ST3241 Categorical Data Analysis
Review II

Introduction

- We have discussed methods for analyzing associations in two-way and three-way tables.
- Now we will use models as the basis of such analysis.
- Models can handle more complicated situations than discussed so far.
- We can also estimate the parameters, which describe the effects in a more informative way.

Components of a GLM

- Random component: Identifies the response variable Y and assumes a probability distribution (Binomial, Poisson, or Multinomial) for it
- Systematic component: Specifies the explanatory variables $x_1, \cdots, x_p$ used as predictors in the model through a linear combination $\eta = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$.
- Link: Describes the functional relation between the systematic component and expected value of the random component: $g(\mu) = \eta$

Some Popular Link Functions

- Identity Link $g(\mu) = \mu$
- Log link $g(\mu) = \log(\mu)$
- Logit link $g(\mu) = \log[\frac{\mu}{1-\mu}]$
- Canonical link: the link function that uses the natural parameter as $g(\mu)$ in the GLM

Linear Probability Model

- To model the effect of X, use ordinary linear regression, by which the expected value of Y is a linear function of X.
- The model
$$\pi(x) = \alpha + \beta x$$
is called a linear probability model.
- Probabilities fall between 0 and 1 but for large of small values of x, the model may predict $\pi(x) < 0$ or $\pi(x) > 1$.
- This model is valid only for a finite range of x values

Logistic Regression Model

- A simple logistic regression model: $\log(\frac{\pi(x)}{1-\pi(x)}) = \alpha + \beta x$
- That is, $\pi(x) = F_0(\alpha + \beta x), F_0(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$ where $F_0(x)$ is the cdf of the logistic distribution. Its pdf is $F_0(x)(1 - F_0(x))$.
- The associated GLM is called the *logistic regression function*.
- Logistic regression models are often referred as *logit* models as the link in this GLM is the *logit* link: $\text{logit}(\pi) = F_0^{-1}(\pi)$

Parameters

- The parameter β determines the rate of increase or decrease of the curve.
- When $\beta > 0$, $\pi(x)$ increases with x.
- When $\beta < 0$, $\pi(x)$ decreases as x increases.
- The magnitude of β determines how fast the curve increases or decreases.
- As $|\beta|$ increases, the curve has a steeper rate of change.

Alternative Binary Links

- In general, a class of models for binary responses can be written as

$$\pi(x) = F(\alpha + \beta x)$$

where F is a cdf for some distribution.

- It is equivalent to use the link function $g(\pi) = F^{-1}(\pi)$.
- The probit link: $g(\pi) = \Phi^{-1}(\pi)$ where $\Phi(x)$ is the cdf of $N(0, 1)$

Poisson Regression

- Random component: a Poisson distribution assumed
- Systematic component: $\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
- Log-link: $g(\mu) = \log(\mu)$

Exponential Family

- The random variable Y has a distribution in the exponential family, if its *p.d.f* (or *p.m.f.*) can be written as

$$f(y;\theta,\phi) = \exp\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\}$$

for some specific function $a(\phi)$, $b(\theta)$ and $c(y,\phi)$.

- The parameter θ is called the *natural parameter* and ϕ is called the *dispersion* (or *scale*) *parameter*.

Examples

- $N(\mu, \sigma^2)$: the canonical link: $g(\mu) = \mu$.
- Binomial(n, p): the canonical link $g(\pi) = \log(\frac{\pi}{1-\pi})$.
- Poisson(λ): the canonical link $g(\lambda) = \log(\lambda)$.

Mean and Variances

- $E(Y) = b'(\theta)$.
- $var(Y) = b''(\theta)a(\phi)$.

Maximum Likelihood Estimates

- ML estimates of β_j's are obtained by solving the likelihood equations using numerical methods.
- The ML estimates $\hat{\beta}_j$'s are approximately normally distributed.
- Thus, a confidence interval for a model parameter β_j equals

$$\hat{\beta}_j \pm z_{\alpha/2} ASE$$

where ASE is the asymptotic standard error of $\hat{\beta}_j$.

Testing For Significance

- To test: $H_0 : \beta_j = 0$.
- z-test: under H_0, $Z = \hat{\beta}_j / ASE \sim N(0,1)$ approximately
- Wald-type test: under H_0, $Z^2 \sim \chi^2_1$ approximately
- The likelihood-ratio test statistic equals

  $$-2\log(L_0/L_1) = -2[\log L_0 - \log L_1] = -2[l_0 - l_1] \sim \chi^2_1 \quad \text{approximately}$$

  under H_0 where L_0 and L_1 are the maximized likelihood functions under H_0 and H_1 respectively
- The score test uses the size of the derivative of the log-likelihood function evaluated at $\beta_j = 0$.

Model Residuals

- Raw residual:$r_i = y_i - \hat{\mu}_i =$ Observed $-$ fitted.
- Pearson residual$=\frac{\text{Oberved-fitted}}{\sqrt{v\hat{a}r(\text{observed})}} = \frac{y_i - \hat{\mu}_i}{\sqrt{v\hat{a}r(y_i)}}$.
- Adjusted residuals: the Pearson residuals divided by its estimated standard error.

A Simple Logistic Regression Model

- For a binary response variable Y and an explanatory variable X, let $\pi(x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x)$.
- The logistic regression model is

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

- Equivalently, the log odds, called the *logit*, has the linear relationship

$$logit[\pi(x)] = \log \frac{\pi(x)}{1 - \pi(x)} = \alpha + \beta x$$

- This equates the logit link function to the linear predictor.

Interpretation of Parameters

- The parameter β determines the rate of increase or decrease of the S-shaped curve.
- The sign of β indicates whether the curve ascends or descends.
- The rate of change increases as $|\beta|$ increases.
- When the model holds with $\beta = 0$, then $\pi(x)$ is identical at all x, so the curve becomes a horizontal straight line, and Y is then independent of X.

Linear Approximation Interpretations

- The slope approaches 0 as the probability approaches 1.0 or 0.
- The steepest slope of the curve occurs at x for which $\pi(x) = 0.5$; that x value is $x = -\alpha/\beta$.
- This value of x is sometimes called the *median effective level* and is denoted by EL_{50}.
- It represents the level at which each outcome has a 50% chance.

Odds Ratio Interpretation

- The odds of a success (i.e. $Y = 1$) at $X = x$ is:

$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta x) = e^{\alpha}(e^{\beta})^{x}$$

- The odds of a success at $X = x + 1$ is:

$$\frac{\pi(x+1)}{1 - \pi(x+1)} = \exp(\alpha + \beta(x+1)) = e^{\alpha}(e^{\beta})^{x+1}$$

- Therefore, the odds ratio:

$$OR = \frac{\pi(x+1)/(1 - \pi(x+1))}{\pi(x)/(1 - \pi(x))} = e^{\beta}$$

- Therefore, β can be considered as a log odds ratio for one unit width increase.

Confidence Interval For Effects

- For a simple logistic regression model:

$$logit[\pi(x)] = \alpha + \beta x$$

  a large sample 95% confidence interval is

$$\hat{\beta} \pm z_{\alpha/2}(ASE)$$

- Exponentiating the endpoints of this interval yields one for e^{β} the odds ratio for a 1-unit increase in X.

Tests of Significance

- To test $H_0 : \beta = 0$.
- z-test: Under H_0, $z = \hat{\beta}/ASE \sim N(0,1)$ approximately
- Wald-type test: Under H_0, $z^2 \sim \chi^2_1$
- likelihood ratio test: T^2 =residual deviance under H_0 -residual deviance under H_1. Under H_0, $T^2 \sim \chi^2_1$ approximately

Estimates of Probability

- The estimated probability that $Y = 1$ at $X = x$ is

$$\hat{\pi}(x) = \frac{\exp(\hat{\alpha} + \hat{\beta}x)}{1 + \exp(\hat{\alpha} + \hat{\beta}x)}$$

- The large sample standard error of the estimated logit is:

$$Var(\hat{\alpha} + \hat{\beta}x) = Var(\hat{\alpha}) + x^2 Var(\hat{\beta}) + 2xCov(\hat{\alpha}, \hat{\beta})$$

- A 95% confidence interval for the true logit is then

$$(\hat{\alpha} + \hat{\beta}x) \pm 1.96 \times \sqrt{Var(\hat{\alpha} + \hat{\beta}x)}$$

- Substituting each endpoint into the inverse transformation $\pi(x) = \exp(\text{logit})/[1 + \exp(\text{logit})]$ gives a corresponding interval for $\pi(x)$.

Model Checking

- Use the residual deviance
- Use the Pearson χ^2-test or the Likelihood ratio test based on the fitted values

Likelihood Ratio Tests for Goodness of Fit

- Let M_0 and M_1 be two competing models.
- Let L_0 and L_1 be the maximized log-likelihoods under the models M_0 and M_1 respectively.
- Similarly, let L_S denote the maximized log likelihood of the saturated model.
- Then the deviances for the models M_0 and M_1 are $G^2(M_0) = -2(L_0 - L_S)$ and $G^2(M_1) = -2(L_1 - L_S)$.

Likelihood Ratio Tests for Goodness of Fit

- Denote the likelihood ratio statistic for testing M_0, given that M_1 holds, by $G^2(M_0|M_1)$.
- Then

$$\begin{aligned} G^2(M_0|M_1) &= -2(L_0 - L_1) = -2(L_0 - L_S) - [-2(L_1 - L_S)] \\ &= G^2(M_0) - G^2(M_1) \end{aligned}$$

- This statistic is large when M_0 fits poorly compared to M_1.
- It has a large sample chi-squared distribution with d.f. equal to the difference between the residual $d.f.$ values for the two models.

Residuals

- Let y_i denote the number of successes for n_i trials at the i-th setting of the explanatory variables.
- Let $\hat{\pi}_i$ denote the predicted probability of success for the model fit.
- Then the Pearson residual for the setting i is:

$$e_i = \frac{y_i - n_i\hat{\pi}_i}{\sqrt{n_i\hat{\pi}_i(1 - \hat{\pi}_i)}}$$

- The Pearson statistic for testing the model fit satisfies

$$\chi^2 = \sum e_i^2$$

Qualitative Predictors

- Suppose the binary response Y has two binary predictors X and Z.
- For the $2 \times 2 \times 2$ contingency table, the model: $\text{logit}(\pi) = \alpha + \beta_1 x + \beta_2 z$ has separate main effects for the two predictors and no interaction effect.
- The variables X and Z in this model are *dummy variables* that indicates categories for the predictors.

Coefficient Interpretations

- At a fixed level z of Z, the effect on the logit of changing from $x = 0$ to $x = 1$ is

$$[\alpha + \beta_1 \times 1 + \beta_2 z] - [\alpha + \beta_1 \times 0 + \beta_2 z] = \beta_1$$

- It equals the log odds ratio between X and Y at $Z = z$.
- Thus $exp(\beta_1)$ describes the conditional odds ratio between X and Y.
- The lack of interaction term in this model implies that the model satisfies the *homogeneous association*.

Conditional Independence

- Conditional independence between X and Y, controlling for Z implies $\beta_1 = 0$.
- The simpler model $logit(\pi) = \alpha + \beta_2 z$ then applies to the three way model.
- One can test whether $\beta_1 = 0$ using a Wald statistic or a likelihood ratio statistic comparing the two models.

ANOVA Type Representations

- A factor having two levels requires only a single dummy variable.
- A factor having I levels requires $I - 1$ dummy variables.
- The model formula $logit(\pi_{ik}) = \alpha + \beta_i^X + \beta_k^Z$ represents the effects of X through parameters $\{\beta_i^X\}$ and the effects of Z through parameters $\{\beta_k^Z\}$.
- This model applies to any number of levels of X and Z.

Notes

- Each factor has as many parameters as it has levels, but one is redundant.
- For instance, if X has I levels, it has $I-1$ non-redundant parameters.
- β_i^X denotes the effects on the logit of being classified in level i of X.
- Conditional independence between X and Y, given Z, corresponds to $\beta_1^X = \beta_2^X = \cdots = \beta_I^X$

Redundancy In Parameters

- To account for the redundancy in parameters, one can set the parameter for the last category to be zero.
- An analogous approach is to set the parameter for the first category to be zero.
- Alternatively, one can impose the restriction $\beta_1^X + \beta_2^X + \cdots + \beta_I^X = 0$

Logit Model for $2 \times 2 \times K$ Tables

- Consider X to be binary and Z is a control variable with K levels.
- In the model $logit(\pi_{ik}) = \alpha + \beta_i^X + \beta_k^Z$ conditional independence exists between X and Y controlling for Z, if $\beta_1^X = \beta_2^X$.
- In such a case, common odds ratio $\exp(\beta_1^X - \beta_2^X)$ for the K partial tables equal 1.
- The CMH statistic used earlier is the efficient score statistic for testing $X - Y$ conditional independence in this model.
- The ML estimate of the common odds ratio $\exp(\beta_1^X - \beta_2^X)$ is an alternative to the Mantel-Haenszel estimator.

Multiple Logistic Regression

- Denote a set of k predictors for a binary response Y by $X_1, X_2, \cdots, X_k$.
- Model for the logit of the probability π that $Y = 1$ generalizes to $\text{logit}(\pi(x)) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$
- The parameters β_i refers to the effect of X_i on the log odds that $Y = 1$, controlling for other Xs.
- e.g. $\exp(\beta_i)$ is the multiplicative effect on the odds of a 1-unit increase in X_i, at fixed levels of other Xs.

Model selection: Elimination

- To select a model, we can use a *backward elimination procedure*, starting with a complex model and successively taking out the terms.
- At each stage, we eliminate the term in the model that has the largest p-value when we test that its parameters equal to zero.
- We test only the highest order terms for each variable.
- It is inappropriate to remove a main effect term if the model contains higher-order interactions involving that term.

Log-linear regression model

Two-way Tables

- Consider an $I \times J$ contingency table that cross-classifies a sample of n subjects on two categorical responses.
- Y_{ij}: observed cell frequency and μ_{ij}: expected cell frequency of the (i, j)-th cell.
- The cell counts Y_{ij} are independent having Poisson(μ_{ij}) distribution.
- Note that, if π_{ij} is the cell probability, then $\mu_{ij} = n\pi_{ij}$.

Various Log-linear Models

- The independence model is

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

for a row effect λ_i^X and a column effect λ_j^Y.

- The saturated model is

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

In $I \times 2$ Table

- Response Y has only 2 levels.
- In row i, the logit for the probability π that $Y = 1$ is:

$$\begin{aligned} \log\left(\frac{\pi_i}{1-\pi_i}\right) = \log\left(\frac{\mu_{i1}}{\mu_{i2}}\right) &= \log \mu_{i1} - \log \mu_{i2} \\ &= (\lambda + \lambda_i^X + \lambda_1^Y) - (\lambda + \lambda_i^X + \lambda_2^Y) \\ &= \lambda_1^Y - \lambda_2^Y \end{aligned}$$

- logit for Y does not depend on the levels of X.

Interpretation of Interaction

- Under the saturated model,there is a direct relationship between log odds ratios and $\{\lambda_{ij}^{XY}\}$ association parameters.
- In a 2×2 table,

$$\begin{aligned}\log\theta &= \log(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}) \\ &= \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}\end{aligned}$$

- The saturated model has as many parameters as it has Poisson observations.
- Thus, it gives a perfect fit.
- The sample odds ratio is the same as the estimated odds ratio based on the fitted values

Test of Independence

- In $I \times J$ tables, only $(I-1)(J-1)$ parameters are non-redundant.
- These *interaction* parameters in the saturated model are coefficients of cross products of $(I-1)$ dummy variables for X with $(J-1)$ dummy variables for Y.
- Tests of independence analyze whether these $(I-1)(J-1)$ parameters equal 0, so they have residual $d.f. = (I-1)(J-1)$.
- The likelihood ratio test based on the residual deviances under the null and full models can be used.

Three-way Tables

- The cell expected frequencies in the contingency table are denoted by $\{\mu_{ijk}\}$.
- Single factor terms $\lambda_i^X, \lambda_j^Y, \lambda_k^Z$ represent marginal distributions.
- Two factor terms $\lambda_{ij}^{XY}, \lambda_{ik}^{XZ}, \lambda_{jk}^{YZ}$ are related to partial associations between two variables conditional to the third variable.
- Three factor terms λ_{ijk}^{XYZ} are related to three-factor interactions.

Various Log-linear Models for 3-way Tables

- The independence model (X, Y, Z):
  $\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$
- The partial association model (XZ, YZ):

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

- The model permits association between X and Z controlling for Y.
- It also permits a $Y - Z$ association, controlling for X.
- It specifies conditional independence between X and Y, controlling for Z.

Various Log-linear Models

- The model (XY, XZ, YZ) permits all three pairs of variables to be conditionally dependent:

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

- For this model, the conditional odds ratios between any two variables are identical at each level of the third variable.
- The saturated model (XYZ):

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

- This model permits the odds ratio between any two variables to vary across levels of the third variable.
- It provides a perfect fit in a three-way table.

Interpreting Model Parameters

- Interpretation of loglinear model parameters refer to their highest order terms.
- Interpretations for the homogeneous association model use the two factor terms to describe associations.
- The two-factor parameters relate directly to conditional odds ratios:

$$\log \theta_{XY(k)} = \log(\frac{\mu_{11k}\mu_{22k}}{\mu_{12k}\mu_{21k}}) = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}$$

which does not depend on k.

Fitting Loglinear Models

- A log-linear model can be fitted to the two or three way table using R or SAS
- Fitted values can be obtained using the fitted equation
- Estimated odds ratios can be obtained using the fitted values or associated estimated parameters

Chi-Square Goodness-of-Fit Tests

- Consider the null hypothesis that the expected frequencies for a three-way table satisfy a given loglinear model.
- The LR and Pearson Chi-square statistics based on the fitted values are:

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}\log(\frac{n_{ijk}}{\hat{\mu}_{ijk}}),$$

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}$$

- The degrees of freedom equals the number of cell counts minus the number of non-redundant parameters in the model.
- The saturated model has $d.f. = 0$.

Residuals

- Cell residuals can be used to study the quality of the log-linear fit.
- They may indicate why a particular model does not fit well or highlight cells that display lack of fit.
- We may use *adjusted residuals* or *Pearson residuals*.
- When the model holds, the adjusted residuals have approximately standard normal distribution.
- So, absolute values of *adjusted residuals* larger than 2 when there are few cells and larger than 3 when there are many cells , indicate lack of fit.

Tests About Partial Association

- Test about partial association by comparing different loglinear models.
- Likelihood ratio test or Pearson chi-squared test can be constructed based on the fitted values
- Likelihood ratio test can also be based on the residual deviance difference between two models

Confidence Intervals For Odds Ratios

- ML estimators of parameters have large sample normal distributions.
- For models in which the highest order terms are two-factor associations, the estimates refer to the conditional log odds ratios.
- One can use the estimates along with their standard errors to construct confidence intervals for true log odds ratios and then exponentiate them to form intervals for odds ratios.

Four-way Tables

- Basic concepts of three-way tables extend readily to multi-way tables.
- We consider a four-way table with variables W, X, Y, and Z.
- Interpretations are simplest when there are no three-factor interaction terms.
- The homogeneous association model is (WX, WY, WZ, XY, XZ, YZ).
- Here each pair of variables is conditionally dependent, with the same odds ratios at each combination of levels of the other two variables.

Four-way Tables

- An absence of a two factor term implies conditional independence for those variables.
- Model (WX, WY, WZ, XZ, YZ) does not contain an $X - Y$ term, so it treats X and Y as conditionally independent at each combination of levels of W and Z.
- A model could contain any of the four possible three factor interaction terms: WXY, WXZ, WYZ, XYZ.
- The saturated model contains all these terms plus a four factor interaction term.

Dissimilarity Index

- For a table of arbitrary dimension with cell counts $\{n_i = np_i\}$ and fitted values $\{\hat{\mu}_i = n\hat{\pi}_i\}$ one can summarize the closeness of the model fit to the sample data by the dissimilarity index

$$D = \sum |n_i - \hat{\mu}_i|/(2n) = \sum |p_i - \hat{\pi}_i|/2$$

- This index takes values between 0 and 1, with smaller values representing a better fit.
- It represents the proportion of sample cases that must move to different cells in order for the model to achieve a perfect fit.

Dissimilarity Index

- The dissimilarity index D estimates a corresponding index Δ that describes model lack-of-fit in the population sampled.
- The value $\Delta = 0$ occurs when the model holds perfectly.
- In that case D overestimates Δ, substantially so for small samples, because of sampling variation.
- When the model does not hold, for sufficiently large n, the goodness-of-fit statistics G^2 and χ^2 will be large, showing lack-of-fit.
- The estimator D then reveals whether the lack of fit suggested by those statistics is important in practical sense.
- $D < 0.03$ suggests that the sample data follow the model quite closely, even though the model is not $perfect$.

Loglinear-Logit Connection

- Consider the loglinear model of homogeneous association in three-way tables

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

- Suppose Y is binary, and we treat it as a response and X and Z as explanatory.

- Let π denote the probability that $Y = 1$, which depends on the levels of X and Z.

Loglinear-Logit Connection

- The logit for Y is

$$\begin{aligned}
\text{logit}(\pi_{ik}) &= (\lambda_1^Y - \lambda_2^Y) + (\lambda_{i1}^{XY} - \lambda_{i2}^{XY}) + (\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ}) \\
&= \alpha + \beta_i^X + \beta_k^Z
\end{aligned}$$

Linear-by-Linear Association

- Consider a two-way table with two ordinal categorical variables X: I levels and Y:J levels
- Assign scores u_i to the I rows and v_j to the J columns.
- We must have $u_1 \leq u_2 \leq \cdots \leq u_I$ and $v_1 \leq v_2 \leq \cdots \leq v_J$ to reflect the category ordering.
- The linear-by-linear association model is

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

- The independence model is the special case $\beta = 0$. The final term represents the deviation from independence.

Interpretations

- The parameter β refers to the direction and strength of association.
- When $\beta > 0$, there is a tendency for Y to increase as X increases.
- When $\beta < 0$, there is a tendency for Y to decrease as X increases.
- When the data display a positive or negative trend, this model usually fits much better than the independence model.

Describing Associations

- For the 2×2 table using the cells intersecting rows a and c with columns b and d, the model has odds ratio equal to

$$\frac{\mu_{ab}\mu_{cd}}{\mu_{ad}\mu_{cb}} = \exp[\beta(u_c - u_a)(v_d - v_b)]$$

- The association is stronger as $|\beta|$ increases.
- For given β pairs of categories that are farther apart have greater differences between their scores and odds ratios farther from 1.

Further Comments

- In practice, the most common choice of scores is $u_i = i$ and $v_j = j$, simply the row and column numbers.
- The odds ratios formed using adjacent rows and adjacent columns are called *local odds ratios*.
- For these unit spaced scores, the local odds ratios simplifies so that e^{β} is the common value of all the local odds ratios.
- Any set of equally-spaced row and column scores has the property of uniform local odds ratios.
- This special case of the model is called *uniform association*.

Ordinal Tests of Independence

- The likelihood ratio test can be constructed based on the residual deviance difference of the two models
- The Wald's statistic provides an alternative to test this hypothesis.

Detecting Ordinal Conditional Association

- A useful model

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \beta u_i v_j + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

  The model is called a *homogeneous linear − by − linear association* model.

- The conditional independence model (XZ, YZ) is the special case of this model with $\beta = 0$.

- Unless this models fits very poorly, the tests comparing this model are more powerful than tests that ignore the ordering.