Chapter 7: Loglinear Models for Contingency Tables

Loglinear Models for Contingency Tables

In Chapter 3, we looked briefly at the Poisson regression model with the log link,

$$\log(\mu) = \alpha + \beta x.$$

We extended this to a multiple regression model (multiple covariates) for Poisson data with the log link:

$$\log(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k.$$

This formulation is useful for developing models that are expressed using multiplicative relationships. After exponentiating the above expression, we obtain

$$\mu = e^{\alpha} \times e^{\beta_1 x_1} \times \cdots \times e^{\beta_k x_k}.$$

The effects of the explanatory variables appear in the product form.

Loglinear Models for Two-Way Tables

Consider an $I \times J$ contingency table where we classify n subjects on two categorical responses, X and Y. Letting $\pi_{ij} = P(X = i, Y = j)$ denote the probability of the (i, j) cell, we say that the rows and columns are independent if for all (i, j)

$$\pi_{ij} = \pi_{i+} \times \pi_{+j},$$

where $\pi_{i+} = P(X = i) = \sum_{j=1}^{J} \pi_{ij}$ and $\pi_{+j} = P(Y = j) = \sum_{i=1}^{I} \pi_{ij}$.
We can rewrite this in terms of the expected values. That means, under independence, the expected count for cell (i, j) is

$$\mu_{ij} = n\pi_{ij} = n\pi_{i+} \times \pi_{+j}.$$

If we take the logarithm, we obtain

$$\log(\mu_{ij}) = \log(n) + \log(\pi_{i+}) + \log(\pi_{+j}).$$

Loglinear formulas are based on the cell means $\{\mu_{ij}\}$ and are useful for Poisson modelling of cell counts.

Independence Model

We express the loglinear model for independence:

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

where λ depends on the sample size, λ_i^X is the additive effect of row i, and λ_j^Y is an additive effect of column j.
The log-linear model does not distinguish between response and explanatory classification of variables. The model treats all variables jointly as responses, counting the number of responses at each combination.

The null hypothesis of independence in the $I \times J$ contingency table is

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}, \; i = 1, \ldots, I, \; j = 1, \ldots, J.$$

This hypothesis is equivalent to the loglinear model holding. The fitted values under independence are

$$\hat{\mu}_{ij} = n\hat{\pi}_{i+}\hat{\pi}_{+j} = n \times \frac{n_{i+}}{n} \times \frac{n_{+j}}{n} = \frac{n_{i+}n_{+j}}{n}.$$

The chi-squared tests using X^2 and G^2 for goodness-of-fit of this model are equivalent to tests of independence for a 2×2 table.

Interpretation of Parameters for 2 × 2 Tables

- When there are two rows and two columns, a constraint such as $\lambda_1^X + \lambda_2^X = 0$ and $\lambda_1^Y + \lambda_2^Y = 0$ must be placed on the parameters to make the model identifiable. Another possible constraint is setting $\lambda^X = 0$ for one category of X, and $\lambda^Y = 0$ for one category of Y. This implies that the λ parameters are not unique and that the interpretation will depend on the constraint.
- However, the differences of coefficients will provide information on the odds that does not depend of the constraint.
- Consider the odds of falling in the first column ($Y = 1$) relative to the second column ($Y = 2$) for the first row ($X = 1$):

$$\frac{\text{pr}(Y = 1|X = 1)}{\text{pr}(Y = 2|X = 1)} = \frac{\pi_{11}/\pi_{1+}}{\pi_{12}/\pi_{1+}} = \frac{\mu_{11}}{\mu_{12}}.$$

The log-odds or logit is

$$\log\left(\frac{\mu_{11}}{\mu_{12}}\right) = \log(\mu_{11}) - \log(\mu_{12}) = (\lambda + \lambda_1^X + \lambda_1^Y) - (\lambda + \lambda_1^X + \lambda_2^Y) = \lambda_1^Y - \lambda_2^Y.$$

The odds and log-odds are the same for both rows under independence. Similarly, the odds or log-odds for the first row would be the same for both columns. The log-odds and hence the differences remain the same for the different parameterizations of the model.

Example: Several thousand children are diagnosed with a certain variety of leukemia each year in the United States. A standard criterion for considering a child with leukemia to be in remission is if doctors cannot see any cancerous cells in the bone marrow when looking through a microscope. A genetic fingerprinting technique (PCR) can detect a much smaller amount of cancer cells. Investigators examined 178 children who appeared to be in remission using the standard criterion. Using the PCR test, traces of cancer were found in 75 of the children. The children were followed for 3 years, and the number of children in both groups suffering relapse was recorded:

	Followup Status		
PCR Status	Relapse	No Relapse	Total
Traces of Cancer	30	45	75
Cancer Free	8	95	103
Total	38	140	178

The null hypothesis is H_0: The two variables PCR test (X) and relapse (Y) are independent, and the alternative hypothesis is H_a: The two variables PCR test and relapse are dependent

Observed frequency		Fitted value under H_0		Log fitted value	
30	45	16.01	58.99	2.7732	4.0774
8	95	21.99	81.01	3.0906	4.3946

Parameter	Set 1	Set 2	Set 3
λ	4.3946	2.7733	3.5839
λ_1^X	-0.3172	0	-0.1586
λ_2^X	0	0.3172	0.1586
λ_1^Y	-1.3041	0	-0.6520
λ_2^Y	0	1.3041	0.6520

Constriant

- Set 1: $\lambda_1^X = 0$, $\lambda_1^Y = 0$
- Set 2: $\lambda_2^X = 0$, $\lambda_2^Y = 0$
- Set 3: $\lambda_1^X + \lambda_2^X = 0$, $\lambda_1^Y + \lambda_2^Y = 0$

Code

```
mydata=array(c(30, 8, 45, 95), dim=c(2, 2), dimnames=list(PCR
=c("traces", "No traces"), relapse=c("Yes", "No")))
 mydata
           relapse
PCR         Yes No
  traces     30 45
  No traces   8 95
 mydata2=as.data.frame(as.table(mydata))
 mydata2
        PCR relapse Freq
1    traces     Yes   30
2 No traces     Yes    8
3    traces      No   45
4 No traces      No   95
```

Fitting of model Set 2

Code

```
 summary(glm(Freq~PCR+relapse, data=mydata2, family=poisson))
Call:
glm(formula = Freq ~ PCR + relapse, family = poisson, data = mydata2)

Deviance Residuals:
     1       2       3       4
 3.114  -3.435  -1.902   1.512

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.7733     0.1845  15.033  < 2e-16 ***
PCRNo traces   0.3172     0.1518   2.090   0.0366 *
relapseNo      1.3041     0.1829   7.129 1.01e-12 ***
---
(Dispersion parameter for poisson family taken to be 1)
    Null deviance: 93.985  on 3  degrees of freedom
Residual deviance: 27.400  on 1  degrees of freedom
AIC: 54.625

Number of Fisher Scoring iterations: 5
```

Fitting of model Set 3

Code

```
contrasts(mydata2$PCR) <- contr.sum(2, contrasts=TRUE)
contrasts(mydata2$relapse) <- contr.sum(2, contrasts=TRUE)

 summary(glm(Freq˜PCR+relapse, data=mydata2, family=poisson))
Call:
glm(formula = Freq ˜ PCR + relapse, family = poisson, data = mydata2)

Deviance Residuals:
     1       2       3       4
 3.114  -3.435  -1.902   1.512

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.58394    0.09223  38.857  < 2e-16 ***
PCR1        -0.15862    0.07590  -2.090   0.0366 *
relapse1    -0.65203    0.09146  -7.129 1.01e-12 ***
---
(Dispersion parameter for poisson family taken to be 1)
    Null deviance: 93.985  on 3  degrees of freedom
Residual deviance: 27.400  on 1  degrees of freedom
AIC: 54.625
```

The Saturated Model

Observe that by definition $\mu_{ij} = n\pi_{ij}$, so

$$\log(\mu_{ij}) = \log(n) + \log(\pi_{i+}) + \log(\pi_{+j}) + \log\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right).$$

Clearly, $\lambda = \log(n)$, $\lambda_i^X = \log(\pi_{i+})$, $\lambda_j^Y = \log(\pi_{+j})$, and

$$\lambda_{ij}^{XY} = \log\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right)$$

is an association term that represents the departure from independence of X and Y. Thus, the general model is

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}.$$

Note that under the independence assumption, $\pi_{ij} = \pi_{i+}\pi_{+j}$, so $\lambda_{ij}^{XY} = \log(1) = 0$. Thus λ_{ij}^{XY} is an interaction term that reflects that the effect of one factor can change depending on the level of the other factor (i.e., a lack of independence).

The Saturated Model

We need to eliminate redundant parameters by placing constraints on λ_{ij}^{XY}. A common set of constraints is

$$\lambda_{11} + \lambda_{12} = \lambda_{21} + \lambda_{22} = \lambda_{11} + \lambda_{21} = \lambda_{12} + \lambda_{22} = 0.$$

Alternatively, we could also set the last term in each row and in each column equal to zero:

$$\lambda_{12} = \lambda_{21} = \lambda_{22} = 0.$$

Without further restriction on the λ_{ij}^{XY} terms, the model is *saturated* in that it fits the $I \times J$ table perfectly. On the other hand, the independence model satisfies all $\lambda_{ij}^{XY} = 0$.

The association terms are related to the log-odds ratio for a 2×2 table:

$$\begin{aligned}
\log(\theta) &= \log\left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right) \\
&= \log(\mu_{11}) + \log(\mu_{22}) - \log(\mu_{12}) - \log(\mu_{21}) \\
&= (\lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}) + (\lambda + \lambda_2^X + \lambda_2^Y + \lambda_{22}^{XY}) \\
&\quad -(\lambda + \lambda_1^X + \lambda_2^Y + \lambda_{12}^{XY}) - (\lambda + \lambda_2^X + \lambda_1^Y + \lambda_{21}^{XY}) \\
&= \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}.
\end{aligned}$$

It is seen that under the independence assumption that means when $\lambda_{11}^{XY} = \lambda_{22}^{XY} = \lambda_{12}^{XY} = \lambda_{21}^{XY} = 0$, $\log(\theta) = 0$, resulting in unit odds ratio.

Example: Saturated Loglinear Model for Leukemia Data

Equivalent Association Parameter Estimates

Association Parameter	Set 1	Set 2	Set 3
λ_{11}^{XY}	2.0690	0	0.5172
λ_{12}^{XY}	0	0	-0.5172
λ_{21}^{XY}	0	0	-0.5172
λ_{22}^{XY}	0	2.0690	0.5172

We verify that

$$\log(\hat{\theta}) = \hat{\lambda}_{11}^{XY} + \hat{\lambda}_{22}^{XY} - \hat{\lambda}_{12}^{XY} - \hat{\lambda}_{21}^{XY} = 2.0690 = \log\left(\frac{30 \times 95}{8 \times 45}\right) = \log(7.9167)$$

Independence Model

The saturated model contains the following nonredundant terms:

- 1 λ (intercept) term
- $(I-1)$ λ_i^X (main effect of X) terms
- $(J-1)$ λ_j^Y (main effect of Y) terms
- $(I-1)(J-1)$ λ_{ij}^{XY} (two-factor interactions of X and Y) terms

Independence Model

- This results in $1 + (I-1) + (J-1) + (I-1)(J-1) = IJ$ nonredundant terms in the saturated model. The model has as many parameters as the number of cells of the $I \times J$ table based on Poisson observations and so it gives perfect fit to the data. Here the error degrees of freedom is zero.
- The saturated model is an example of a *hierarchical model*, because it includes all lower order terms of the variables that are contained in the higher order terms of the model.
- The significance and practical interpretation of lower order terms depends on how the variables are coded. If one uses a hierarchical model, the results for the higher order terms remain the same no matter how the variables are coded.
- The interpretation of lower order terms depends on the coding of the higher order terms. Consequently, we typically limit our interpretation to that of the highest order terms in a hierarchical model.

Loglinear models for three-way tables

- Loglinear models for three-way tables can incorporate different independence and association patterns for the categorical variables X, Y, and Z.
- The expected cell frequencies in the $I \times J \times K$ table are denoted by $\{\mu_{ijk}\}$, and the observed cell frequencies are denoted by $\{n_{ijk}\}$.
- The saturated model is denoted by (XYZ) and has the form

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}.$$

The model is made identifiable by imposing constraints on the parameters such as

$$\begin{gathered}\sum_i \lambda_i^X = \sum_i \lambda_{ij}^{XY} = \sum_i \lambda_{ik}^{XZ} = \sum_i \lambda_{ijk}^{XYZ} = 0, \\ \sum_j \lambda_j^Y = \sum_j \lambda_{ij}^{XY} = \sum_j \lambda_{jk}^{YZ} = \sum_j \lambda_{ijk}^{XYZ} = 0, \\ \sum_k \lambda_k^Z = \sum_k \lambda_{ik}^{XZ} = \sum_k \lambda_{jk}^{YZ} = \sum_k \lambda_{ijk}^{XYZ} = 0.\end{gathered}$$

Models that include subsets of terms from the saturated model represent different forms of independence and association in the table. We will use only hierarchical models where the presence of a higher-order term in the model implies that all the lower order terms are present.

Loglinear Models and Independence Structure

Terms such as λ_{ij}^{XY} are used to indicate partial association between two variables. The terms included in the model will determine the nature of the association in the model.

- The *homogeneous association model* has the form

  $$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

  This model includes terms permitting X-Y, X-Z, and Y-Z associations. We denote this model (XY, XZ, YZ).

- If we omit one of the two-factor terms, we allow for conditional independence between two of the variables. Consider the model of the form

  $$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

  This model includes terms permitting X-Z and Y-Z associations. Thus, X and Y are conditionally independent, controlling for Z. For $2 \times 2 \times K$ tables, this model corresponds to the hypothesis tested using the Cochran-Mantel-Haenszel statistic in Chapters 2 and 4. This model is denoted by (XZ, YZ).

- If we include only one two-factor term, we obtain the model

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}.$$

  This model is denoted by (XY, Z). In this model, X and Z are marginally independent as are Y and Z.

- The model where X, Y, and Z are mutually independent is denoted by (X, Y, Z) and has the form

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z.$$

- If there is no term involving a particular variable, then the marginal probability for that variable is uniform. For instance, X and Y are marginally independent of Z and Z is uniform in the model

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}.$$

Interpreting model parameters

Consider the saturated model:

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}.$$

The three-factor term pertains to a three-factor interaction. This term permits the odds ratio between any two variables to vary across levels of the third factor. If we omit this term, we obtain the homogeneous association model:

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

The two-factor terms are partial association terms and relate to the conditional odds ratios. For a $2 \times 2 \times K$ table, consider any of the models that contain the X-Y term. The conditional odds ratio for X and Y given Z is

$$\theta_{XY(k)} = \frac{\mu_{11k}\mu_{22k}}{\mu_{12k}\mu_{21k}}.$$

For the homogeneous association model, we take the logarithm of the odds ratio and obtain

$$\begin{array}{rcl}\log(\theta_{XY(k)}) & = & \log(\mu_{11k}) + \log(\mu_{22k}) - \log(\mu_{12k}) - \log(\mu_{21k}) \\ & = & \lambda + \lambda_1^X + \lambda_1^Y + \lambda_k^Z + \lambda_{11}^{XY} + \lambda_{1k}^{XZ} + \lambda_{1k}^{YZ} \\ & & +\lambda + \lambda_2^X + \lambda_2^Y + \lambda_k^Z + \lambda_{22}^{XY} + \lambda_{2k}^{XZ} + \lambda_{2k}^{YZ} \\ & & -(\lambda + \lambda_1^X + \lambda_2^Y + \lambda_k^Z + \lambda_{12}^{XY} + \lambda_{1k}^{XZ} + \lambda_{2k}^{YZ} \\ & & +\lambda + \lambda_2^X + \lambda_1^Y + \lambda_k^Z + \lambda_{21}^{XY} + \lambda_{2k}^{XZ} + \lambda_{1k}^{YZ}) \\ & = & \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}.\end{array}$$

The right-hand side does not depend on k, so the odds-ratio is the same at each level of Z. Similarly, the X-Z odds ratio is the same at each level of Y, and the Y-Z odds ratio is the same at each level of X.

Job satisfaction example

The following table gives the three-way cross-classification of quality of management (M), supervisor's job satisfaction (S), and worker's job satisfaction (W) for a random sample of 715 workers selected from Danish industry. Since all three variables are response variables, we use loglinear models to investigate the patterns of association among the three variables.

Quality of Management	Supervisor's Job Satisfaction	Worker's Job Satisfaction		
		Low	High	
Bad	Low	103	87	190
	High	32	42	74
Good	Low	59	109	168
	High	78	205	283

The following models were fit to the data:

- the saturated model (MSW)
- the homogeneous association model (MS,MW,SW)
- the best model with two interactions (MS,MW)
- the best model with one interaction (MS,W)
- the complete independence model (M,S,W)

We compare the fitted values to the observed data for several loglinear models. The saturated model (MSW) provides the observed data. 1 $\equiv$ MSW, 2 $\equiv$ (MS,MW,SW), 3 $\equiv$ (MS,MW), 4 $\equiv$(MS,W), 5 $\equiv$ (M,S,W)

Fitted Values for Loglinear Models

Quality of Mgmt	Superv. Job Sat.	Worker's Job Sat.	Loglinear Model 1	2	3	4	5
Bad	Low	Low	103	102.3	97.2	72.3	50.3
	Low	High	87	87.7	92.8	117.7	81.9
	High	Low	32	32.7	37.8	28.2	50.1
	High	High	42	41.3	36.2	45.8	81.7
Good	Low	Low	59	59.7	51.0	63.9	85.9
	Low	High	109	108.3	117.0	104.1	139.9
	High	Low	78	77.3	86.0	107.7	85.7
	High	High	205	205.7	197.0	175.3	139.5

- The saturated model (MSW) fits the data exactly.
- The homogeneous association model (MS, MW, SW) has fitted values close to the observed frequencies.
- As the models become simpler, their fit appears worse.

We also form a table of the estimated odds ratios that correspond to the various models fit in the above table. Consider, for instance, the model (MS, MW). To compute the odds ratio for the M-S conditional association, we use the fitted odds ratio for either of the two levels of W:

$$\frac{97.2 \times 86.0}{37.8 \times 51.0} = \frac{92.8 \times 197.0}{117.0 \times 36.2} = 4.33.$$

The marginal odds ratio is found from the fitted M-S marginal table:

$$\frac{(97.2 + 92.8) \times (86.0 + 197.0)}{(37.8 + 36.2) \times (51.0 + 117.0)} = 4.33.$$

Estimated Odds Ratios for Loglinear Models

	Conditional Association			Marginal Association		
Model	M-S	M-W	S-W	M-S	M-W	S-W
(M,S,W)	1	1	1	1	1	1
(MS,W)	4.33	1	1	4.33	1	1
(MS,MW)	4.33	2.40	1	4.33	2.40	1.33
(MS,MW,SW)	4.04	2.11	1.47	4.33	2.40	1.86
*(MSW)*Level 1	4.26	2.19	1.55	4.33	2.40	1.86
*(MSW)*Level 2	3.90	2.00	1.42			

Inference for Loglinear Models

We can use methods developed for generalized linear models to carry out statistical inference for loglinear models. A good-fitting loglinear model provides information about the nature of association among categorical response variables.

Chi-Squared Goodness-of-Fit Tests

We can test the goodness of fit (lack of fit) of a loglinear model for a three-way contingency table by comparing the cell fitted values to the observed counts. The likelihood-ratio (G^2) and Pearson chi-squared statistics (X^2) are

$$G^2 = 2\sum n_{ijk}\log\left(\frac{n_{ijk}}{\hat{\mu}_{ijk}}\right), \quad X^2 = \sum \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}.$$

The degrees of freedom (df) equal the number of cell counts minus the number of nonredundant parameters in the model. The saturated model has zero d.f.

The following table presents the results of testing the goodness of fit of several loglinear models for the worker satisfaction data.

- The best fitting model (other than the saturated model) is the homogeneous association model (MS, MW, SW).
- All simpler models except (MS, MW) have very poor fit to the data.

Goodness-of-Fit Tests for Loglinear Models

Model	G^2	X^2	df	p-value
(M,S,W)	118.0	128.09	4	0.000
(MS,W)	35.60	35.62	3	0.000
(MS,MW)	5.39	5.41	2	0.068
(MS,SW)	71.90	70.88	2	0.000
(MW,SW)	19.71	19.88	2	0.000
(MS,MW,SW)	0.06	0.06	1	0.799
(MSW)	0.0	0.0	0	$--$

A Dissimilarity Index

When the sample size is large, often small effects can be statistically significant but not of much practical importance. In loglinear models, it is useful to use a measure of closeness of the model fit to the sample data that is not affected by sample size. Suppose that a table has cell counts $n_i (= n\pi_i)$ and fitted values $\hat{\mu}_i (= n\hat{\pi}_i)$. We define the *dissimilarity index* as

$$D = \frac{1}{2n} \sum |n_i - \hat{\mu}_i|,$$

where n is the total frequency. We have $0 \leq D \leq 1$ where D represents the proportion of cases that must move to other cells to obtain a perfect fit. A very small value of D indicates a good fit for the model.

Example: The following table presents the dissimilarity index for some of the models fit to the worker satisfaction data.

Model	D
(M,S,W)	0.17244
(MS,MW)	0.03862
(MS,SW)	0.07335
(MS,MW,SW)	0.00412
(MSW)	0.00000

Loglinear Residuals

The residuals are useful for examining the quality of fit for a model. Often they will show how a model fits poorly and highlight cells that exhibit lack of fit.
We can use the adjusted residuals for the Poisson regression model:

$$r_{a,ijk} = \frac{(n_{ijk} - \hat{\mu}_{ijk})}{\sqrt{\hat{\mu}_{ijk}(1 - h_{ijk})}}$$

where h_{ijk} is the leverage of the observation in the (i, j, k) cell. The standard normal distribution serves as a guide to the size of the adjusted residuals.

We look at the adjusted residuals for the fit of the model (MS, MW) of conditional independence of S and W at each level of M and also for the poorly fitting model (MS, SW) of conditional independence of M and W at each level of S.

$\widehat{\mu} \equiv$ fitted count, $r_a \equiv$ adjusted residual

Adjusted Residuals for Loglinear Models

Quality of Mgm	Superv. Job Sat.	Worker Job Sat.	Obs. Count	(MS,MW,SW) $\widehat{\mu}$	(MS,MW,SW) r_a	(MS,MW) $\widehat{\mu}$	(MS,MW) r_a	(MS,SW) $\widehat{\mu}$	(MS,SW) r_a
Bad	Low	Low	103	102.3	.255	97.2	1.601	86.0	3.62
	Low	High	87	87.7	-.255	92.8	-1.601	104.0	-3.62
	High	Low	32	32.7	-.255	37.8	-1.601	22.8	-2.60
	High	High	42	41.3	.255	36.2	1.601	51.2	2.60
Good	Low	Low	59	59.7	-.255	51.0	1.687	76.0	3.62
	Low	High	109	108.3	.255	117.0	-1.687	92.0	3.62
	High	Low	78	77.3	.255	86.0	-1.687	87.2	-2.60
	High	High	205	205.7	-.255	197.0	1.687	195.8	2.60

- Notice the small adjusted residuals for the (MS, MW, SW) model. Since the residual $df = 1$, there is one nonredundant residual. We notice that $X^2 = 0.065 = 0.255^2$, the square of any of the residuals.
- The adjusted residuals are larger for the (MS, MW) model. Here there are $df = 2$ and two nonredundant residuals. The goodness-of-fit statistic is $X^2 = 5.41 = 1.601^2 + 1.687^2$.
- The adjusted residuals for the (MS, SW) indicate definite problems with fit of this model. The goodness-of-fit statistic is $X^2 = 19.88 = 3.62^2 + 2.60^2$.

Tests about Partial Associations

We can conduct tests for partial associations by comparing different loglinear models. For instance, in the homogeneous association model (MS, MW, SW), the null hypothesis of no partial association between worker satisfaction and supervisor satisfaction corresponds to the λ^{SW} term equals zero. We can test this by comparing the (MS, MW, SW) model to the (MS, MW) using the likelihood ratio test.
Recall that we can form the likelihood ratio statistic for comparing nested models by taking the differences of their deviances:

$$G^2 = -2(L_0 - L_1).$$

We use the χ^2 critical value with degrees of freedom equalling the difference in the degrees of freedom for the two models.

Thus, the test statistic for testing $\lambda^{SW} = 0$ in the model (MS, MW, SW) is the difference

$$\begin{aligned} G^2[(MS, MW)|(MS, MW, SW)] &= G^2(MS, MW) - G^2(MS, MW, SW) \\ &= 5.39 - 0.06 = 5.33, \end{aligned}$$

with $df = 2 - 1$. The p-value equals 0.021. We thus reject the null hypothesis of no partial association at level 0.05 and conclude that the (MS, MW, SW) is the better model.

For this example, if we look at the other models with no partial association between two of the variables, the results of the corresponding likelihood ratio test are even stronger.

- For $2 \times 2 \times K$ tables, the test statistic $G^2[(XZ, YZ)|(XY, XZ, YZ)]$ for partial association is directly comparable to the Cochran-Mantel-Haenszel test. If the sample sizes in the 2×2 tables are reasonably large, the CMH statistic and the likelihood ratio statistic of conditional independence given homogeneous association will be similar. Both are compared to a χ^2_{K-1} critical value.
- Similarly the likelihood ratio test for the fit of the homogeneous association model is directly comparable to the Breslow-Day statistic.
- The sample size can affect inferences in a three-way table. The test is more likely to detect an effect of a given size as the sample size increases.
  - For small sample sizes, reality may be more complex than the simplest model that passes a goodness of fit test.
  - For large sample sizes, statistically significant effects may be weak and unimportant.
- Confidence intervals are often useful for assessing the importance of results.

Confidence Intervals for Odds Ratios

We can use computer software to obtain estimates of loglinear model parameters. For models where the highest-order terms are two-factor interactions, the estimates refer to conditional log-odds ratios. We use the large-sample normal distributions of the parameter estimates to form confidence intervals for true log odds ratios and then exponentiate them to obtain confidence intervals for the odds ratios.

Assume that the (MS, MW, SW) model holds for the worker satisfaction example. We will estimate the conditional odds ratio between worker satisfaction and supervisor satisfaction.

- From the output, $\hat{\lambda}_{11}^{SW} = 0.3847$ with $SE = 0.1667$. Then $\hat{\lambda}_{12}^{SW} = \hat{\lambda}_{21}^{SW} = \hat{\lambda}_{22}^{SW} = 0$. The conditional log odds ratio is $\lambda_{11}^{SW} + \lambda_{22}^{SW} - \lambda_{12}^{SW} - \lambda_{21}^{SW}$. Thus, $\hat{\lambda}_{11}^{SW}$ estimates the conditional log odds ratio.

  A 95% confidence interval for the conditional log odds ratio is

$$0.3847 \pm 1.96 \times 0.1667 = 0.3847 \pm .3267$$

  and a 95% confidence interval for the conditional odds ratio is

$$(e^{.05797}, e^{.7114}) = (1.0597, 2.0369)$$

The Loglinear-Logit Connection

Loglinear models for contingency tables have all variables as response variables whereas logit models have a binary response variable which depends on a set of explanatory variables. To help interpret a loglinear model, it is sometimes useful to construct an equivalent logit model.

Consider the homogeneous association model:

$$\log\mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_i^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

We will suppose that Y is a binary variable and treat it as a response. We let X and Z be considered as explanatory variables. Let π be the probability that $Y = 1$. The logit for Y is

$$\begin{aligned}
\text{logit}(\pi) &= \log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{P(Y=1|X=i,Z=k)}{P(Y=2|X=i,Z=k)}\right) \\
&= \log\left(\frac{\mu_{i1k}}{\mu_{i2k}}\right) = \log(\mu_{i1k}) - \log(\mu_{i2k}) \\
&= \lambda + \lambda_i^X + \lambda_1^Y + \lambda_i^Z + \lambda_{i1}^{XY} + \lambda_{ik}^{XZ} + \lambda_{1k}^{YZ} \\
&\quad -(\lambda + \lambda_i^X + \lambda_2^Y + \lambda_i^Z + \lambda_{i2}^{XY} + \lambda_{ik}^{XZ} + \lambda_{2k}^{YZ}) \\
&= (\lambda_1^Y - \lambda_2^Y) + (\lambda_{i1}^{XY} - \lambda_{i2}^{XY}) + (\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ}).
\end{aligned}$$

For the homogeneous association model, we showed that

$$\text{logit}(\pi) = (\lambda_1^Y - \lambda_2^Y) + (\lambda_{i1}^{XY} - \lambda_{i2}^{XY}) + (\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ}).$$

The first term is a constant and does not depend on i or k. The second term depends on the level i of X. The third term depends on the level k of Z. Thus, the logit can be written

$$\text{logit}(\pi) = \alpha + \beta_i^X + \beta_k^Z.$$

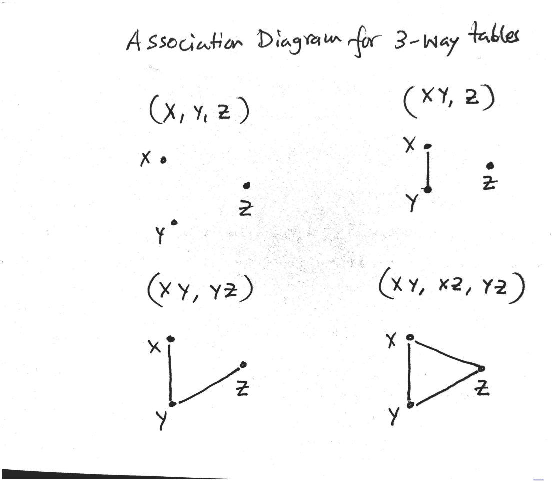
- When Y is binary, the loglinear model with homogeneous association is equivalent to the above logit model.
- When X is also binary, this logit model and the loglinear model (XY, XZ, YZ) have equal odds ratios between X and Y at each of the K levels of Z. The G^2 or X^2 goodness-of-fit statistics are an alternative way to test for a common odds ratio.
- When we derived the logit model corresponding to the (XY, XZ, YZ) loglinear model, the λ_{ik}^{XZ} terms cancelled out. Thus, a similar derivation for the (XY, YZ) model would also lead to the same logit model. However, the loglinear model that has the same fit as the logit model is the one containing a general interaction term for relationships among the explanatory variables. The logit model does not describe relationships among the explanatory variables, so it allows a general interaction pattern.
- The following table presents some equivalent loglinear and logit models when Y is a binary response variable.

Equivalent Models for Three-Way Tables

Loglinear	Logit
(Y, XZ)	α
(XY, XZ)	$\alpha + \beta_i^X$
(YZ, XZ)	$\alpha + \beta_k^Z$
(XY, YZ, XZ)	$\alpha + \beta_i^X + \beta_k^Z$
(XYZ)	$\alpha + \beta_i^X + \beta_k^Z + \beta_{ik}^{XZ}$

Association Graphs and Collapsibility

We can represent associations in hierarchical loglinear models by using association graphs. Each variable in the model is represented by a point (vertex) in the graph. Two points in the graph are connected if the two variables are partially associated. For three-dimensional tables, we will draw some of the possible association diagrams.



The association diagram of a model is helpful for interpreting the association patterns in the model.

- If two vertices are directly connected by an edge, the model contains the partial association term between the two variables corresponding to the two vertices.
- For the model (X, Y, Z), there is no partial association term, so there is no edge in the association graph.
- For the model (X, Y, Z), the mean is $\log(\mu_{i,j,k}^{xyz}) = \lambda_0 + \lambda_i^x + \lambda_j^y + \lambda_k^z$. Take any two variables, say X and Z. For this model, the variables X and Z are marginally independent. Under this independence model,

$$\begin{aligned} \mu_{2+2}^{xyz} &= \exp(\lambda_0 + \lambda_2^x + \lambda_2^y + \lambda_2^z) + \exp(\lambda_0 + \lambda_2^x + 0 + \lambda_2^z) \\ &= \exp(\lambda_0 + \lambda_2^x + \lambda_2^z)\{1 + \exp(\lambda_2^y)\}. \end{aligned}$$

Similarly, $\mu_{1+1}^{xyz} = \exp(\lambda_0 + \lambda_2^y) + \exp(\lambda_0) = \exp(\lambda_0)\{1 + \exp(\lambda_2^y)\}$, $\mu_{2+1}^{xyz} = \exp(\lambda_0 + \lambda_2^x + \lambda_2^y) + \exp(\lambda_0 + \lambda_2^x) = \exp(\lambda_0 + \lambda_2^x)\{1 + \exp(\lambda_2^y)\}$ and $\mu_{1+2}^{xyz} = \exp(\lambda_0 + \lambda_2^y + \lambda_2^z) + \exp(\lambda_0 + \lambda_2^z) = \exp(\lambda_0 + \lambda_2^z)\{1 + \exp(\lambda_2^y)\}$. Using the above expressions we obtain

$$\frac{\mu_{2+2}^{xyz}\mu_{1+1}^{xyz}}{\mu_{1+2}^{xyz}\mu_{2+1}^{xyz}} = 1.$$

- For this model, any two variables are conditionally independent. For example, X and Z are conditionally independent of Y. If Y has only two categories, for each category of Y, 1 and 2,

$$\frac{\mu_{212}^{xyz}\mu_{111}^{xyz}}{\mu_{112}^{xyz}\mu_{211}^{xyz}} = 1 \text{ and } \frac{\mu_{222}^{xyz}\mu_{121}^{xyz}}{\mu_{122}^{xyz}\mu_{221}^{xyz}} = 1.$$

- For the model (XY, YZ), there are two partial association terms, XY and YZ, so there are two edges X-Y and Y-Z in the association graph.

- If two vertices are not directly connected by any edge but connected through other vertices, then the two vertices (variables) are conditionally independent.
- For instance, for the model (XZ, YZ), the variables X and Y are not connected directly, indicating that they are conditionally independent, given Z.
- For the model (XY, Z), the variable Z is completely disconnected from X and Y, indicating that it is unconditionally independent of X and Y.
- The model (XY, XZ, YZ) has the same diagram as the saturated model (XYZ) so no independence conditions are implied by its diagram.

We have seen for many 3-way tables that the association in the partial tables given the levels of the third variable differ from the marginal association. However, for some models certain marginal and partial associations are equal. A table is **collapsible** over a variable if the conditional associations given that variable is the same as the marginal association when that variable is ignored. Association diagrams can be used to determine *collapsibility conditions* for a table.

For three-way tables, the $X - Y$ marginal and partial odds are equal if either Z and X are conditionally independent, or if Z and Y are conditionally independent. These conditions correspond to the loglinear models (XY, YZ) and (XY, XZ). The association diagrams for these two models are

$$X\text{—}Y\text{—}Z \text{ and } Y\text{—}X\text{—}Z,$$

but not the model (XZ, YZ) with graph $X\text{—}Z\text{—}Y$.
In terms of the association diagram, the model is not collapsible over Z if Z is connected between X and Y in the diagram.

Example: In the worker satisfaction data, consider the model (MS, MW) with association diagram

$$S \text{—} M \text{—} W.$$

The three-way table is collapsible over S or W, but not over M. This is shown by the following excerpt from slide 15:

Estimated Odds Ratios for Loglinear Models

	Conditional Association			Marginal Association		
Model	$M-S$	$M-W$	$S-W$	$M-S$	$M-W$	$S-W$
(MS,MW)	4.33	2.40	1	4.33	2.40	1.33
(MS,MW,SW)	4.04	2.11	1.47	4.33	2.40	1.86

Four-way Contingency Tables

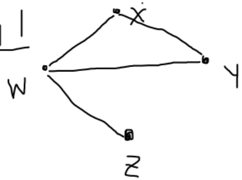
The methods that we developed for three-way tables extend naturally to higher order tables. We will use hierarchical models as before. We interpret interaction effects as reflecting differences in lower order effects at difference of control levels.
Sometimes association diagrams will help in the interpretation of a model. They are constructed in the same way as for three-way tables. Terms that involve three effects will be represented by three-way connections. The interpretation is similar to that for three-way tables.

- Two terms are independent if they are not connected at all in the diagram.
- They are conditionally independent given a set of other variables is they are connected only through a path that passes through that set of variables.

In some cases, several models for four-dimensional tables will have the same association diagram. Thus, these models have the same interpretation in terms of marginal and conditional independence.

Example: Association Diagram for the Models (WXY, WZ) and (WX, WY, WZ, XY)

Fig 1



The variables X and Y are jointly modeled since they are connected to each other. They are both connected to W, but not to Z. We could view the four-dimensional table as a three-way table with factors $X \times Y$, W, and Z, where $X \times Y$ is conditionally independent of Z given W.

Is the four-way table collapsible with respect to Z?

Example–Stillbirth and Premature Birth among Australians

Stillbirth is the death of a fetus at any time after the twentieth week of pregnancy. A premature birth is the live birth of a child from the twentieth until the thirty-seventh week of pregnancy. The following variables were recorded in a study of stillbirth in the Australian state of Queensland:

- Birth status(B)–stillbirth or live birth
- Gender(G)–male or female
- Gestational age(A)–$\leq 24,\ 25-28,\ 29-32,\ 33-36,\ 37-41$ weeks
- Race(R)–Aborigine or white

Various loglinear models were fit to the data to ascertain the complexity of the needed model.

Model	*df*	G^2	*p*-value	AIC
(A, B, R, G)	32	6848.9	<0.001	6864.9
(AB, AR, AG, BR, BG, RG)	17	48.5	<0.001	94.5
(ABR, ABG, BRG, ARG)	4	3.2	0.52	76.5
(ABR, ABG, ARG)	5	3.6	0.61	73.6
(ABR, ARG, BG)	9	7.1	0.63	69.2
(ARG, AB, BR, BG)	13	16.4	0.23	70.4
(ARG, AB, BG)	14	19.0	0.17	71.0

Graphical Models

Graphical models are a subclass of loglinear models that have interpretations in terms of conditional independence. These models are determined by their two-factor interactions. Define a *clique* to be a maximal subset of points such that all are connected. A graphical model includes the highest order interaction of all the variables within a clique. For instance, if the XY, YZ, XZ are all in the model, the XYZ interaction must be included in the graphical model.

In this figure, the association diagram represents the models (XW, WYZ) and (XW, WY, YZ, WZ). The graphical model corresponding to this graph must contain WYZ, and thus, (XW, WYZ) is the graphical model corresponding to this graph.

In this figure, the association diagram represents the models (XWZ, WYZ), (XWZ, WY, YZ), (WYZ, XW, XZ), and (XW, XZ, WY, XZ, YZ). The graphical model corresponding to this graph must contain XWZ and WYZ. Thus, the graphical model is (XWZ, WYZ) [1].

---

[1] Graphical Models with R by Søren Højsgaard, David Edwards, and Steffen Lauritzen.

A Loglinear Model for Linear-by-Linear Association

Loglinear models provide a convenient method for modeling linear dependence in contingency tables. We assign scores $\{u_i\}$ to the I rows and $\{v_j\}$ to the J columns. To show the ordering, we let $u_1 \leq u_2 \leq \cdots \leq u_I$ and $v_1 \leq v_2 \leq \cdots \leq v_J$. A one-parameter model for association in a two-way table with ordered row and column categories is the model for *linear-by-linear (or L × L) association*:

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \beta u_i v_j.$$

- This is a special case of the saturated model where $\lambda_{ij}^{XY} = \beta u_i v_j$.
- The rows and columns are independent when $\beta = 0$.
- When $\beta \neq 0$, the deviation of $\log(\mu_{ij})$ from independence is given by $\beta u_i v_j$.
- For a fixed level of X, the deviation is linear in the Y scores.
- When $\beta > 0$, the association is positive and the greatest deviation occurs in upper left and lower right corners in the table.
- When $\beta < 0$, the association is negative and the greatest deviation occurs in upper right and lower left corners in the table.

A useful interpretation for the parameter β is obtained by looking at the odds ratios of partial 2×2 tables formed from rows $r < s$ and columns $c < d$:

$$\log\left(\frac{\mu_{rc}\mu_{sd}}{\mu_{rd}\mu_{sc}}\right) = \beta(u_s - u_r)(v_d - v_c).$$

- For adjacent rows and columns with indices one apart, the log odds ratio equals β. This is called *uniform association*.
- This log odds ratio is larger when
    - (i) $|\beta|$ is larger
    - (ii) the rows (r, s) and columns (c, d) are farther apart.

We can test the null hypothesis of independence $H_0 : \beta = 0$ versus an alternative of linear-by-linear association by forming the likelihood ratio statistic

$$G^2 = G^2(I) - G^2(L \times L)$$

which has under H_0 a chi-squared distribution with one d.f. When linear-by-linear association is present, this test is more powerful than the general test of independence for a $I \times J$ table.

Next consider the dataset on the mental health status of a sample of 1660 young New York residents in midtown Manhattan classified by their parents' socioeconomic status (SES). It is given in Srole et al. (1978, p. 289) [2].

---

[2] Srole, L., Langner, T. S., Michael, S. T., Kirkpatrick, P., Opler, M. K., and Rennie, T. A. C. (1978). Mental Health in the Metropolis: The Midtown Manhattan Study. New York: NYU Press.

Let’s consider the mental impairement data

Code

```
 mydata=array(c(64, 57, 57, 72, 36, 21,  94, 94, 105, 141, 97, 71, 58,
 54, 65, 77, 54, 54, 46, 40, 60, 94, 78, 71), dim=c(6, 4),
 dimnames=list(SES=c(1:6), Mental_impairement=c("Well", "Mild",
 "Moderate", "Impaired")))
> mydata
   Mental_impairement
SES Well Mild Moderate Impaired
  1   64   94       58       46
  2   57   94       54       40
  3   57  105       65       60
  4   72  141       77       94
  5   36   97       54       78
  6   21   71       54       71
mydata2=as.data.frame(as.table(mydata))
```

Let’s consider the mental impairement data

Code

```
out.ind=glm(Freq~as.factor(SES)+Mental_impairement,poisson,data=mydata2)
 mu=rep(1:6, 4)
 nu=rep(1:4, each=6)

# Linear by linear fit
 out.linbylin=glm(Freq~as.factor(SES)+Mental_impairement+mu:nu,poisson,
data=mydata2)

 summary(out.ind)

 summary(out.linbylin)
```

Further Models for Association

The linear-by-linear association model uses one degree of freedom to express the dependence relation. It also supposes that there are prespecified row and column scores. We now consider some models for association that are more flexible than the model for linear-by-linear association (also called uniform association).

Row Effects Model

We consider the situation where the rows are unordered, but the columns have scores $v_1 \leq v_2 \leq \cdots \leq v_J$. We replace the ordered terms $\{\beta u_i\}$ in $\beta u_i v_j$ term of the linear-by-linear association model by the unordered terms $\{\mu_i\}$ to obtain

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \mu_i v_j.$$

We need constraints on the parameters such as $\lambda_I^X = \lambda_J^Y = \mu_I = 0$. This model adds $I - 1$ degrees of freedom to the independence model so that the degrees of freedom for this model are $(I-1)(J-1) - (I-1) = (I-1)(J-2)$. The row effects model treats the rows as nominal and the columns as ordinal with known scores.

We can interpret the parameters by looking at the odds ratios of partial 2×2 tables formed from rows $r < s$ and columns $c < d$:

$$\log\left(\frac{\mu_{rc}\mu_{sd}}{\mu_{rd}\mu_{sc}}\right) = (\mu_s - \mu_r)(v_d - v_c).$$

The log odds ratio is proportional to the distance between the columns with the constant of proportionality being $\mu_s - \mu_r$. If we consider rows $s = i + 1$ and $r = i$ and assume that the scores for the columns are equally spaced one unit apart, the local log odds ratio equals $\mu_{i+1} - \mu_i$. Since this depends on i, the local log odds ratios are not constant as in the uniform association model and depend on the row.

Revisit the mental impairement data

Code

```
# Row effects model
out.roweffect=glm(Freq~as.factor(SES)+Mental_impairement+
as.factor(SES):nu, poisson,data=mydata2)

summary(out.roweffect)
```

Column Effects Model

We consider the situation where the columns are unordered, but the rows have scores $u_1 \leq u_2 \leq \cdots \leq u_I$. We replace the ordered terms $\{\beta v_j\}$ in the $\beta u_i v_j$ term of the linear-by-linear association model by the unordered terms $\{\nu_i\}$ to obtain

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + u_i \nu_j.$$

We need constraints on the parameters such as $\lambda_I^X = \lambda_J^Y = \nu_J = 0$. This model adds $J - 1$ degrees of freedom to the independence model so that the degrees of freedom for this model are $(I-1)(J-1) - (J-1) = (I-2)(J-1)$. The column effects model treats the columns as nominal and the rows as ordinal with known scores.

We can interpret the model using the local odds ratio in a manner analogous to that for the row effect model. If we consider columns $c = j + 1$ and $d = j$ and assume that the scores for the rows are equally spaced one unit apart, the local log odds ratio equals $\nu_{j+1} - \nu_j$. Since this depends on i, the local log odds ratios are not constant as in the uniform association model and depend on the column.

Row and Column Effects Model

A generalization of both the row effects model and the column effects model that allows for both row effects and column effects in the local odds ratio is row and column effects model. We consider the situation where the rows are ordered with scores $u_1 \leq u_2 \leq \cdots \leq u_I$, and the columns have scores $v_1 \leq v_2 \leq \cdots \leq v_J$. We replace the ordered terms $\{\beta u_i v_j\}$ in the linear-by-linear association model by $\{\mu_i v_j + \nu_j u_i\}$ to obtain

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \mu_i v_j + \nu_j u_i.$$

We need constraints on the parameters such as $\lambda_I^X = \lambda_J^Y = \mu_I = \nu_J = 0$. This model adds $I-1+J-1$ degrees of freedom to the independence model so that the degrees of freedom for this model are $(I-2)(J-2)$. The row effects model treats both the rows and the columns as ordinal with known scores.

We can interpret the model using the local odds ratio in a manner analogous to that for the row effects and column effects model. If we consider rows $s = i + 1$ and $r = i$ and columns $c = j + 1$ and $d = j$ and columns and assumes the the scores are equally spaced one unit apart, the local log odds ratio equals $\mu_{i+1} - \mu_i + \nu_{j+1} - \nu_j$. Since this depends on i, the local log odds ratios are not constant as in the uniform association model and depend on both the row and the column.

These provide a hierarchy of increasingly complex models ranging from the independence model up to the saturated model. We can use the deviances to form likelihood ratio statistics to compare nested models.