# Chapter 2: Describing Contingency Tables - II 

Dipankar Bandyopadhyay

Department of Biostatistics,<br>Virginia Commonwealth University

BIOS 625: Categorical Data \& GLM
[Acknowledgements to Tim Hanson and Haitao Chu]

- In describing a relationship between categorical variables $X$ and $Y$, one should be aware of possible confounding between $(X, Y)$ and another variable $Z$.
- Your text gives an example on studying the relationship between spousal secondhand smoke and lung cancer. Obvious variables to control for (i.e. stratify by) are age, socioeconomic status, smoke exposure elsewhere, etc.
- For now, let's assume $Z$ is also categorical with $K$ levels. The outcome is now $n_{i j k}$, the number out of $n$ such that $(X, Y, Z)=(i, j, k)$.


### 2.3.1 \& 2.3.2 Partial Tables \& Death Penalty Example

- Initially consider $K$ different $2 \times 2$ partial tables of counts; for $Z=k$ we have

$$
\begin{array}{c|cc} 
& Y=1 & Y=2 \\
\hline X=1 & n_{11 k} & n_{12 k} \\
X=2 & n_{21 k} & n_{22 k}
\end{array}
$$

- Each table may have a different association between $X$ and $Y$, perhaps estimated by $\hat{\theta}_{X Y(k)}=n_{11 k} n_{22 k} /\left[n_{12 k} n_{21 k}\right]$, and this association will usually change with levels of $Z=k$.
- The marginal table ignores the role of $Z$ and collapses the table:

$$
\begin{array}{c|cc} 
& Y=1 & Y=2 \\
\hline X=1 & n_{11+} & n_{12+} \\
X=2 & n_{21+} & n_{22+}
\end{array}
$$

The association in this marginal table may be similar to that observed for some levels of $Z=k$, or not.

## Death Penalty Example

- Here's a $2 \times 2 \times 2$ contingency table on the data of $n=674$ convicted murder cases in Florida from 1976 to 1987.

| Victim's Race | Defendant's Race | Death penalty | No death penalty |
| :--- | :--- | :---: | :---: |
| White | White | 53 | 414 |
|  | Black | 11 | 37 |
| Black | White | 0 | 16 |
|  | Black | 4 | 139 |

- We are interested in the association between $X=$ defendant race (black or white) and $Y=$ death penalty (yes or no).
- Let's look at the table collapsed over the victim's race:

| Defendant's Race | Death penalty | No death penalty |
| :--- | :---: | :---: |
| White | 53 | 430 |
| Black | 15 | 176 |

- The probability of a death penalty is estimated to be $53 / 484=0.11$ versus 0.08 for white versus black defendants. Ignoring the victim's race leads us to believe that whites are more likely to get the death penalty.
- However, when we stratify by the victim's race, these probabilities are 0.11 and 0.23 (white versus black defendants) for white victims and 0.00 and 0.03 for black victims. In both cases black defendants are more likely to be given the death penalty.
- This illustrates the importance of adjusting for concomitant, often confounding variables (victim's race) that may be associated with both the response (death penalty) and a predictor (defendant's race).
- This is an example of Simpson's Paraox. This happens because whites strongly tend to kill whites \& although less strong, blacks tend to kill blacks. Let $D=$ death penalty, $D_{w} \& D_{b}$ defendant, and $V_{b} \&$ $V_{w}$ victim.

$$
\begin{aligned}
P\left(D \mid D_{w}\right) & =P\left(D \mid D_{w}, V_{b}\right) P\left(V_{b} \mid D_{w}\right)+P\left(D \mid D_{w}, V_{w}\right) P\left(V_{w} \mid D_{w}\right) \\
& \doteq 0.00(0.03)+0.11(0.97)=0.11 \\
P\left(D \mid D_{b}\right) & =P\left(D \mid D_{b}, V_{b}\right) P\left(V_{b} \mid D_{b}\right)+P\left(D \mid D_{b}, V_{w}\right) P\left(V_{w} \mid D_{b}\right) \\
& \doteq 0.03(0.75)+0.23(0.25)=0.08
\end{aligned}
$$

- For example, $P\left(D \mid D_{w}\right)$ is a weighted average of $P\left(D \mid D_{w}, V_{b}\right)$ and $P\left(D \mid D_{w}, V_{w}\right)$ and the conditional weights $P\left(V_{b} \mid D_{w}\right)=0.03$ and $P\left(V_{w} \mid D_{w}\right)=0.97$ favor the larger value because whites tend to kill whites.


Figure: Illustration of Simpson's Paradox

### 2.3.3 Conditional and Marginal Odds Ratios

- Consider a $2 \times 2 \times K$ table. Within a fixed level $k$ of $Z$, let the odds ratio for $(X, Y)$ be

$$
\theta_{X Y(k)}=\frac{\mu_{11 k} \mu_{22 k}}{\mu_{12 k} \mu_{21 k}}
$$

where $\mu_{i j k}=n \pi_{i j k}$ is the expected cell number in the table. Of course these can be estimated by replacing expected frequencies by MLEs $\hat{\mu}_{i j k}$.

- There are $k=1, \ldots, K$ of these conditional odds ratios. The marginal odds ratio from the collapsed table is given by

$$
\theta_{X Y}=\frac{\mu_{11+} \mu_{22+}}{\mu_{12+} \mu_{21+}}
$$

- Now consider $I \times J \times K$ tables. If

$$
P(X=i, Y=j \mid Z=k)=P(X=i \mid Z=k) P(Y=j \mid Z=k)
$$

for all $i=1, \ldots, I$ and $j=1, \ldots, J$ then $X$ is (conditionally) independent of $Y$ given $Z=k$. We write $X \perp Y \mid Z=k$.

- If this holds for all $k=1, \ldots, K$ then $X$ is independent of $Y$ given $Z$, $X \perp Y \mid Z$. This is equivalent to (2.8) on page 52 .
- Assuming a single multinomial applies to all counts $\left[n_{i j k}\right]_{/ \times J \times K}$, independence implies that

$$
\frac{\pi_{i j k}}{\pi_{++k}}=\frac{\pi_{i+k}}{\pi_{++k}} \times \frac{\pi_{+j k}}{\pi_{++k}}
$$

or $\pi_{i j k} \pi_{++k}=\pi_{i+k} \pi_{+j k}$.

### 2.3.4 Conditional and Marginal Independence

Note: Conditional independence does not imply marginal independence!

- That is, $X \perp Y \mid Z$ does not imply $X \perp Y$. This is true in general for any $(X, Y, Z)$. For $2 \times 2 \times K$ tables, $X \perp Y \mid Z$ if and only if $\theta_{X Y(k)}=1$ for $k=1, \ldots, K$; i.e. if the relative rates of success do not change with levels of $Z$.
- Example: The following is a stratified table containing the (virtually always unknown) $\mu_{i j k}$ where $i=1,2$ indicates treatment, $j=1,2$ indicates outcome, and $k=1,2$ indicates clinic.

|  |  | Success | Failure |
| :--- | :---: | :---: | :---: |
| Clinic 1 | Treatment A | 18 | 12 |
|  | Treatment B | 12 | 8 |
| Clinic 2 | Treatment A | 2 | 8 |
|  | Treatment B | 8 | 32 |

- Here $\theta_{X Y(1)}=\theta_{X Y(2)}=1: X$ and $Y$ are conditionally independent within a clinic. We conclude $X$ and $Y$ are not associated.
- However, when we examine the marginal table

|  | Success | Failure |
| :---: | :---: | :---: |
| Treatment A | 20 | 20 |
| Treatment B | 20 | 40 |

we obtain $\theta_{X Y}=2$, the odds of success are twice as great with treatment A instead of B . What is happening here?

- Loosely: Clinic 1 has a better overall success rate $\left(P\left(S \mid C_{1}\right)=0.6\right)$ than clinic $2\left(P\left(S \mid C_{2}\right)=0.2\right)$ - perhaps clinic 1 serves a more vital population. Also, clinic 1 tends to use treatment $A$ more than $B$. So the collapsed results are weighted by clinic 1 's more frequent use of $A$ and better success rate.
- Bottom line: it does not matter which treatment you receive, but you should try to get into clinic 1!


### 2.3.5 Homogeneous Association

- When $\theta_{X Y(1)}=\theta_{X Y(2)}=\cdots=\theta_{X Y(K)}$ the association between $X$ and $Y$ is the same for each fixed value of $Z=k$. This is called homogeneous association.
- If additionally, $\theta_{X Y(k)}=1$ for each $Z=k$ then $X \perp Y \mid Z$.
- Example: $X=$ smoking (yes, no), $Y=$ lung cancer (yes, no), and $Z=$ age $(<45,45-65,>65)$. If $\theta_{X Y(1)}=1.2, \theta_{X Y(2)}=3.9$, and $\theta_{X Y(3)}=8.8$, then the association between smoking and lung cancer strengthens with age. $X$ and $Y$ are conditionally dependent on age $Z$.
- The Cochran-Mantel-Haenszel statistics tests $H_{0}: X \perp Y \mid Z$, coming up in Section 6.4.


### 2.3.6 Collapsibility: Identical Conditional and Marginal Associations

- Collapsibility of Odds Ratio: When $\theta_{X Y(k)}$ is identical at every level $k$ of $Z$, that value equals $\theta_{X Y}$ if either $Z$ and $X$ are conditionally independent or if Z and Y are conditionally independent.
- Collapsibility of Difference of Proportions (or Relative Risk): When $\pi_{1}-\pi_{2}$ (or $\frac{\pi_{1}}{\pi_{2}}$ ) is same at every level $k$ of $Z$, that value equals the corresponding marginal measure if $Z$ is independent of $X$ in the marginal $X Z$ table or if $Z$ is conditionally independent of $Y$ given $X$.


### 2.4.1 Odds Ratios in $I \times J$ tables

- For $2 \times 2$ tables, $\theta$ summarizes the association between $X$ and $Y$. For larger two-dimensional tables, i.e. $I \times J$ tables, we need to generalize this idea. There are $(I-1)(J-1)$ local odds ratios

$$
\theta_{i j}=\frac{\pi_{i j} \pi_{i+1, j+1}}{\pi_{i, j+1} \pi_{i+1, j}} \text { for } i=1, \ldots, I-1 \text { and } j=1, \ldots, J-1
$$

- IS THIS RIGHT? $\theta_{i j}$ is relative odds of $Y=j$ versus $Y=j+1$ when $X=i$ versus $X=i+1$. All possible odds ratios each $2 \times 2$ table obtained from any two of the $\binom{1}{2}$ rows and any two $\binom{J}{2}$ columns from the $I \times J$ table can be obtained from the $(I-1)(J-1)$ local odds ratios $\left\{\theta_{i j}\right\}$.

For example, say $I=J=3$. Then there are $(3-1)(3-1)=4$ local odds ratios $\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}$.
Homework problem - not graded: Obtain

$$
\theta=\frac{P(Y=3 \mid X=2) / P(Y=1 \mid X=2)}{P(Y=3 \mid X=3) / P(Y=1 \mid X=3)}
$$

as a function of

$$
\begin{aligned}
\theta_{11} & =\frac{P(Y=1 \mid X=1) / P(Y=2 \mid X=1)}{P(Y=1 \mid X=2) / P(Y=2 \mid X=2)}, \\
\theta_{12} & =\frac{P(Y=2 \mid X=1) / P(Y=3 \mid X=1)}{P(Y=2 \mid X=2) / P(Y=3 \mid X=2)}, \\
\theta_{21} & =\frac{P(Y=1 \mid X=2) / P(Y=2 \mid X=2)}{P(Y=1 \mid X=3) / P(Y=2 \mid X=3)}, \\
\theta_{22} & =\frac{P(Y=2 \mid X=2) / P(Y=3 \mid X=2)}{P(Y=2 \mid X=3) / P(Y=3 \mid X=3)}
\end{aligned}
$$

- When $X$ and $Y$ are ordinal, we often examine odds of cumulative probabilities of the form

$$
\theta=\frac{P(Y \leq 3 \mid X \leq 1) / P(Y>3 \mid X \leq 1)}{P(Y \leq 3 \mid X>1) / P(Y>3 \mid X>1)}
$$

- For example if $Y$ is the answer to "We should increase funding to public schools" (strongly disagree, disagree, ambivalent, agree, strongly agree) and $X$ is education level (high school, undergraduate, graduate degree), this would be the odds of a random subject not agreeing for more money for schools given the subject has a high school versus these odds with a college degree.
- Of these types of odds are the same across a table,

$$
\theta=\frac{P(Y \leq j \mid X \leq i) / P(Y>j \mid X \leq i)}{P(Y \leq j \mid X>i) / P(Y>j \mid X>i)}
$$

for all $i$ and $j$, then $\theta$ is termed a global odds ratio. It is a single number that summarizes association in an $I \times J$ table.

### 2.4.4 Ordinal Trends: Concordant and Discordant Pairs

- Another single statistic that summarizes association for ordinal $(X, Y)$ uses the idea of concordant and discordant pairs. Consider:

|  | Job satisfaction |  |  |
| :---: | :---: | :---: | :---: |
| Age | Not | Fairly | Very |
| Satisfied | Satisfied | Satisfied |  |
| $<30$ | 34 | 53 | 88 |
| $30-50$ | 80 | 174 | 304 |
| $>50$ | 29 | 75 | 172 |

- Job satisfaction tends to increase with age. How to summarize this association?
- One measure of positive association is the probability of concordance.
- Consider two independent, randomly drawn individuals $\left(X_{1}, Y_{1}\right)$ and $\left(X_{2}, Y_{2}\right)$. This pair is concordant if either $X_{1}<X_{2}$ and $Y_{1}<Y_{2}$ simultaneously, or $X_{1}>X_{2}$ and $Y_{1}>Y_{2}$ simultaneously. An example would be (30-50, Fairly Satisfied) and ( $>50$, Very Satisfied). This indicates some measure of increased satisfaction with age.
- The probability of concordance $\Pi_{C}$ is

$$
\begin{aligned}
& P\left(X_{2}>X_{1}, Y_{2}>Y_{1} \text { or } X_{2}<X_{1}, Y_{2}<Y_{1}\right) \\
= & P\left(X_{2}>X_{1}, Y_{2}>Y_{1}\right)+P\left(X_{2}<X_{1}, Y_{2}<Y_{1}\right)
\end{aligned}
$$

- Using iterated expectation we can easily show

$$
\begin{aligned}
\Pi_{c} & =P\left(X_{2}>X_{1}, Y_{2}>Y_{1}\right)+P\left(X_{2}<X_{1}, Y_{2}<Y_{1}\right) \\
& =2 \sum_{i=1}^{\prime} \sum_{j=1}^{J} \pi_{i j}\left(\sum_{h=i+1}^{\prime} \sum_{k=j+1}^{J} \pi_{h k}\right)
\end{aligned}
$$

- Similarly, the probability of discordance $\Pi_{d}$ is given by

$$
\begin{aligned}
\Pi_{d} & =P\left(X_{1}>X_{2}, Y_{1}<Y_{2}\right)+P\left(X_{1}<X_{2}, Y_{1}>Y_{2}\right) \\
& =2 \sum_{i=1}^{\prime} \sum_{j=1}^{J} \pi_{i j}\left(\sum_{h=i+1}^{\prime} \sum_{k=1}^{j-1} \pi_{h k}\right)
\end{aligned}
$$

### 2.4.5 Ordinal Measure of Association: $\gamma$

- For pairs that are untied on both variables (i.e. they do not share the same salary or satisfaction categories), the probability of concordance is $\Pi_{c} /\left(\Pi_{c}+\Pi_{d}\right)$ and the probability of discordance is $\Pi_{d} /\left(\Pi_{c}+\Pi_{d}\right)$.
- The difference in these is the gamma statistic

$$
\gamma=\frac{\Pi_{c}-\Pi_{d}}{\Pi_{c}+\Pi_{d}}
$$

We have $-1 \leq \gamma \leq 1$. $\gamma=1$ only if $\Pi_{c}=1$, all pairs are perfectly concordant. (What would this force on the $\left\{\pi_{i j}\right\}$ ?)

- For the job satisfaction data,
$\hat{\gamma}=(C-D) /(C+D)=(99566-73943) /(99566+73943)=0.148$,
a weak, positive association between job satisfaction and age.


## Modifications to $\gamma$

- $\hat{\gamma}$ ignores ties. Kendall's $\hat{\tau}_{b}$ corrects for ties and has same interpretation; as with $\gamma,-1 \leq \tau_{b} \leq 1$.
- Stuart's $\hat{\tau}_{c}$ is Kendall's $\hat{\tau}_{b}$ corrected for sample size.
- Somer's $D, D(C \mid R)$ and $D(R \mid C)$ are asymmetric versions of $\hat{\tau}_{b}$, looking at either the column variable or the row variable as the dependent outcome.


## Polychoric correlation

- Another measure of association for two ordinal variables.
- For ordinal $(X, Y)$, we can envision underlying latent continuous variables $\left(Z_{1}, Z_{2}\right)$ that determine $(X, Y)$ according to cutoffs.

$$
X=i \Leftrightarrow \alpha_{i-1}<Z_{1}<\alpha_{i}
$$

and

$$
Y=j \Leftrightarrow \beta_{j-1}<Z_{2}<\beta_{j}
$$

where

$$
-\infty=\alpha_{0}<\alpha_{1}<\ldots<\alpha_{I-1}<\alpha_{I}=\infty
$$

and

$$
-\infty=\beta_{0}<\beta_{1}<\ldots<\beta_{J-1}<\alpha_{J}=\infty
$$

## Polychoric continued...

- Assume that $\left(Z_{1}, Z_{2}\right)$ are bivariate normal $N_{2}\left(\mathbf{O}_{2}, \Sigma\right)$, where $\Sigma=\left[\begin{array}{ll}1 & \rho \\ \rho & 1\end{array}\right]$; then there are $1+(I-1)+(J-1)$ parameters to estimate: $\rho, \boldsymbol{\alpha}=\left(\alpha_{1}, \alpha_{2}, \ldots, \alpha_{I-1}\right)$ and $\boldsymbol{\beta}=\left(\beta_{1}, \beta_{2}, \ldots, \beta_{J-1}\right)$.
- The parameter $\rho$ is called the polychoric correlation between $X$ and $Y$ , and can be estimated via maximum likelihood.


## A measure of nominal outcomes $\lambda_{Y \mid X}$

- Treat $Y$ as an outcome and $X$ as a predictor.
- If we ignore $X$, our best prediction for $Y$ is the $j$ with largest marginal $\pi_{+j}$, denoted $\pi_{m}=\max _{j=1, \ldots, J}\left\{\pi_{+j}\right\}$. The error in this choice is $1-\pi_{m}$.
- If we know $X=i$, then our best prediction for $Y$ is the $j$ with the largest $\pi_{i j}$, denoted $\pi_{c \mid i}=\max _{j=1, \ldots, J}\left\{\pi_{i j}\right\}$. The error in this choice is $1-\sum_{i=1}^{\prime} \pi_{c \mid i}=1-\pi_{c}$ when all $X^{\prime} s$ are weighted the same.
- The measure of association proposed by Goodman and Kruskall (1954) is the reduction in error when considering $X$ in the prediction of $Y$ vs. ignoring $X$ :

$$
\lambda_{Y \mid X}=\frac{\left(1-\pi_{m}\right)-\left(1-\pi_{c}\right)}{1-\pi_{m}}=\frac{\pi_{c}-\pi_{m}}{1-\pi_{m}}
$$

- SAS calls this $\lambda(C \mid R)$ or $\lambda(R \mid C)$, depending on whether the row or the column is the outcome variable.
- $\lambda_{Y \mid X}$ gives the proportion of error in predicting $Y$ that can be eliminated by using a known value of $X: 0 \leq \lambda_{Y \mid X} \leq 1$.

