Describing Contingency Tables (Ch.2)

Jaehyoung Hong

Section

- Section 2.1 : Basic terminology and notation
- Section 2.2 : Comparing groups
- Section 2.3 : Association
- Section 2.4 : Extension for nominal and ordinal multicategory variables

Contingency table describes frequency of X and Y

Y : Response variable 결과

	Myocardial Infarction		
	Fatal Attack	Nonfatal Attack	No Attack
Placebo	18	171	10,845
Aspirin	5	99	10,933

X : Explanatory variable

실험 설정

Cell

Distribution of contingency table

	Myocardial Infarction		
	Fatal Attack	Nonfatal Attack	No Attack
Placebo	18	171	10,845
Aspirin	5	99	10,933

N : Total number

Joint distribution $\pi_{11} = 18/N$ $\pi_{12} = 171/N$ …… $\pi_{12} = 10933/N$

Marginal distribution $\boxed{\pi_{i+} = \sum_j \pi_{ij}}$ $\boxed{\pi_{+j} = \sum_i \pi_{ij}}$ $\sum_i \pi_{i+} = \sum_j \pi_{+j} = \sum_i \sum_j \pi_{ij} = 1.0.$

Conditional distribution $\pi_{1|1} = \pi_{11}/(\pi_{11} + \pi_{12} + \pi_{13})$

$\{\pi_{1|i}, \ldots, \pi_{J|i}\}$ Conditional distribution of Y at category i of X

Sensitivity and Specificity

Breast Cancer	Diagnosis of Test: Positive	Diagnosis of Test: Negative	Total
Yes	0.82	0.18	1.0
No	0.01	0.99	1.0

Correct Diagnosis

$\pi_{11} = 0.82$: Has disease & Detect it → Sensitivity

$\pi_{22} = 0.99$: No disease & Detect it → Specificity

Notation

	Column		
Row	1	2	Total
1	π_{11} $(\pi_{1\|1})$	π_{12} $(\pi_{2\|1})$	π_{1+} (1.0)
2	π_{21} $(\pi_{1\|2})$	π_{22} $(\pi_{2\|2})$	π_{2+} (1.0)
Total	π_{+1}	π_{+2}	1.0

Independence

Independence

$$\pi_{ij} = \pi_{i+}\pi_{+j} \qquad \text{for } i = 1, \ldots, I \quad \text{and} \quad j = 1, \ldots, J.$$

Property

$$\pi_{j|i} = \pi_{ij}/\pi_{i+} = (\pi_{i+}\pi_{+j})/\pi_{i+} = \pi_{+j} \qquad \text{for } i = 1, \ldots, I.$$

$\therefore \pi_{j|1} = \cdots = \pi_{j|I}$ for $j = 1, \ldots J$ if independent

Homogeneity

Poisson, Binomial, and Multinomial Sampling

Condition	Sampling	Probability
Nothing is fixed	$\{Y_{ij}\}$ as indep.Poisson with $\{\mu_{ij}\}$	$\prod_i \prod_j \exp(-\mu_{ij}) \mu_{ij}^{n_{ij}} / n_{ij}!\,.$
Total n is fixed	Multinomial sampling	$[n!/(n_{11}! \cdots n_{IJ}!)] \prod_i \prod_j \pi_{ij}^{n_{ij}}\,.$
Row totals fixed (Y is indep when X is fixed)	Multinomial sampling for each row	$\frac{n_i!}{\prod_j n_{ij}!} \prod_j \pi_{j\mid i}^{n_{ij}}\,.$

$$n_i = n_{i+} = \sum_j n_{ij}$$

Example

Seat-Belt Use	Result of Crash: Fatality	Result of Crash: Nonfatality
Yes		
No		

Nothing is fixed Poisson random variables with unknown means $\{\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}\}$.

200 random sample **Multinomial variables with n=200 trials and unknown joint prob** $\{\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}\}$.

Cross-sectional

100 random sample with fatality
100 random sample with nonfatality
⇒ Column sum is fixed : Binomial for each column

Case-Control (Retrospective)

100 random sample with Seat-Belt
100 random sample without Seat-Belt
⇒ Row sum is fixed : Binomial for each row

Cohort (Perspective)

Comparing two groups : Difference, Relative risk, and Odd ratio

	Success	Failure
Group1	$\pi_{1\|1}$	$\pi_{2\|1}$
Group2	$\pi_{1\|2}$	$\pi_{2\|2}$

simpler notation π_i for $\pi_{1|i}$.

	Definition	Property	Usage
Difference of Proportions	$\pi_1 - \pi_2$	Zero if independent	
Relative Risk (RR)	π_1/π_2	1 if independent	Survival rate (0.010 & 0.001 vs 0.410 &0.401)
Odds ratio	$\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$	1 if independent	Orientation / Constant multiplication invariant is needed

$$\theta = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}.$$ Cross-product ratio

Odds ratio for case-control study

TABLE 2.5 Cross-Classification of Smoking by Lung Cancer

	Lung Cancer	
Smoker	Cases	Controls
Yes	688	650
No	21	59
Total	709	709

When smoking, Lung cancer increased?

Difference of proportion and RR has no-meaning (Since row-sum is not meaningful)

$$\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \frac{P(Y=1|X=1)/P(Y=2|X=1)}{P(Y=1|X=2)/P(Y=2|X=2)}$$

$$= \frac{P(X=1|Y=1)/P(X=2|Y=1)}{P(X=1|Y=2)/P(X=2|Y=2)} \cdot = \frac{(688/709)/(21/709)}{(650/709)/(59/709)} = \frac{688 \times 59}{650 \times 21} = 3.0.$$

Odd is meaningful
(Invariant under orientation)

Relationship between Odds and RR

$$\text{odds ratio} = \text{relative risk}\left(\frac{1-\pi_2}{1-\pi_1}\right).$$

If π_1 and π_2 both small : Odd = RR

Odd can be interpreted as RR

TABLE 2.1 Cross-Classification of Aspirin Use and Myocardial Infarction

	Myocardial Infarction		
	Fatal Attack	Nonfatal Attack	No Attack
Placebo	18	171	10,845
Aspirin	5	99	10,933

RR = 1.82

Odd = 1.83

for having heart attack

When ‘covariates’ matter : Partial table

TABLE 2.6 Death Penalty Verdict by Defendant’s Race and Victims’ Race

Victims’ Race	Defendant’s Race	Death Penalty Yes	Death Penalty No	Percent Yes
White	White	53	414	11.3
	Black	11	37	22.9
Black	White	0	16	0.0
	Black	4	139	2.8
Total	White	53	430	11.0
	Black	15	176	7.9

Covariate : Factor that you must normalize
→ Victim (Z)

<Partial table>

<Marginal table>

Not normalizing, but ‘ignoring’

Partial table and marginal table can give different result

TABLE 2.6 Death Penalty Verdict by Defendant's Race and Victims' Race

Victims' Race	Defendant's Race	Death Penalty Yes	Death Penalty No	Percent Yes
White	White	53	414	11.3
	Black	11	37	22.9
Black	White	0	16	0.0
	Black	4	139	2.8
Total	White	53	430	11.0
	Black	15	176	7.9

Partial table (W) : Black dependant have 22.9-11.3 = 11.6% more death penalty

Partial table (B) : Black dependant have 2.8-0.0 = 2.8% more death penalty

Marginal table : Black dependant have 11.0-7.9 = 2.1% less death penalty

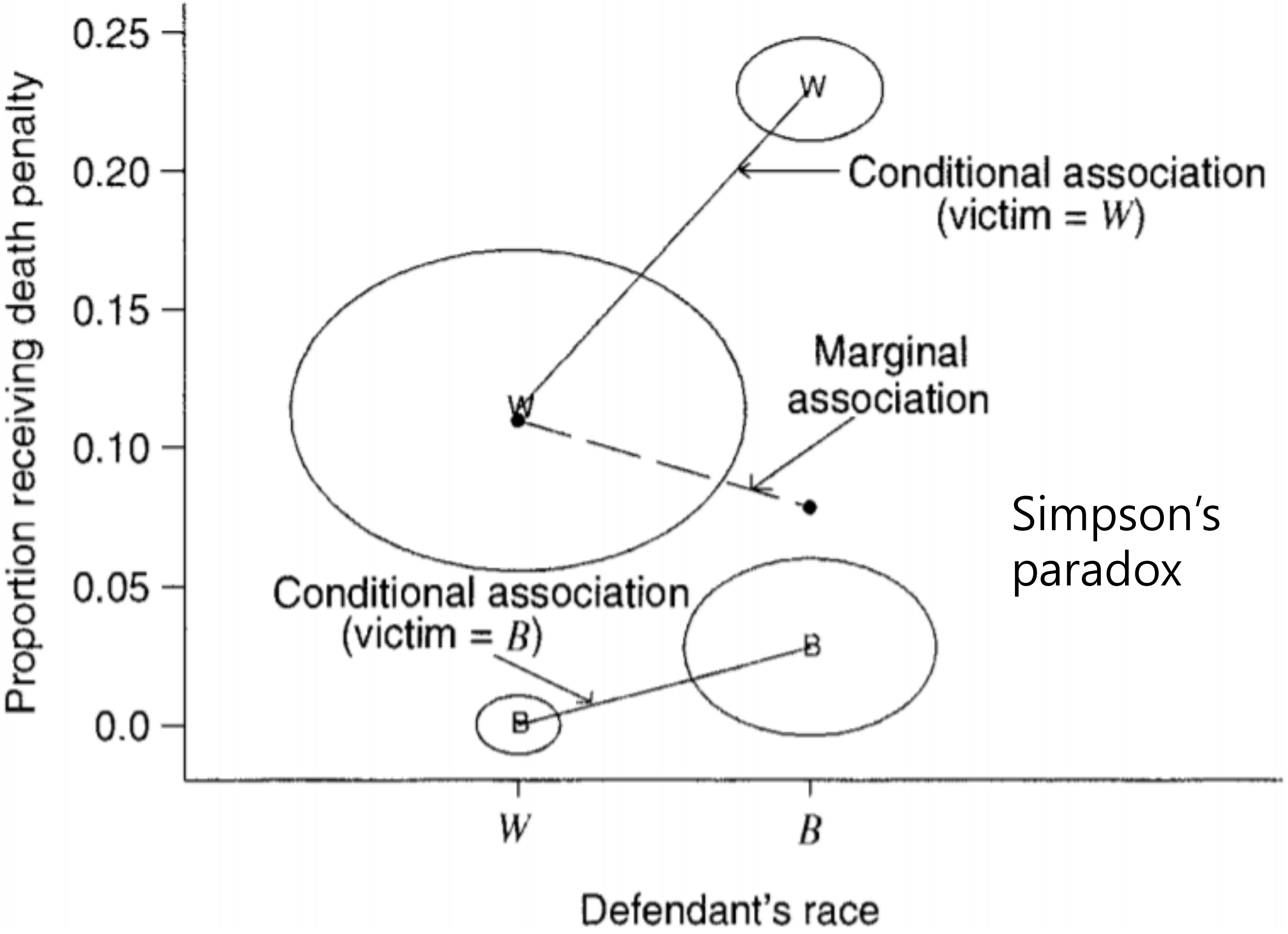
Why? 1) Victim race ~ Defendant race : $odd = \frac{467\times143}{48\times16} = 87$

2) Victim race ~ Death penalty

Partial table and marginal table can give different result

TABLE 2.6 Death Penalty Verdict by Defendant's Race and Victims' Race

Victims' Race	Defendant's Race	Death Penalty Yes	Death Penalty No	Percent Yes
White	White	53	414	11.3
	Black	11	37	22.9
Black	White	0	16	0.0
	Black	4	139	2.8
Total	White	53	430	11.0
	Black	15	176	7.9



Simpson's paradox

X축 : Explanatory
Y축 : Response
Center : Covariate
Circle size : number of pair between explanatory and covariate

Conditional and marginal odds ratios

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

Odd for *k*-th partial table
(Conditional odds)

$\neq$

$$\theta_{XY} = \frac{\mu_{11+}\,\mu_{22+}}{\mu_{12+}\,\mu_{21+}}$$

Odd for marginal table

TABLE 2.6 Death Penalty Verdict by Defendant's Race and Victims' Race

Victims' Race	Defendant's Race	Death Penalty Yes	Death Penalty No	Percent Yes
White	White	53	414	11.3
	Black	11	37	22.9
Black	White	0	16	0.0
	Black	4	139	2.8
Total	White	53	430	11.0
	Black	15	176	7.9

$$\hat{\theta}_{XY(1)} = \frac{53 \times 37}{414 \times 11} = 0.43.$$

$$\hat{\theta}_{XY(2)} = (0 \times 139)(16 \times 4) = 0$$

$$\hat{\theta}_{XY} = (53 \times 176)/(430 \times 15) = 1.45.$$

Conditional and marginal independence

Conditionally independent at level *k* of Z

$$P(Y=j\,|\,X=i, Z=k) = P(Y=j\,|\,Z=k), \qquad \text{for all } i, j.$$

Conditionally independent given Z

Independent at every level of Z

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

$$= \pi_{i+k} P(Y=j\,|\,Z=k) = \pi_{i+k} P(Y=j, Z=k)/P(Z=k)\;.$$

$$\Rightarrow \quad \pi_{ijk} = \pi_{i+k}\pi_{+jk}/\pi_{++k} \qquad \text{for all } i,\, j, \text{ and } k.$$

$\neq$ Marginal independence, since $\pi_{ij+} = \pi_{i++}\pi_{+j+} \neq \pi_{ij+} = \sum_k (\pi_{i+k}\pi_{+jk}/\pi_{++k}).$

Conditional and marginal independence

Clinic	Treatment	Response: Success	Response: Failure
1	A	18	12
	B	12	8
2	A	2	8
	B	8	32
Total	A	20	20
	B	20	40

$$\theta_{XY(1)} = \frac{18 \times 8}{12 \times 12} = 1.0, \qquad \theta_{XY(2)} = \frac{2 \times 32}{8 \times 8} = 1.0.$$

Indep conditionally

$$\theta_{XY} = (20 \times 40)/(20 \times 20) = 2.0$$

Not Indep marginally

Why? $\theta_{XZ} = (18 \times 8)/(12 \times 2) = 6.0.$ Clinic 1 tends to use Treat A

Homogenous association

A $2 \times 2 \times K$ table has *homogeneous XY association* when

$$\theta_{XY(1)} = \theta_{XY(2)} = \cdots = \theta_{XY(K)}.$$

XZ, YZ also homogeneous

Not homogeneous, but have trend among Z

$$\theta_{XY(1)} > \theta_{XY(2)} > \theta_{XY(3)}$$

Z is called effect modifier

TABLE 2.6 Death Penalty Verdict by Defendant's Race and Victims' Race

Victims' Race	Defendant's Race	Death Penalty Yes	Death Penalty No	Percent Yes
White	White	53	414	11.3
	Black	11	37	22.9
Black	White	0	16	0.0
	Black	4	139	2.8
Total	White	53	430	11.0
	Black	15	176	7.9

Ordinal trends : concordant and discordant pairs

TABLE 2.8 Cross-Classification of Job Satisfaction by Income

Income (dollars)	Job Satisfaction: Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied
< 15,000	1	3	10	6
15,000–25,000	2	3	10	7
25,000–40,000	1	6	14	12
> 40,000	0	1	9	11

Concordant pair

Tied pair

Discordant pair

$$C = 1(3 + 10 + 7 + 6 + 14 + 12 + 1 + 9 + 11) + 3(10 + 7 + 14 + 12 + 9 + 11) + 10(7 + 12 + 11) + 2(6 + 14 + 12 + 1 + 9 + 11) + 3(14 + 12 + 9 + 11) + 10(12 + 11) + 1(1 + 9 + 11) + 6(9 + 11) + 14(11) = 1331.$$

$$> \quad D = 3(2 + 1 + 0) + 10(2 + 3 + 1 + 6 + 0 + 1) + \cdots + 12(0 + 1 + 9) = 849.$$

Low income, low satisfaction

Ordinal trends : concordant and discordant pairs

$$C = 1(3 + 10 + 7 + 6 + 14 + 12 + 1 + 9 + 11)$$
$$+ 3(10 + 7 + 14 + 12 + 9 + 11) + 10(7 + 12 + 11)$$
$$+ 2(6 + 14 + 12 + 1 + 9 + 11) + 3(14 + 12 + 9 + 11)$$
$$+10(12 + 11) + 1(1 + 9 + 11) + 6(9 + 11) + 14(11) = 1331.$$

$$D = 3(2 + 1 + 0) + 10(2 + 3 + 1 + 6 + 0 + 1) + \cdots + 12(0 + 1 + 9) = 849.$$

Probability ver.

$$\Pi_c = 2\sum_i \sum_j \pi_{ij}\left(\sum_{h>i} \sum_{k>j} \pi_{hk}\right), \qquad \Pi_d = 2\sum_i \sum_j \pi_{ij}\left(\sum_{h>i} \sum_{k<j} \pi_{hk}\right).$$

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d},$$

Property는 correlation과 동일

$$\hat{\gamma} = (1331 - 849)/(1331 + 849) = 0.221.$$