

Modelos Loglineares para Tabelas de Contingência de Dimensões Superiores

- Os conceitos básicos para modelos loglineares com tabelas de contingência de três dimensões podem ser estendidos para tabelas de múltiplas dimensões ou dimensões de ordem superiores.
- Modelos loglineares que têm:
 - termos de ordem mais alta de dois fatores têm associações homogêneas: duas variáveis condicionalmente associadas têm as mesmas razões de chance em cada combinação de níveis das outras variáveis.
 - A ausência de um termo de dois fatores implica independência condicional para essas variáveis.
 - Um termo de três fatores permite associações condicionais heterogêneas, em que a associação entre qualquer par dessas três variáveis varia entre as categorias da terceira variável, em quaisquer configurações das outras variáveis.

EXEMPLO - Acidentes

- Os dados apresentados a seguir correspondem a 68694 passageiros de automóveis e pequenos caminhões envolvidos no estado de Maine em 1991. A tabela classifica os passageiros por :
 - G** – gênero (**m** – masculino e **f** - feminino)
 - L** – localização do acidente (**u** –urbana e **r** –rural)
 - S** – uso de cinto de segurança (**s**-sim e **n**-não)
 - I** – ferimento (**s**-sim e **n**-não)

Gênero	Local	Cinto de Segurança	Ferimentos		Proporção de Passageiros.c/ Ferimentos
			Não	Sim	
Feminino	Urbana	Não	7287	996	0,12
		Sim	11587	759	0,06
	Rural	Não	3246	973	0,23
		Sim	6134	757	0,11
Masculino	Urbana	Não	10381	812	0,07
		Sim	10969	380	0,03
	Rural	Não	6123	1084	0,15
		Sim	6693	513	0,07

Para investigar a complexidade de modelos necessários, vamos considerar os modelos:

- (**G, L, S, I**) contendo os termos com um único fator;
- (**GL, GS, GI, LS, LI, SI**) contendo os termos com um único fator e todos os termos com dois fatores
- (**GLS, GLI, GSI, LSI**) contendo os termos com um fator, com dois fatores e todos os termos com três fatores

Modelo Saturado:

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{il}^{XW} + \lambda_{jk}^{YZ} + \lambda_{jl}^{YW} + \lambda_{kl}^{ZW} + \lambda_{ijk}^{XYZ} + \lambda_{ijl}^{XYW} + \lambda_{ikl}^{XZW} + \lambda_{jkl}^{YZW} + \lambda_{ijkl}^{XYZW}$$

Maximum Likelihood Analysis of Variance			
Source	DF	Chi-Square	Pr > ChiSq
genero	1	1.94	0.1640
local	1	287.69	<.0001
genero*local	1	85.50	<.0001
cintoseg	1	49.03	<.0001
genero*cintoseg	1	299.04	<.0001
local*cintoseg	1	4.51	0.0336
genero*local*cintoseg	1	1.74	0.1865
dano	1	26365.03	<.0001
genero*dano	1	383.56	<.0001
local*dano	1	728.70	<.0001
genero*local*dano	1	2.93	0.0867
cintoseg*dano	1	851.26	<.0001
genero*cintoseg*dano	1	0.06	0.7989
local*cintoseg*dano	1	2.42	0.1197
genero*local*cintoseg*dano	1	1.32	0.2497
Likelihood Ratio	0	.	.

Modelo com todas as Interações Triplas: (GLS, GLI, GSI, LSI)

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{il}^{XW} + \lambda_{jk}^{YZ} + \lambda_{jl}^{YW} + \lambda_{kl}^{ZW} + \lambda_{ijk}^{XYZ} + \lambda_{ijl}^{XYW} + \lambda_{ikl}^{XZW} + \lambda_{jkl}^{YZW} + \lambda_{ijkl}^{XYZW}$$

Maximum Likelihood Analysis of Variance			
Source	DF	Chi-Square	Pr > ChiSq
genero	1	1.88	0.1707
local	1	298.09	<.0001
genero*local	1	85.85	<.0001
cintoseg	1	48.23	<.0001
genero*cintoseg	1	298.44	<.0001
local*cintoseg	1	3.81	0.0510
genero*local*cintoseg	1	14.79	0.0001
dano	1	26403.79	<.0001
genero*dano	1	382.97	<.0001
local*dano	1	738.01	<.0001
genero*local*dano	1	2.24	0.1349
cintoseg*dano	1	852.66	<.0001
genero*cintoseg*dano	1	0.04	0.8382
local*cintoseg*dano	1	3.05	0.0808
Likelihood Ratio	1	1.33	0.2496

Modelo com todas as Interações Duplas: (GL, GS, GI, LS, LI, SI)

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{il}^{XW} + \lambda_{jk}^{YZ} + \lambda_{jl}^{YW} + \lambda_{kl}^{ZW} + \lambda_{ijk}^{XYZ} + \lambda_{ijl}^{XYW} + \lambda_{ikl}^{XZW} + \lambda_{jkl}^{YZW} + \lambda_{ijkl}^{XYZW}$$

Maximum Likelihood Analysis of Variance			
Source	DF	Chi-Square	Pr > ChiSq
genero	1	1.41	0.2343
local	1	308.40	<.0001
genero*local	1	169.50	<.0001
cintoseg	1	53.42	<.0001
genero*cintoseg	1	860.14	<.0001
dano	1	27211.68	<.0001
genero*dano	1	394.37	<.0001
local*dano	1	784.96	<.0001
cintoseg*dano	1	868.68	<.0001
Likelihood Ratio	5	23.35	0.0003

Modelo com todos os termos de um fator: (G, L, S, I)

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_l^W$$

Maximum Likelihood Analysis of Variance			
Source	DF	Chi-Square	Pr > ChiSq
genero	1	395.29	<.0001
local	1	4435.96	<.0001
cintoseg	1	688.74	<.0001
dano	1	30092.43	<.0001
Likelihood Ratio	11	2792.77	<.0001

Modelo

G²

GI

p-value

- 1 (G, L, S, I)
- 2 (GL, GS, GI, LS, LI, SI)
- 3 (GLS, GLI, GSI, LSI)

1.33

Hipóteses:

H₀) O modelo (GLS, GLI, GSI, LSI) se ajusta, ou seja,

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{il}^{XW} + \lambda_{jk}^{YZ} + \lambda_{jl}^{YW} + \lambda_{kl}^{ZW} + \lambda_{ijk}^{XYZ} + \lambda_{ijl}^{XYW} + \lambda_{ikl}^{XZW} + \lambda_{jkl}^{YZW} + \lambda_{ijkl}^{XYZW}$$

H₁) O modelo saturado se ajusta, ou seja,

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{il}^{XW} + \lambda_{jk}^{YZ} + \lambda_{jl}^{YW} + \lambda_{kl}^{ZW} + \lambda_{ijk}^{XYZ} + \lambda_{ijl}^{XYW} + \lambda_{ikl}^{XZW} + \lambda_{jkl}^{YZW} + \lambda_{ijkl}^{XYZW}$$

Temos que deviance é

$$G_2 = 1,33$$

e considerando uma distribuição qui-quadrado com $v = 16 - 15 = 1$ gl temos que:

$$p - \text{valor} = P(\chi^2 \geq 1,33) = 0,2496 \Rightarrow \text{Não rejeitar } H_0$$

Portanto, o modelo de associação homogênea (GLS, GLI, GSI, LSI) se ajusta.

- Investigar os resultados da adição de um único termo de três fatores ao modelo (GL,GS,LS,GI,LI,SI).

- Os quatros modelos possíveis são:

- (GLS, GI, IL, IS)
- (GIL, GS, IS, LS)
- (GIS, GL, LS, IL)
- (ILS, GI, GL, GS)

Modelo (GLS, GL, LI, SI)

Maximum Likelihood Analysis of Variance				
Source	DF	Chi-Square	Pr > ChiSq	
genero	1	1.44	0.2298	
local	1	314.48	<.0001	
genero*local	1	179.89	<.0001	
cintoseg	1	52.01	<.0001	
genero*cintoseg	1	857.25	<.0001	
local*cintoseg	1	32.51	<.0001	
genero*local*cintoseg	1	15.87	<.0001	
dano	1	27202.65	<.0001	
genero*dano	1	399.28	<.0001	
local*dano	1	789.90	<.0001	
cintoseg*dano	1	873.28	<.0001	
Likelihood Ratio	4	7.46	0.1133	

Modelo (GLS, GL, LI, SI, I)

Maximum Likelihood Analysis of Variance				
Source	DF	Chi-Square	Pr > ChiSq	
genero	1	1.94	0.1638	
local	1	293.94	<.0001	
genero*local	1	90.88	<.0001	
dano	1	27074.47	<.0001	
genero*dano	1	395.73	<.0001	
local*dano	1	786.90	<.0001	
genero*local*dano	1	4.78	0.0289	
cintoseg	1	53.39	<.0001	
genero*cintoseg	1	860.11	<.0001	
local*cintoseg	1	27.47	<.0001	
cintoseg*dano	1	868.72	<.0001	
Likelihood Ratio	4	18.57	0.0010	

Modelo (GLS, GL, LI, SI, I, S)

Maximum Likelihood Analysis of Variance				
Source	DF	Chi-Square	Pr > ChiSq	
genero	1	1.77	0.1839	
cintoseg	1	53.55	<.0001	
genero*cintoseg	1	293.92	<.0001	
dano	1	26597.79	<.0001	
genero*dano	1	379.78	<.0001	
cintoseg*dano	1	855.10	<.0001	
genero*cintoseg*dano	1	0.50	0.4778	
local	1	308.42	<.0001	
genero*local	1	169.51	<.0001	
local*cintoseg	1	27.51	<.0001	
local*dano	1	785.02	<.0001	
Likelihood Ratio	4	22.85	0.0001	

Modelo (GLS, GL, LI, SI, I, S, L)

Maximum Likelihood Analysis of Variance				
Source	DF	Chi-Square	Pr > ChiSq	
local	1	305.94	<.0001	
cintoseg	1	51.06	<.0001	
local*cintoseg	1	3.12	0.0772	
dano	1	27203.37	<.0001	
local*dano	1	737.76	<.0001	
cintoseg*dano	1	866.00	<.0001	
local*cintoseg*dano	1	2.72	0.0992	
genero	1	1.42	0.2333	
genero*local	1	169.51	<.0001	
genero*cintoseg	1	860.14	<.0001	
genero*dano	1	394.36	<.0001	
Likelihood Ratio	4	20.63	0.0004	

Modelo	G ²	GI	p-value	Diferença	GI	P-value
2 (GI, GL, GS, IL, IS, LS)	23.35	5	0.0003			
a (GLS, GI, LI, SI)						
b (GLI, GS, LS, SI)						
c (GSI, GL, LS, LI)						
d (LSI, GL, GS, GI,)						
3 (GIL, GIS, GLS, ILS)	1.33	1	0.2496			

Assim, por exemplo, ao comparar o modelo (GLS, GI, LI, SI), que mostrou se ajustar aos dados, com o modelo com todos os três fatores temos:

• Hipóteses:

H_0) O modelo (GLS, GI, LI, SI) se ajusta, ou seja,
 $\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_l^W + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{il}^{XW} + \lambda_{jk}^{YZ} + \lambda_{jl}^{YW} + \lambda_{kl}^{ZW} + \lambda_{ijk}^{XYZ} (M_0)$

H_1) O modelo (GLS, GLI, GSI, LSI) se ajusta, ou seja,
 $\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_l^W + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{il}^{XW} + \lambda_{jk}^{YZ} + \lambda_{jl}^{YW} + \lambda_{kl}^{ZW} + \lambda_{ijk}^{XYZ} + \lambda_{ijl}^{XYW} + \lambda_{ikl}^{XZW} + \lambda_{jkl}^{YZW} (M_1)$

Temos que a deviance parcial é

$$\text{Deviance parcial} = -2[L_0 - L_1] = \text{deviance}_0 - \text{deviance}_1 = 7,46 - 1,33 = 6,13$$

e considerando uma distribuição qui-quadrado com

$$v = gl_1 - gl_0 = 15 - 12 = 3 \text{ gl}$$


temos que:

$$p - \text{valor} = P(\chi^2 \geq 6,13) = 0,1054 \Rightarrow \text{Não rejeitar } H_0$$

Portanto, o modelo M_0 – (GLS, GI, LI, SI) se ajusta melhor que o modelo M_1 .

■ Modelo. (GLS, GI, LI, SI)

- cada par de variáveis é condicionalmente dependente e, em cada categoria de I, a associação entre G e L ou entre G e S ou entre L e S varia entre as categorias da variável restante .
- Como I não ocorre em um termo de três fatores, a razão de chances condicional entre I e cada variável é a mesma em cada combinação de categorias das outras duas variáveis.

- 
- Como I não ocorre em um termo de três fatores, a razão de chances condicional entre I e cada variável é a mesma em cada combinação de categorias das outras duas variáveis.
 - a razão de chances condicional SI estimada é

$$\lambda_{kl}^{SI} = -0,2035 \Rightarrow \hat{\theta}_{SI|GL} = \exp(4 \times \lambda_{kl}^{SI}) = \exp(4 \times -0,2035) = \mathbf{0,4431}$$

- *As chances de ferimentos para os passageiros que usavam cinto de segurança eram menos da metade das chances para os passageiros que não os usavam, para cada combinação de gênero e localização.*