Universidade de Brasília
IE - Departamento de Estatística

Modelos Log-lineares

Tabelas tridimensionais

Maria Teresa Leão Costa

Modelos Log-lineares para Tabelas Tridimensionais

- Considere uma tabela de contingência tridimensional contendo a classificação cruzada das variáveis X, Y e Z.
  - tabelas parciais - distribuição de X e Y para cada nível de Z, por exemplo.
  - tabelas marginais
- Designando as probabilidades conjuntas das células por:

$$\pi_{ijk}, \quad i=1,\ldots,I, \quad j=1,\ldots,J, \quad k=1,\ldots,K$$

$$onde:$$

$$\sum_i \sum_j \sum_k \pi_{ijk} = 1$$

- Modelos log-lineares podem representar vários padrões de independência e associação.
- Termos de associação de dois fatores descrevem a odds ratio condicional entre as variáveis.

Tipos de Independência

- **Independência Mútua**:
  - As três variáveis são mutuamente independentes quando:

$$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}, \; para\ todo\ i, j\ e\ k$$

  - Na escala logarítmica independência mútua é equivalente ao seguinte modelo log-linear:

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

- **Independência Conjunta**:
  - A variável Y é conjuntamente independente de ***X*** e ***Z*** quando:

$$\pi_{ijk} = \pi_{i+k}\pi_{+j+}, \text{ para todo } i, j \text{ e } k$$

  - O modelo log-linear equivalente é dado por:

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}$$

  - Definições similares aplica-se para X ser conjuntamente independente das variáveis Y e para Z ser conjuntamente independente de X e Y.

- **Independência Marginal**:
  - As variáveis X e Y são marginalmente independentes se:

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

$$i = 1, 2, ..., I \qquad j = i = 1, 2, ..., J$$

Tipos de Independências

- Se X e Y são independentes na tabela parcial para a k-ésima categoria de Z, então X e Y são condicionalmente independentes no k-ésimo nível de Z.
- Seja $\pi_{ij|k} = \pi_{ijk} / \pi_{++k}, i = 1, \ldots, I, j = 1, \ldots, J$ a distribuição conjunta de X e Y no k-ésimo nível de Z. Então independência condicional no k-ésimo nível de Z é dado por:

$$\pi_{ij|k} = \pi_{i+|k}\pi_{+j|k} \quad \text{para todo } i \text{ e } j$$

Independência Condicional e Marginal

- Vimos que associação parcial pode ser bem diferente da associação marginal. Além disso, vimos que independência condicional de X e Y dado Z não implica em independência marginal de X e Y.
- Ambas as independências condicional e marginal se verificam se um dos tipos de independências estudas anteriormente se aplicam.

Ausência de Interação entre os Três Fatores

- Nos modelos log-lineares descritos anteriormente existem três, dois e um par de variáveis condicionalmente independentes.
- O termo λ_{ij}^{XY} indicam que as variáveis são condicionalmente dependentes.
- Um modelo que permita que todas as três variáveis sejam condicionalmente dependentes é dado por:

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

Modelos Log-Lineares para Tabelas Tridimensionais - Modelos Hierárquicos

- A soma dos parâmetros para qualquer índice é igual a zero. Isto é,

$$\sum_i \lambda_i^X = \sum_j \lambda_j^Y = \sum_k \lambda_k^Z = \sum_i \lambda_{ij}^{XY} = \sum_j \lambda_{ij}^{XY} = \cdots = \sum_k \lambda_{ijk}^{XYZ} = 0$$

- O modelo log-linear geral para uma tabela tridimensional é:

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

Modelos Log-Lineares para Tabelas Tridimensionais - Modelos Hierárquicos

- Os preditores lineares para os modelos tem uma estrutura análoga aos modelos de fatoriais de análise de variância.

- O termo com índice simples são análogos ao efeitos principais, os termos com duplo índice são análogos a interação entre dois fatores.

Modelos Log-Lineares para Tabelas Tridimensionais - Modelos Hierárquicos

- Alguns modelos log-lineares para tabelas tridimensionais

Modelo Log – linear	*Símbolo*
$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$	(X, Y, Z)
$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$	(XY, Z)
$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$	(XY, YZ)
$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ}$	(XY, YZ, XZ)
$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ} + \lambda_{ijk}^{XYZ}$	(XYZ)

Modelos Log-Lineares para Tabelas Tridimensionais - Modelos Hierárquicos

- Os modelos log-lineares apresentados na tabela anterior são chamados de modelos hierárquicos. Isto significa se o modelo contém efeitos de ordem maior ele deve também conter efeitos de ordem menor. Por exemplo, quando o modelo contém λ_{ij}^{XY} ele deve conter λ_i^X e λ_j^Y .

Modelos Log-Lineares para Tabelas Tridimensionais - Interpretação dos Parâmetros

- Para interpretar modelos log-lineares, descreveremos as associações marginais e parciais usando as razões de chances. A tabela marginal $\{\pi_{ij+}\}$ usa um conjunto de (I-1) (J-1) razões de chances tal que.

$$\theta_{ij}^{XY} = \frac{\pi_{ij+}\pi_{i+1,j+1,+}}{\pi_{i+1,j,+}\pi_{i,j+1,+}}, 1 \le i \le I-1, 1 \le j \le J-1$$

Modelos Log-Lineares para Tabelas Tridimensionais - Interpretação dos Parâmetros

- Para interpretar modelos log-lineares, descreveremos as associações marginais e parciais usando as razões de chances. A tabela marginal $\{\pi_{ij+}\}$ usa um conjunto de (I-1) (J-1) razões de chances tal que.

$$\theta_{ij}^{XY} = \frac{\pi_{ij+}\pi_{i+1,j+1,+}}{\pi_{i+1,j,+}\pi_{i,j+1,+}}, 1 \le i \le I-1, 1 \le j \le J-1$$

Modelos Log-Lineares para Tabelas Tridimensionais - Interpretação dos Parâmetros

- Para um nível fixo k de Z, a razão de chances correspondente é dada por:

$$\theta_{ij(k)}^{XY} = \frac{\pi_{ijk}\pi_{i+1,j+1,k}}{\pi_{i,j+1,k}\pi_{i+1,j,k}}, 1 \le i \le I-1, 1 \le j \le J-1$$

- Descreve a associação condicional X e Y. Similarmente, as associações condicionais Xe Z é descrita por (I-1)(K-1) razões de chances $\{\theta_{i(j)k}\}$ para cada nível de J de Y

Tridimensionais - Interpretação dos Parâmetros

- E as associações condicionais Y e Z é descrita por (J-1)(K-1) razões de chances $\{\theta_{(i)jk}\}$ para cada nível de I de X .
- Os parâmetros do modelo log-linear são funções das razões de chances condicionais. As relações são mais simples para uma variável binária. Para tabelas 2x2x2, substituindo-se o modelo saturado na razão de chances condicional obteremos:

$$\lambda_{111}^{XYZ} = \frac{1}{8}\log\left(\frac{\theta_{11(1)}}{\theta_{11(2)}}\right) = \frac{1}{8}\log\left(\frac{\theta_{1(1)1}}{\theta_{1(2)1}}\right) = \frac{1}{8}\log\left(\frac{\theta_{(1)11}}{\theta_{(2)11}}\right)$$

Tridimensionais - Interpretação dos Parâmetros

- Para a restrição de soma zero sobre $\{\lambda_{ijk}^{XYZ}\}$. Cada λ_{ijk}^{XYZ} é igual a zero quando a razão de chances entre as duas variáveis é a mesma para cada nível da terceira variável. E portanto,

$$\lambda_{11}^{XY} = \frac{1}{4}\log\theta_{11(k)}$$