

Tabelas Tridimensionais

Unidade II

Introdução

- Ao estudar o efeito de uma variável exploratória X sobre uma variável resposta Y , deve-se controlar um conjunto de variáveis que podem influenciar esta relação (*covariáveis*).
- Caso contrário, um efeito observado de X em Y pode refletir simplesmente as associações das covariáveis em X e Y .
- Uma estratégia é usar algum mecanismo que mantenha tais variáveis constantes enquanto se estuda o efeito de X sobre Y .

Introdução

- A inclusão de variáveis controle na análise exige que realizemos uma análise multivariada.
- Inicialmente será considerado o caso em que apenas uma variável controle é introduzida.
- O principal objetivo é analisar a associação entre duas variáveis categorizadas X e Y , controlando-se o efeito de uma variável confundidora Z .

Tabelas Parciais

- Suponha que dividamos a tabela de contingência tridimensional em tabelas de contingência bidimensionais separadamente para cada nível da variável controle Z .
- Estas tabelas bidimensionais que mostram a relação de X e Y para cada nível de Z , são chamadas de **tabelas parciais**. Elas mostram o efeito de X sobre Y controlando-se Z .
- As tabelas parciais removem o efeito de Z , pois este efeito mantém-se constante.

Tabela Marginal

- A tabela de contingência bidimensional obtida pela combinação das tabelas parciais é chamada de **tabela marginal X-Y**.
- Cada casela na **tabela marginal** é a soma das frequências das mesmas caselas nas tabelas parciais.
- A **tabela marginal** ao invés de controlar Z, ignora-o. A tabela marginal não contém nenhuma informação sobre Z. Ela simplesmente relaciona X com Y.

Exemplo - Pena de Morte

Veredicto de Pena de Morte segundo
Raça do Réu e Raça da Vítima

Raça Da Vítima	Raça da Réu	Veredicto de Pena de Morte		% de Sim
		Sim	Não	
Branca	Branca	53	414	11,3
	Negra	11	37	22,9
Negra	Branca	0	16	0,0
	Negra	4	139	2,8
Total	Branca	53	430	11,0
	Negra	15	176	7,9

Tabelas Parciais

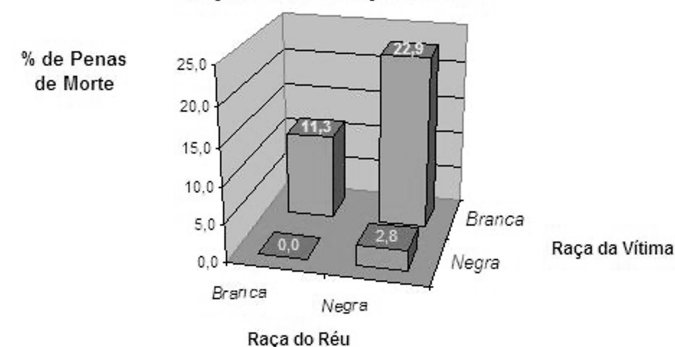
Tabela Marginal

Fonte: M.L. Radelet and G.L. Pierce, *Flórida Law Rev.* 43:1-34(1991).
Reimpresso com permissão de *Flórida Law Review*.

Associação Condicional e Associação Parcial

- As associações nas tabelas parciais são chamadas de **associações condicionais**, porque elas se referem ao efeito de X sobre Y condicionado a um nível fixo de Z.
- As associações condicionais nas tabelas parciais podem ser bastante diferentes das associações nas tabelas marginais.
- Analisar apenas uma tabela marginal de uma tabela de contingência multidimensional, pode conduzir a erros de interpretação.

Distribuição dos Veredictos de Pena de Morte segundo
Raça do Réu e Raça da Vítima



Exemplo - Pena de Morte

- Estudou-se o efeito da raça do réu sobre a pena de morte, tratando-se a raça da vítima como uma variável controle.
- Quando a vítima era branca a pena de morte foi imposta 22,9 - 11,3 = 11,6 % mais freqüente a negros do que a brancos.
- Quando a vítima era negra, a pena de morte foi imposta 2,8 % mais freqüente a negros do que brancos.
- Portanto, controlando-se a raça da vítima a porcentagem de penas de morte foi maior em negros do que em brancos.

Exemplo - Pena de Morte

- No entanto, ignorando-se a raça da vítima a penas de morte foi imposta 11,0 - 7,9 = 3,1 % mais freqüente a brancos do que a negros.
- Quando os resultados nas tabelas parciais são diferentes do resultado na tabela marginal temos um paradoxo, conhecido como *paradoxo de Simpson*.

Razão de Chances Marginal e Condicional

- Pode-se descrever as associações marginais e condicionais usando a razão de chances.
- Suponha uma tabela 2x2xK, onde K denota o número de níveis da variável controle. Seja $\{n_{ijk}\}$ as freqüências observadas e $\{\mu_{ijk}\}$ as freqüências esperadas.
- Para um nível k de Z seja:

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

- $\theta_{XY(k)}$ descreve a **associação condicional de X-Y**
- As razões de chances para as K tabelas parciais são denominadas de razão de chances condicionais de X-Y.

Razão de Chances Marginal e Condicional

- As razões de chances condicionais podem ser diferentes das *razões de chances marginais*, na qual a variável controle é ignorada.
- As X-Y tabelas marginais têm frequências esperadas $\{\mu_{ij+} = \sum_k \mu_{ijk}\}$.
- A *razão de chances marginal* é dada por:

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

- Substituindo-se as frequências esperadas pelas observadas teremos as estimativas de $\theta_{XY(k)}$ e de θ_{XY}

Razão de Chances Marginal e Condicional - Exemplo

- Do exemplo da pena de morte temos que a razão de chances da primeira e da segunda tabelas parciais são dadas por:

$$\hat{\theta}_{XY(1)} = \frac{53 \times 37}{414 \times 11} = 0,43, \quad \hat{\theta}_{XY(2)} = \frac{0 \times 139}{16 \times 4} = 0$$

- A estimativa da razão de chances para a raça do réu e pena de morte, desprezando a raça da vítima é

$$\theta_{XY} = \frac{53 \times 176}{430 \times 15} = 1,45$$

Independência Condicional e Independência Marginal

- Considere a relação entre X e Y controlada por Z.
 - Se X e Y são independentes para cada tabela parcial, então **X e Y são condicionalmente independente dado Z**.
 - Todas as razões de chances condicionais entre X e Y são iguais a 1.
- Independência condicional de X e Y dado Z, não implica em independência marginal de X e Y.
 - Isto é, quando as razões de chances entre X e Y são iguais a 1 para cada nível de Z, a razão de chances marginal pode ser diferente de 1.

Independência Condicional e Independência Marginal - Exemplo

- Considere o exemplo hipotético na qual:
 - Y = resposta (sucesso e falha),
 - X = tipo de droga (A, B)
 - Z = clínica (1, 2).

Clínica	Droga (tratamento)	Resposta	
		sucesso	fracasso
1	A	18	12
	B	12	8
2	A	2	8
	B	8	32
Total	A	20	20
	B	20	40

Independência Condicional e Independência Marginal

- A **associação condicional** entre X e Y para os dois níveis de Z é descrita pelas razões de chances:

$$\theta_{XY(1)} = \frac{18 \times 8}{12 \times 12} = 1 \quad \theta_{XY(2)} = \frac{2 \times 32}{8 \times 8} = 1$$

- Dado a clínica, a resposta e o tratamento são condicionalmente independentes.
- A **associação marginal** é dada pela seguinte razão de chances:

$$\theta_{XY} = \frac{20 \times 40}{20 \times 20} = 2$$

- Então as variáveis não são marginalmente independentes.

Independência Condicional e Independência Marginal

- Cometeríamos um erro em estudar apenas a associação marginal concluindo que o tratamento A produz mais sucesso do que o tratamento B.
- Indivíduos dentro de uma particular clínica são provavelmente mais homogêneos do que na amostra toda. Portanto, a resposta independe do tratamento em cada clínica.

Associação Homogênea

- Existe associação homogênea entre X e Y em uma tabela $2 \times 2 \times K$ quando:

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

- As razões de chances condicionais entre X e Y são idênticas para cada nível de Z. Então o efeito de X sobre Y é o mesmo para cada nível de Z, e um único número descreve as associações condicionais.
- Independência condicional de X e Y é um caso especial na qual cada razão de chance condicional é igual a 1.

Associação Homogênea

- Quando as razões de chances condicionais X-Y são idênticas para cada nível de Z, a mesma propriedade se verifica para outras associações. Por exemplo, as razões de chances condicionais entre os dois níveis de X e os dois níveis de Z são idênticas para cada nível de Y.
- Quando existe associação homogênea dizemos que não existe interação entre duas variáveis sobre uma terceira variável.

Associação Homogênea

- Quando não existe associação homogênea, a razão de chances condicional para qualquer par de variáveis muda ao longo dos níveis da terceira variável.

Exemplo:

Para X = fuma (sim, não);

Y = câncer de pulmão (sim, não)

Z = idade (< 45, 45-65, > 65),

Suponha que:

$$\theta_{XY(1)} = 1,2 \quad \theta_{XY(2)} = 2,8 \quad \theta_{XY(3)} = 6,2$$

- Então o fumo tem efeito fraco sobre o câncer de pulmão em pessoas jovens, mas o efeito aumenta consideravelmente com a idade.

Métodos de Cochran-Mantel-Haenszel

- Teste de independência condicional.
- Para tabelas 2x2xK, a hipótese nula:

H_0) X e Y são condicionalmente independentes, dado Z

significa que a razão de chances entre X e Y é igual a 1 em cada tabela parcial, isto é,

$$H_0) \theta_{XY(k)} = 1 \quad \text{para todo } k$$

Métodos de Cochran-Mantel-Haenszel

- Considere **K** tabelas 2x2 onde a **k-ésima** tabela tem frequências:

X	Y		Total
	1	2	
1	n_{11k}	n_{12k}	n_{1+k}
2	n_{21k}	n_{22k}	n_{2+k}
Total	n_{+1k}	n_{+2k}	n_{++k}

- Na tabela parcial **k**, as totais da linhas são $\{n_{1+k}, n_{2+k}\}$, e os totais das colunas são $\{n_{+1k}, n_{+2k}\}$.
- Como todos estes totais são fixos, todos os esquemas de amostragem fornecem uma distribuição hipergeométrica para a frequência na primeira linha e primeira coluna n_{11k} .
- O teste estatístico utiliza esta casela em cada tabela parcial.

Teste de Cochran-Mantel-Haenszel

- Sobre a hipótese nula, a média e a variância de n_{11k} são dadas por:

$$\mu_{11K} = E(n_{11k}) = \frac{n_{1+k} n_{+1k}}{n_{++k}}$$

$$Var(n_{11k}) = \frac{n_{1+k} n_{2+k} n_{+1k} n_{+2k}}{n_{++k}^2 (n_{++k} - 1)}$$

- Quando $\theta_{XY(k)}$ excede 1 na tabela parcial **k**, esperamos encontrar $(n_{11k} - \mu_{11k}) > 0$. A estatística do teste combina estas diferenças através de todas as **K** tabelas.

Teste de Cochran-Mantel-Haenszel

- A estatística do teste é dada por:

$$CMH = \frac{[\sum_k (n_{11k} - \mu_{11k})]^2}{\sum_k Var(n_{11k})}$$

- A estatística de Cochran-Mantel-Haenszel, para grandes amostras, tem distribuição aproximadamente qui-quadrado com 1 grau de liberdade.

Teste de Cochran-Mantel-Haenszel

- A estatística CMH assume grandes valores quando o numerador é consistentemente positivo ou negativo para todas as tabelas.
 - Quando a razão de chances excede um em todas as tabelas parciais, a soma destas diferenças tende a ser um relativamente grande e positivo;
 - Quando a razão de chances é menor do que um em cada tabela, a soma das diferenças é relativamente grande e negativa.
- O teste não é apropriado quando as associações mudam drasticamente entre as tabelas parciais, ou seja, quando as diferenças ocorrem em direções opostas e de magnitudes semelhantes.

Exemplo - Câncer de Pulmão

- Em um estudo sobre associação entre fumo e câncer de pulmão realizado em algumas cidades chinesas, observou-se o seguinte resultado:

Cidade	Fumo	Câncer de Pulmão		Razão de Chances
		Sim	Não	
Beijing	Fumante	126	100	2,2
	Não Fumante	35	61	
Shanghai	Fumante	908	688	2,14
	Não Fumante	497	807	
Shenyang	Fumante	913	747	2,18
	Não Fumante	336	598	

Exemplo - Câncer de Pulmão

- Testamos a hipótese de independência condicional entre fumo e câncer de pulmão para cada cidade,

*H₀) Fumo e câncer de pulmão são **condicionalmente independentes** para cada cidade*

*H₀) Fumo e câncer de pulmão são **condicionalmente dependentes** para cada cidade*

Estimação da Razão de Chances Comum

- Quando a associação é estável entre as tabelas parciais, podemos estimar um valor comum para as K razões de chances.
- Em uma tabela $2 \times 2 \times K$, suponha que $\theta_{XY(1)} = \dots = \theta_{XY(K)}$.

O estimador de Mantel-Haenszel da razão de chances comum é dado por:

$$\hat{\theta}_{MH} = \frac{\sum_K \left(\frac{n_{11K} n_{22K}}{n_{++K}} \right)}{\sum_K \left(\frac{n_{12K} n_{21K}}{n_{++K}} \right)}$$

Teste de Homogeneidade das Razões de Chances

- Pode-se testar a hipótese de que as razões de chances entre X e Y são as mesmas para cada nível de Z, isto é:

$$H_0: \theta_{XY(1)} = \dots = \theta_{XY(K)}$$

- Sejam $\{\hat{\mu}_{11K}, \hat{\mu}_{12K}, \hat{\mu}_{21K}, \hat{\mu}_{22K}\}$ as frequências esperadas estimadas na k -ésima tabela marginal sob a hipótese de igualdade da razão de chances.
- A estatística do teste, denominada **estatística de Breslow-Day** é dada por:

$$X_{BD}^2 = \sum \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}$$

e tem distribuição qui-quadrado com $k-1$ g.l.