mu_{jk} = E(Y_{jk})$  for different designs

		Y			
		1	2		
X	1	Poisson	$\mu_{11}$	$\mu_{12}$	(Poisson mean)
		Double Dichotomy	$np_{11}$	$np_{12}$	(table prob sums to 1)
		Prospective	$n_1 p_1$	$n_1(1 - p_1)$	(row prob sums to 1)
		Case Control	$n_1 \pi_1$	$n_2 \pi_2$	(col prob sums to 1)
	2	Poisson	$\mu_{21}$	$\mu_{22}$	
		Double Dichotomy	$np_{21}$	$n(1 - p_{11} - p_{12} - p_{21})$	
	Prospective	$n_2 p_2$	$n_2(1 - p_2)$		
	Case Control	$n_1(1 - \pi_1)$	$n_2(1 - \pi_2)$		



# Log-linear models

---

- Often, when you are not really sure how you want to model the data (conditional on the total, conditional on the rows or conditional on the columns), you can treat the data as if they are Poisson (the most general model) and use log-linear models to explore relationships between the row and column variables.
- The most general model for a  $(2 \times 2)$  table is a Poisson model (4 non-redundant expected cell counts).
- Since the expected cell counts are always positive, we model  $\mu_{jk}$  as an exponential function of row and column effects:

$$\mu_{jk} = \exp(\mu + \lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY})$$

where

$$\lambda_j^X = j^{th} \text{ row effect}$$

$$\lambda_k^Y = k^{th} \text{ column effect}$$

$$\lambda_{jk}^{XY} = \text{interaction effect in } j^{th} \text{ row, } k^{th} \text{ column}$$

- Equivalently, we can write the model as a log-linear model:

$$\log(\mu_{jk}) = \mu + \lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}$$

- Treating the 4 expected cell counts as non-redundant, we can write the model for  $\mu_{jk}$  as a function of at most 4 parameters. However, in this model, there are 9 parameters,

$$\mu, \lambda_1^X, \lambda_2^X, \lambda_1^Y, \lambda_2^Y, \lambda_{11}^{XY}, \lambda_{12}^{XY}, \lambda_{21}^{XY}, \lambda_{22}^{XY},$$

but only four expected cell counts  $\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}$ .

- Thus, we need to put constraints on the  $\lambda$ 's, so that only four are non-redundant.
- We will use the 'reference cell' constraints, in which we set any parameter with a '2' in the subscript to 0, i.e.,

$$\lambda_2^X = \lambda_2^Y = \lambda_{12}^{XY} = \lambda_{21}^{XY} = \lambda_{22}^{XY} = 0,$$

leaving us with 4 unconstrained parameters

$$\mu, \lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY}$$

as well as 4 expected cell counts:

$$[\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}]$$

# Expected Cell Counts for the Model

---

- Again, the model for the expected cell count is written as

$$\mu_{jk} = \exp(\mu + \lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY})$$

- In particular, given the constraints, we have:

$$\mu_{11} = \exp(\mu + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY})$$

$$\mu_{12} = \exp(\mu + \lambda_1^X)$$

$$\mu_{21} = \exp(\mu + \lambda_1^Y)$$

$$\mu_{22} = \exp(\mu)$$

# Regression Framework

- In terms of a regression framework, you write the model as

$$\begin{bmatrix} \log(\mu_{11}) \\ \log(\mu_{12}) \\ \log(\mu_{21}) \\ \log(\mu_{22}) \end{bmatrix} = \begin{bmatrix} \mu + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY} \\ \mu + \lambda_1^X \\ \mu + \lambda_1^Y \\ \mu \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \lambda_1^X \\ \lambda_1^Y \\ \lambda_{11}^{XY} \end{bmatrix}$$

- i.e., you create dummy or indicator variables for the different categories.

$$\log(\mu_{jk}) = \mu + I(j = 1)\lambda_1^X + I(k = 1)\lambda_1^Y + I[(j = 1), (k = 1)]\lambda_{11}^{XY}$$

where

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is not true} \end{cases} .$$

- For example,

$$\begin{aligned}\log(\mu_{21}) &= \mu + I(2 = 1)\lambda_1^X + I(1 = 1)\lambda_1^Y + I[(2 = 1), (1 = 1)]\lambda_{11}^{XY} \\ &= \mu + 0 \cdot \lambda_1^X + 1 \cdot \lambda_1^Y + 0 \cdot \lambda_{11}^{XY} \\ &= \mu + \lambda_1^Y\end{aligned}$$

# Interpretation of the $\lambda$ 's

---

- We can solve for the  $\lambda$ 's in terms of the  $\mu_{jk}$ 's.

$$\log(\mu_{22}) = \mu$$

$$\begin{aligned}\log(\mu_{12}) - \log(\mu_{22}) &= (\mu + \lambda_1^X) - \mu \\ &= \lambda_1^X\end{aligned}$$

$$\begin{aligned}\log(\mu_{21}) - \log(\mu_{22}) &= (\mu + \lambda_1^Y) - \mu \\ &= \lambda_1^Y\end{aligned}$$

# Odds Ratio

---

$$\begin{aligned}\log(OR) &= \log \frac{\mu_{11}\mu_{22}}{\mu_{21}\mu_{12}} \\ &= \log(\mu_{11}) + \log(\mu_{22}) - \log(\mu_{21}) - \log(\mu_{12}) \\ &= (\mu + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}) + \mu \\ &\quad - (\mu + \lambda_1^Y) - (\mu + \lambda_1^X) \\ &= \lambda_{11}^{XY}\end{aligned}$$

Important: the main parameter of interest is the log odd ratio, which equals  $\lambda_{11}^{XY}$  in this model.



- 
- The model with the 4 parameters

$$\mu, \lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY}$$

is called the ‘saturated model’ since it has as many free parameters as possible for a  $(2 \times 2)$  table which has the four expected cell counts  $\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}$ .

- Also, you will note that Agresti uses different constraints for the log-linear model, namely

$$\sum_{j=1}^2 \lambda_j^X = 0,$$

$$\sum_{k=1}^2 \lambda_k^X = 0,$$

and

$$\sum_{j=1}^2 \lambda_{jk}^{XY} = 0 \quad \text{for } k = 1, 2$$

and

$$\sum_{k=1}^2 \lambda_{jk}^{XY} = 0 \quad \text{for } j = 1, 2$$

- 
- Agresti's model is just a different parameterization for the 'saturated model'. I think the one we are using (Reference Category) is a little easier to work with.
  - The log-linear model as we have written it, makes no distinction between what margins are fixed by design, and what margins are random.
  - Again, when you are not really sure how you want to model the data (conditional on the total, conditional on the rows or conditional on the columns) or which model is appropriate, you can use log-linear models to explore the data.

# Parameters of interest for different designs and the MLE's

---

- For all sampling plans, we are interested in testing independence:

$$H_0: OR = 1.$$

- As shown earlier for the log-linear model, the null is

$$H_0: \lambda_{11}^{XY} = \log(OR) = 0.$$

- Depending on the design, some of the parameters of the log-linear model are actually fixed by the design.
- However, for all designs, we can estimate the parameters (that are not fixed by the design) with a Poisson likelihood, and get the MLE's of the parameters for all designs.
- This is because the kernel of the log-likelihood for any of these design is the same

# The different designs

## Random Counts

- To derive the likelihood, note that

$$P(Y_{jk} = n_{jk} | \text{Poisson}) = \frac{e^{-\mu_{jk}} \mu_{jk}^{n_{jk}}}{n_{jk}!}$$

- Thus, the full likelihood is

$$L = \prod_j \prod_k \frac{e^{-\mu_{jk}} \mu_{jk}^{n_{jk}}}{n_{jk}!}$$

- Or,

$$l = \sum_j \sum_k -\mu_{jk} + \sum_j \sum_k n_{jk} \log \mu_{jk} + K$$

- Or, in terms of the kernel, the Poisson log-likelihood is

$$l^* = -\mu_{++} + y_{11} \log(\mu_{11}) + y_{12} \log(\mu_{12}) + y_{21} \log(\mu_{21}) + y_{22} \log(\mu_{22})$$

- Consider the log-linear model

$$\mu_{jk} = \exp[\mu + \lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]$$

- Then, substituting this in the log-likelihood, we get

$$\begin{aligned} \log[L(\mu, \lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY})] &= \\ -\mu_{++} + \sum_{j=1}^2 \sum_{k=1}^2 y_{jk} [\mu + \lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}] &= \\ -\mu_{++} + \mu y_{++} + \sum_{j=1}^2 \lambda_j^X y_{j+} + \sum_{k=1}^2 \lambda_k^Y y_{+k} + \sum_{j=1}^2 \sum_{k=1}^2 y_{jk} \lambda_{jk}^{XY} &= \\ -\mu_{++} + \mu y_{++} + \lambda_1^X y_{1+} + \lambda_1^Y y_{+1} + \lambda_{11}^{XY} y_{11} & \end{aligned}$$

since we constrained all  $\lambda$  terms to be 0 with a subscript equal to 2.

- Note, here, that the likelihood is a function of the parameters

$$(\mu, \lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY})$$

and the random variables

$$(y_{++}, y_{1+}, y_{+1}, y_{11})$$

- The random variables

$$(y_{++}, y_{1+}, y_{+1}, y_{11})$$

are called sufficient statistics, i.e., all the information from the data in the likelihood are contained in the sufficient statistics

- In particular, when taking derivatives of the log-likelihood to find the MLE, we will be solving for the estimate of  $(\mu, \lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY})$  as a function of the sufficient statistics  $(y_{++}, y_{1+}, y_{+1}, y_{11})$

# Example

Cold incidence among French Skiers (Pauling, Proceedings of the national Academy of Sciences, 1971).

## OUTCOME

		COLD	NO COLD	Total
T R E A T M E N T	VITAMIN C	17	122	139
	NO VITAMIN C	31	109	140
	Total	48	231	279



# Poisson log-linear model Model

- For the Poisson likelihood, we write the log-linear model for the expected cell counts as:

$$\begin{bmatrix} \log(\mu_{11}) \\ \log(\mu_{12}) \\ \log(\mu_{21}) \\ \log(\mu_{22}) \end{bmatrix} = \begin{bmatrix} \mu + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY} \\ \mu + \lambda_1^X \\ \mu + \lambda_1^Y \\ \mu \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \lambda_1^X \\ \lambda_1^Y \\ \lambda_{11}^{XY} \end{bmatrix}$$

- We will use this in SAS Proc Genmod to obtain the estimates

# SAS PROC GENMOD

---

```
data one;
input vitc cold count;
cards;
1 1 17
1 2 122
2 1 31
2 2 109
;
run;
```

```
proc genmod data=one;
  class vitc cold; /* Class automatically create */
                  /* constraints, i.e., dummy */
                  /* variables */
  model count = vitc cold vitc*cold / /* can put interaction terms in */
                  link=log dist = poi; /* directly */
run;
```

```
/* SELECTED OUTPUT */
```

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT		1	4.6913	0.0958	2398.9532	0.0000
VITC	1	1	0.1127	0.1318	0.7308	0.3926
VITC	2	0	0.0000	0.0000	.	.
COLD	1	1	-1.2574	0.2035	38.1575	0.0000
COLD	2	0	0.0000	0.0000	.	.
VITC*COLD	1 1	1	-0.7134	0.3293	4.6934	0.0303
VITC*COLD	1 2	0	0.0000	0.0000	.	.
VITC*COLD	2 1	0	0.0000	0.0000	.	.
VITC*COLD	2 2	0	0.0000	0.0000	.	.
SCALE		0	1.0000	0.0000	.	.

# Estimates

---

- From the SAS Output, the Estimates are:

$$\hat{\mu} = 4.6913$$

$$\hat{\lambda}_1^{VITC} = 0.1127$$

$$\hat{\lambda}_1^{COLD} = -1.2574$$

$$\lambda_{11}^{VITC,COLD} = \log(OR) = -0.7134$$

- The OR the “regular” way is

$$\log(OR) = \log\left(\frac{17 \cdot 109}{31 \cdot 122}\right) = \log(0.499) = -0.7134$$

# Double Dichotomy

- For the double dichotomy in which the data follow a multinomial, we first rewrite the log-likelihood

$$l^* = -\mu_{++} + y_{11}\log(\mu_{11}) + y_{12}\log(\mu_{12}) + y_{21}\log(\mu_{21}) + y_{22}\log(\mu_{22})$$

in terms of the expected cell counts, and the  $\lambda$ 's :

$$\mu_{11} = np_{11} = \exp(\mu + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY})$$

$$\mu_{12} = np_{12} = \exp(\mu + \lambda_1^X)$$

$$\mu_{21} = np_{21} = \exp(\mu + \lambda_1^Y)$$

$$\mu_{22} = n(1 - p_{11} - p_{12} - p_{21}) = \exp(\mu)$$

- Recall, the multinomial is a function of 3 probabilities

$$(p_{11}, p_{12}, p_{21})$$

since  $p_{22} = 1 - p_{11} - p_{12} - p_{21}$ .

- Adding up the  $\mu_{jk}$ 's in terms of the  $np_{jk}$ 's, it is pretty easy to see that

$$\mu_{++} = \sum_{j=1}^2 \sum_{k=1}^2 \mu_{jk} = n$$

(fixed by design), so that the first term in the log-likelihood,  $-\mu_{++} = -n$  is not a function of the unknown parameters for the multinomial.

- Then, the multinomial probabilities can be written as

$$p_{jk} = \frac{\mu_{jk}}{n} = \frac{\mu_{jk}}{\mu_{++}}$$

- We can also write  $\mu_{++}$  in terms of the  $\lambda$ 's,

$$\begin{aligned}\mu_{++} &= \sum_{j=1}^2 \sum_{k=1}^2 \mu_{jk} = \\ & \sum_{j=1}^2 \sum_{k=1}^2 \exp[\mu + \lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}] = \\ & \exp[\mu] \sum_{j=1}^2 \sum_{k=1}^2 \exp[\lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]\end{aligned}$$

- Then, we can rewrite the multinomial probabilities as

$$\begin{aligned}
 p_{jk} &= \frac{\mu_{jk}}{\mu_{++}} \\
 &= \frac{\exp[\mu + \lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]}{\sum_{j=1}^2 \sum_{k=1}^2 \exp[\mu + \lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]} \\
 &= \frac{\exp[\mu] \exp[\lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]}{\exp[\mu] \sum_{j=1}^2 \sum_{k=1}^2 \exp[\lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]} \\
 &= \frac{\exp[\lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]}{\sum_{j=1}^2 \sum_{k=1}^2 \exp[\lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]},
 \end{aligned}$$

which is not a function of  $\mu$



# The Multinomial

---

- We see that these probabilities do not depend on the parameter  $\mu$ .
- In particular, for the multinomial, there are only three free probabilities

$$(p_{11}, p_{12}, p_{21})$$

and three parameters.

$$(\lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY}).$$

- These probabilities could also have been determined by noting that, conditioning on the table total  $n = Y_{++}$ , the Poisson random variables follow a conditional multinomial,

$$(Y_{11}, Y_{12}, Y_{21} | Y_{++} = y_{++}) \sim \text{Mult}(y_{++}, p_{11}, p_{12}, p_{21})$$

with

$$p_{jk} = \frac{\mu_{jk}}{\mu_{++}}$$

which we showed above equals

$$p_{jk} = \frac{\exp[\lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]}{\sum_{j=1}^2 \sum_{k=1}^2 \exp[\lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]},$$

and is not a function of  $\mu$ .

# Obtaining MLE's

---

Thus, to obtain the MLE's for  $(\lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY})$ , we have 2 choices:

1. We can maximize the Poisson likelihood.
2. We can maximize the conditional multinomial likelihood.
  - If the data are from a double dichotomy, the multinomial likelihood is not a function of  $\mu$ . Thus, if you use a Poisson likelihood to estimate the log-linear model when the data are multinomial, the estimate of  $\mu$  really is not of interest.
  - We will use this in SAS Proc Catmod to obtain the estimates using the multinomial likelihood.

# Multinomial log-linear model Model

- For the Multinomial likelihood in SAS Proc Catmod, we write the log-linear model for the three probabilities  $(p_{11}, p_{12}, p_{21})$  as:

$$p_{11} = \frac{\exp(\lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY})}{\exp(\lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}) + \exp(\lambda_1^X) + \exp(\lambda_1^Y) + 1}$$

$$p_{12} = \frac{\exp(\lambda_1^X)}{\exp(\lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}) + \exp(\lambda_1^X) + \exp(\lambda_1^Y) + 1}$$

$$p_{21} = \frac{\exp(\lambda_1^Y)}{\exp(\lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}) + \exp(\lambda_1^X) + \exp(\lambda_1^Y) + 1}$$

- Note that the denominator in each probability is

$$\sum_{j=1}^2 \sum_{k=1}^2 \exp[\lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}]$$

For  $j = k = 2$  in this sum, we have the constraint that  $\lambda_2^X = \lambda_2^Y = \lambda_{22}^{XY} = 0$  so that

$$\exp[\lambda_2^X + \lambda_2^Y + \lambda_{22}^{XY}] = e^0 = 1$$

- Using SAS Proc Catmod, we make the design matrix equal to the combinations of  $(\lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY})$  found in the exponential function in the numerators:

$$\begin{bmatrix} \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY} \\ \lambda_1^X \\ \lambda_1^Y \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1^X \\ \lambda_1^Y \\ \lambda_{11}^{XY} \end{bmatrix}$$

# SAS PROC CATMOD

---

```
data one;
input vitc cold count;
cards;
1 1 17
1 2 122
2 1 31
2 2 109
;
run;

proc catmod data=one;
  model vitc*cold = ( 1 1 1,      /* 1st col = lambda^V */
                    1 0 0,      /* 2nd col = lambda^C */
                    0 1 0 );    /* 3rd col = lambda^VC */
weight count;
run;
```

/\* SELECTED OUTPUT \*/

Response Profiles

Response	vitc	cold
1	1	1
2	1	2
3	2	1
4	2	2

Analysis of Maximum Likelihood Estimates

Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Model	1	0.1127	0.1318	0.73	0.3926
	2	-1.2574	0.2035	38.16	<.0001
	3	-0.7134	0.3293	4.69	0.0303

# Estimates

- From the SAS Output, the Estimates are:

$$\hat{\lambda}_1^{VITC} = 0.1127$$

$$\hat{\lambda}_1^{COLD} = -1.2574$$

$$\lambda_{11}^{VITC,COLD} = \log(OR) = -0.7134$$

Which is the same as for the Poisson Log-Linear Model and

$$e(-0.7134) = 0.49$$

which is the estimate obtained from PROC FREQ

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits
Case-Control (Odds Ratio)	0.4900	0.2569 0.9343



# Prospective Study

- Now, suppose the data are from a prospective study, or, equivalently, we condition on the row totals of the  $(2 \times 2)$  table. We know that, conditional on the row totals  $n_1 = Y_{1+}$  and  $n_2 = Y_{2+}$  are fixed, and the total sample size is  $n_{++} = n_1 + n_2$ .
- Further, we are left with a likelihood that is a product of two independent row binomials.

$$(Y_{11}|Y_{1+} = y_{1+}) \sim \text{Bin}(y_{1+}, p_1)$$

where

$$p_1 = P[Y = 1|X = 1] = \frac{\mu_{11}}{\mu_{1+}} = \frac{\mu_{11}}{\mu_{11} + \mu_{12}};$$

and

$$(Y_{21}|Y_{2+} = y_{2+}) \sim \text{Bin}(y_{2+}, p_2)$$

where

$$p_2 = P[Y = 1|X = 2] = \frac{\mu_{21}}{\mu_{2+}} = \frac{\mu_{21}}{\mu_{21} + \mu_{22}}$$

- And the conditional binomials are independent.

- Conditioning on the rows, the log-likelihood kernel is

$$l^* = -(n_1+n_2)+y_{11}\log(n_1p_1)+y_{12}\log(n_1(1-p_1))+y_{21}\log(n_2p_2)+y_{22}\log(n_2(1-p_2))$$

- What are  $p_1$  and  $p_2$  in terms of the  $\lambda$ 's?
- The probability of success for row 1 is

$$\begin{aligned} p_1 &= \frac{\mu_{11}}{\mu_{11} + \mu_{12}} \\ &= \frac{\exp(\mu + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY})}{\exp(\mu + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}) + \exp(\mu + \lambda_1^X)} \\ &= \frac{\exp(\mu + \lambda_1^X) \exp(\lambda_1^Y + \lambda_{11}^{XY})}{\exp(\mu + \lambda_1^X) [\exp(\lambda_1^Y + \lambda_{11}^{XY}) + 1]} \\ &= \frac{\exp(\lambda_1^Y + \lambda_{11}^{XY})}{1 + \exp(\lambda_1^Y + \lambda_{11}^{XY})} \end{aligned}$$

- The probability of success for row 2 is

$$\begin{aligned}
 p_2 &= \frac{\mu_{21}}{\mu_{21} + \mu_{22}} \\
 &= \frac{\exp(\mu + \lambda_1^Y)}{\exp(\mu + \lambda_1^Y) + \exp(\mu)} \\
 &= \frac{\exp(\mu) \exp(\lambda_1^Y)}{\exp(\mu) [\exp(\lambda_1^Y) + 1]} \\
 &= \frac{\exp(\lambda_1^Y)}{1 + \exp(\lambda_1^Y)}
 \end{aligned}$$

- Now, conditional on the row totals (as in a prospective study), we are left with two free probabilities  $(p_1, p_2)$ , and the conditional likelihood is a function of two free parameters  $(\lambda_1^Y, \lambda_{11}^{XY})$ .

# Logistic Regression

- Looking at the previous pages, the conditional probabilities of  $Y$  given  $X$  from the log-linear model follow a logistic regression model:

$$\begin{aligned} p_x &= P[Y = 1 | X^* = x^*] \\ &= \frac{e^{[\lambda_1^Y + \lambda_{11}^{XY} x^*]}}{e^{[\lambda_1^Y + \lambda_{11}^{XY} x^*]} + 1} \\ &= \frac{e^{[\beta_0 + \beta_1 x^*]}}{1 + e^{[\beta_0 + \beta_1 x^*]}} \end{aligned}$$

where

$$x^* = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{if } x = 2 \end{cases} .$$

and

$$\beta_0 = \lambda_1^Y$$

and

$$\beta_1 = \lambda_{11}^{XY}$$

- 
- From the log-linear model, we had that  $\lambda_{11}^{XY}$  is the log-odds ratio, which we know from the logistic regression, is  $\beta_1$ .
  - Note, the intercept in a logistic regression with  $Y$  as the response is the main effect of  $Y$  in the log-linear model:

$$\beta_0 = \lambda_1^Y$$

- The conditional probability  $p_x$  is not a function of  $\mu$  or  $\lambda_1^X$ .

# Obtaining MLE's

---

- Thus, to obtain the MLE's for  $(\lambda_1^Y, \lambda_{11}^{XY})$ , we have 3 choices:
  1. We can maximize the Poisson likelihood.
  2. We can maximize the conditional multinomial likelihood.
  3. We can maximize the row product binomial likelihood using a logistic regression package.
- If the data are from a prospective study, the product binomial likelihood is not a function of  $\mu$  or  $\lambda_1^X$ .
- Thus, if you use a Poisson likelihood to estimate the log-linear model when the data are from a prospective study, the estimate of  $\mu$  or  $\lambda_1^X$  really are not of interest.

# Revisiting the Cold Vitamin C Example

---

- We will let the 'covariate'  $X = \text{TREATMENT}$  and 'outcome'  $Y = \text{COLD}$ .
- We will use SAS Proc Logistic to get the MLES of the intercept

$$\beta_0 = \lambda_1^Y = \lambda_1^{\text{COLD}}$$

and log-odds ratio

$$\beta_1 = \lambda_{11}^{XY} = \lambda_{11}^{\text{VITC,COLD}}$$

# SAS PROC LOGISTIC

---

```
data one;
input vitc cold count;
if vitc=2 then vitc=0;
if cold=2 then cold=0;
cards;
1 1 17
1 2 122
2 1 31
2 2 109
;
run;
```

```
proc logistic data=one descending; /* descending model pr(Y=1) */
  model cold = vitc / rl ;          /* rl gives 95 % CI for OR */
  freq count;                       /* tells SAS how many subjects */
                                     /* each record in dataset represent */
run;
```



```
/* SELECTED OUTPUT */
```

### Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.2574	0.2035	38.1575	<.0001
vitc	1	-0.7134	0.3293	4.6934	0.0303

### Wald Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
vitc	1.0000	0.490	0.257	0.934

## Estimates

- From the SAS Output, the Estimates are:

$$\hat{\beta}_0 = \hat{\lambda}_1^{COLD} = -1.2574$$

$$\hat{\beta}_1 = \lambda_{11}^{VITC,COLD} = \log(OR) = -0.7134$$

- ~~Which are the same as for the Poisson and Multinomial Log-Linear Models.~~

# Recap

- Except for combinatorial terms that are not function of any unknown parameters, using  $\mu_{jk}$  from the previous table, the kernel of the log-likelihood for any of these design can be written as

$$l^* = -\mu_{++} + y_{11}\log(\mu_{11}) + y_{12}\log(\mu_{12}) + y_{21}\log(\mu_{21}) + y_{22}\log(\mu_{22})$$

- In this likelihood, the table total  $\mu_{++}$  is actually known for all designs,

Double Dichotomy	$n$
Prospective	$n_1 + n_2$
Case Control	$n_1 + n_2$

except for the Poisson, in which

$$\mu_{++} = E(Y_{++})$$

is a parameter that must be estimated (i.e., the sum of  $JK$  independent Poisson random variables).

# Recap

---

## Key Points:

- We have introduced Log-linear models
- We have defined a parameter in the model to represent the OR
- We do not have an “outcome” per se
- If you can designate an outcome, you minimize the number of parameters estimated
- You should feel comfortable writing likelihoods, If not, you have 3 weeks to gain the comfort
- Expect the final exam to have at least one likelihood problem