**Modelos Log-Lineares
para
Análise de Dados Categorizados**

Isildinha Marques dos Reis

Dissertação apresentada ao Instituto de Matemática e
Estatística da Universidade de São Paulo para obtenção
do grau de mestre em Estatística

Área de Concentração: Estatística

Orientador: Prof. Dr. Julio da Motta Singer

São Paulo, Fevereiro de 1989

Aos meus pais Pompilio e Angelina

e ao meu companheiro Bilu

Agradecimentos

*“ . . . e aprendi que se depende sempre
de tanta muita diferente gente,
toda pessoa sempre é as marcas
das lições diárias de outras tantas pessoas . . . ”*

Luiz Gonzaga Jr.

Ao Prof. Dr. Julio da Motta Singer, nosso orientador.

Aos professores Carlos Daniel Mimoso Paulino, Maria Cristina Martín e
Antonio Carlos Pedroso de Lima.

Aos companheiros do Departamento de Epidemiologia da Faculdade de Saúde
Pública da USP, em especial às professoras Nilza e Maria Regina.

E a todos os amigos que nos incentivaram.

Índice

Capítulo 1 - Introdução	1
Capítulo 2 - Modelos Estruturais Log-Lineares	18
2.1 - Modelos log-lineares para dados com distribuição Multinomial	19
2.2 - Modelos log-lineares para dados com distribuição Produto de Multinomiais	31
2.3 - Relação entre as diferentes formulações para modelos estruturais log-lineares	38
2.4 - Modelos log-lineares generalizados	49
2.5 - Uma classificação para modelos estruturais log-lineares	57
Capítulo 3 - Métodos de Estimação e Testes de Hipóteses	65
3.1 - O método de Máxima Verossimilhança	66
3.2 - O método de Mínimos Quadrados Generalizados	75
Capítulo 4 - Aplicações Práticas	81
4.1 - Pacotes computacionais para ajuste de modelos estruturais log-lineares	81
4.2 - Exemplos ilustrativos	84
Referências Bibliográficas	108
Anexo A - Programas LOGLIN.AJU e LOGLIN.HIP	114
Anexo B - Listagens de comandos de programas utilizados nos exemplos	128

Capítulo 1

Introdução

Em diversas áreas do conhecimento, como Economia, Sociologia, Psicologia, Biologia, Medicina, Epidemiologia ou Saúde Pública, é comum a coleta de dados multivariados envolvendo variáveis categorizadas cujos níveis ou categorias são empregados na descrição ou identificação das unidades amostrais em estudo. As diferentes formas de coleta podem englobar:

- 1) dados históricos provenientes de uma certa população de interesse;
- 2) dados de levantamentos amostrais com seleção aleatória;
- 3) dados experimentais com alocação aleatória de unidades experimentais a tratamentos;
- 4) combinação das duas formas anteriores (2 e 3).

Quanto à apresentação, esse tipo de dados pode ser expresso na forma de matrizes de dados individuais ou tabelas de contingência multidimensionais. Na forma de matriz, cada linha está associada a uma unidade amostral e cada coluna a uma variável; em geral, uma coluna é utilizada para identificar as unidades amostrais. Essa informação, irrelevante para a análise estatística, pode ser útil na verificação de algum dado que possa parecer duvidoso. Em uma tabela de contingência cada variável corresponde a uma dimensão; a frequência apresentada em cada casela (ou cela) da tabela é a contagem das unidades amostrais com mesma descrição (perfil multivariado) em termos das variáveis categorizadas envolvidas. A tabela de contingência é uma forma condensada de apresentação dos dados. Como exemplo consideremos os dados analisados por Benício(1983), correspondentes à 20.878 nascimentos ocorridos em 31 maternidades do município de São Paulo em 1978. Cada nascimento foi classificado segundo cinco variáveis categorizadas: Peso do Recém Nascido (PRN), Status em relação à Assitência Pré-Natal (APN), Nível de Escolaridade da mãe (NE), Peso Pré-Gestacional da mãe (GPP) e Hábito de Fumar da mãe (HF). Esses dados poderiam ser apresentados em uma matriz de dados individuais como a esquematizada na Tabela 1.1 .

Tabela 1.1 - Dados individuais de 20.878 nascimentos ocorridos no município de São Paulo em 1978

Nascimento	PRN	APN	NE	GPP	HF
1	$\leq 2,5\text{kg}$	ausente	baixo	$<50\text{kg}$	sim
2	$\leq 2,5\text{kg}$	ausente	alto	$<50\text{kg}$	não
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
20.878	$> 2,5\text{kg}$	presente	alto	$\geq 50\text{kg}$	não

Uma forma alternativa para a apresentação desses dados é a tabela de contingência abaixo.

Tabela 1.2 - Classificação de 20.878 nascimentos ocorridos no município de São Paulo em 1978

APN	NE	GPP	HF	PRN		Total	
				≤ 2,5kg	> 2,5kg		
Ausente	Baixo	<50kg	Sim	67	249	316	
			Não	36	261	297	
		≥50kg	Sim	113	518	631	
			Não	45	721	766	
	Alto	<50kg	Sim	10	29	39	
			Não	8	23	31	
		≥50kg	Sim	5	47	52	
			Não	4	71	75	
	Presente	Baixo	<50kg	Sim	202	1.325	1.527
				Não	179	2.001	2.180
≥50kg			Sim	219	2.655	2.874	
			Não	249	5.665	5.914	
Alto		<50kg	Sim	59	525	584	
			Não	65	991	1.056	
		≥50kg	Sim	77	1.408	1.485	
			Não	93	2.958	3.051	
Total				1.431	19.447	20.878	

Quanto à escala de mensuração, as variáveis categorizadas podem ser classificadas como ORDINAIS ou NOMINAIS conforme suas categorias sejam ordenáveis ou não. Além disso, variáveis quantitativas (contínuas ou discretas) podem ser transformadas em variáveis categorizadas; por exemplo, em alguns estudos pode ser de interesse classificar indivíduos como hipertensos ou não hipertensos ao invés de utilizar os valores da pressão arterial medidos em mmHg, ou ainda, classificá-los como sendo de nível sócio-econômico baixo, médio ou alto em função de sua renda total.

Uma outra dimensão de interesse na classificação das variáveis que definem uma tabela de contingência ou as colunas de uma matriz de dados individuais é a distinção entre variáveis RESPOSTAS e variáveis EXPLICATIVAS. As variáveis RESPOSTAS (variáveis dependentes) são aleatórias e descrevem o que foi observado ou ocorreu com as unidades amostrais durante o experimento ou a pesquisa em consideração. As variáveis EXPLICATIVAS (variáveis independentes ou fatores) são fixas (por planejamento ou condicionamento) e as combinações de seus níveis definem estratos, populações ou subpopulações às quais as unidades amostrais pertencem. A classificação de uma variável como variável resposta ou variável explicativa depende fundamentalmente do esquema de amostragem e do objetivo da pesquisa. Em função dessa classificação podemos definir dois tipos básicos de tabelas de contingência (ou matriz de dados individuais):

I) tabelas em que todas as variáveis são respostas;

II) tabelas em que algumas variáveis são explicativas e outras são respostas.

Com o tipo I, o interesse frequentemente é dirigido para o estudo da associação entre as diversas variáveis respostas, i.e., para a verificação da existência de algum grau de dependência estocástica entre essas variáveis. O termo “associação” neste caso, tem significado semelhante ao termo “correlação”, empregado no estudo da relação entre variáveis estocásticas quantitativas (ver Bhapkar & Koch (1968a) por exemplo). Quando existe associação entre as variáveis respostas, em geral é de interesse avaliar o grau da associação e nesse sentido diversas medidas de associação — definidas por meio de funções apropriadas das probabilidades correspondentes às diversas caselas — tem sido propostas; por exemplo, coeficientes de associação, medidas de concordância como a estatística Kappa, quocientes de probabilidades (“odds”) ou funções de quocientes de probabilidades como a razão de produtos cruzados (“cross-product ratio” ou “odds ratio”) (ver Goodman & Kruskal (1954, 1959, 1963, 1972), Altham (1970a, 1970b) e Bishop et al. (1975, cap. 11) por exemplo).

Em tabelas do tipo II, o objetivo, em geral, é estudar os efeitos das variáveis explicativas na distribuição conjunta das variáveis respostas ou sobre alguma característica específica dessas distribuições, como uma medida de associação entre as variáveis respostas (ver Bhapkar & Koch (1968a) por exemplo). Um caso particular de grande importância é aquele em que se considera apenas uma variável como resposta e as demais como variáveis explicativas. Neste caso, como em Análise de Variância (ANOVA), o objetivo é estudar a influência isolada ou combinada das variáveis explicativas na distribuição da variável resposta, através de comparações entre as distribuições de probabilidades correspondentes às subpopulações (definidas através da combinação dos níveis das variáveis explicativas). Nessa situação, em geral o interesse recai sobre alguma característica específica das distribuições, como a resposta média se a variável resposta for ordinal e se for possível associar pesos ("scores") às diferentes categorias, a proporção de indivíduos com determinado perfil de resposta ou alguma outra função das probabilidades correspondentes.

Se a análise visar propósitos inferenciais é necessário considerar modelos probabilísticos para os dados. Nesse sentido, os principais modelos probabilísticos que podem ser associados às tabelas de contingência são: Produto de distribuições de Poisson, distribuição Multinomial, Produto de distribuições Multinomiais, distribuição Hipergeométrica, Produto de distribuições Hipergeométricas e distribuição Binomial Negativa.

Quando todas as frequências da tabela de contingência são variáveis aleatórias, um modelo comumente empregado é aquele que associa uma distribuição de Poisson a cada contagem. Em termos gerais, a frequência em cada casela (variável resposta) refere-se à observação de algum evento raro em uma certa medida de exposição fixada a priori, como, tempo, volume, área, etc ... Por hipótese, o número de ocorrências do evento em cada uma das caselas é gerado por processos de Poisson independentes. Nesta situação não se tem conhecimento a priori do total de observações a ser obtido e um modelo global para a tabela é um Produto de distribuições de Poisson. As caselas são definidas pela com-

binação dos níveis das demais “variáveis” em estudo, tratadas como variáveis explicativas. Consequentemente, as caselas representam as subpopulações de interesse.

Como um exemplo considere a Tabela 1.3 apresentada e analisada por Koch et al. (1985, p. 168). Os dados epidemiológicos se referem a casos novos de Melanoma, ocorridos entre 1969 e 1971, para indivíduos do sexo masculino e raça branca segundo diferentes faixas etárias e região de residência nos EUA (norte ou sul). São também apresentadas as estimativas da população exposta ao risco e da taxa de incidência desse tipo de câncer para cada casela (subpopulação). Um dos objetivos dessa pesquisa era estudar a influência da faixa etária e da região de residência na taxa de incidência de Melanoma. A identificação de um caso novo de Melanoma pode ser considerada como um evento raro; as distintas subpopulações, definidas pela combinação dos níveis de idade e região de residência, podem ser supostas independentes. Esses dois aspectos aliados ao objetivo da pesquisa permitem associar aos dados um modelo probabilístico correspondente a um Produto de distribuições de Poisson, com 12 subpopulações.

Tabela 1.3 - Número de casos novos de Melanoma para indivíduos do sexo masculino e raça branca no período de 1969 a 1971 nos Estados Unidos, estimativa da população exposta ao risco e da taxa de incidência

Idade (anos)	Região de residência	No. de casos de Melanoma	População em risco	Taxa de incidência (x 1.000.000)
< 35	Norte	61	2.880.262	21,2
< 35	Sul	64	1.074.246	59,6
35 a 44	Norte	76	564.535	134,6
35 a 44	Sul	75	220.407	340,3
45 a 54	Norte	98	592.983	165,3
45 a 54	Sul	68	198.119	343,2
55 a 64	Norte	104	450.740	230,7
55 a 64	Sul	63	134.084	469,9
65 a 74	Norte	63	270.908	232,6
65 a 74	Sul	45	70.708	636,4
≥ 75	Norte	80	161.850	494,3
≥ 75	Sul	27	34.233	788,7

A distribuição Multinomial constitui um modelo probabilístico adequado para situações onde apenas o total geral da tabela é fixado no planejamento da pesquisa. Tudo se passa como se uma única amostra aleatória de tamanho igual ao “total geral” da tabela tenha sido selecionada de uma população supostamente infinita, com probabilidades constantes (e positivas) para toda unidade populacional com as mesmas características. O modelo Multinomial também pode ser derivado do modelo Poisson por condicionamento na soma das frequências das caselas (total geral da tabela); ver Bishop et al. (1975, cap. 3 e 13), Haberman (1978, p. 61–63, 134–135, 143) ou Plackett (1974), por exemplo.

Sob o modelo Multinomial as variáveis categorizadas envolvidas são todas consideradas como variáveis respostas. Cada unidade amostral é classificada de maneira única em uma das distintas categorias de resposta, indexadas por $j = 1, \dots, r$, obtidas pela combinação dos níveis dessas variáveis. No caso de mais de uma variável resposta, cada categoria de resposta j corresponde a um perfil multivariado de respostas. Esse tipo de delineamento é conveniente para verificar hipóteses de independência, associação parcial ou associação condicional entre as variáveis envolvidas (ver Everitt (1977, p. 74–75) por exemplo).

Para ilustração, considere a Tabela 1.2 apresentada anteriormente. Um dos objetivos daquela pesquisa foi estudar a intensidade da relação (associação) entre as variáveis citadas. Em função desse objetivo, todas as variáveis podem ser encaradas como respostas e um modelo Multinomial com 32 categorias de resposta pode ser associado àquela tabela.

Para outro exemplo considere os dados (hipotéticos) da Tabela 1.4, onde 50 pessoas adultas selecionadas aleatoriamente são classificadas segundo Sexo (masculino ou feminino) e Hábito de Fumar (“fumante” ou “não fumante”). Suponha que estamos interessados em saber se há associação entre Sexo e Hábito de Fumar. Neste exemplo, as duas variáveis envolvidas são respostas e um modelo probabilístico adequado é o modelo Multinomial com $n = 50$ e 4 categorias de resposta, definidas pela combinação dos níveis dessas variáveis.

Tabela 1.4 - Distribuição de pessoas adultas segundo sexo e hábito de fumar

Sexo	Hábito de fumar		Total
	Sim	Não	
Masculino	12	15	27
Feminino	10	13	23
Total	22	28	50

Alternativamente, observe a Tabela 1.5. Os dados analisados por Grizzle, Starmer & Koch (1969), referem-se à comparação de 4 tipos de cirurgia para o tratamento de úlcera duodenal. Os pacientes submetidos a cada tipo de cirurgia são classificados segundo a severidade de uma sequela das mesmas. As diferentes cirurgias se caracterizam pela porcentagem de tecido estomacal removida e a severidade da sequela descreve a extensão de um possível efeito da cirurgia. Considerando como objetivo do estudo a análise de uma possível associação entre severidade da sequela e tipo de cirurgia, apenas o total geral da tabela deve ser tomado como fixo e um modelo Multinomial, com 12 categorias de resposta, pode ser associado a esses dados.

Tabela 1.5 - Número de pacientes segundo tipo de cirurgia e severidade de uma sequela

Cirurgia	Severidade da sequela			Total
	Nenhuma	Pouca	Moderada	
V+D (0%)	61	28	7	96
V+A (25%)	68	23	13	104
V+H (50%)	58	40	12	110
GR (75%)	53	38	16	107
Total	240	129	48	417

Se o esquema de amostragem adotado em uma determinada pesquisa for tal que fixe a priori alguns totais marginais relativos a combinações de níveis de algumas variáveis, um modelo amplamente empregado corresponde à distribuição Produto de Multinomiais. As variáveis para as quais foram fixados os totais marginais são consideradas variáveis explicativas; as combinações de seus níveis definem subpopulações independentes. As demais variáveis são consideradas variáveis respostas e as combinações de seus níveis geram as categorias de respostas.

Como exemplo, mostramos na Tabela 1.6 dados apresentados por Bartlett (1935). Esses dados são de um experimento conduzido para investigar o crescimento de plantas submetidas a diferentes tratamentos. As variáveis envolvidas são dicotômicas: os níveis da variável *A* correspondem à sobrevivência da planta (“sobrevive” ou “não sobrevive”), os níveis da variável *B* referem-se à época do plantio (“na primavera” ou “outras estações”) e os da variável *C* ao comprimento do corte (“longo” ou “curto”). As diferentes combinações dos níveis das variáveis *B* e *C* definem 4 subpopulações para as quais foi observada a distribuição da variável *A*, considerada como resposta. Uma distribuição Produto de Multinomiais (Binomiais, no caso) com 4 subpopulações e 2 categorias de resposta pode ser considerada como um modelo adequado para a geração das frequências dessa tabela.

Tabela 1.6 - Número de plantas

Comprimento do corte	Época do plantio	Sobrevivência da planta		Total
		Sobrevive	Não sobrevive	
Longo	Outras estações	156	84	240
Longo	Na primavera	84	156	240
Curto	Outras estações	107	133	240
Curto	Na primavera	31	209	240

Os dados apresentados na Tabela 1.7, retirados de Koch et al. (1985, p. 121), são de um estudo sobre infecções urinárias. Pacientes com diagnóstico complicado ou não complicado, submetidos a 3 tratamentos (A, B e C) são observados com relação à cura. Questões de interesse incluem i) a verificação da existência de interação entre tratamento e tipo de diagnóstico relativamente à proporção de curados; ii) a comparação dos tratamentos para cada tipo de diagnóstico. O processo de obtenção dos dados e os objetivos sugeridos acima permitem associar aos dados um modelo probabilístico Produto de Multinomiais com 6 subpopulações, obtidas pela combinação dos níveis de Tipo de diagnóstico e Tratamento, e 2 categorias de respostas, “curado” e “não curado”.

Tabela 1.7 - Número de pacientes

Tipo de diagnóstico	Tratamento	Curados	Não curados
Complicado	A	78	20
	B	101	11
	C	68	46
Não complicado	A	40	5
	B	54	5
	C	34	6

Os dados mostrados na Tabela 1.8 foram obtidos de um estudo sobre arteriosclerose (Strong et al. (1969)) apresentado por Williams & Grizzle (1971, p. 38–43). Essa tabela de contingência é um exemplo com duas variáveis respostas dicotômicas, Ocorrência de Infarto e Sequela Miocárdica, e duas variáveis explicativas, Idade (2 categorias) e a combinação de Localidade e Raça (3 categorias). Os totais marginais correspondentes às combinações das categorias das duas variáveis explicativas foram fixados no planejamento da pesquisa. Suponha que seja de interesse estudar os efeitos das variáveis explicativas em relação à razão

de produtos cruzados (*), tomada como medida de associação. Neste sentido um modelo probabilístico adequado é a distribuição Produto de Multinomiais com 6 subpopulações e 4 categorias de respostas (combinação das categorias das duas variáveis respostas).

Tabela 1.8 - Casos de doença cardíaca coronariana

		New Orleans (branca)		Oslo		New Orleans (negra)	
Idade 35-54		Infarto não sim		Infarto não sim		Infarto não sim	
Sequela miocárdica	não	19	34	13	11	14	15
	sim	22	20	9	16	16	7
Idade 54-69		Infarto não sim		Infarto não sim		Infarto não sim	
Sequela miocárdica	não	21	60	15	38	4	17
	sim	39	26	66	55	17	4

(*) Em tabelas 2×2 , a razão de produtos cruzados θ ("cross-product ratio" ou "odds ratio") é definida por:

$$\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

onde π_{ij} é a probabilidade associada a casela (i, j) da tabela de contingência $(i, j = 1, 2)$. Essa medida de associação assume valores de 0 a ∞ , com:

$\theta = 1$ indicando independência e,

$\theta \neq 1$ indicando algum grau de associação, positiva se $\theta > 1$ e negativa se $\theta < 1$.

Outras considerações sobre a razão de produtos cruzados podem ser encontradas em Bishop et al. (1975, cap. 2, p. 14-15 e 18), e Fienberg (1981, p. 16-19) por exemplo.

Mesmo que a amostragem tenha sido realizada na forma mais conveniente para aplicação da distribuição Multinomial, se os objetivos envolverem a explicação do comportamento de uma variável (ou mais de uma) em função de um conjunto de variáveis explicativas, é possível resolver o problema através do condicionamento em totais marginais de interesse. Essa estratégia de condicionamento tem justificativa baseada no fato de que se as contagens da tabela foram geradas segundo o modelo Multinomial, i.e., com total geral fixado por planejamento, então a distribuição condicional das frequências dados os totais marginais de interesse corresponde a um modelo Produto de Multinomiais (ver Bishop et al. (1975, cap. 13) e Fienberg (1981, p. 167–170), por exemplo).

Por exemplo, com os dados da Tabela 1.2 poderíamos estar interessados em verificar o efeito das variáveis APN, NE, GPP e HF no Peso do Recém-Nascido (PRN). Para tal objetivo as primeiras variáveis deveriam ser consideradas como variáveis explicativas e os totais relativos às combinações de seus níveis deveriam ser considerados fixos, definindo 16 subpopulações; a variável Peso do Recém-Nascido, com dois níveis, seria a variável resposta. Neste caso, um modelo probabilístico apropriado é o modelo Produto de Multinomiais que pode ser encarado como resultante do modelo Multinomial original por condicionamento nos totais marginais.

Como um outro exemplo, os dados da Tabela 1.5 tem sido utilizados por alguns autores, Agresti (1984, p. 80–83) ou Koch et al. (1985, p. 219) por exemplo, como provenientes de um esquema Produto de Multinomiais com 4 subpopulações, designadas pelas diversas cirurgias (variável explicativa), e 3 categorias de resposta, correspondentes aos níveis da variável Severidade da sequela (variável resposta). Segundo essa abordagem, o objetivo poderia ser a comparação das 4 distribuições da variável resposta ou de alguma característica específica das mesmas, como a probabilidade de sequela com intensidade moderada.

Por outro lado, convém notar que o modelo Produto de Multinomiais não exige um modelo Multinomial original. Um modelo original mais geral é o Produto de distribuições de Poisson independentes, do qual também se obtém o modelo Multinomial por condicionamento. Além disso, os estimadores de máxima verossimilhança das frequências esperadas nas caselas são geralmente os mesmos quer o modelo probabilístico utilizado seja Poisson, Multinomial ou Produto de Multinomiais; isto decorre do fato de que tais estimadores para um determinado modelo satisfazem as restrições naturais impostas pelo outro.

Alguns planejamentos amostrais podem ainda sugerir a distribuição Hipergeométrica ou o Produto de distribuições Hipergeométricas como modelo de geração dos dados. A distribuição Hipergeométrica se caracteriza por considerar fixos todos os totais marginais da tabela de contingência; assim, todas as variáveis envolvidas são consideradas explicativas. Essa distribuição também pode ser derivada de uma distribuição Multinomial ou de uma distribuição Produto de Multinomiais ou ainda de um Produto de distribuições de Poisson, através do condicionamento nos totais marginais livres. O teste exato de Fisher para tabelas 2×2 é baseado nessa distribuição (ver Bishop et al. (1975, p. 364, 448–450) por exemplo). Em estudos epidemiológicos do tipo caso-controle, também é frequente a aplicação da distribuição Hipergeométrica; neste caso os totais marginais relativos às categorias “caso” e “controle” são fixados a priori e os totais relativos à variável de exposição são fixados a posteriori por condicionamento (ver Breslow (1980, p. 125–128), Fienberg (1981, p. 31 e seção 7.5) por exemplo).

Aspectos técnicos sobre modelos probabilísticos para dados categorizados e sobre condicionamento podem ser encontrados em Birch (1963), Plackett (1974), Haberman (1974, 1978), Bishop et al. (1975, cap. 13) e Fienberg (1981, p. 167–170) entre outros.

Diante de uma tabela de contingência (ou matriz de dados individuais) para a qual admitimos um determinado modelo probabilístico, uma estratégia de análise consiste em propor modelos estatísticos simples que permitam entender a estrutura dos dados. Esses

modelos, que denominaremos de **MODELOS ESTRUTURAIS**, devem possibilitar a verificação das hipóteses estatísticas formuladas a partir dos objetivos do estudo. Os objetivos da pesquisa, além de influírem na escolha de modelos probabilísticos adequados aos dados, podem ser utilizados na sugestão de possíveis modelos estruturais envolvendo os parâmetros desses modelos probabilísticos. Na realidade, os objetivos da pesquisa indicam relações entre os parâmetros dos modelos probabilísticos, que podem ser traduzidas através da imposição de restrições sobre os mesmos. Em geral, o objetivo da análise é encontrar um modelo estrutural simples, com parâmetros interpretáveis, e que seja suficiente para explicar uma parcela significativa da variabilidade dos dados.

Embora muitos tipos de modelos tenham sido considerados na literatura (ver Fienberg (1981), Grizzle et al. (1969), Haber (1985), Koch et al. (1985) por exemplo), uma classe de modelos estruturais que tem se revelado muito útil é a dos chamados modelos log-lineares. Referências básicas sobre modelos log-lineares são Bishop et al. (1975) e Imrey et al. (1981, 1982). Essa classe de modelos é adequada para uma variedade de esquemas amostrais (Poisson, Multinomial, Produto de Multinomiais, por exemplo). Particularmente modelos log-lineares com parâmetros e suposições semelhantes aos empregados em modelos de ANOVA são prontamente interpretáveis e seus parâmetros se relacionam com razões de produtos cruzados, medida de associação usada em Medicina, Epidemiologia, Química, Genética, etc... . Esse tipo de modelo log-linear é muito empregado quando se deseja estudar a estrutura de associação de múltiplas variáveis respostas ou quando o interesse reside no efeito conjunto e/ou isolado de variáveis explicativas sobre uma ou mais variáveis respostas.

Por exemplo, considere a Tabela 1.4 sob o modelo Multinomial. Seja π_{ij} a probabilidade de uma pessoa adulta ser classificada segundo o i -ésimo Sexo e j -ésima categoria de Hábito de fumar ($i, j = 1, 2$); o vetor de probabilidades correspondente à tabela é dado por $\pi = (\pi_{11} \ \pi_{12} \ \pi_{21} \ \pi_{22})'$ com a restrição natural $\sum_i \sum_j \pi_{ij} = 1$. Um modelo.

estrutural útil para a análise da associação entre Sexo (S) e Hábito de fumar (H) é o modelo estrutural log-linear:

$$\log \pi_{ij} = u + u_i^S + u_j^H + u_{ij}^{SH} \quad i, j = 1, 2 \quad (1.1)$$

com as restrições:

$$\sum_i u_i^S = \sum_j u_j^H = \sum_i u_{ij}^{SH} = \sum_j u_{ij}^{SH} = 0$$

Consequentemente os parâmetros correspondem às seguintes combinações lineares dos $\log \pi_{ij}$:

$$\begin{aligned} u &= \frac{1}{4} \sum_i \sum_j \log \pi_{ij} \\ u_i^S &= \frac{1}{2} \sum_j \log \pi_{ij} - u \\ u_j^H &= \frac{1}{2} \sum_i \log \pi_{ij} - u \\ u_{ij}^{SH} &= \log \pi_{ij} - u_i^S - u_j^H - u . \end{aligned}$$

As restrições associadas ao modelo estrutural log-linear servem para definir univocamente os parâmetros a partir dos $\log \pi_{ij}$, i.e., são restrições de identificabilidade. Note que em termos das probabilidades $\{\pi_{ij}\}$ o modelo (1.1) pode ser escrito como:

$$\pi_{ij} = \exp(u + u_i^S + u_j^H + u_{ij}^{SH}) \quad i, j = 1, 2 \quad (1.2)$$

com as mesmas restrições dadas acima. Nesse caso, a razão de produtos cruzados é dada por :

$$\begin{aligned} \theta &= \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} \\ &= \frac{\exp(u + u_1^S + u_1^H + u_{11}^{SH}) \exp(u + u_2^S + u_2^H + u_{22}^{SH})}{\exp(u + u_1^S + u_2^H + u_{12}^{SH}) \exp(u + u_2^S + u_1^H + u_{21}^{SH})} \\ &= \exp(u_{11}^{SH} + u_{22}^{SH} - u_{12}^{SH} - u_{21}^{SH}) \\ &= \exp(4 u_{11}^{SH}) \end{aligned}$$

pois de (1.1) sabemos que $u_{11}^{SH} = u_{22}^{SH} = -u_{12}^{SH} = -u_{21}^{SH}$. Dessa forma, $u_{11}^{SH} = \frac{1}{4} \log \theta$; portanto, $u_{11}^{SH} = 0$ ocorre se e só se $\theta = 1$ o que corresponde à independência entre as

variáveis. Logo um modelo estrutural log-linear que impõe uma relação de independência entre as duas variáveis é:

$$\log \pi_{ij} = u + u_i^S + u_j^H \quad i, j = 1, 2 \quad (1.3)$$

com $\sum_i u_i^S = \sum_j u_j^H = 0$.

O modelo (1.3) é uma formulação alternativa do modelo:

$$\pi_{ij} = \pi_{i.} \pi_{.j} \quad (1.4)$$

com $\pi_{i.} = \sum_j \pi_{ij}$ e $\pi_{.j} = \sum_i \pi_{ij}$, que também estabelece a independência entre as duas variáveis. Para verificar esse fato, observe que de (1.3) temos:

$$\pi_{i.} = \exp(u + u_i^S) [\exp(u_1^H) + \exp(u_2^H)];$$

$$\pi_{.j} = \exp(u + u_j^H) [\exp(u_1^S) + \exp(u_2^S)].$$

Da restrição natural $\sum_i \sum_j \pi_{ij} = 1$ temos:

$$\pi_{..} = \sum_i \sum_j \pi_{ij} = \exp(u) [\exp(u_1^H) + \exp(u_2^H)] [\exp(u_1^S) + \exp(u_2^S)] = 1$$

Consequentemente,

$$\begin{aligned} \pi_{i.} \pi_{.j} &= \exp(u + u_i^S) [\exp(u_1^H) + \exp(u_2^H)] \exp(u + u_j^H) [\exp(u_1^S) + \exp(u_2^S)] \\ &= \exp(u + u_i^S + u_j^H) \{ \exp(u) [\exp(u_1^H) + \exp(u_2^H)] [\exp(u_1^S) + \exp(u_2^S)] \} \\ &= \exp(u + u_i^S + u_j^H) = \pi_{ij}. \end{aligned}$$

Verificamos, portanto, que o parâmetro u_{ij}^{SH} em (1.1) é uma medida da associação entre as duas variáveis, sendo que $u_{ij}^{SH} = 0$ implica $\pi_{ij} = \pi_{i.} \pi_{.j}$, ou seja a independência entre as duas variáveis estudadas. Observamos também que esse parâmetro está diretamente relacionado com a razão de produtos cruzados. Essa relação dos parâmetros do modelo

com a independência entre as variáveis envolvidas torna os modelos log-lineares bastante atrativos (ver Bishop et al. (1975, p. 24) por exemplo).

No exemplo apresentado, empregamos uma formulação utilizada por muitos autores como Bishop et al. (1975), Haberman (1978), Fienberg (1981) ou Agresti (1984). Outros autores, como Imrey et al. (1981, 1982) ou , Koch et al. (1985), empregam formulações matriciais equivalentes. Neste trabalho, procuramos estudar as diferentes formulações para modelos estruturais LOG-LINEARES, com o intuito de esclarecer as similaridades existentes entre elas. Restringir-nos-emos a situações para as quais se supõe que os modelos probabilísticos Multinomial ou Produto de Multinomiais descrevem adequadamente a geração de dados. No Capítulo 2 apresentamos as diversas formulações e vários exemplos são abordados, enfatizando a definição e interpretação dos parâmetros em cada caso. Mostramos também a equivalência entre as diversas formulações e uma forma de classificação dos modelos estruturais log-lineares. No Capítulo 3 descrevemos as duas principais metodologias clássicas para análise de dados categorizados: Máxima Verossimilhança e Mínimos Quadrados Generalizados (ou Ponderados, segundo a terminologia usual). Com relação a essa variação na terminologia o leitor pode consultar Seber (1977, p. 62) por exemplo. O Capítulo 4 é dedicado a aplicações práticas, ressaltando-se alguns pacotes computacionais disponíveis para a análise de dados categorizados. No anexo A apresentamos um programa para ajuste de modelos log-lineares pelos dois procedimentos inferenciais citados acima; esse programa foi desenvolvido utilizando o módulo CM do pacote científico SOC (EMBRAPA) e pode ser implantado em microcomputadores de 16 bits ou em qualquer máquina para a qual um compilador C esteja disponível.

Capítulo 2

Modelos Estruturais Log-Lineares

Neste capítulo apresentamos diversas formulações para o modelo estrutural log-linear, considerando os modelos probabilísticos Multinomial e Produto de Multinomiais. A título de ilustração indicaremos vários modelos para os exemplos apresentados no Capítulo 1. Na Seção 2.1 apresentamos as formulações usualmente empregadas para modelos estruturais log-lineares nos casos em que a distribuição Multinomial é adequada como modelo probabilístico de geração dos dados. Na Seção 2.2 mostramos a utilização de funções “logito” para modelagem (estrutural) log-linear em situações onde um modelo probabilístico Produto de Multinomiais é adequado. Na Seção 2.3 mostramos a equivalência entre as formulações apresentadas nas Seções 2.1 e 2.2, dando alguns exemplos. A Seção 2.4 aborda modelos estruturais log-lineares generalizados, que engloba a classe dos modelos

estudados neste trabalho. A Seção 2.5 é dedicada a uma forma de classificação dos modelos log-lineares; essa classificação permite a identificação de modelos para os quais é possível a estimação direta das frequências esperadas através de cálculos consideravelmente mais simplificados.

2.1 - Modelos log-lineares para dados com distribuição Multinomial

Considerando r como o número total de perfis de resposta multivariados (número total de caselas da tabela obtida pelo cruzamento das variáveis categorizadas em estudo) a função de probabilidade para o modelo Multinomial é dada por:

$$P\{\mathbf{n}|\mathbf{n}, \boldsymbol{\pi}\} = n! \prod_{j=1}^r \frac{\pi_j^{n_j}}{n_j!} \quad (2.1.1)$$

com as restrições naturais: $0 < \pi_j < 1$ e $\sum_j \pi_j = 1$ para $j = 1, \dots, r$; $\mathbf{n} = (n_1 \ n_2 \ \dots \ n_r)'$, n_j é a frequência observada correspondente a categoria de resposta j ; $\boldsymbol{\pi} = (\pi_1 \ \pi_2 \ \dots \ \pi_r)'$, π_j é a probabilidade associada à categoria de resposta j , i.e., é a probabilidade de cada elemento da população selecionado aleatoriamente ser classificado na categoria de resposta j ; e $n = \sum_j n_j$ é o tamanho da amostra.

Um modelo estrutural log-linear para dados com essa estrutura probabilística relaciona o logaritmo do vetor de probabilidades associadas às caselas a um conjunto de parâmetros através de uma função linear.

Em geral modelos estruturais log-lineares para dados com distribuição Multinomial podem ser escritos na forma:

$$\log \boldsymbol{\pi} = (\mathbf{1}_r \ \mathbf{X}) \begin{pmatrix} \lambda \\ \boldsymbol{\beta} \end{pmatrix} = \mathbf{1}_r \lambda + \mathbf{X} \boldsymbol{\beta} \quad (2.1.2)$$

onde $\mathbf{X}$ ($r \times t$), é uma matriz conhecida de posto completo $t \leq (r - 1)$ que especifica o modelo estrutural, $\boldsymbol{\beta}$ ($t \times 1$) é o vetor de parâmetros (desconhecidos) do modelo estrutural proposto e $\lambda = -\log[\mathbf{1}_r' \exp(\mathbf{X} \boldsymbol{\beta})]$ é uma constante que padroniza $\boldsymbol{\pi}$ de forma a satisfazer a restrição natural $\mathbf{1}_r' \boldsymbol{\pi} = 1$ ($\mathbf{1}_r$ ($r \times 1$) é um vetor com todos os elementos iguais a um).

Esta formulação matricial é uma alternativa para a formulação polinomial (discriminada por componentes) utilizada no capítulo anterior. Muitos autores preferem empregar a formulação polinomial na expressão de modelos estruturais log-lineares, principalmente quando a matriz de especificação $\mathbf{X}$ define modelos hierárquicos fatoriais semelhantes aos utilizados em Análise de Variância, como em (1.1) e (1.3).

Note que de (2.1.2) temos:

$$\begin{aligned}\pi &= \exp(\log \pi) = \exp(\mathbf{1}_r \lambda + \mathbf{X}\beta) \\ &= \mathbf{D}_{\mathbf{1}_r \exp \lambda} \exp(\mathbf{X}\beta) \\ &= \mathbf{D}_{\mathbf{1}_r \exp(-\lambda)}^{-1} \exp(\mathbf{X}\beta)\end{aligned}$$

onde $\mathbf{D}_{\mathbf{1}_r \exp \lambda}$ denota uma matriz diagonal com os elementos do vetor $\mathbf{1}_r \exp(\lambda)$ na diagonal principal. Como $\lambda = -\log[\mathbf{1}_r' \exp(\mathbf{X}\beta)]$ temos:

$$\exp(-\lambda) = \mathbf{1}_r' \exp(\mathbf{X}\beta) \quad \text{e} \quad \mathbf{1}_r \exp(-\lambda) = \mathbf{1}_r \{\mathbf{1}_r' \exp(\mathbf{X}\beta)\} = \eta.$$

Consequentemente uma expressão alternativa para o modelo log-linear (2.1.2) é dada por:

$$\pi = \mathbf{D}_{\eta}^{-1} \exp(\mathbf{X}\beta) \quad (2.1.3)$$

Os elementos do vetor η tem a função de padronizar os elementos de $\exp(\mathbf{X}\beta)$ de forma que satisfaçam a restrição natural associada ao vetor de probabilidades π . O modelo expresso nessa forma é conveniente para a obtenção dos estimadores de máxima verossimilhança, como veremos na Seção 3.1.

Alternativamente, (2.1.2) pode ser expresso através da imposição de restrições da forma:

$$\mathbf{W} \log \pi = \mathbf{0}_{(r-t)} \quad (2.1.4)$$

onde $\mathbf{W} ((r-t) \times r)$, com posto completo $(r-t)$, é uma matriz ortogonal a $[\mathbf{1}_r, \mathbf{X}]$. Imrey et al. (1981, 1982) denominam (2.1.4) de *formulação através de restrições* ("constraint formulation") e a expressão (2.1.2) de *formulação através de equações de liberdade* ("freedom

equation formulation"). A forma (2.1.4) é frequentemente usada no contexto de redução de modelos. Exemplificaremos sua aplicação no Capítulo 4.

Em seguida apresentaremos exemplos de modelos log-lineares para algumas das situações práticas descritas no Capítulo 1. Os modelos serão expressos tanto através da formulação matricial quanto através da formulação polinomial que é mais comum (ver Goodman(1970, 1971, 1978), Bishop et al.(1975), Haberman(1978), Fienberg(1981) e Agresti(1984), por exemplo).

Inicialmente consideremos a Tabela 1.4 sob o modelo probabilístico Multinomial com $\pi = (\pi_{11} \ \pi_{12} \ \pi_{21} \ \pi_{22})'$. O modelo estrutural log-linear (1.3) que considera independência entre as variáveis Sexo e Hábito de fumar pode ser expresso na forma (2.1.2) com

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix}$$

$$\boldsymbol{\beta} = (u_1^S \ u_1^H)'$$

e

$$\lambda = -\log[\mathbf{1}_4' \exp(\mathbf{X}\boldsymbol{\beta})] = u .$$

Nesse caso temos:

$$\begin{pmatrix} \log \pi_{11} \\ \log \pi_{12} \\ \log \pi_{21} \\ \log \pi_{22} \end{pmatrix} = \mathbf{1}_4 \lambda + \mathbf{X}\boldsymbol{\beta} = \begin{pmatrix} u + u_1^S + u_1^H \\ u + u_1^S - u_1^H \\ u - u_1^S + u_1^H \\ u - u_1^S - u_1^H \end{pmatrix} .$$

Como outro exemplo, consideremos a situação da Tabela 1.6 e vamos admitir que apenas o total geral ($n = 960$) tenha sido fixado no planejamento amostral. As 3 variáveis dicotômicas: Sobrevivência da planta (A), Época de plantio (B) e Comprimento do corte (C), devem ser consideradas como variáveis respostas e um modelo probabilístico apropriado corresponde a (2.1.1) com $n = 960$, $r = 8$ e $\pi = (\pi_{111} \ \pi_{211} \ \pi_{121} \ \pi_{221} \cdots \pi_{122} \ \pi_{222})'$, onde $0 < \pi_{ijk} < 1$ e $\sum_i \sum_j \sum_k \pi_{ijk} = 1$. Se o interesse for estudar a associação entre as

variáveis, um possível modelo estrutural log-linear, escrito na forma matricial, corresponde a (2.1.2) com

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{pmatrix} \quad (2.1.6)$$

$$\boldsymbol{\beta} = (u_1^A \ u_1^B \ u_1^C \ u_{11}^{AB} \ u_{11}^{AC} \ u_{11}^{BC} \ u_{111}^{ABC})'$$

e

$$\lambda = -\log[\mathbf{1}'_8 \exp(\mathbf{X}\boldsymbol{\beta})] = u .$$

Alternativamente, na forma polinomial, esse modelo é expresso como:

$$\log \pi_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC} \quad i, j, k = 1, 2 \quad (2.1.7)$$

com as restrições:

$$\sum_i u_i^A = \sum_j u_j^B = \dots = \sum_i u_{ij}^{AB} = \sum_j u_{ij}^{AB} = \dots = \sum_k u_{ijk}^{ABC} = 0 \quad (2.1.8)$$

Em um caso mais geral em que A tem I níveis, B tem J níveis e C tem K níveis, numa analogia com Análise de Variância, os parâmetros associados teriam a seguinte interpretação:

- $u = \frac{1}{IJK} \sum_i \sum_j \sum_k \log \pi_{ijk}$: média dos $\log \pi_{ijk}$ (“média geral”);
- $u_i^A = \frac{1}{JK} \sum_j \sum_k \log \pi_{ijk} - u$: desvio da média dos $\log \pi_{ijk}$ correspondentes às caselas com variável A no nível i em relação a u (“efeito principal de A ”);
- $u_j^B = \frac{1}{IK} \sum_i \sum_k \log \pi_{ijk} - u$: desvio da média dos $\log \pi_{ijk}$ correspondentes às caselas com variável B no nível j em relação a u (“efeito principal de B ”);
- $u_k^C = \frac{1}{IJ} \sum_i \sum_j \log \pi_{ijk} - u$: desvio da média dos $\log \pi_{ijk}$ correspondentes às caselas com variável C no nível k em relação a u (“efeito principal de C ”);

$u_{ij}^{AB} = \frac{1}{K} \sum_k \log \pi_{ijk} - u_i^A - u_j^B - u$: desvio da média dos $\log \pi_{ijk}$ correspondentes às caselas com variáveis A e B nos níveis i e j respectivamente em relação a $(u_i^A + u_j^B + u)$ (“interação entre A e B ”);

$u_{ik}^{AC} = \frac{1}{J} \sum_j \log \pi_{ijk} - u_i^A - u_k^C - u$: desvio da média dos $\log \pi_{ijk}$ correspondentes às caselas com variáveis A e C nos níveis i e k respectivamente em relação a $(u_i^A + u_k^C + u)$ (“interação entre A e C ”);

$u_{jk}^{BC} = \frac{1}{I} \sum_i \log \pi_{ijk} - u_j^B - u_k^C - u$: desvio da média dos $\log \pi_{ijk}$ correspondentes às caselas com variáveis B e C nos níveis i e j respectivamente em relação a $(u_j^B + u_k^C + u)$ (“interação entre B e C ”); e

$u_{ijk}^{ABC} = \log \pi_{ijk} - u_{ij}^{AB} - u_{ik}^{AC} - u_{jk}^{BC} - u_i^A - u_j^B - u_k^C - u$: desvio do $\log \pi_{ijk}$ em relação a $(u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_i^A + u_j^B + u_k^C + u)$ (“interação entre A , B e C ”).

Neste caso, $I = J = K = 2$.

É importante notar que a linguagem empregada em Análise de Variância para a identificação dos parâmetros tem sido amplamente utilizada neste tipo de modelo log-linear, dada a similaridade de notação (ver Goodman (1970, 1978), Bishop et al. (1975), Haberman (1978), Fienberg (1981) entre outros); os parâmetros referentes a uma única variável são chamados “efeitos principais” e os parâmetros envolvendo duas ou mais variáveis são chamados de “interações”. No entanto, quando a tabela de contingência multidimensional envolve somente variáveis respostas como neste caso, a interpretação do termo “interação” é distinta da empregada em Análise de Variância (*). Neste caso os termos referidos por “interação” correspondem a diferentes formas de associação entre as variáveis. Então, ao se formular uma hipótese de interação nula, algum grau de independência estará sendo admitido. Por exemplo, na situação citada anteriormente “interação de segunda ordem nula” quer dizer que uma medida de associação entre duas respostas quaisquer é

(*) Segundo Winer (1971, p. 309 e 318) “interação é um efeito atribuível a combinação de variáveis, acima e além daquele que pode ser predito a partir das variáveis consideradas individualmente”.

constante nos níveis da terceira variável resposta. O leitor poderá consultar, por exemplo, Bhapkar & Koch (1968a,b), Goodman (1970), Bishop et al. (1975, p. 33, 36–37), Everitt (1977, p. 84–86) ou Haberman (1978, p. 185–187, 199–201) para maiores detalhes.

Outro aspecto a considerar é que a formulação polinomial do modelo log-linear geralmente é super-parametrizada e necessita vir acompanhada de restrições que tornem os parâmetros escolhidos identificáveis. Por exemplo, em (2.1.7) os parâmetros não são identificáveis. As restrições (2.1.8) tornam o modelo identificável e, conseqüentemente, passível de estimação. Contudo, o mesmo objetivo pode ser atingido por outros conjuntos de restrições linearmente independentes, desde que sejam definidos por funções paramétricas não identificáveis (ver Paulino (1988), por exemplo).

O modelo estrutural (2.1.6) (ou (2.1.7) com as restrições (2.1.8)) é um exemplo de modelo saturado, ou seja, um modelo onde o número de parâmetros livres é igual ao número de caselas da tabela de contingência menos o número de restrições impostas pelo modelo probabilístico. O número de parâmetros de modelos saturados não permite uma interpretação simples pois não reduz a dimensão do problema; em geral as estratégias de análise envolvem a procura de modelos não saturados, preferencialmente com um número bem reduzido de parâmetros, que expliquem grande parte da variabilidade dos dados. Para apresentar um exemplo de modelo estrutural log-linear não saturado, consideremos novamente os dados da Tabela 1.6 sob um modelo probabilístico Multinomial. Vamos admitir que o termo u_{ijk}^{ABC} no modelo (2.1.6) seja zero (interação de segunda ordem nula). Dessa forma, fica estabelecido um modelo que considera “associação parcial” entre cada par de variáveis; i.e., as variáveis A (Sobrevivência da planta), B (Época de plantio) e C (Comprimento do corte) são tais que a associação entre quaisquer duas delas não depende do nível da terceira. O modelo estrutural log-linear correspondente pode ser expresso na forma (2.1.2) com,

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 \end{pmatrix} \quad (2.1.9)$$

$$\boldsymbol{\beta} = (u_1^A \ u_1^B \ u_1^C \ u_{11}^{AB} \ u_{11}^{AC} \ u_{11}^{BC})'$$

e

$$\lambda = -\log[\mathbf{1}_8' \exp(\mathbf{X}\boldsymbol{\beta})] = u$$

ou equivalentemente na forma:

$$\log \pi_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} \quad i, j, k = 1, 2 \quad (2.1.10)$$

com as restrições apropriadas obtidas de (2.1.8).

Se pudermos admitir também que o termo u_{ij}^{AB} seja nulo, o novo modelo incorpora a independência entre as variáveis A e B para todo nível da variável C , i.e., a independência condicional das variáveis A e B dado qualquer nível da variável C . Tal modelo estrutural log-linear não saturado pode ser especificado por (2.1.2) com

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{pmatrix} \quad (2.1.11)$$

$$\boldsymbol{\beta} = (u_1^A \ u_1^B \ u_1^C \ u_{11}^{AC} \ u_{11}^{BC})'$$

e

$$\lambda = -\log[\mathbf{1}_8' \exp(\mathbf{X}\boldsymbol{\beta})] = u ;$$

na forma polinomial, temos:

$$\log \pi_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ik}^{AC} + u_{jk}^{BC} \quad i, j, k = 1, 2 \quad (2.1.12)$$

com as restrições apropriadas obtidas de (2.1.8).

Fazendo $u_{ijk}^{ABC} = u_{ij}^{AB} = u_{ik}^{AC} = 0$ obtem-se um modelo log-linear não saturado que estabelece independência entre a variável A e o par de variáveis (B, C) admitindo a dependência entre as variáveis B e C . A matriz de especificação desse modelo poderia ser obtida, por exemplo, do modelo saturado (2.1.6) excluindo-se as colunas 4, 5 e 7. O modelo que representa a completa independência das três variáveis, também pode ser obtido de (2.1.6) excluindo-se, neste caso, as colunas 4, 5, 6 e 7 que correspondem aos parâmetros de “interação”.

É interessante notar que em muitos casos, a literatura clássica sobre modelos log-lineares (Bishop et al. (1975), Goodman (1978), Haberman (1978) ou Fienberg (1981) por exemplo) utiliza a formulação polinomial para modelos log-lineares em termos das frequências esperadas ($m = n\pi$). Por exemplo, para a mesma situação descrita acima, se o interesse for descrever as frequências esperadas nas caselas (m), ao invés das probabilidades π , um modelo estrutural log-linear possível, expresso na forma polinomial, é:

$$\log m_{ijk} = u^* + u_i^A + u_j^B + u_k^C + u_{ik}^{AC} + u_{jk}^{BC} \quad i, j, k = 1, 2 \quad (2.1.13)$$

com as restrições apropriadas obtidas de (2.1.8). Aqui m_{ijk} denota a frequência esperada correspondente à casela (i, j, k) . Note que a restrição natural $\sum_i \sum_j \sum_k \pi_{ijk} = 1$ implica $\sum_i \sum_j \sum_k m_{ijk} = n$. Como decorrência de (2.1.13) temos:

$$\begin{aligned} u^* &= \frac{1}{8} \sum_i \sum_j \sum_k \log m_{ijk} = \frac{1}{8} \sum_i \sum_j \sum_k \log(n \pi_{ijk}) \\ &= \frac{1}{8} \sum_i \sum_j \sum_k (\log n + \log \pi_{ijk}) \\ &= \log n + \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk} \\ &= \log n + u. \end{aligned}$$

$$\beta = (u_1^C \ u_2^C \ u_3^C \ u_1^S \ u_2^S \ u_{11}^{CS} \ u_{12}^{CS} \ u_{21}^{CS} \ u_{22}^{CS} \ u_{31}^{CS} \ u_{32}^{CS})'$$

e

$$\lambda = -\log[\mathbf{1}'_{12} \exp(\mathbf{X}\beta)] = u.$$

Na forma polinomial, esse modelo é dado por:

$$\log \pi_{ij} = u + u_i^C + u_j^S + u_{ij}^{CS} \quad i = 1, \dots, 4; \ j = 1, \dots, 3 \quad (2.1.15)$$

com as restrições:

$$\sum_i u_i^C = \sum_j u_j^S = \sum_i u_{ij}^{CS} = \sum_j u_{ij}^{CS} = 0 \quad (2.1.16)$$

A partir dessa formulação, os parâmetros podem ser interpretados como:

$$\begin{aligned} u &= \frac{1}{12} \sum_i \sum_j \log \pi_{ij} && : \text{média dos } \log \pi_{ij}; \\ u_i^C &= \frac{1}{3} \sum_j \log \pi_{ij} - u && : \text{desvio da média dos } \log \pi_{ij} \text{ correspondentes ao } i\text{-ésimo} \\ &&& \text{tipo de cirurgia, } i = 1, 2, 3, 4, \text{ em relação a } u; \\ u_j^S &= \frac{1}{4} \sum_i \log \pi_{ij} - u && : \text{desvio da média dos } \log \pi_{ij} \text{ correspondentes à } j\text{-ésima} \\ &&& \text{severidade da sequela, } j = 1, 2, 3, \text{ em relação a } u; \\ u_{ij}^{CS} &= \log \pi_{ij} - u_i^C - u_j^S - u && : \text{desvio do } \log \pi_{ij} \text{ (} i=1,2,3,4 \text{ e } j=1,2,3 \text{) em relação} \\ &&& \text{a } (u_i^C + u_j^S + u). \end{aligned}$$

Um modelo estrutural log-linear não saturado, para uma situação em que os efeitos de Tipo de cirurgia e Severidade da sequela podem ser considerados independentes, é obtido de (2.1.14) eliminando-se as colunas relativas aos efeitos de interação de sua matriz de

especificação. Tal modelo corresponde a (2.1.2) com

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 0 \\ -1 & -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & -1 & -1 \end{pmatrix} \quad (2.1.17)$$

$$\boldsymbol{\beta} = (u_1^C \ u_2^C \ u_3^C \ u_1^S \ u_2^S)'$$

e

$$\lambda = -\log[\mathbf{1}_{12}' \exp(\mathbf{X}\boldsymbol{\beta})] = u .$$

Na forma polinomial (2.1.19) é expresso por:

$$\log \pi_{ij} = u + u_i^C + u_j^S \quad i = 1, \dots, 4; \ j = 1, \dots, 3 \quad (2.1.18)$$

com as restrições apropriadas obtidas de (2.1.16).

Um modelo estrutural log-linear que incorpora a ordinalidade das duas variáveis é dado por (2.1.2) com

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1,5 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & -1 & -1,5 \\ 0 & 1 & 0 & 1 & 0 & 0,5 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & -1 & -0,5 \\ 0 & 0 & 1 & 1 & 0 & -0,5 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0,5 \\ -1 & -1 & -1 & 1 & 0 & -1,5 \\ -1 & -1 & -1 & 0 & 1 & 0 \\ -1 & -1 & -1 & -1 & -1 & 1,5 \end{pmatrix} \quad (2.1.19)$$

$$\boldsymbol{\beta} = (u_1^C \ u_2^C \ u_3^C \ u_1^S \ u_2^S \ \gamma)'$$

e

$$\lambda = -\log[\mathbf{1}'_{12} \exp(\mathbf{X}\beta)] = u$$

ou, alternativamente por:

$$\log \pi_{ij} = u + u_i^C + u_j^S + \gamma(a_i - \bar{a})(b_j - \bar{b}) \quad i = 1, \dots, 4; j = 1, \dots, 3 \quad (2.1.20)$$

com $a_i = i$, $b_j = j$, $\bar{a} = \frac{1}{4} \sum_i a_i$, $\bar{b} = \frac{1}{3} \sum_j b_j$ e restrições $\sum_i u_i^C = \sum_j u_j^S = 0$.

O modelo (2.1.19), cuja formulação polinomial é (2.1.20), é um exemplo de modelo de associação uniforme. Goodman (1979) atribui essa denominação a modelos log-lineares que estabelecem um valor comum para as razões de produtos cruzados locais ("odds ratio" definidos para linhas e colunas adjacentes). Neste caso, para a Tabela 1.5 podem ser definidas 6 razões de produtos cruzados locais que são da forma:

$$\theta_{ij} = \frac{\pi_{ij} \pi_{i+1,j+1}}{\pi_{i,j+1} \pi_{i+1,j}} \quad i = 1, 2, 3; j = 1, 2.$$

Empregando a expressão (2.1.20) obtém-se:

$$\theta_{ij} = \exp(\gamma) \quad \text{para } i = 1, 2, 3 \text{ e } j = 1, 2.$$

Portanto, esse modelo corresponde a uma situação com associação uniforme entre cirurgias adjacentes e severidades da seqüela também adjacentes. Para maiores detalhes ver Agresti (1984, p. 78).

Embora no caso de modelos Multinomiais seja mais comum propor modelos lineares para $\log \pi$ ou $\log m$, em muitas situações, por razões de facilidade de interpretação ou até mesmo computacionais, pode ser de interesse propor estruturas lineares para funções da forma $\mathbf{A} \log \pi$ onde a matriz $\mathbf{A}$ deve satisfazer certas restrições. Veremos nas Seções 2.3 e 2.4 que modelos lineares para essas funções são equivalentes aos modelos log-lineares para o vetor π .

2.2 - Modelos log-lineares para dados com distribuição Produto de Multinomiais

O objetivo da análise de dados sob o modelo Produto de Multinomiais consiste geralmente em estudar o efeito das variáveis explicativas em um conjunto de variáveis respostas. Nesse sentido o primeiro passo da análise consiste em classificar as variáveis em respostas ou explicativas levando em conta o planejamento amostral e os objetivos da pesquisa.

Em geral os dados podem ser dispostos na forma de uma tabela de contingência bidimensional $s \times r$, como indicado na Tabela 2.2.1. As linhas da tabela referem-se às subpopulações definidas pelas *combinações dos níveis das variáveis explicativas* e as colunas correspondem às categorias de resposta obtidas pelas *combinações dos níveis das variáveis respostas*. Cada n_{ij} denota a frequência do j -ésimo *perfil de respostas* na i -ésima subpopulação. Essa disposição de dados é muito semelhante àquela utilizada em ANOVA; as linhas definem os tratamentos e as colunas definem as distribuições a serem comparadas.

Tabela 2.2.1 - Frequência das respostas para as s subpopulações

Subpopulação	Categoria de resposta				Total
	1	2	...	r	
1	n_{11}	n_{12}	...	n_{1r}	n_1
2	n_{21}	n_{22}	...	n_{2r}	n_2
...	...	...	...	...	...
...	...	...	...	...	...
s	n_{s1}	n_{s2}	...	n_{sr}	n_s

O vetor $\mathbf{n}_i = (n_{i1} \ n_{i2} \ \cdots \ n_{ir})'$, com $\sum_j n_{ij} = n_i$, tem, por hipótese, distribuição Multinomial com parâmetros n_i e $\boldsymbol{\pi}_i = (\pi_{i1} \ \cdots \ \pi_{ir})'$, que representam respectivamente o tamanho da amostra e o vetor de probabilidades associadas às categorias de resposta para a i -ésima subpopulação.

Se as s amostras dessas subpopulações forem selecionadas de forma aleatória e independentemente, um modelo probabilístico associado à tabela pode ser um produto de s distribuições Multinomiais independentes cuja função de probabilidade é dada por:

$$P\{\mathbf{n}|\mathbf{n}, \boldsymbol{\pi}\} = \prod_{i=1}^s n_i! \prod_{j=1}^r \frac{\pi_{ij}^{n_{ij}}}{n_{ij}!} \quad i = 1, \dots, s; \ j = 1, \dots, r \quad (2.2.1)$$

com as restrições naturais: $0 < \pi_{ij} < 1$ e $\sum_j \pi_{ij} = 1$ para $i = 1, \dots, s$ e $j = 1, \dots, r$, $\mathbf{n} = (\mathbf{n}'_1 \ \cdots \ \mathbf{n}'_s)'$ é o vetor de frequências observadas e $\boldsymbol{\pi} = (\boldsymbol{\pi}'_1 \ \cdots \ \boldsymbol{\pi}'_s)'$ é o vetor de probabilidades associadas às caselas.

Em geral modelos estruturais log-lineares para dados com distribuição Produto de Multinomiais são convenientemente expressos na forma:

$$\mathbf{A} \log \boldsymbol{\pi} = \mathbf{X}_L \boldsymbol{\beta} \quad (2.2.2)$$

onde $\mathbf{A} [s(r-1) \times sr]$ é uma matriz de constantes conhecidas, ortocomplementar a $\mathbf{U} = [\mathbf{I}_s \otimes \mathbf{1}'_r]$ (i.e., $\mathbf{A}\mathbf{U}' = \mathbf{0}$), que define funções de $\log \boldsymbol{\pi}$ de interesse, $\mathbf{X}_L [s(r-1) \times t]$, de posto completo $t \leq s(r-1)$, é uma matriz conhecida que especifica o modelo estrutural e $\boldsymbol{\beta} (t \times 1)$ é o vetor de parâmetros (desconhecidos) do modelo estrutural proposto. $\mathbf{I}_s (s \times s)$ representa uma matriz identidade e o símbolo $\otimes$ indica o produto de Kronecker à direita, i.e., $\mathbf{A} \otimes \mathbf{B} = \{a_{ij}\mathbf{B}\}$.

Em muitas aplicações práticas convém considerar o vetor de logitos múltiplos $\log \boldsymbol{\pi}^*$, onde

$$\boldsymbol{\pi}^* = (\boldsymbol{\pi}_1^{*'}, \dots, \boldsymbol{\pi}_s^{*'})' \quad \text{com} \quad \boldsymbol{\pi}_i^{*'} = \left(\frac{\pi_{i1}}{\pi_{ir}} \ \cdots \ \frac{\pi_{i(r-1)}}{\pi_{ir}} \right) \quad (2.2.3)$$

que podem ser expressos na forma $\mathbf{A} \log \boldsymbol{\pi}$ com $\mathbf{A} = \{\mathbf{I}_s \otimes [\mathbf{I}_{(r-1)}, -\mathbf{1}_{r-1}]\}$.

É interessante notar que podemos obter π a partir de π^* . De (2.2.2), podemos escrever:

$$\log \pi_{ij}^* = \log \left(\frac{\pi_{ij}}{\pi_{ir}} \right) = \mathbf{x}_{L,ij}' \boldsymbol{\beta}$$

onde $\mathbf{x}_{L,ij}$ ($t \times 1$) é um vetor formado pelos elementos da linha correspondentes à casela (i,j) da matriz $\mathbf{X}_L$. Dessa expressão segue que

$$\pi_{ij} = \pi_{ir} \exp(\mathbf{x}_{L,ij}' \boldsymbol{\beta}) \quad \text{para } j = 1, \dots, (r-1) \text{ e } i = 1, \dots, s$$

Das restrições naturais dadas em (2.2.1) temos que:

$$\sum_{j=1}^{(r-1)} \pi_{ij} + \pi_{ir} = \sum_{j=1}^{(r-1)} \pi_{ir} \exp(\mathbf{x}_{L,ij}' \boldsymbol{\beta}) + \pi_{ir} = 1 \quad \text{para } i = 1, \dots, s$$

o que implica:

$$\pi_{ir} = \frac{1}{1 + \sum_{j=1}^{(r-1)} \exp(\mathbf{x}_{L,ij}' \boldsymbol{\beta})} \quad \text{e} \quad \pi_{ij} = \frac{\exp(\mathbf{x}_{L,ij}' \boldsymbol{\beta})}{1 + \sum_{j=1}^{(r-1)} \exp(\mathbf{x}_{L,ij}' \boldsymbol{\beta})}$$

para $j = 1, \dots, (r-1)$ e $i = 1, \dots, s$.

Outras funções de π que também tem sido usadas são os logitos adjacentes:

$$\log \left(\frac{\pi_{ij}}{\pi_{i(j+1)}} \right) \quad i = 1, \dots, s; j = 1, \dots, (r-1) \quad (2.2.4)$$

que podem ser expressos na forma $\mathbf{A} \log \pi$ com $\mathbf{A} = \{\mathbf{I}_{s(r-1)} \otimes [1 \ -1]\}$ (ver Fienberg (1981, p. 111), por exemplo).

Ressaltamos ainda que, alternativamente, (2.2.2) pode também ser expresso através da imposição de restrições da forma:

$$\mathbf{W} [\mathbf{A} \log \pi] = \mathbf{0}_{[s(r-1)-t]} \quad (2.2.5)$$

onde $\mathbf{W}$ $\{[s(r-1)-t] \times s(r-1)\}$, com posto completo $[s(r-1)-t]$, é uma matriz de contrastes ortogonal a $\mathbf{X}_L$, i.e., $\mathbf{W} \mathbf{X}_L = \mathbf{0}$ (ver Imrey et al. (1981,1982) ou Koch et al. (1985, p. 80) por exemplo).

Como veremos na Seção 2.3, também podemos expressar o modelo estrutural log-linear para situações em que a distribuição Produto de Multinomiais é adequada, em uma formulação semelhante àquela empregada no caso Multinomial.

Independentemente da forma escolhida para expressar um modelo estrutural log-linear, no caso de tabelas de contingência com distribuição Produto de Multinomiais, convém salientar dois importantes aspectos: um ligado às restrições do modelo probabilístico (Produto de Multinomiais) e outro ligado a interpretação dos parâmetros do modelo estrutural. Podemos verificar que o número de restrições naturais é igual ao número de subpopulações. Por outro lado, a interpretação dos parâmetros neste caso é semelhante à empregada em ANOVA (ver por exemplo, Winer (1971, p. 316-319)).

O modelo log-linear na formulação logito é especialmente apropriado para o caso particular de uma única variável resposta e uma ou mais variáveis explicativas. Como exemplo considere os dados da Tabela 1.6. Podemos associar a essa tabela uma distribuição Produto de Multinomiais na forma (2.2.1) com $s = 4$ subpopulações independentes, definidas pela combinação das categorias das variáveis explicativas B (Época do plantio) e C (Comprimento do corte), $r = 2$ categorias de resposta, correspondendo aos níveis da variável A (Sobrevivência da planta) e $\pi = (\pi_{111} \ \pi_{211} \ \pi_{121} \ \pi_{221} \ \cdots \ \pi_{122} \ \pi_{222})'$, onde $0 < \pi_{ijk} < 1$ ($i, j, k = 1, 2$) e $\sum_i \pi_{ijk} = 1$ para $j, k = 1, 2$. Façamos $A = \{I_4 \otimes [1 \ -1]\}$ para definir o vetor de logitos

$$\log \pi^* = A \log \pi = (L_{11} \ L_{21} \ L_{12} \ L_{22})',$$

onde $L_{jk} = \log\left(\frac{\pi_{1jk}}{\pi_{2jk}}\right)$ para $j, k = 1, 2$. Então um modelo estrutural log-linear aditivo é dado por (2.2.2) com

$$X_L = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \quad (2.2.6)$$

e

$$\beta = (w \ w_1^B \ w_1^C)'$$

Expressando esse modelo na forma polinomial temos:

$$L_{jk} = \log \left(\frac{\pi_{1jk}}{\pi_{2jk}} \right) = w + w_j^B + w_k^C \quad j, k = 1, 2 \quad (2.2.7)$$

com as restrições: $\sum_j w_j^B = \sum_k w_k^C = 0$. Os parâmetros podem ser definidos como se segue:

$$\begin{aligned} w &= \frac{1}{4} \sum_j \sum_k \log \left(\frac{\pi_{1jk}}{\pi_{2jk}} \right) : \text{média dos logitos esperados;} \\ w_j^B &= \frac{1}{2} \sum_k \log \left(\frac{\pi_{1jk}}{\pi_{2jk}} \right) - w : \text{desvio da média dos logitos com variável } B \text{ no nível } j \text{ em} \\ &\quad \text{relação a } w; \\ w_k^C &= \frac{1}{2} \sum_j \log \left(\frac{\pi_{1jk}}{\pi_{2jk}} \right) - w : \text{desvio da média dos logitos com variável } C \text{ no nível } k \text{ em} \\ &\quad \text{relação a } w. \end{aligned}$$

Em analogia com ANOVA podemos interpretar w_j^B como sendo o efeito principal do j -ésimo nível da variável B em relação ao logito L_{jk} , e w_k^C o efeito principal do k -ésimo nível da variável C em relação ao logito L_{jk} . Empregando as restrições, também verificamos que: $w_2^B = -w_1^B$ e $w_2^C = -w_1^C$. Além disso, neste caso particular em que a variável resposta apresenta dois níveis, note que L_{jk} corresponde a uma medida da relação entre a chance de “sobreviver” e a chance de “não sobreviver” e consequentemente traz subsídios para a interpretação do resultado da análise.

Como outro exemplo consideremos os dados apresentados na Tabela 1.5 envolvendo duas variáveis ordinais. Um possível modelo probabilístico de geração de dados corresponde à distribuição Produto de Multinomiais definida por (2.2.1) com $s = 4$ subpopulações (diferentes cirurgias), $r = 3$ categorias de resposta (diferentes severidades de uma sequela) e $\boldsymbol{\pi} = (\pi_{11} \ \pi_{12} \ \pi_{13} \ \cdots \ \pi_{41} \ \pi_{42} \ \pi_{43})'$ onde $0 < \pi_{ij} < 1$ ($i = 1, \dots, 4$; $j = 1, \dots, 3$) e $\sum_j \pi_{ij} = 1$ para todo $i = 1, \dots, 4$. Um modelo estrutural log-linear saturado envolvendo os logitos múltiplos:

$$L_{i1} = \log \left(\frac{\pi_{i1}}{\pi_{i3}} \right) \quad \text{e} \quad L_{i2} = \log \left(\frac{\pi_{i2}}{\pi_{i3}} \right) \quad \text{para } i = 1, \dots, 4$$

pode ser escrito na forma (2.2.2) com:

$$\mathbf{A} = \{\mathbf{I}_4 \otimes [\mathbf{I}_2, -\mathbf{1}_2]\}$$

$$\mathbf{X}_L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.2.8)$$

e

$$\boldsymbol{\beta} = (u_1 \quad v_1 \quad \alpha_2 \quad \beta_2 \quad \alpha_3 \quad \beta_3 \quad \alpha_4 \quad \beta_4)'$$

onde u_1 é o valor esperado do logito correspondente a nenhuma severidade versus severidade moderada para a cirurgia $(V + D)$ tomada como referência, v_1 é o valor esperado do logito correspondente a pouca severidade versus severidade moderada para a cirurgia $(V + D)$ tomada como referência e α_i, β_i , para $i = 2, 3, 4$, são as variações médias nesses logitos devidas às demais cirurgias.

A parametrização definida em (2.2.8) é conhecida por “parametrização da casela de referência” pois os parâmetros α_i e β_i representam variações médias nos logitos correspondentes às cirurgias $(V + A)$, $(V + H)$ e GR em relação aos logitos correspondentes à cirurgia $(V + D)$, tomada como referência.

Na forma polinomial (2.2.8) pode ser escrito como:

$$L_{i1} = \log \left(\frac{\pi_{i1}}{\pi_{i3}} \right) = u_1 + \alpha_i, \quad L_{i2} = \log \left(\frac{\pi_{i2}}{\pi_{i3}} \right) = v_1 + \beta_i \quad i = 1, \dots, 4 \quad \text{com} \quad \alpha_1 = \beta_1 = 0 \quad (2.2.9)$$

Um modelo estrutural linear não saturado para os mesmos logitos que considera a estrutura ordinal da variável resposta corresponde a (2.2.2) com

$$\mathbf{A} = \{\mathbf{I}_4 \otimes [\mathbf{I}_2, -\mathbf{1}_2]\}$$

$$\mathbf{X}_L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix} \quad (2.2.10)$$

e

$$\boldsymbol{\beta} = (u_1 \quad v_1 \quad \beta_2 \quad \beta_3 \quad \beta_4)'$$

Na forma polinomial (2.2.10) fica reduzido a:

$$L_{i1} = \log\left(\frac{\pi_{i1}}{\pi_{i3}}\right) = u_1 + 2\beta_i, \quad L_{i2} = \log\left(\frac{\pi_{i2}}{\pi_{i3}}\right) = v_1 + \beta_i \quad i = 1, \dots, 4 \quad \text{com } \beta_1 = 0 \quad (2.2.11)$$

onde u_1 é o valor esperado do logito correspondente a nenhuma severidade versus severidade moderada para a cirurgia $(V + D)$ tomada como referência, v_1 é o valor esperado do logito correspondente a pouca severidade versus severidade moderada para a cirurgia $(V + D)$ tomada como referência e β_i , para $i = 2, 3, 4$, são as variações médias nesses logitos devidas às demais cirurgias. Este modelo é conhecido na literatura como “modelo de quocientes de razões de riscos adjacentes iguais” (ver Imrey et al. (1981, 1982), Semenza et al. (1983) ou Singer (1987)). Esses autores mostram que para quaisquer duas cirurgias, i e k com $i \neq k$, a imposição de que os quocientes de razões de riscos adjacentes sejam iguais é satisfeita pelo modelo acima. Os efeitos das cirurgias (β_i) em relação aos logitos referentes a categorias de resposta mais afastadas $\left(\log\left(\frac{\pi_{i1}}{\pi_{i3}}\right)\right)$ são múltiplos dos efeitos para os logitos correspondentes a categorias mais próximas $\left(\log\left(\frac{\pi_{i2}}{\pi_{i3}}\right)\right)$.

2.3 - Relação entre as diferentes formulações para modelos estruturais log-lineares

Uma das finalidades desta seção é mostrar que embora as formulações utilizadas na Seções 2.1 e 2.2 para modelos estruturais log-lineares sejam aparentemente distintas, na realidade elas são equivalentes e ambas podem ser utilizadas tanto em situações onde o modelo probabilístico é Multinomial quanto naquelas onde é Produto de Multinomiais.

Inicialmente apresentamos uma generalização da formulação (2.1.2), que também pode ser utilizada para o caso de dados com distribuição Produto de Multinomiais. Assim, um modelo log-linear para dados supostamente gerados por essa distribuição pode ser expresso na forma:

$$\log \pi = \{[\mathbf{I}_s \otimes \mathbf{1}_r], \mathbf{X}\} \begin{pmatrix} \lambda \\ \beta \end{pmatrix} = [\mathbf{I}_s \otimes \mathbf{1}_r] \lambda + \mathbf{X} \beta \quad (2.3.1)$$

onde $\lambda = (\lambda_1, \dots, \lambda_s)'$ com $\lambda_i = -\log[1'_r \exp(\mathbf{X}_i \beta)]$ para $i = 1, \dots, s$, $\mathbf{X}$ ($sr \times t$), de posto completo $t \leq s(r-1)$, é uma matriz conhecida que especifica o modelo estrutural e β ($t \times 1$) é o vetor de parâmetros (desconhecidos) do modelo estrutural proposto. Os valores $\lambda_1, \dots, \lambda_s$ são constantes correspondentes às s subpopulações cuja função é padronizar π para que satisfaça às restrições naturais dadas por $\mathbf{U}' \pi = \mathbf{1}_s$, com $\mathbf{U}' = [\mathbf{I}_s \otimes \mathbf{1}'_r]$. As linhas de $\mathbf{X}$ devem ser tais que $\mathbf{U}' \mathbf{X} = \mathbf{0}$.

Como exemplo, considere a Tabela 1.6 sob o modelo Produto de Multinomiais na forma (2.2.1) com $s = 4$ subpopulações, $r = 2$ categorias de respostas e vetor de probabilidades $\pi = (\pi_{111} \pi_{211} \pi_{121} \pi_{221} \dots \pi_{122} \pi_{222})'$, onde $0 < \pi_{ijk} < 1$ ($i, j, k = 1, 2$) e $\sum_i \pi_{ijk} = 1$ para todo $j, k = 1, 2$. As subpopulações são definidas pela combinação das categorias das variáveis B e C (indexadas por j e k , respectivamente) e as categorias de respostas correspondem às categorias da variável A (indexadas por i). Um modelo estrutural log-

linear saturado para π é dado pela forma matricial (2.3.1) com

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 \end{pmatrix} \quad (2.3.2)$$

$$\boldsymbol{\beta} = \left(v_{A(1)} \ v_{A(1)}^{B(1)} \ v_{A(1)}^{C(1)} \ v_{A(1)}^{BC(11)} \right)'$$

e

$$\boldsymbol{\lambda} = (\lambda_1 \ \lambda_2 \ \lambda_3 \ \lambda_4)' \quad \text{onde} \quad \lambda_h = -\log[1_2' \exp(\mathbf{X}_h \boldsymbol{\beta})] \quad \text{para} \quad h = 1, \dots, 4.$$

Na forma polinomial esse modelo é expresso por:

$$\log \pi_{ijk} = v^{BC(jk)} + v_{A(i)} + v_{A(i)}^{B(j)} + v_{A(i)}^{C(k)} + v_{A(i)}^{BC(jk)} \quad i, j, k = 1, 2 \quad (2.3.3)$$

com as restrições:

$$\begin{aligned} \sum_i v_{A(i)} &= \sum_i v_{A(i)}^{B(j)} = \sum_i v_{A(i)}^{C(k)} = \sum_i v_{A(i)}^{BC(jk)} = 0 \\ \sum_j v_{A(i)}^{B(j)} &= \sum_k v_{A(i)}^{C(k)} = \sum_j v_{A(i)}^{BC(jk)} = \sum_k v_{A(i)}^{BC(jk)} = 0 \\ \text{e } v^{BC(11)} &= \lambda_1, \ v^{BC(21)} = \lambda_2, \ v^{BC(12)} = \lambda_3 \text{ e } v^{BC(22)} = \lambda_4. \end{aligned} \quad (2.3.4)$$

Como decorrência das restrições (2.3.4) temos que:

$$v_{A(i)} = \frac{1}{4} \sum_j \sum_k \log \pi_{ijk} - \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk} \quad \text{é o desvio entre a média dos } \log \pi_{ijk} \text{ com variável resposta no nível } i \text{ e a média geral dos } \log \pi_{ijk};$$

$$v_{A(i)}^{B(j)} = \frac{1}{2} \sum_k \log \pi_{ijk} - \frac{1}{4} \sum_j \sum_k \log \pi_{ijk} - \frac{1}{4} \sum_i \sum_k \log \pi_{ijk} + \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk} \\ \text{é o efeito da variável } B, \text{ nível } j, \text{ na categoria de resposta } i;$$

$$v_{A(i)}^{C(k)} = \frac{1}{2} \sum_j \log \pi_{ijk} - \frac{1}{4} \sum_j \sum_k \log \pi_{ijk} - \frac{1}{4} \sum_i \sum_j \log \pi_{ijk} + \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk} \\ \text{é o efeito da variável } C, \text{ nível } k, \text{ na categoria de resposta } i;$$

$$v_{A(i)}^{BC(jk)} = \log \pi_{ijk} - \frac{1}{2} \sum_i \log \pi_{ijk} - \frac{1}{2} \sum_k \log \pi_{ijk} + \frac{1}{4} \sum_j \sum_k \log \pi_{ijk} \\ + \frac{1}{4} \sum_i \sum_k \log \pi_{ijk} + \frac{1}{4} \sum_i \sum_j \log \pi_{ijk} - \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk}$$

é o efeito de interação entre as variáveis B e C , níveis j e k respectivamente, na categoria de resposta i .

Um modelo log-linear não saturado para π que considera nulo o efeito de interação entre as variáveis B e C sobre a variável A corresponde a (2.3.1) com

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \end{pmatrix} \quad (2.3.5)$$

$$\beta = \begin{pmatrix} v_{A(1)} & v_{A(1)}^{B(1)} & v_{A(1)}^{C(1)} \end{pmatrix}'$$

e

$$\lambda = (\lambda_1 \lambda_2 \lambda_3 \lambda_4)' \quad \text{onde} \quad \lambda_h = -\log[\mathbf{1}_2' \exp(\mathbf{X}_h \beta)] \quad \text{para} \quad h = 1, \dots, 4.$$

Este modelo pode ser diretamente obtido de (2.3.2) eliminando-se a última coluna da matriz $\mathbf{X}$ (coluna correspondente ao parâmetro considerado nulo).

Na forma polinomial esse modelo é expresso por:

$$\log \pi_{ijk} = v^{BC(jk)} + v_{A(i)} + v_{A(i)}^{B(j)} + v_{A(i)}^{C(k)} \quad i, j, k = 1, 2 \quad (2.3.6)$$

com as restrições apropriadas obtidas de (2.3.4).

Equivalentemente, (2.3.1) também pode ser reescrita na forma exponencial:

$$\pi = \pi(\beta) = \mathbf{D}_\eta^{-1} \exp(\mathbf{X}\beta) \quad (2.3.7)$$

onde $\mathbf{D}_\eta^{-1}$ indica a inversa de uma matriz diagonal com os elementos de

$$\eta = \{[\mathbf{I}_s \otimes \mathbf{1}_r \mathbf{1}_r'] \exp(\mathbf{X}\beta)\}$$

na diagonal. A matriz D_{η}^{-1} tem como finalidade padronizar $\exp(X\beta)$ fazendo com que os elementos de $\pi(\beta)$ satisfaçam as restrições naturais mencionadas acima. Esta forma é útil na obtenção de estimadores de máxima verossimilhança.

Observemos também que assim como o vetor de probabilidades π pode ser expresso como $\pi = (\pi'_1, \dots, \pi'_s)'$, o mesmo ocorre com a matriz X , i.e., $X = [X'_1, \dots, X'_s]'$; portanto (2.3.7) é uma forma compacta para escrever $\pi_i = \exp(X_i\beta) / [1'_r \exp(X_i\beta)]$, $i = 1, \dots, s$ com X_i correspondendo à matriz que gera π_i a partir de β .

Em algumas situações práticas, pode ser necessário empregar uma ou outra formulação em função dos pacotes computacionais disponíveis para a obtenção de estimativas. Demonstramos, a seguir, a equivalência entre as formulações (2.2.2) e (2.3.1) (Haber (1985), Carlos Daniel Mimoso Paulino (comunicação pessoal)):

(2.3.1)→(2.2.2) Multiplicando ambos os membros de (2.3.1) à esquerda por uma matriz A [$s(r-1) \times sr$], com posto completo $s(r-1)$ e satisfazendo $AU' = A[I_s \otimes 1_r] = 0$, temos:

$$A \log \pi = A[I_s \otimes 1_r]\lambda + AX\beta = AX\beta$$

Assim obtemos o resultado apresentado em (2.2.2), ou seja :

$$A \log \pi = X_L \beta$$

com $X_L = AX$ e $AU' = 0$.

(2.2.2)→(2.3.1) Consideremos agora o modelo dado por (2.2.2) que essencialmente corresponde a um sistema linear não homogêneo em $\log \pi$. Da teoria sobre sistemas lineares (ver Searle (1982, cap. 9), por exemplo) sabemos que a solução geral desse sistema é dada por :

$$\log \pi = \mu + T\beta$$

onde μ é solução geral do sistema homogêneo $A \log \pi = 0$ e $T\beta$ é uma solução particular do sistema não homogêneo (2.2.2). Como A tem posto completo, a matriz AA' é não singular. Assim,

$$A \log \pi = X_L \beta = (AA')(AA')^{-1} X_L \beta$$

que implica

$$A [\log \pi - A'(AA')^{-1} X_L \beta] = 0.$$

Então $\log \pi = A'(AA')^{-1} X_L \beta$ é uma solução particular do sistema não homogêneo (2.2.2). Por outro lado, soluções do sistema homogêneo $A \log \pi = 0$ estão no espaço ortocomplementar às linhas da matriz A . Como $A[I_s \otimes 1_r] = 0$; pela hipótese sobre A , e as colunas de $[I_s \otimes 1_r]$ são linearmente independentes (l.i.) então $[I_s \otimes 1_r]$ é base do espaço de soluções do sistema. Logo $\mu = [I_s \otimes 1_r] \lambda$, $\lambda = (\lambda_1 \cdots \lambda_s)'$, $\lambda_i \in \mathbb{R}$, $i = 1, \dots, s$, é solução geral do sistema homogêneo $A \log \pi = 0$. Portanto a solução geral do sistema não homogêneo (2.2.2) é :

$$\log \pi = [I_s \otimes 1_r] \lambda + A'(AA')^{-1} X_L \beta$$

que corresponde à formulação (2.3.1) sendo $X = A'(AA')^{-1} X_L$ a matriz de especificação ■

Sendo comum especificar as matrizes X e X_L com 0's 1's e -1's, podemos ainda considerar que a relação entre as respectivas matrizes X e X_L é dada por $X_L = k_1 AX$ ou $X = k_2 A'(AA')^{-1} X_L$, onde k_1 e k_2 são escalares que podem ser incorporados aos respectivos vetores de parâmetros. Detalhes poderão ser examinados através dos exemplos que consideramos a seguir.

Exemplo 2.3.1

Como exemplo da equivalência entre as formulações (2.2.2) e (2.3.1), mostraremos como o modelo (2.2.6) pode ser obtido a partir do modelo log-linear para as probabilidades nas caselas (2.3.5) e vice-versa. Partindo de (2.3.5) (formulação do tipo (2.3.1)) e multiplicando à esquerda por $A = \{I_4 \otimes [1 \ -1]\}$, obtemos:

$$X_L = 2 \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{pmatrix}$$

ou seja, a matriz de especificação dada em (2.2.6) multiplicada pelo escalar 2. Incorporando esse escalar ao vetor de parâmetros temos que os parâmetros em (2.2.6) correspondem aos parâmetros em (2.3.5) multiplicados por 2.

Reciprocamente, a partir de (2.2.6) (formulação do tipo (2.2.2)) podemos reescrever o modelo na formulação (2.3.1) com

$$X = A'(AA')^{-1}X_L \quad \text{onde} \quad A = \{I_4 \otimes [1 \ -1]\}$$

Como $AA' = \{I_4 \otimes [1 \ -1]\} \{I_4 \otimes [1 \ -1]'\} = 2 I_4$ segue que

$$(AA')^{-1} = \frac{1}{2} I_4$$

$$\text{e} \quad A'(AA')^{-1} = \{I_4 \otimes [1 \ -1]'\} \frac{1}{2} I_4 = \frac{1}{2} \{I_4 \otimes [1 \ -1]'\}$$

o que implica

$$X = A'(AA')^{-1}X_L = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

Essa matriz de especificação corresponde à matriz dada em (2.3.5) multiplicada pelo escalar $\frac{1}{2}$. Incorporando esse fator ao vetor de parâmetros, podemos concluir que os parâmetros em (2.3.5) correspondem aos parâmetros em (2.2.6) multiplicados por $\frac{1}{2}$.

Empregando as formulações polinomiais relativas a esses modelos também podemos verificar o mesmo resultado. Definindo $L_{jk} = \log \pi_{1jk} - \log \pi_{2jk}$, $j, k = 1, 2$, de (2.3.6) segue que:

$$\begin{aligned} L_{jk} &= \log \left(\frac{\pi_{1jk}}{\pi_{2jk}} \right) = \log \pi_{1jk} - \log \pi_{2jk} \\ &= (v^{BC(jk)} + v_{A(1)} + v_{A(1)}^{B(j)} + v_{A(1)}^{C(k)}) - (v^{BC(jk)} + v_{A(2)} + v_{A(2)}^{B(j)} + v_{A(2)}^{C(k)}) \\ &= 2 v_{A(1)} + 2 v_{A(1)}^{B(j)} + 2 v_{A(1)}^{C(k)} \end{aligned}$$

Em analogia com (2.2.7) temos que a relação entre os parâmetros é:

$$w = 2 v_{A(1)} ; \quad w_j^B = 2 v_{A(1)}^{B(j)} ; \quad w_k^C = 2 v_{A(1)}^{C(k)}$$

Para demonstrar a primeira igualdade note que:

$$\begin{aligned} w &= \frac{1}{4} \sum_j \sum_k \log \left(\frac{\pi_{1jk}}{\pi_{2jk}} \right) \\ &= \frac{1}{4} \sum_j \sum_k (\log \pi_{1jk} - \log \pi_{2jk}) \\ &= \frac{1}{4} \sum_j \sum_k \log \pi_{1jk} - \frac{1}{4} \sum_j \sum_k \log \pi_{2jk} \\ &= 2 \left(\frac{1}{8} \sum_j \sum_k \log \pi_{1jk} - \frac{1}{8} \sum_j \sum_k \log \pi_{2jk} \right) \\ &= 2 \left(\frac{2}{8} \sum_j \sum_k \log \pi_{1jk} - \frac{1}{8} \sum_j \sum_k \log \pi_{1jk} - \frac{1}{8} \sum_j \sum_k \log \pi_{2jk} \right) \\ &= 2 v_{A(1)} . \end{aligned}$$

As demais igualdades podem ser demonstradas de forma análoga.

Exemplo 2.3.2

Para mais um exemplo, consideremos novamente os dados da Tabela 1.6. Suponha agora que a distribuição Multinomial dada por (2.1.1) com $n = 960$ e $r = 8$ seja um modelo probabilístico apropriado de geração dos dados. Consideremos ainda o modelo (2.1.11) escrito na formulação (2.1.2), caso particular da formulação (2.3.1), e a matriz $A = X'_0$, onde X_0 é a matriz de especificação do modelo log-linear saturado dado em (2.1.6). Multiplicando ambos os membros do modelo correspondente a (2.1.11) pela matriz $A = X'_0$ à esquerda, obtemos:

$$X'_0 \log \pi = X'_0 1_8 \lambda + X'_0 X \beta = X_L \beta$$

onde

$$X'_0 X = X_L = 8 \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Assim, (2.1.11) pode ser escrito na formulação (2.2.2) com $A = X'_0$, $X_L = X'_0 X$ e o mesmo vetor de parâmetros β . Observe ainda que $X'_0 \log \pi$ traduz as combinações lineares dos elementos de $\log \pi$ que definem os parâmetros do modelo saturado. Reciprocamente, a partir da formulação (2.2.2) podemos obter a forma tradicional (2.3.1). Como $AA' = X'_0 X_0 = 8 I_7$ temos

$$(AA')^{-1} = \frac{1}{8} I_7 \quad \text{e} \quad A'(AA')^{-1} = \frac{1}{8} X_0 .$$

Assim, utilizando o resultado apresentado na página 42, temos:

$$X = A'(AA')^{-1} X_L = \frac{1}{8} X_0 X_L = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 \\ -1 & -1 & -1 & 1 & 1 \end{pmatrix}$$

que corresponde à matriz X dada em (2.1.11) .

Exemplo 2.3.3

Como vimos na Seção 2.2 (p. 34–36), quando o esquema de amostragem permite associar aos dados o modelo probabilístico Produto de Multinomiais, é comum utilizar uma formulação que considera modelos lineares para o vetor de logitos (múltiplos, se for o caso). Embora de uma maneira artificial, essa forma também pode ser empregada para situações em que a distribuição Multinomial é adequada. Como exemplo, consideremos os dados da Tabela 1.5 sob o modelo Multinomial. Um modelo log-linear de interesse é aquele dado por (2.1.14) que está na forma (2.1.2), correspondente a um caso particular da formulação (2.3.1):

$$\log \pi = \mathbf{1}_{12} \lambda + X \beta .$$

Multiplicando-se ambos os membros dessa expressão pela matriz $A = [I_{11}, -I_{11}]$ à esquerda, obtemos:

$$A \log \pi = A \mathbf{1}_{12} \lambda + A X \beta = X_L \beta .$$

uma vez que $A \mathbf{1}_{12} = 0$ e $A X = X_L$. Consequentemente, o modelo log-linear saturado expresso através da formulação (2.2.2), em termos de logitos múltiplos, obtido a partir de (2.1.14) tem como matriz de especificação:

$$X_L = \begin{pmatrix} 2 & 1 & 1 & 2 & 1 & 0 & -1 & -1 & -1 & -1 & -1 \\ 2 & 1 & 1 & 1 & 2 & -1 & 0 & -1 & -1 & -1 & -1 \\ 2 & 1 & 1 & 0 & 0 & -2 & -2 & -1 & -1 & -1 & -1 \\ 1 & 2 & 1 & 2 & 1 & -1 & -1 & 0 & -1 & -1 & -1 \\ 1 & 2 & 1 & 1 & 2 & -1 & -1 & -1 & 0 & -1 & -1 \\ 1 & 2 & 1 & 0 & 0 & -1 & -1 & -2 & -2 & -1 & -1 \\ 1 & 1 & 2 & 2 & 1 & -1 & -1 & -1 & -1 & 0 & -1 \\ 1 & 1 & 2 & 1 & 2 & -1 & -1 & -1 & -1 & -1 & 0 \\ 1 & 1 & 2 & 0 & 0 & -1 & -1 & -1 & -1 & -2 & -2 \\ 0 & 0 & 0 & 2 & 1 & 0 & -1 & 0 & -2 & -1 & -2 \\ 0 & 0 & 0 & 1 & 2 & -1 & -2 & -1 & -2 & -1 & -2 \end{pmatrix} \quad (2.3.8)$$

e vetor de parâmetros

$$\beta = (u_1^C \ u_2^C \ u_3^C \ u_1^S \ u_2^S \ u_{11}^{CS} \ u_{12}^{CS} \ u_{21}^{CS} \ u_{22}^{CS} \ u_{31}^{CS} \ u_{32}^{CS})'$$

Reciprocamente, vamos considerar o modelo dado por (2.3.8). Como

$$\mathbf{A}\mathbf{A}' = [\mathbf{I}_{11}, -\mathbf{1}_{11}] [\mathbf{I}_{11}, -\mathbf{1}_{11}]' = \mathbf{I}_{11} + \mathbf{1}_{11}\mathbf{1}_{11}'$$

segue que

$$(\mathbf{A}\mathbf{A}')^{-1} = \mathbf{I}_{11} - \frac{1}{12} \mathbf{1}_{11}\mathbf{1}_{11}'$$

$$\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1} = [\mathbf{I}_{11}, -\mathbf{1}_{11}]' \left[\mathbf{I}_{11} - \frac{1}{12} \mathbf{1}_{11}\mathbf{1}_{11}' \right] = \begin{pmatrix} \mathbf{I}_{11} \\ \mathbf{0}_{11}' \end{pmatrix} - \frac{1}{12} \begin{pmatrix} \mathbf{1}_{11}\mathbf{1}_{11}' \\ \mathbf{1}_{11}' \end{pmatrix}$$

o que implica

$$\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{X}_L = \begin{pmatrix} \mathbf{X}_L - \mathbf{1}_{11}\mathbf{1}_{11}' \\ -\mathbf{1}_{11}' \end{pmatrix}$$

que corresponde à matriz $\mathbf{X}$ dada em (2.1.14). Portanto o modelo “logito” (2.3.8) é uma formulação alternativa para o modelo log-linear (2.1.14).

Note que $\mathbf{A} \log \boldsymbol{\pi}$ define o vetor de logitos múltiplos $\mathbf{L}$ (11×1) com elementos

$$L_{ij} = \log \left(\frac{\pi_{ij}}{\pi_{43}} \right) = \log \pi_{ij} - \log \pi_{43} \quad i = 1, \dots, 4, \quad j = 1, \dots, 3, \quad ij \neq 43.$$

De (2.1.15) e (2.1.6) (forma polinomial do modelo (2.1.14)) temos:

$$L_{ij} = \log \left(\frac{\pi_{ij}}{\pi_{43}} \right) = \log \pi_{ij} - \log \pi_{43} = (u + u_i^C + u_j^S + u_{ij}^{CS}) - (u + u_4^C + u_3^S + u_{43}^{CS})$$

ou seja, a forma polinomial correspondente ao modelo (2.3.8) é dada por:

$$L_{ij} = u_i^C + u_j^S + u_{ij}^{CS} - (u_4^C + u_3^S + u_{43}^{CS}) \quad i = 1, \dots, 4, \quad j = 1, \dots, 3, \quad ij \neq 43 \quad (2.3.9)$$

com as restrições: $\sum_i u_i^S = \sum_j u_j^S = \sum_i u_{ij}^{CS} = \sum_j u_{ij}^{CS} = 0$.

Como a formulação polinomial é também muito empregada, mostraremos, a seguir, a equivalência entre as formulações polinomiais (2.1.15) e (2.3.9):

i) Somando-se os dois membros de (2.3.9) em relação aos índices i e j , temos:

$$L_{..} = \sum_i \sum_j L_{ij} = -12 (u_4^C + u_3^S + u_{43}^{CS})$$

o que implica

$$(u_4^C + u_3^S + u_{43}^{CS}) = -\frac{L_{..}}{12} = -\frac{L_{11} + L_{12} + L_{13} + \cdots + L_{41} + L_{42}}{12}.$$

Como $L_{ij} = \log\left(\frac{\pi_{ij}}{\pi_{43}}\right) = \log \pi_{ij} - \log \pi_{43} = l_{ij} - l_{43}$ podemos escrever:

$$(u + u_4^C + u_3^S + u_{43}^{CS}) = -\frac{(l_{..} - l_{43}) - 11 l_{43}}{12} = l_{43} - \frac{l_{..}}{12} = l_{43} - u,$$

com $l_{..} = \sum_i \sum_j l_{ij} = \sum_i \sum_j \log \pi_{ij}$, que corresponde à expressão de $(u + u_4^C + u_3^S + u_{43}^{CS})$ obtida de (2.1.15) e (2.1.16).

ii) Somando-se (2.3.9) em relação ao índice j , temos:

$$u_i^C = \frac{L_{i.}}{3} + (u_4^C + u_3^S + u_{43}^{CS}) = \frac{L_{i1} + L_{i2} + L_{i3}}{3} + (u_4^C + u_3^S + u_{43}^{CS}).$$

Substituindo L_{ij} por $(l_{ij} - l_{43})$ e também utilizando o resultado obtido no item i) temos:

$$\begin{aligned} u_i^C &= \frac{l_{i1} + l_{i2} + l_{i3} - 3 l_{43}}{3} + l_{43} - u \\ &= \frac{l_{i1} + l_{i2} + l_{i3}}{3} - l_{43} + l_{43} - u \\ &= \frac{l_{i.}}{3} - u \end{aligned}$$

que corresponde à expressão de u_i^C obtida de (2.1.15) e (2.1.16). Analogamente podemos mostrar que:

$$u_j^S = \frac{L_{.j}}{4} + (u_4^C + u_3^S + u_{43}^{CS}) = \frac{l_{.j}}{4} - u$$

corresponde à definição de u_j obtida de (2.1.15) e (2.1.16) e que

$$u_{ij}^{CS} = L_{ij} - u_i^C - u_j^S + (u_4^C + u_3^S + u_{43}^{CS}) = l_{ij} - u_i^C - u_j^S - u$$

corresponde à definição de u_{ij}^{CS} obtida de (2.1.15) e (2.1.16).

2.4 - Modelos log-lineares generalizados

Os MODELOS LOG-LINEARES GENERALIZADOS constituem uma classe mais ampla de modelos estruturais. Essa classe é motivada pelo interesse em modelar linearmente funções logarítmicas de razões de probabilidades ou chances (ver Haber (1985)). Um modelo log-linear generalizado é uma relação da forma:

$$\mathbf{A}^* \log \boldsymbol{\pi} = \mathbf{X}_L^* \boldsymbol{\beta}^* \quad (2.4.1)$$

onde $\boldsymbol{\pi}$ ($sr \times 1$) é o vetor de probabilidades desconhecidas correspondente a uma distribuição Produto de Multinomiais, $\mathbf{A}^*$ ($u \times sr$) de posto completo $u \leq s(r-1)$, é uma matriz de constantes conhecidas, $\mathbf{X}_L^*$ ($u \times k$) é uma matriz também conhecida que especifica o modelo estrutural de posto completo $k \leq u$ e $\boldsymbol{\beta}^*$ ($k \times 1$) é o vetor de parâmetros correspondente.

Note que o modelo log-linear generalizado (2.4.1) é uma generalização do modelo log-linear, definido por (2.2.2), porque a matriz de constantes utilizada na definição de funções de $\log \boldsymbol{\pi}$ pode ter $u \leq s(r-1)$ linhas e pode não ser ortogonal a $\mathbf{U}' = [\mathbf{I}_s \otimes \mathbf{1}_r]$.

Haber (1985) mostra que se $\mathbf{A}^* \mathbf{U}' = \mathbf{0}$ o modelo (2.4.1) é equivalente a um modelo log-linear da forma

$$\log \boldsymbol{\pi} = \mathbf{T} \boldsymbol{\gamma} \quad (2.4.2)$$

onde $\mathbf{T}$ [$sr \times (sr - u + k)$] é uma matriz cujas colunas geram o espaço vetorial ortogonal ao espaço vetorial gerado pelas colunas de $\mathbf{A}^{*'} \mathbf{Y}$. As colunas da matriz $\mathbf{Y}$ ($u \times (u - k)$) geram o espaço ortogonal ao espaço gerado pelas colunas da matriz $\mathbf{X}_L^*$.

Consequentemente todo modelo da forma (2.4.1) satisfazendo a condição acima pode ser analisado através da metodologia específica para modelos log-lineares. Nesse sentido convém expressar o modelo (2.4.1) numa forma análoga à (2.3.1), ou seja, na forma:

$$\log \pi = [\mathbf{I}_s \otimes \mathbf{1}_r] \lambda + \mathbf{W} \delta$$

onde λ é um vetor de constantes, $\mathbf{W}$ é uma matriz conhecida que especifica o modelo estrutural e δ é o vetor de parâmetros.

Primeiramente utilizando um argumento semelhante àquele empregado na Seção 2.3 notemos que o modelo (2.4.1) pode ser expresso como:

$$\log \pi = \mu^* + \mathbf{T}^* \beta^*$$

onde μ^* é a solução geral de $\mathbf{A}^* \log \pi = \mathbf{0}$ e $\mathbf{T}^* \beta^*$ é uma solução particular do sistema não homogêneo (2.4.1). Como pode ser visto facilmente, $\mathbf{T}^* \beta^* = \mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{X}_L^* \beta^*$ é uma solução particular nesse caso. Além disso, da teoria de sistemas lineares (Searle (1982, cap.9), por exemplo) a solução geral de $\mathbf{A}^* \log \pi = \mathbf{0}$ é da forma:

$$\{\mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{A}^* - \mathbf{I}\} \gamma$$

onde γ é um vetor de constantes arbitrárias. Consequentemente podemos escrever:

$$\log \pi = \{\mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{A}^* - \mathbf{I}\} \gamma + \mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{X}_L^* \beta^* \quad (2.4.3)$$

Note que as sr colunas de $\{\mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{A}^* - \mathbf{I}\}$ são ortogonais às colunas de $\mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{X}_L^*$. Além disso elas são linearmente dependentes (l.d.) pois

$$\text{posto}\{\mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{A}^* - \mathbf{I}\} = sr - u .$$

Para verificar esta última afirmação, note que o posto de $\{\mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{A}^* - \mathbf{I}\}$ é igual ao posto de $\mathbf{I} - \{\mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{A}^*\}$ que é idempotente. Logo

$$\begin{aligned} \text{posto}\{\mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{A}^* - \mathbf{I}\} &= \text{posto}\{\mathbf{I} - \mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{A}^*\} \\ &= \text{tr}\{\mathbf{I} - \mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{A}^*\} \\ &= sr - \text{tr}\{(\mathbf{A}^* \mathbf{A}^{*'}) (\mathbf{A}^* \mathbf{A}^{*'})^{-1}\} \\ &= sr - \text{tr}\{\mathbf{I}_u\} = sr - u . \end{aligned}$$

Esse fato implica a não identificabilidade dos parâmetros de (2.4.3). Para eliminar esse problema basta substituir $\{A^{*'}(A^*A^{*'})^{-1}A^* - I\}$ por uma matriz cujas colunas constituam uma base do espaço gerado pelas colunas de $\{A^{*'}(A^*A^{*'})^{-1}A^* - I\}$. Em particular, como queremos escrever o modelo numa forma análoga à (2.3.1) essa base deve conter os vetores correspondentes às colunas de $[I_s \otimes 1_r]$. Mais explicitamente, temos que determinar uma matriz Z ($sr \times (sr - u - s)$), com colunas linearmente independentes (l.i.) ortogonais à $[I_s \otimes 1_r]$ e à $A^{*'}(A^*A^{*'})^{-1}X_L^*$, de modo que $\{[I_s \otimes 1_r], Z\}$ seja a base procurada. Consequentemente podemos expressar o modelo (2.4.3) na forma:

$$\log \pi = [I_s \otimes 1_r]\lambda + Z\alpha + A^{*'}(A^*A^{*'})^{-1}X_L^*\beta^* \quad (2.4.4).$$

Assim, todo modelo log-linear generalizado dado por (2.4.4) com $A^*U' = A^*[I_s \otimes 1_r] = 0$ pode ser expresso como um modelo log-linear usual na forma (2.3.1) com

$$X = [Z, A^{*'}(A^*A^{*'})^{-1}X_L^*]$$

e

$$\beta = \begin{pmatrix} \alpha \\ \beta^* \end{pmatrix}.$$

Exemplo 2.4.1

Consideremos uma situação para a qual o modelo Produto de Multinomiais, envolvendo s subpopulações e $r = 4$ categorias de respostas, seja um modelo adequado. Admitamos que o interesse seja modelar linearmente a função $A^*\log \pi$ definida por $A^* = [I_s \otimes (1 \ -1 \ -1 \ 1)]$. Neste caso, temos que determinar $Z = [I_s \otimes (v_1, v_2)]$ onde v_1 e v_2 são l.i. e ortogonais à $[I_s \otimes 1_4]$ e à $A^{*'}(A^*A^{*'})^{-1}X_L^*$. O conjunto de vetores ortogonais à A^* é da forma $[I_s \otimes (x \ y \ w \ z)']$ onde $x - y - w + z = 0$, o que implica $x = y + w - z$, ou seja, $[I_s \otimes ((y + w - z) \ y \ w \ z)']$. Dentre esses, aqueles ortogonais à $[I_s \otimes 1_4]$ devem satisfazer $(y + w - z) + y + w + z = 0$, o que implica $y = -w$, ou seja,

devem ser da forma $[\mathbf{I}_s \otimes (-z \ -w \ w \ z)']$. Assim, basta tomar, por exemplo, $w = 1$ e $z = 0$ definindo $\mathbf{v}_1$ e $w = 0$ e $z = 1$ definindo $\mathbf{v}_2$ para termos

$$\mathbf{Z} = \left[\mathbf{I}_s \otimes \begin{pmatrix} 0 & -1 \\ -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \right]$$

ou $w = z = -1$ para $\mathbf{v}_1$ e $w = -z = 1$ para $\mathbf{v}_2$ para termos

$$\mathbf{Z} = \left[\mathbf{I}_s \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix} \right] .$$

Caso o interesse seja modelar a função $\mathbf{A}^* \log \boldsymbol{\pi}$ definida por $\mathbf{A}^* = [\mathbf{I}_s \otimes (1 \ 0 \ 0 \ -1)]$, o conjunto de vetores ortogonais à $\mathbf{A}^*$ deve ser da forma $[\mathbf{I}_s \otimes (x \ y \ w \ z)']$ onde $x - z = 0$, o que implica $x = z$, ou seja, tal conjunto é da forma: $[\mathbf{I}_s \otimes (z \ y \ w \ z)']$. Os vetores também ortogonais à $[\mathbf{I}_s \otimes \mathbf{1}_r]$ devem satisfazer $z + y + w + z = 0$, o que implica $y = -(w + 2z)$. Então o conjunto de vetores que são ortogonais à $[\mathbf{I}_s \otimes \mathbf{1}_4]$ e à $\mathbf{A}^{*'}(\mathbf{A}^* \mathbf{A}^*)^{-1} \mathbf{X}_L^*$ é da forma $[\mathbf{I}_s \otimes (z \ -(w + 2z) \ w \ z)']$. Consequentemente uma possível matriz $\mathbf{Z} = [\mathbf{I}_s \otimes (\mathbf{v}_1, \mathbf{v}_2)]$, onde $\mathbf{v}_1$ e $\mathbf{v}_2$ são l.i., é

$$\mathbf{Z} = \left[\mathbf{I}_s \otimes \begin{pmatrix} 0 & -1 \\ -1 & 1 \\ 1 & 1 \\ 0 & -1 \end{pmatrix} \right] .$$

O vetor $\mathbf{v}_1$ é obtido fazendo $w = 1$ e $z = 0$ e o vetor $\mathbf{v}_2$ fazendo $w = 1$ e $z = -1$.

Exemplo 2.4.3

Para um exemplo mais prático, considere os dados da Tabela 1.8 que envolve duas variáveis respostas. A essa tabela podemos associar a distribuição Produto de Multinomiais na forma (2.2.1) com $s = 6$ subpopulações (combinação dos níveis das duas variáveis explicativas), $r = 4$ categorias de resposta (combinação dos níveis das variáveis respostas: Ocorrência de Infarto e Sequela Miocárdica) e $\boldsymbol{\pi} = (\pi_{1111} \ \pi_{2111} \ \pi_{1211} \ \pi_{2211} \ \pi_{1121} \ \cdots \ \pi_{1123} \ \pi_{2123} \ \pi_{1223} \ \pi_{2223})'$ onde $0 < \pi_{ijkl} < 1$ e $\sum_i \sum_j \pi_{ijkl} = 1$ para $i, j, k = 1, 2$ e

$l = 1, 2, 3$. Para cada combinação de níveis das variáveis explicativas (Idade e Combinação de localidade e raça indexadas respectivamente por k e l) temos uma subtabela da forma dada abaixo com subtotal fixo por planejamento (n_{ijkl} representa a frequência observada na casela (i, j, k, l) e $n_{..kl} = \sum_i \sum_j n_{ijkl}$ representa o subtotal da subtabela indexada por kl).

	Sequela miocárdica		Infarto	
	não	sim	não	sim
não	n_{11kl}	n_{21kl}		
sim	n_{12kl}	n_{22kl}		
	$n_{..kl}$			

Neste caso, uma função de interesse possível é razão de produtos cruzados definida, para cada subtabela, por :

$$\theta_{kl} = \frac{\pi_{11kl} \pi_{22kl}}{\pi_{12kl} \pi_{21kl}} = \frac{n_{11kl} n_{22kl}}{n_{12kl} n_{21kl}}$$

Os efeitos das variáveis explicativas sobre as funções θ_{kl} podem ser estudados através de diversos modelos estruturais. Um modelo log-linear que considera nulo o efeito de interação entre as variáveis explicativas (idade x localidade-raça) com respeito à medida de associação θ é dado por (2.4.1) com

$$\mathbf{A}^* = \{\mathbf{I}_6 \otimes [1 \ -1 \ -1 \ 1]\}$$

$$\mathbf{X}_L^* = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix} \quad (2.4.5)$$

e

$$\boldsymbol{\beta}^* = (u \ \alpha_1 \ \beta_1 \ \beta_2)'$$

ou na forma polinomial por:

$$\log \theta_{kl} = u + \alpha_k + \beta_l \quad k = 1, 2 \text{ e } l = 1, \dots, 3 \quad (2.4.6)$$

com as restrições: $\sum_k \alpha_k = \sum_l \beta_l = 0$.

Para uma situação em que a variável indexada por k tem K níveis e a variável indexada por l tem L níveis, os parâmetros podem ser definidos e interpretados por:

$$u = \frac{1}{KL} \sum_k \sum_l \log \theta_{kl} \quad : \text{ média dos } \log \theta_{kl};$$

$$\alpha_k = \frac{1}{L} \sum_l \log \theta_{kl} - u \quad : \text{ desvio da média dos } \log \theta_{kl} \text{ correspondentes ao } k\text{-ésimo grupo etário, } k = 1, 2, \text{ em relação a } u;$$

$$\beta_l = \frac{1}{K} \sum_k \log \theta_{kl} - u \quad : \text{ desvio da média dos } \log \theta_{kl} \text{ correspondentes à } l\text{-ésima combinação de localidade e raça, } l = 1, 2, 3, \text{ em relação a } u.$$

Neste caso $K = 2$ e $L = 3$. A matriz de especificação do modelo (2.4.5) é igual à utilizada para experimentos fatoriais em ANOVA. Empregando também a linguagem de ANOVA, podemos interpretar os parâmetros α_k e β_l como sendo respectivamente os efeitos principais das variáveis Idade e Combinação de localidade e raça. Esse modelo estabelece que o efeito dessas variáveis explicativas sobre o $\log \theta_{kl}$ é aditivo.

Utilizando resultados do exemplo anterior podemos reescrever o modelo (2.4.5) na forma (2.4.4) com

$$\mathbf{Z} = \left[\mathbf{I}_6 \otimes \begin{pmatrix} 0 & -1 \\ -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \quad \text{ou} \quad \mathbf{Z} = \left[\mathbf{I}_s \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix} \right]$$

por exemplo. Consequentemente o modelo (2.4.5) estará também expresso na formulação (2.3.1).

Embora não consideremos neste trabalho a classe de modelos estruturais log-lineares generalizados em detalhe, como ilustração, apresentaremos a seguir dois modelos dessa

classe que não são modelos log-lineares. Aos leitores interessados nesse tipo de modelos sugerimos ver Koch et al. (1985) ou Haber (1985) por exemplo.

Consideremos a Tabela 1.7. Combinando-se os níveis de Tipo de diagnóstico (“complicado” e “não complicado”) e Tratamento (A, B e C) definem-se 6 grupos de pacientes. Podemos assumir que os dados seguem uma distribuição Produto de Multinomiais com função de probabilidades dada por (2.2.1) com $s = 6$ subpopulações, $r = 2$ categorias de resposta (“curado” e “não curado”) e $\pi = (\pi_{111} \ \pi_{112} \ \pi_{121} \ \pi_{122} \ \cdots \ \pi_{231} \ \pi_{232})'$ com $0 < \pi_{ijk} < 1$ ($i, k = 1, 2$; $j = 1, 2, 3$) e $\sum_k \pi_{ijk} = 1$ para todo $i = 1, 2$ e $j = 1, 2, 3$. Uma característica das distribuições de resposta que pode ser de interesse é a proporção de indivíduos com determinado perfil de resposta, por exemplo proporção de “curados”, e um possível modelo saturado que dá ênfase a esse aspecto corresponde a (2.4.1) com $A^* = \{I_6 \otimes [1 \ 0]\}$ definindo o vetor de probabilidades de cura,

$$X_L^* = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} \quad (2.4.7)$$

e

$$\beta^* = (v \ v_1^D \ v_1^T \ v_2^T \ v_{11}^{DT} \ v_{12}^{DT})'$$

Na forma polinomial esse modelo pode ser escrito como:

$$\log \pi_{ij1} = v + v_i^D + v_j^T + v_{ij}^{DT} \quad i = 1, 2; \ j = 1, 2, 3 \quad (2.4.8)$$

com as restrições: $\sum_i v_i^D = \sum_j v_j^T = \sum_i v_{ij}^{DT} = \sum_j v_{ij}^{DT} = 0$

Como decorrência dessas restrições temos:

$$v = \frac{1}{6} \sum_i \sum_j \log \pi_{ij1}, \quad v_i^D = \frac{1}{3} \sum_j \log \pi_{ij1} - v, \quad v_j^T = \frac{1}{2} \sum_i \log \pi_{ij1} - v,$$

e $v_{ij}^{DT} = \log \pi_{ij1} - v_i^D - v_j^T - v.$

Deste modo, em analogia com ANOVA, os parâmetros podem ser interpretados como se segue:

- v : média dos $\log \pi_{ij1}$;
- v_i^D : efeito principal do i -ésimo diagnóstico no $\log \pi_{ij1}$;
- v_j^T : efeito principal do j -ésimo tratamento no $\log \pi_{ij1}$;
- v_{ij}^{DT} : interação entre diagnóstico e tratamento, níveis i e j , no $\log \pi_{ij1}$.

O modelo estrutural (2.4.7) não é log-linear pois a matriz $A^* = \{I_6 \otimes [1 \ 0]\}$ não é ortogonal a $U' = [I_6 \otimes 1_2]$.

Outro modelo estrutural para a mesma função definida por A^* , que considera iguais as taxas de cura para pacientes com diagnóstico “não complicado” (independentemente do tratamento a que foram submetidos) e para pacientes com diagnóstico “complicado” e tratamento B é dado por (2.4.1) com

$$A^* = \{I_6 \otimes [1 \ 0]\}$$

$$X_L^* = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad (2.4.9)$$

$$\beta^* = (\gamma_1 \ \gamma_2 \ \gamma_3)'$$

onde γ_1 é o valor esperado para $\log \pi_{ij1}$ correspondente a diagnóstico “não complicado” ou tratamento B e diagnóstico “complicado”, γ_2 é a variação devida ao tratamento A para o diagnóstico complicado e γ_3 é a variação devida ao tratamento C para o diagnóstico complicado. Os parâmetros γ_2 e γ_3 são interpretados como efeitos diferenciais, indicando o quanto aumenta (ou diminui) o $\log \pi_{ij1}$ em relação ao valor da casela de referência quando pacientes com diagnóstico “complicado” são submetidos aos tratamentos A e C respectivamente.

Para efeito de comparação, note que um modelo log-linear para os mesmos dados pode ser obtido tomando A^* ortogonal à $U' = [I_6 \otimes 1_2]$. Por exemplo, basta tomar $A^* = \{I_6 \otimes [1 \ -1]\}$ e X_L^* obtido de (2.4.9). Note que a interpretação dos parâmetros seria distinta, uma vez que os “efeitos” mencionados seriam sobre os logaritmos das razões entre probabilidade de “curados” e de “não curados” e não mais sobre o logaritmo da probabilidade de cura. Os resultados das duas análises podem ser completamente diferentes como pode ser constatado da análise dos dados da Tabela 1.7 apresentada por Koch et al. (1985, p. 120–128, 207–208).

2.5 - Uma classificação para modelos estruturais log-lineares

A classificação de modelos log-lineares em *hierárquicos* e *não hierárquicos* é de especial importância na obtenção das frequências esperadas, pois para os modelos hierárquicos podemos facilmente determiná-las a partir das estatísticas suficientes, não sendo necessário o cálculo intermediário das estimativas dos parâmetros associados a esses modelos. A todo modelo log-linear hierárquico corresponde um conjunto de estatísticas suficientes. Para alguns modelos hierárquicos as frequências esperadas são funções explícitas das estatísticas suficientes, enquanto que para outros necessita-se de um procedimento iterativo para sua obtenção a partir do conjunto de estatísticas suficientes. Bishop et al. (1975, cap.3) apresentam algumas regras para a distinção entre esses dois tipos de modelos hierárquicos e descrevem o procedimento iterativo IPF (Ajuste Proporcional Iterativo) desenvolvido por Deming & Stephan (1940) para o cálculo de tais frequências quando não existem formas explícitas. Um exemplo elementar de situação onde existe forma explícita é o de modelos de independência em tabelas $s \times r$, para os quais o teste de Qui-quadrado de Pearson é apropriado.

Um modelo log-linear é classificado como sendo hierárquico se obedece à seguinte regra: “ Se um parâmetro correspondente a um grupo de s variáveis é incluído no modelo então todos os parâmetros correspondentes a qualquer subgrupo das s variáveis devem também estar no modelo. Por outro lado, se algum parâmetro é excluído do modelo, todos os parâmetros de ordem mais alta que são seus parentes^(*) também devem ser excluídos do modelo. (ver Bishop et al. (1975, p. 34, 67–68), Haberman (1978, p. 160–161), Agresti (1984, p. 38–43) por exemplo). Essencialmente isto equivale a dizer, por exemplo, que se a interação AB for incluída em um modelo, então os efeitos principais de A e B também devem ser incluídos. Por outro lado, se um dos efeitos citados é excluído do modelo (B , por exemplo), quaisquer outros parâmetros de maior ordem que o envolva também devem ser excluídos do modelo (AB , por exemplo).

Dessa forma um modelo que inclui um ou mais parâmetros de “interação” é denominado hierárquico se todos os parâmetros de ordem mais baixa tendo os mesmos índices são também incluídos. Os modelos log-lineares (2.1.6), (2.1.9), (2.1.11), (2.1.13), (2.1.14), (2.1.17), (2.2.6), (2.2.8), (2.4.5), (2.3.3), (2.3.5) e (2.3.8), apresentados nas seções anteriores, são exemplos de modelos hierárquicos.

Todo parâmetro de um modelo hierárquico pode ser escrito como uma combinação linear dos logaritmos das frequências esperadas ou equivalentemente, logaritmos das probabilidades das caselas (ver exemplo (2.3.2)). Cada parâmetro pode ainda ser expresso como uma função da razão dos produtos cruzados (ver Bishop et al. (1975, p. 17–18, 34–35), Haberman (1978, p. 180–181)). Por exemplo, de (2.1.7) e (2.1.8) obtemos:

$$u = \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk}$$

(*) Dois parâmetros, um com r índices (r variáveis) e outro com s índices ($r > s$), são PARENTES se o conjunto r contém todo o conjunto de s índices (Bishop et al. (1975, p.34)). u_{ij}^{12} e u_i^1 ; ou u_{ijkl}^{1234} , u_{ikl}^{134} , u_{kl}^{34} são exemplos de parâmetros parentes .

$$\begin{aligned}
u_{11}^{AB} &= \frac{1}{2} \sum_k \log \pi_{11k} - u_1^A - u_1^B - u \\
&= \frac{1}{2} \sum_k \log \pi_{11k} - \left(\frac{1}{4} \sum_j \sum_k \log \pi_{1jk} - u \right) - \left(\frac{1}{4} \sum_i \sum_k \log \pi_{i1k} - u \right) - u \\
&= \frac{1}{2} \sum_k \log \pi_{11k} - \frac{1}{4} \sum_j \sum_k \log \pi_{1jk} - \frac{1}{4} \sum_i \sum_k \log \pi_{i1k} + u \\
&= \frac{\log \pi_{111} + \log \pi_{112}}{2} - \frac{\log \pi_{111} + \log \pi_{121} + \log \pi_{112} + \log \pi_{122}}{4} \\
&\quad - \frac{\log \pi_{111} + \log \pi_{112} + \log \pi_{211} + \log \pi_{212}}{4} + \frac{\log \pi_{111} + \dots + \log \pi_{222}}{8} \\
&= \frac{1}{8} [(\log \pi_{111} + \log \pi_{221} - \log \pi_{211} - \log \pi_{211}) + (\log \pi_{112} + \log \pi_{221} - \log \pi_{122} - \log \pi_{212})]
\end{aligned}$$

ou seja,

$$u_{11}^{AB} = \frac{1}{8} \left[\log \frac{\pi_{111} \pi_{221}}{\pi_{121} \pi_{211}} + \log \frac{\pi_{112} \pi_{222}}{\pi_{122} \pi_{212}} \right] = \frac{1}{8} \log(\theta^{(1)} \theta^{(2)})$$

onde $\theta^{(k)}$ é o produto cruzado para o k -ésimo nível da variável C . Análogamente podemos demonstrar que $u_{111}^{ABC} = \log(\theta^{(1)} / \theta^{(2)})$.

Imrey et al. (1981) apresentam uma definição de modelos log-lineares hierárquicos em termos da formulação matricial: “Um modelo log-linear é hierárquico se um conjunto C de colunas de X_i , $i = 1, \dots, s$, pode ser escrito como $C = \bigcup_l C_l$, onde cada C_l consiste de colunas definindo um modelo fatorial completo para os elementos de n_i , $i = 1, \dots, s$ ”. Em outras palavras, a matriz X não pode conter colunas relativas a interações se não incluir as colunas referentes aos efeitos principais correspondentes.

Além disso, parametrizações de modelos log-lineares hierárquicos, semelhantes às encontradas em Análise de Variância de experimentos fatoriais, facilitam a construção das matrizes de especificação desses modelos. Por exemplo, observemos a matriz X dada em (2.1.14) que especifica um modelo log-linear saturado hierárquico para a Tabela 1.5 sob um modelo probabilístico Multinomial. As colunas de 1 a 5 referem-se aos parâmetros

correspondentes a “efeitos principais” das variáveis, u_1^C , u_2^C , u_3^C , u_1^S , u_2^S ; as demais colunas aos parâmetros relativos a associação entre Tipo de cirurgia e Severidade da sequela, u_{11}^{CS} , u_{12}^{CS} , u_{21}^{CS} , u_{22}^{CS} , u_{31}^{CS} e u_{32}^{CS} . Como em modelos fatoriais de ANOVA essas últimas colunas podem ser obtidas multiplicando-se elementos correspondentes de determinadas colunas de “efeitos principais”; por exemplo, multiplicando-se os elementos correspondentes das colunas (1) e (4), coeficientes relativos a u_1^C e u_1^S , obtem-se os valores da coluna (6) que são os coeficientes relativos ao parâmetro u_{11}^{CS} . Um outro exemplo ilustrativo desse fato pode ser visto em (2.1.7). A coluna da matriz $\mathbf{X}$ relativa ao parâmetro u_{111}^{ABC} (coluna (7)) pode ser obtida pela multiplicação dos elementos de uma coluna referente a “efeito principal” pelos elementos correspondentes a uma coluna referente a “interação” de outras duas variáveis (colunas (1) e (6) por exemplo); ou, simplesmente, pela multiplicação dos elementos das colunas relativas aos “efeitos principais” (colunas (1), (2) e (3)).

Outra peculiaridade dos modelos log-lineares hierárquicos é a possibilidade de tirar conclusões sobre modelos hierárquicos definidos sob modelos probabilísticos Produto de Multinomiais a partir de modelos definidos sob esquemas Multinomiais. Consideremos por exemplo os dados da Tabela 1.6 e vamos supor que as variáveis B e C possam ser consideradas explicativas e a variável A como resposta. Um modelo log-linear hierárquico de interesse é aquele dado por (2.2.6) ou (2.2.7). Mostraremos que esse modelo, proposto sob o esquema Produto de Multinomiais, pode ser obtido do modelo (2.1.9) ou (2.1.10), proposto sob o esquema Multinomial. Empregando o resultado demonstrado na Seção 2.3, podemos verificar que tomando $\mathbf{X}_L = \mathbf{A}\mathbf{X}\mathbf{B}$, com $\mathbf{A} = \{\mathbf{I}_4 \otimes [1 \quad -1]\}$, $\mathbf{X}$ dado em (2.1.9) e

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

obtemos a matriz de especificação correspondente ao modelo (2.2.6) multiplicada por 2.

Multiplicando, à direita, o vetor de parâmetros correspondente ao modelo (2.1.9) pela matriz $\mathbf{B}'$, obtemos o vetor de parâmetros correspondente ao modelo (2.2.6), multiplicado por $\frac{1}{2}$. Assim, os parâmetros em (2.2.6) correspondem àqueles parâmetros do modelo (2.1.9) que envolvem a variável resposta (A), multiplicados por 2. O produto $\mathbf{A} \mathbf{X}$ gera colunas com elementos iguais a zero. A matriz $\mathbf{B}$ tem a finalidade de selecionar as colunas não nulas desse produto e sua correspondente matriz transposta seleciona os parâmetros em β que envolvem a variável resposta A . O número de colunas da matriz $\mathbf{B}$ é igual a esse número de parâmetros.

Empregando a forma polinomial, essa correspondência entre modelos também é facilmente evidenciada. Podemos escrever o logito L_{jk} a partir de (2.1.10) como sendo:

$$\begin{aligned} L_{jk} &= \log \left(\frac{\pi_{1jk}}{\pi_{2jk}} \right) = \log \pi_{1jk} - \log \pi_{2jk} \\ &= (u_1^A - u_2^A) + (u_{1j}^{AB} - u_{2j}^{AB}) + (u_{1k}^{AC} - u_{2k}^{AC}) \\ &= 2 u_1^A + 2 u_{1j}^{AB} + 2 u_{1k}^{AC} \quad \text{para } j, k = 1, 2 \end{aligned}$$

Pode ser mostrado que a relação entre esses parâmetros e os parâmetros do modelo (2.2.7) é:

$$w = 2 u_1^A, \quad w_j^B = 2 u_{1j}^{AB} \quad \text{e} \quad w_k^C = 2 u_{1k}^{AC}.$$

Por exemplo, verificando a primeira igualdade temos:

$$\begin{aligned} 2 u_1^A &= 2 \left(\frac{1}{4} \sum_j \sum_k \log \pi_{1jk} - \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk} \right) \\ &= \frac{2}{4} \sum_j \sum_k \log \pi_{1jk} - \frac{1}{4} \left(\sum_j \sum_k \log \pi_{1jk} + \sum_j \sum_k \log \pi_{2jk} \right) \\ &= \frac{1}{4} \sum_j \sum_k \log \pi_{1jk} - \frac{1}{4} \sum_j \sum_k \log \pi_{2jk} = \frac{1}{4} \left(\sum_j \sum_k \log \pi_{1jk} - \sum_j \sum_k \log \pi_{2jk} \right) \\ &= \frac{1}{4} [(\log \pi_{111} + \log \pi_{112} + \log \pi_{121} + \log \pi_{122}) - (\log \pi_{211} + \log \pi_{212} + \log \pi_{221} + \log \pi_{222})] \\ &= \frac{1}{4} (L_{11} + L_{12} + L_{21} + L_{22}) = \frac{1}{4} \sum_j \sum_k L_{jk} = w. \end{aligned}$$

Analogamente podemos mostrar as demais igualdades.

Ao se obter o modelo linear para os logitos L_{jk} (2.2.6) ou (2.2.7) a partir de (2.1.9) ou (2.1.10) a restrição $\sum_i \pi_{ijk} = 1$, para $j, k = 1, 2$, foi incluída, transformando o modelo Multinomial de geração de dados ($r = 8$, $n = 960$) no modelo Produto de Multinomiais ($r = 2$, $s = 4$) por condicionamento em relação às marginais definidas pelas variáveis B e C (época de plantio e comprimento do corte). Os parâmetros do modelo (2.1.9) passaram a ter a mesma interpretação que aqueles relativos ao modelo (2.2.6) ou (2.2.7). Mais explicitamente, $2 u_1^A$ que corresponde a definição de w passou a ser interpretado como a média dos logitos, $2 u_{1j}^{AB} = w_j^B$ como o efeito do nível j da variável B no logito L_{jk} e $2 u_{1k}^{AC} = w_k^C$ como o efeito do nível k da variável C no logito L_{jk} .

Também os modelos (2.1.9) e (2.3.5) guardam estreita relação entre si embora sejam associados a distintos modelos probabilísticos de geração dos dados. Considerando suas expressões na forma polinomial, (2.1.10) e (2.3.6) respectivamente, mostraremos a relação entre eles. Somando ambos os membros das expressões (2.1.10) e (2.3.6) em relação ao índice i e empregando as restrições respectivamente associadas a esses modelos obtemos:

$$u + u_j^B + u_k^C + u_{jk}^{BC} = v^{BC(jk)}.$$

As relações entre os parâmetros são obtidas abaixo. De (2.1.10) temos:

i)

$$u_i^A = \frac{1}{4} \sum_j \sum_k \log \pi_{ijk} - u = \frac{1}{4} \sum_j \sum_k \log \pi_{ijk} - \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk}$$

que corresponde a definição de $v_{A(i)}$ obtida a partir de (2.3.6);

ii)

$$\begin{aligned} u_{ij}^{AB} &= \frac{1}{2} \sum_k \log \pi_{ijk} - u_i^A - u_j^B - u \\ &= \frac{1}{2} \sum_k \log \pi_{ijk} - \frac{1}{4} \sum_j \sum_k \log \pi_{ijk} - \frac{1}{4} \sum_i \sum_k \log \pi_{ijk} + \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk} \end{aligned}$$

que corresponde a definição de $v_{A(i)}^{B(j)}$ obtida a partir de (2.3.6);

iii)

$$\begin{aligned} u_{ij}^{AC} &= \frac{1}{2} \sum_j \log \pi_{ijk} - u_i^A - u_k^C - u \\ &= \frac{1}{2} \sum_j \log \pi_{ijk} - \frac{1}{4} \sum_j \sum_k \log \pi_{ijk} - \frac{1}{4} \sum_i \sum_j \log \pi_{ijk} + \frac{1}{8} \sum_i \sum_j \sum_k \log \pi_{ijk} \end{aligned}$$

que corresponde a definição de $v_{A(i)}^{C(k)}$ obtida a partir de (2.3.6).

Salienta-se, entretanto, que a interpretação dos parâmetros para os dois modelos é completamente distinta, muito embora do ponto de vista algébrico ocorram as igualdades demonstradas anteriormente. Note que embora $u_{ij}^{AB} = v_{A(i)}^{B(j)}$, u_{ij}^{AB} é uma medida de associação entre as variáveis A e B , enquanto $v_{A(i)}^{B(j)}$ mede o efeito da variável B na distribuição da variável resposta A .

Os resultados mencionados nos dois exemplos anteriores são muito úteis sobretudo do ponto de vista computacional. O programa computacional BMDP/4F (ver Dixon (1985)), por exemplo, ajusta modelos log-lineares hierárquicos ao vetor $\log m$ tratando todas as variáveis como variáveis respostas, o que implica associar aos dados o modelo Multinomial. A partir das relações discutidas acima é possível utilizar esse programa para inferência sobre modelos hierárquicos sob esquema probabilístico Produto de Multinomiais. Isto será exemplificado no Capítulo 4.

Todo modelo log-linear que não satisfaz o princípio hierárquico é chamado de modelo log-linear não hierárquico. Um exemplo é:

$$\log m_{ij} = u + u_j^2 + u_{ij}^{12}$$

com, $\sum_j u_j^2 = \sum_i u_{ij}^{12} = \sum_j u_{ij}^{12} = 0$. Esse modelo admite associação entre as variáveis 1 e 2 mas força a igualdade entre os níveis da variável 1 quanto a média dos logaritmos das frequências esperadas (ver Agresti (1984, p. 43)). Esse modelo não é hierárquico pois

o parâmetro u_i^1 foi excluído do modelo embora o parâmetro u_{ij}^{12} tenha sido incluído no modelo.

Os modelos não hierárquicos são muito utilizados quando pelo menos uma das variáveis é ordinal, pois esse tipo de modelo possibilita incorporar a estrutura de ordem de tais variáveis. É o caso do modelo (2.1.20). Ele corresponde ao modelo hierárquico (2.1.15) com o termo geral de associação, u_{ij}^{CS} , substituído por $\gamma(a_i - \bar{a})(b_j - \bar{b})$. Esse modelo denominado “modelo de associação uniforme” além de incluir a ordinalidade das duas variáveis Tipo de cirurgia e Severidade da seqüela apresenta um número reduzido de parâmetros. O modelo (2.2.10) também é um exemplo de modelo não hierárquico.

Na forma matricial, as matrizes de especificação de modelos log-lineares não hierárquicos são semelhantes às matrizes utilizadas em modelos de Análise de Variância não hierárquicos ou em modelos de Regressão, cujos elementos podem ser quaisquer valores reais. Como exemplo, veja o modelo (2.1.19).

Capítulo 3

Métodos de Estimação e Testes de Hipóteses

Neste capítulo discutiremos os dois principais métodos clássicos para estimação dos parâmetros de modelos estruturais log-lineares: Máxima Verossimilhança (MV) e Mínimos Quadrados Generalizados (MQG). O modelo probabilístico considerado corresponde à distribuição Produto de Multinomiais; no entanto, os resultados apresentados também se aplicam ao modelo Multinomial, que constitui um caso particular dessa distribuição. A Seção 3.1 é dedicada ao método de Máxima Verossimilhança e a Seção 3.2 aborda a metodologia de Mínimos Quadrados Generalizados.

Ainda dentro da abordagem clássica é possível utilizar o método de Regressão Funcional Assintótica (RFA) discutido em Imrey et al. (1981, 1982), por exemplo. Este método consiste de duas etapas: na primeira obtêm-se as estimativas de MV dos parâmetros de

um modelo log-linear inicial, e na segunda emprega-se o método de MQG para reduzir esse modelo a um modelo de interesse com menor número de parâmetros. O método RFA tem maior interesse no ajuste de modelos mais gerais que não apresentam estrutura log-linear principalmente quando as frequências amostrais podem comprometer a aplicação do método de MQG diretamente. Por essa razão esse método não será objeto de discussão neste trabalho. Aos leitores interessados nessa metodologia recomendamos também consultar Koch et al. (1985).

3.1 - O método de Máxima Verossimilhança

Inicialmente indicaremos os principais passos para a obtenção dos estimadores de MV. Maiores detalhes além de outras referências podem ser obtidos em Imrey et al. (1981,1982).

Consideremos a função de verossimilhança (2.2.1) associada à distribuição Produto de Multinomiais e vamos admitir o modelo estrutural (2.3.7). O núcleo de $\log \mathbf{P}$ é igual a:

$$\begin{aligned} \mathbf{n}' \log \pi &= \mathbf{n}' \log \pi(\beta) \\ &= \mathbf{n}' \log [D_{\eta}^{-1} \exp(\mathbf{X}\beta)] = \sum_i \sum_j n_{ij} \log \pi_{ij}(\beta) \end{aligned} \quad (3.1.1)$$

onde: $\mathbf{n}' = (\mathbf{n}'_1, \dots, \mathbf{n}'_s)$, com $\mathbf{n}'_i = (n_{i1}, \dots, n_{ir})$, para $i = 1, \dots, s$, e

$$\pi_{ij}(\beta) = \frac{\exp(\mathbf{x}'_{ij}\beta)}{\sum_{j=1}^r \exp(\mathbf{x}'_{ij}\beta)} = \frac{\exp(\sum_{k=1}^t x_{ijk}\beta_k)}{\sum_{j=1}^r \exp(\sum_{k=1}^t x_{ijk}\beta_k)}$$

com $\mathbf{x}'_{ij} = (x_{ij1} \ x_{ij2} \ \dots \ x_{ijt})$ representando a linha da matriz $\mathbf{X}$ associada à j -ésima categoria de resposta da i -ésima subpopulação. Desenvolvendo (3.1.1), temos:

$$\begin{aligned} \mathbf{n}' \log \pi &= \sum_i \sum_j n_{ij} \log \pi_{ij}(\beta) \\ &= \sum_i \sum_j n_{ij} \log [\exp(\mathbf{x}'_{ij}\beta) / \sum_j \exp(\mathbf{x}'_{ij}\beta)] \\ &= \sum_i \sum_j n_{ij} \mathbf{x}'_{ij}\beta - \sum_i \sum_j n_{ij} \log [\sum_j \exp(\mathbf{x}'_{ij}\beta)] \\ &= \sum_i \sum_j n_{ij} \sum_k x_{ijk}\beta_k - \sum_i n_i \log [\sum_j \exp(\sum_k x_{ijk}\beta_k)] \end{aligned} \quad (3.1.2)$$

Então:

$$\begin{aligned}
 \frac{\partial}{\partial \beta_l} (\mathbf{n}' \log \pi) &= \frac{\partial}{\partial \beta_l} \left(\sum_i \sum_j n_{ij} \sum_k x_{ijk} \beta_k - \sum_i n_i \log \left[\sum_j \exp \left(\sum_k x_{ijk} \beta_k \right) \right] \right) \\
 &= \sum_i \sum_j n_{ij} x_{ijl} - \sum_i n_i \frac{\sum_j x_{ijl} \exp \left(\sum_k x_{ijk} \beta_k \right)}{\sum_j \exp \left(\sum_k x_{ijk} \beta_k \right)} \\
 &= \sum_i \sum_j n_{ij} x_{ijl} - \sum_i n_i \sum_j x_{ijl} \pi_{ij}(\beta) \\
 &= \mathbf{x}'_l \mathbf{n} - \mathbf{x}'_l \mathbf{m} \quad l = 1, \dots, t
 \end{aligned}$$

com $\mathbf{x}'_{l(1 \times sr)} = (\mathbf{x}'_{1l}, \dots, \mathbf{x}'_{sl})$ representando a coluna da matriz $\mathbf{X}$ associada ao l -ésimo parâmetro e $\mathbf{m} = \mathbf{m}(\beta) = \mathbf{D}_{\pi(\beta)} [\mathbf{I}_s \otimes \mathbf{1}_r \mathbf{1}'_r] \mathbf{n}$ é o vetor de frequências esperadas sob o modelo (2.3.7). Logo a derivada de $\mathbf{n}' \log \pi$ em relação a β pode ser escrita na forma compacta:

$$\frac{\partial}{\partial \beta} (\mathbf{n}' \log \pi) = \mathbf{X}' \mathbf{n} - \mathbf{X}' \mathbf{m} \quad (3.1.3)$$

Também podemos calcular :

$$\begin{aligned}
 \frac{\partial^2}{\partial \beta_l \partial \beta_m} (\mathbf{n}' \log \pi) &= \frac{\partial}{\partial \beta_m} \left(\sum_i \sum_j n_{ij} x_{ijl} - \sum_i n_i \frac{\sum_j x_{ijl} \exp \left(\sum_k x_{ijk} \beta_k \right)}{\sum_j \exp \left(\sum_k x_{ijk} \beta_k \right)} \right) \\
 &= - \sum_i n_i \left[\frac{\sum_j x_{ijl} x_{ijm} \exp \left(\sum_k x_{ijk} \beta_k \right)}{\left[\sum_j \exp \left(\sum_k x_{ijk} \beta_k \right) \right]^2} - \frac{\sum_j x_{ijl} \exp \left(\sum_k x_{ijk} \beta_k \right) \sum_j x_{ijm} \exp \left(\sum_k x_{ijk} \beta_k \right)}{\left[\sum_j \exp \left(\sum_k x_{ijk} \beta_k \right) \right]^2} \right] \\
 &= - \sum_i n_i \left[\sum_j x_{ijl} x_{ijm} \pi_{ij}(\beta) - \left(\sum_j x_{ijl} \pi_{ij}(\beta) \right) \left(\sum_j x_{ijm} \pi_{ij}(\beta) \right) \right] \\
 &= - \sum_i n_i \left[\mathbf{x}'_{il} \text{DIAG}(\pi_i(\beta)) \mathbf{x}_{im} - \mathbf{x}'_{il} \pi_i(\beta) \pi_i(\beta)' \mathbf{x}_{im} \right] \\
 &= - \sum_i n_i \mathbf{x}'_{il} \left[\text{DIAG}(\pi_i(\beta)) - \pi_i(\beta) \pi_i(\beta)' \right] \mathbf{x}_{im} \quad l = 1, \dots, t
 \end{aligned}$$

Consequentemente :

$$\frac{\partial^2}{\partial \beta \partial \beta'} (\mathbf{n}' \log \pi) = - \sum_i n_i \mathbf{X}'_i \left[\text{DIAG}(\pi_i(\beta)) - \pi_i(\beta) \pi_i(\beta)' \right] \mathbf{X}_i \quad (3.1.4)$$

onde $\mathbf{X}_i$ ($r \times t$) é a sub-matriz de $\mathbf{X} = [\mathbf{X}'_1 \cdots \mathbf{X}'_s]$ corresponde a i -ésima subpopulação e $\mathbf{DIAG}(\boldsymbol{\pi}_i)$ é uma matriz diagonal com os elementos do vetor $\boldsymbol{\pi}_i$ na diagonal principal.

O estimador de MV $\mathbf{b}$ de $\boldsymbol{\beta}$ pode ser obtido como a solução do conjunto de equações:

$$\frac{\partial \log \mathbf{P}}{\partial \boldsymbol{\beta}} = \mathbf{0}_t$$

que corresponde à solução de:

$$\mathbf{X}'\hat{\mathbf{m}} - \mathbf{X}'\mathbf{n} = \mathbf{0}_t \quad (3.1.5)$$

onde $\hat{\mathbf{m}} = \mathbf{m}(\mathbf{b}) = \mathbf{D}_{\boldsymbol{\pi}(\mathbf{b})} [\mathbf{I}_s \otimes \mathbf{1}_r \mathbf{1}'_r] \mathbf{n}$ é o vetor de estimativas das frequências esperadas sob o modelo (2.3.7). Aspectos relacionados com a resolução dessas equações serão discutidos no final desta seção.

A matriz de Informação de Fisher correspondente é dada por:

$$\mathbf{I} = -\mathbf{E} \left(\frac{\partial^2 \log \mathbf{P}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right) = \sum_{i=1}^s n_i \mathbf{X}'_i [\mathbf{DIAG}(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i \boldsymbol{\pi}'_i] \mathbf{X}_i$$

o que assegura que as soluções de (3.1.5) são pontos de máximo uma vez que essa matriz é positiva definida.

De resultados bem conhecidos sobre estimadores de MV (ver Andersen (1974), Bishop et al. (1975, cap. 14), Imrey et al. (1981) por exemplo), segue que sob o modelo (2.3.7), $\sqrt{n}(\mathbf{b} - \boldsymbol{\beta})$ tem distribuição normal multivariada assintótica com matriz de covariância dada por:

$$\text{Var}_A(\mathbf{b}) = \left\{ \sum_{i=1}^s n_i \mathbf{X}'_i [\mathbf{DIAG}(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i \boldsymbol{\pi}'_i] \mathbf{X}_i \right\}^{-1} \quad (3.1.6)$$

onde o índice $_A$ de $\text{Var}_A(\mathbf{b})$ indica valor assintótico. Um estimador consistente para $\text{Var}_A(\mathbf{b})$ pode ser obtido substituindo-se $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i(\boldsymbol{\beta})$ por $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i(\mathbf{b})$ em (3.1.6):

$$\widehat{\text{Var}}_A(\mathbf{b}) = \left\{ \sum_{i=1}^s n_i \mathbf{X}'_i [\mathbf{DIAG}(\hat{\boldsymbol{\pi}}_i) - \hat{\boldsymbol{\pi}}_i \hat{\boldsymbol{\pi}}'_i] \mathbf{X}_i \right\}^{-1} \quad (3.1.7)$$

A matriz de covariância assintótica de $\hat{\pi}$ pode ser obtida aplicando-se o método DELTA baseado na expansão de $\hat{\pi}$, expresso como função de $\mathbf{b}$ em série linear de Taylor (ver Bishop et al. (1975, seção 14.6), Agresti (1984, apêndice C) por exemplo); então:

$$\text{Var}_A(\hat{\pi}) = [\mathbf{H}(\pi)] [\text{Var}_A(\mathbf{b})] [\mathbf{H}(\pi)]' \quad (3.1.8)$$

onde:

$$[\mathbf{H}(\pi)]' = \{ \mathbf{X}'_1 [\text{DIAG}(\pi_1) - \pi_1 \pi'_1], \dots, \mathbf{X}'_s [\text{DIAG}(\pi_s) - \pi_s \pi'_s] \}.$$

Um estimador consistente de $\text{Var}_A(\hat{\pi})$ pode ser obtido substituindo-se π_i por $\hat{\pi}_i$ em (3.1.8). Dessa forma,

$$\widehat{\text{Var}}_A(\hat{\pi}) = [\mathbf{H}(\hat{\pi})] [\widehat{\text{Var}}_A(\mathbf{b})] [\mathbf{H}(\hat{\pi})]' \quad (3.1.9)$$

O ajuste do modelo log-linear (2.3.7) pode ser testado pela estatística da Razão de Verossimilhança de Wilks :

$$Q_L = 2 \sum_i \sum_j n_{ij} \log \left(\frac{n_{ij}}{\hat{m}_{ij}} \right) \quad (3.1.10)$$

baseada no modelo probabilístico (2.2.1) onde $\hat{m}_{ij}$ são os elementos do vetor de estimativas das frequências esperadas $\hat{\mathbf{m}}$, apresentado em (3.1.5); alternativamente também pode ser utilizada a estatística de Pearson :

$$Q_P = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \quad (3.1.11)$$

Sob o modelo (2.3.7), Q_L e Q_P têm distribuição Qui-Quadrado assintótica com $[s(r-1)-t]$ graus de liberdade (ver Imrey et al. (1981) e Koch et al. (1985, cap. 3) por exemplo).

Testes de hipóteses sobre combinações lineares dos parâmetros podem ser obtidos usando qualquer uma dessas estatísticas, pelo contraste do modelo inicial com um modelo reduzido induzido pela hipótese testada (H_0). Como em Análise de Regressão o teste se baseia na diferença entre os valores das estatísticas de ajuste para os dois modelos

e, admitindo H_0 verdadeira, essa nova estatística tem distribuição Qui-Quadrado com graus de liberdade correspondendo ao número de parâmetros testados (ver Montgomery & Peck (1982, Seção 4.4), por exemplo). Alternativamente, também pode ser utilizada a estatística de Wald :

$$Q_C = (\mathbf{C} \mathbf{b})' [(\mathbf{C} (\mathbf{X}' \mathbf{V}_F^{-1} \mathbf{X})^{-1} \mathbf{C})^{-1}] (\mathbf{C} \mathbf{b}) \quad (3.1.12)$$

onde $\mathbf{C}$ ($c \times t$), de constantes conhecidas e posto completo c ($c \leq t$), é a matriz que define a combinação linear de interesse. Admitindo $\mathbf{H}_0 : \mathbf{C} \boldsymbol{\beta} = 0$ verdadeira, Q_C tem distribuição assintótica Qui-Quadrado com c graus de liberdade.

Intervalos ou regiões de confiança aproximados para parâmetros, ou contagens ajustadas podem ser obtidos da maneira usual, empregando (3.1.7) ou (3.1.9) conforme o caso (ver Montgomery & Peck (1982, seção 4.3), por exemplo).

Para muitos modelos log-lineares, as equações de verossimilhança (3.1.5) apresentam formas explícitas, o que permite determinar diretamente as estimativas de MV das frequências esperadas nas caselas ($\hat{m}$) como funções das estatísticas suficientes. Por exemplo, no caso de 3 variáveis, com exceção do modelo sem a interação de segunda ordem, todos os demais modelos hierárquicos apresentam as frequências esperadas nas caselas como forma fechada das estatísticas suficientes; os modelos de completa independência também apresentam soluções explícitas. Bishop et al. (1975, Seção 3.4) apresentam regras para detectar se existem estimativas diretas para um determinado modelo e também apresentam um procedimento para a obtenção dessas estimativas. Para outros modelos não existem estimativas diretas e as equações (3.1.5) apresentam estrutura não linear; nesses casos, métodos iterativos como o de Ajuste Proporcional Iterativo (Iterative Proportional Fitting) e o de Newton-Raphson, esquematizados a seguir, podem ser empregados para determinar $\mathbf{b}$, $\hat{\mathbf{m}}$ e $\hat{\pi}$. Para outros detalhes sobre o método de MV sugerimos consultar, por exemplo, Bishop et al. (1975, cap. 3), Imrey et al. (1981) e Koch et al. (1985, cap. 3).

Ajuste Proporcional Iterativo

Quando o modelo log-linear é hierárquico, as estimativas de MV das frequências esperadas nas caselas ($\hat{m}$) podem ser obtidas empregando-se o método de sucessivos ajustamentos proporcionais, originalmente introduzido por Deming & Stephan (1940) e utilizado por diversos autores, entre eles Bishop et al. (1975, seção 3.5), Fienberg (1981) e Agresti (1984).

O método se baseia no conjunto minimal de estatísticas suficientes correspondente ao modelo log-linear hierárquico de interesse. Bishop et al. (1975, Seção 3.3) apresentam um procedimento simples, estabelecido a partir de resultados devidos a Birch (1963), que permite identificar facilmente o conjunto minimal de estatísticas suficientes relativo a um determinado modelo. Esse procedimento consiste dos seguintes passos: i) para cada variável estudada procurar o parâmetro do modelo de maior ordem que a envolve; ii) o conjunto de estatísticas suficientes minimais são os totais marginais correspondentes a esses parâmetros. Por exemplo, no caso do modelo log-linear hierárquico (2.1.10), esses parâmetros são u_{ij}^{AB} , u_{ik}^{AC} e u_{jk}^{BC} ; conseqüentemente, as estatísticas suficientes minimais são $n_{ij.}$, $n_{i.k}$ e $n_{.jk}$; no caso do modelo que admite independência completa entre as três variáveis (A , B e C), o conjunto de estatísticas suficientes é $\{n_{i.}, n_{.j.} \text{ e } n_{..k}\}$; e, para o modelo (2.1.13) esse conjunto é $\{n_{ij.} \text{ e } n_{i.k}\}$.

O procedimento iterativo tem início com um conjunto arbitrário de estimativas para as frequências esperadas nas caselas ($m^{(0)}$). É comum tomar $m^{(0)} = 1$, em todas as caselas, independentemente do modelo a ser ajustado. Os ciclos são constituídos de tantos passos quanto o número de estatísticas suficientes. As estimativas para as frequências esperadas nas caselas correspondentes a cada passo de um ciclo são obtidas multiplicando-se as estimativas do passo anterior por fatores apropriados. Esses fatores são funções das estatísticas suficientes correspondentes às frequências observadas e às ajustadas no

passo anterior. Uma regra de parada é definida em termos das estimativas das frequências das caselas de ciclos consecutivos e de forma a garantir o grau de precisão desejado. O procedimento continua até que seja satisfeita a regra de parada estabelecida.

Como exemplo, considere o modelo (2.1.10). Não é possível escrever $\hat{m}_{ijk}$ (frequência ajustada na casela (i, j, k)) como uma expressão explícita das estatísticas suficientes $n_{ij\cdot}$, $n_{i\cdot k}$, e $n_{\cdot jk}$. Sabemos, entretanto, que os valores esperados estimados por esse modelo são funções dessas estatísticas e devem, segundo Birch (1963), satisfazer:

$$\begin{aligned}\hat{m}_{ij\cdot} &= n_{ij\cdot}, & i, j &= 1, 2, \\ \hat{m}_{i\cdot k} &= n_{i\cdot k}, & i, k &= 1, 2, \\ \hat{m}_{\cdot jk} &= n_{\cdot jk}, & j, k &= 1, 2.\end{aligned}$$

Neste caso, o $(r + 1)$ -ésimo ciclo é constituído por:

$$\begin{aligned}\text{1o. passo : } m_{ijk}^{(3r+1)} &= m_{ijk}^{(3r)} \frac{n_{ij\cdot}}{m_{ij\cdot}^{(3r)}} \\ \text{2o. passo : } m_{ijk}^{(3r+2)} &= m_{ijk}^{(3r+1)} \frac{n_{i\cdot k}}{m_{i\cdot k}^{(3r+1)}} \\ \text{3o. passo : } m_{ijk}^{(3r+3)} &= m_{ijk}^{(3r+2)} \frac{n_{\cdot jk}}{m_{\cdot jk}^{(3r+2)}}\end{aligned}$$

Sejam $m_{ijk} = 1$ para $i, j, k = 1, 2$, as estimativas iniciais. Então, para o primeiro ciclo, temos:

$$\begin{aligned}\text{1o. passo : } m_{ijk}^{(1)} &= m_{ijk}^{(0)} \frac{n_{ij\cdot}}{m_{ij\cdot}^{(0)}} \\ \text{2o. passo : } m_{ijk}^{(2)} &= m_{ijk}^{(1)} \frac{n_{i\cdot k}}{m_{i\cdot k}^{(1)}} \\ \text{3o. passo : } m_{ijk}^{(3)} &= m_{ijk}^{(2)} \frac{n_{\cdot jk}}{m_{\cdot jk}^{(2)}}\end{aligned}$$

O ciclo constituído de 3 passos é repetido até que

$$|m_{ijk}^{(3r+3)} - m_{ijk}^{(3r)}| < \delta, \quad \delta > 0 \quad \text{para } i, j, k = 1, 2 \text{ e } r = 0, 1, \dots$$

Note que as frequências ajustadas tendem a se aproximar das frequências observadas em termos das estatísticas suficientes a cada ciclo. Resultados relativos à aplicação deste algoritmo para o modelo (2.1.10) podem ser encontrados em Bishop et al. (1975, p. 87).

Por outro lado, se existem estimativas diretas, o procedimento calcula essas estimativas em apenas um ciclo. É o caso, por exemplo, do modelo de independência completa; considerando três variáveis, podemos determinar $\hat{m}_{ijk}$ a partir das estatísticas suficientes empregando diretamente a solução explícita:

$$\hat{m}_{ijk} = \frac{n_{i..} n_{.j.} n_{..k}}{N^2}$$

Bishop et al. (1975, seções 3.5 e 3.6) detalham o uso deste algoritmo, inclusive para situações em que existem estimativas diretas, i.e., quando (3.1.5) apresenta solução explícita.

Algoritmo de Newton-Raphson

Quando as equações (3.1.5) não apresentam soluções explícitas, o procedimento de Newton-Raphson pode ser usado para a obtenção das estimativas de MV $\mathbf{b}$ do vetor de parâmetros do modelo log-linear proposto.

O método de Newton-Raphson é um procedimento utilizado para encontrar ponto de máximo de uma função. Para usá-lo devemos admitir um valor inicial e aproximar a função na vizinhança desse valor por um polinômio de segundo grau. O valor que maximiza esse polinômio é o segundo valor utilizado. Em seguida, a função é novamente aproximada por um polinômio de segundo grau na vizinhança desse segundo valor e um terceiro valor é obtido. Assim, uma sequência de valores é gerada e, essa sequência converge para o ponto de máximo se a função for razoavelmente bem comportada.

Mais especificamente, seja $g(\boldsymbol{\beta})$ uma função com argumentos reais $\boldsymbol{\beta}$ e suponha que queremos encontrar a estimativa $\mathbf{b}$ de $\boldsymbol{\beta}$ que maximiza a função $g(\boldsymbol{\beta})$. Sejam $\mathbf{q}$ ($t \times 1$) e $\mathbf{H}$ ($t \times t$) respectivamente com elementos q_l e h_{lm} , $l, m = 1, \dots, t$, onde:

$$q_l = \frac{\partial g(\boldsymbol{\beta})}{\partial \beta_l} \quad \text{e} \quad h_{lm} = \frac{\partial^2 g(\boldsymbol{\beta})}{\partial \beta_l \partial \beta_m}$$

Em cada etapa do procedimento iterativo, $g(\beta)$ é aproximada na vizinhança da estimativa obtida na etapa anterior pelos termos até segunda ordem de sua expansão em série de Taylor. Assim, na etapa $v = 0, 1, 2, \dots$, podemos escrever:

$$Q(\beta)^{(v)} \approx g(b^{(v)}) + q^{(v)}(\beta - b^{(v)}) + \frac{1}{2}(\beta - b^{(v)})' H^{(v)}(\beta - b^{(v)})$$

onde $q^{(v)}$ e $H^{(v)}$ são funções de $b^{(v)}$. Derivando $Q(\beta)^{(v)}$ em relação a β e igualando o resultado a zero, obtemos:

$$\frac{\partial Q^{(v)}}{\partial \beta} = q^{(v)} + H^{(v)}(\beta - b^{(v)}) = 0_t$$

Resolvendo essa expressão em termos de β , obtem-se para a $(v + 1)$ -ésima iteração na sequência de aproximações para b :

$$b^{(v+1)} = b^{(v)} - (H^{(v)})^{-1} q^{(v)} \quad (3.1.12)$$

onde H deve ser não singular.

Para conjuntos de dados que obedecem a um modelo probabilístico Produto de Multinomiais, com função de verossimilhança P dada em (2.2.1), podemos identificar:

$g(\beta) = \log P(\beta) = \log P(n | n, \pi(\beta))$, onde $\pi(\beta)$ é dado por (2.3.7);

$q^{(v)}$ com elementos $q_l^{(v)} = x_l' n - x_l' m^{(v)}$ onde $m^{(v)} = D_{\pi^{(v)}} [I_s \otimes 1_r 1_r'] n$

e x_l' definido em (3.1.4); e

$H^{(v)}$ com elementos

$$h_{lm}^{(v)} = - \sum_{i=1}^s n_i \{ x_{il}' [D_{\pi_i^{(v)}} - \pi_i^{(v)} \pi_i'^{(v)}] x_{im} \}.$$

Como H é negativa definida, $\log P(\beta)$ é estritamente côncava e está garantida a existência de b finito e único que maximiza $\log P$.

O vetor $\pi^{(v)}$, $v = 0, 1, 2, \dots$, é obtido de $b^{(v)}$ através de (2.3.7), e é usado para obter $b^{(v+1)}$ através da expressão (3.1.12), que por sua vez, é usado para obter $\pi^{(v+1)}$ e assim sucessivamente.

As iterações prosseguem até que

$$| \mathbf{b}_l^{(v+1)} - \mathbf{b}_l^{(v)} | < c, \quad l = 1, \dots, t$$

onde c ($c > 0$) é o grau de precisão pré-estabelecido.

O processo iterativo pode ser iniciado com um vetor $\mathbf{b}^{(0)}$, formado por valores arbitrários. É comum tomar as estimativas de MQG como valores iniciais $\mathbf{b}^{(0)}$ pois sua consistência garante a convergência (em probabilidade) do processo iterativo. Em geral, a convergência de $\mathbf{b}^{(v)}$ e $\boldsymbol{\pi}^{(v)}$ para as estimativas de MV $\mathbf{b}$ e $\hat{\boldsymbol{\pi}}$ de $\boldsymbol{\beta}$ e $\boldsymbol{\pi}$ respectivamente é atingida rapidamente.

3.2 - O Método de Mínimos Quadrados Generalizados

O método de Mínimos Quadrados Generalizados introduzido por Grizzle, Starmer & Koch (1969) para análise de dados categorizados pode ser utilizado para uma variedade de modelos. Em particular, esse método pode ser empregado na estimação de parâmetros de modelos log-lineares, como alternativa ao método de MV. Nesta seção citamos os principais resultados no caso geral, particularizando-os para a classe de modelos estruturais log-lineares. O leitor interessado em maiores detalhes deve consultar Koch et al. (1985).

No caso geral, essa metodologia pode ser aplicada a qualquer modelo estrutural da forma :

$$\mathbf{F}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}$$

onde $\boldsymbol{\pi}$ ($v \times 1$) é o vetor de parâmetros do modelo probabilístico correspondente, $\mathbf{X}$ ($u \times t$), $u < v$, é a matriz de especificação do modelo estrutural, com elementos conhecidos e de posto completo $t \leq u$, $\boldsymbol{\beta}$ ($t \times 1$) é o vetor de parâmetros a serem estimados e $\mathbf{F}(\boldsymbol{\pi}) = [F_1(\boldsymbol{\pi}) \cdots F_u(\boldsymbol{\pi})]'$, $u < v$, é um vetor de funções de $\boldsymbol{\pi}$ satisfazendo as seguintes propriedades:

- i) $F(\pi)$ têm derivadas parciais contínuas até 2a. ordem, com respeito aos elementos do vetor π ;
- ii) os elementos de $F(\pi)$ são linearmente independentes e também independentes das restrições naturais associadas a π .

No caso de modelos log-lineares, $F(\pi) = A \log \pi$ onde $A (u \times v)$, $u < v$, é uma matriz de constantes conhecidas, ortogonal à matriz formada pelas restrições naturais associadas ao modelo probabilístico. Em geral, A define logitos múltiplos.

O vetor de probabilidades associado ao modelo probabilístico Produto de Multinomiais, $\pi = (\pi_1 \cdots \pi_s)' (sr \times 1)$, onde $\pi_i = (\pi_{i1} \cdots \pi_{ir})$, com restrições naturais $0 < \pi_{ij} < 1$ e $\sum_j \pi_{ij} = 1$ para $i = 1, \dots, s$ e $j = 1, \dots, r$, tem como estimador de máxima verossimilhança irrestrito o vetor de proporções amostrais

$$p = (p_1 \cdots p_s)' \quad (3.2.1)$$

onde $p_i = (p_{i1} \cdots p_{ir})$, com $p_{ij} = \frac{n_{ij}}{n_i}$ (ver Rao(1973, p. 355-363) por exemplo) .

A matriz de covariância de p é dada por:

$$V(\pi)_{sr \times sr} = \text{DIAG}[V_1(\pi_1), V_2(\pi_2), \dots, V_s(\pi_s)] \quad (3.2.2)$$

onde $\text{DIAG}[V_1(\pi_1), V_2(\pi_2), \dots, V_s(\pi_s)]$ indica uma matriz diagonal em blocos com $V_i(\pi_i)$, $i = 1, \dots, s$ na diagonal principal e $V_i(\pi_i)_{r \times r} = (\text{DIAG}(\pi_i) - \pi_i \pi_i') / n_i$ é a matriz de covariância correspondente a p_i . Um estimador consistente de $V(\pi)$ é:

$$V(p)_{sr \times sr} = \text{DIAG}[V_1(p_1), V_2(p_2), \dots, V_s(p_s)] \quad (3.2.3)$$

onde $V_i(p_i)_{r \times r} = (\text{DIAG}(p_i) - p_i p_i') / n_i$.

Um estimador do vetor de funções $F(\pi)$ é dado por:

$$F_{u \times u} = F(p) = [F_1(p) \cdots F_u(p)]' \quad u \leq s(r-1) \quad (3.2.4)$$

e sua correspondente matriz de covariância assintótica é:

$$\Sigma_{F_{u \times u}} = V_F(\pi) = H(\pi) V(\pi) H(\pi)' \quad (3.2.5)$$

onde $H(\pi) = \left. \frac{\partial F(Z)}{\partial Z} \right|_{Z=\pi}$.

Como consequência do Teorema Limite Central, para amostras de tamanho suficientemente grande, F tem distribuição aproximadamente Normal Multivariada com parâmetros:

$$E_A\{F\} = F(\pi) \quad e \quad V_A\{F\} = \Sigma_F \quad (3.2.6)$$

onde o índice A de E_A e V_A indica valor assintótico. Um estimador consistente da matriz de covariância Σ_F é dado por:

$$V_{F_{u \times u}} = V_F(p) = H(p) V(p) H(p)' \quad (3.2.7)$$

A matriz V_F é não singular uma vez que as funções $F_m(p)$, $m = 1, \dots, u$, são independentes entre si e incluem as restrições naturais associadas ao vetor π .

Quando se considera um modelo estrutural log-linear da forma (2.2.2), temos:

$$F(\pi) = A \log \pi, \quad F = A \log p \quad e \quad H(p) = A D_p^{-1}$$

onde D_p^{-1} indica a inversa de uma matriz diagonal com elementos do vetor p na diagonal.

Para a obtenção das estimativas dos parâmetros pode ser empregado o método de MQG que consiste em minimizar a forma quadrática

$$Q = (F - X\beta)' V_F^{-1} (F - X\beta) \quad (3.2.8)$$

Derivando Q em relação a β ($t \times 1$) e igualando o resultado a zero é fácil mostrar que o estimador de MQG de β é dado por

$$\mathbf{b} = (\mathbf{X}' \mathbf{V}_F^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}_F^{-1} \mathbf{F} \quad (3.2.9)$$

com uma matriz de covariância assintótica que pode ser consistentemente estimada por:

$$\mathbf{V}_b = (\mathbf{X}' \mathbf{V}_F^{-1} \mathbf{X})^{-1} \quad (3.2.10)$$

Se os tamanhos das amostras $\{n_i\}$, $i = 1, \dots, s$, forem suficientemente grandes, sob o modelo (2.2.2), o estimador de MQG $\mathbf{b}$ terá assintoticamente uma distribuição Normal Multivariada, com parâmetros

$$\mathbf{E}_A\{\mathbf{b}\} = \beta \quad \text{e} \quad \mathbf{V}_A\{\mathbf{b}\} = (\mathbf{X}' \Sigma_F^{-1} \mathbf{X})^{-1}$$

Para amostras grandes verifica-se que os estimadores de MQG e MV tem distribuição aproximadamente normal com média em torno de um valor comum e que a razão entre as variâncias converge para um (ver Agresti (1984, p. 228) por exemplo). Consequentemente esses estimadores estão na classe dos estimadores BAN ("Best Asymptotically Normal") e nesse sentido são assintoticamente equivalentes.

O ajuste do modelo (2.2.2) pode ser verificado através da estatística residual de MQG derivada da expressão (3.2.8):

$$Q = (\mathbf{F} - \mathbf{X} \mathbf{b})' \mathbf{V}_F^{-1} (\mathbf{F} - \mathbf{X} \mathbf{b}) \quad (3.2.11)$$

Sob a hipótese $H_0 : \mathbf{F}(\pi) = \mathbf{X}\beta$, Q tem uma distribuição assintótica Qui-Quadrado com $(u - t)$ graus de liberdade (correspondendo à diferença entre o número de elementos em $\mathbf{F}$ e o número de parâmetros no modelo estrutural). Outras estatísticas (ver Koch et al. (1985, p. 69-72, 87), por exemplo) que podem ser utilizadas para testar o ajuste do modelo são:

i) a estatística de Wald

$$Q_W = \mathbf{F}' \mathbf{W}' (\mathbf{W} \mathbf{V}_F \mathbf{W})^{-1} \mathbf{W} \mathbf{F} \quad (3.2.12)$$

onde $\mathbf{W}$ é uma matriz $((u - t) \times u)$ ortocomplementar a $\mathbf{X}$, e

ii) a estatística de Neyman

$$Q_N = (\mathbf{p} - \bar{\boldsymbol{\pi}})' \mathbf{D}_n \mathbf{D}_p^{-1} (\mathbf{p} - \bar{\boldsymbol{\pi}}) \quad (3.2.13)$$

onde $(\mathbf{p} - \bar{\boldsymbol{\pi}}) = \mathbf{V}(\mathbf{p}) [\mathbf{H}(\mathbf{p})]' \mathbf{V}_F^{-1} (\mathbf{F} - \mathbf{X} \mathbf{b})$ e $\bar{\boldsymbol{\pi}}$ é o estimador de $\boldsymbol{\pi}$ obtido pelo critério do Qui-Quadrado Mínimo Modificado. A igualdade entre essas estatísticas, foi mostrada por Bhapkar (1966) e por essa razão $\mathbf{b}$ também é chamado de estimador Qui-Quadrado Mínimo Modificado.

Muitas hipóteses sobre os efeitos das variáveis explicativas ou genericamente sobre combinações lineares dos elementos de $\boldsymbol{\beta}$, podem ser expressas na forma

$$H_0 : \mathbf{C} \boldsymbol{\beta} = \mathbf{0} \quad (3.2.14)$$

onde $\mathbf{C}$ é uma matriz $(c \times t)$ de constantes conhecidas e posto c , $c \leq t$.

Se admitimos que o modelo $\mathbf{F}(\boldsymbol{\pi}) = \mathbf{X} \boldsymbol{\beta}$ caracteriza adequadamente o vetor $\mathbf{F}(\boldsymbol{\pi})$, uma estatística adequada para verificar uma hipótese da forma (3.2.14) é a estatística de Wald:

$$Q_c = (\mathbf{C} \mathbf{b})' [(\mathbf{C} (\mathbf{X}' \mathbf{V}_F^{-1} \mathbf{X})^{-1} \mathbf{C})^{-1}] (\mathbf{C} \mathbf{b}) \quad (3.2.15)$$

Sob H_0 , Q_c tem distribuição assintótica Qui-Quadrado com c graus de liberdade (a expressão entre colchetes em (3.2.15) é a estimativa da matriz de covariância assintótica de $\mathbf{C} \mathbf{b}$). A aceitação da hipótese formulada implica que o modelo inicialmente proposto poderia ser simplificado (ou reduzido) a partir das restrições impostas por (3.2.14).

Uma forma alternativa de testar H_0 é através da diferença entre duas estatísticas de ajuste da forma (3.2.12) associadas a modelos “hierárquicos” (ou “encaixados”). Pode-se

verificar que:

$$Q_c = Q_{W_c} - Q_W \quad (3.2.16)$$

onde Q_{W_c} é o valor da estatística de Wald para o modelo reduzido induzido por H_0 ; Q_W é o valor da estatística de Wald para o modelo inicialmente proposto.

As estatísticas de Wald (Q_W e Q_C), a de Neyman (Q_N) e a residual de MQG (Q) são também assintoticamente equivalentes à estatística da Razão de Verossimilhança (ver Koch et al. (1985, pág. 64), por exemplo).

Valores esperados para $F(\pi)$ podem ser obtidos com base no modelo estrutural ajustado:

$$\hat{F} = Xb = X(X'V_F^{-1}X)^{-1}X'V_F^{-1}F \quad (3.2.17)$$

A matriz de covariância correspondente pode ser consistentemente estimada por:

$$V_{\hat{F}} = X(X'V_F^{-1}X)^{-1}X' \quad (3.2.18)$$

Os valores preditos $\hat{F} = Xb$ são melhores estimadores do que os originais $F = F(p)$. Os primeiros expressam as relações estruturais do modelo ajustado e se baseiam nos dados de toda a tabela. Os estimadores F são obtidos de dados parciais da tabela e como decorrência apresentam maiores variâncias.

Quando a função F de interesse é uma função complexa das contagens das caselas ou do vetor de probabilidades a análise por MQG é mais simples do que a análise por MV. Além disso, em geral, é mais fácil escrever programas computacionais para a obtenção das estimativas dos parâmetros e testes de ajuste empregando a metodologia de MQG do que pelo método de MV (Agresti (1984, p. 228)). A maior desvantagem do método de MQG sobre o método de MV é que o primeiro é mais restritivo quanto ao tamanho da amostra necessário para as aproximações através do Teorema Limite Central (ver Koch et al. (1985, p. 188), por exemplo).

Capítulo 4

Aplicações Práticas

O objetivo deste capítulo é ilustrar a aplicação dos procedimentos apresentados no capítulo anterior e informar sobre alguns recursos computacionais disponíveis. Embora não pretendamos fazer uma descrição exaustiva dos pacotes existentes, na Seção 4.1 apresentamos algumas alternativas usualmente disponíveis em muitos centros computacionais. Na Seção 4.2 apresentamos alguns exemplos de ajuste de modelos estruturais log-lineares.

4.1 - Pacotes computacionais para ajuste de modelos estruturais log-lineares

Diversos programas para equipamentos de grande porte e também para microcomputadores estão disponíveis para a análise de dados categorizados através de modelos log-lineares. Algumas alternativas são:

CATMOD - Procedimento do pacote SAS(1985) para equipamento de grande porte, desenvolvido para modelagem de dados categorizados. Possibilita o ajuste de modelos lineares, log-lineares e regressões logísticas (caso particular de modelo log-linear, ver Bishop et al.(1975, p. 357-360 por exemplo) e análise de medidas repetidas às frequências das caselas ou à combinações lineares dessas frequências, por ambos os métodos de estimação: MV e MQG.

BMDP/4F - Este programa do pacote **BMDP** (Dixon (1985)), disponível para computadores e microcomputadores, ajusta modelos log-lineares hierárquicos para as frequências esperadas nas caselas. A estimação dos parâmetros é feita pelo método de MV, utilizando o algoritmo de Ajuste Proporcional Iterativo.

HILOGLINEAR e **LOGLINEAR** - Procedimentos do pacote **SPSS^X** (Norusis (1986a)) para computadores de grande porte. O primeiro, disponível também para microcomputadores (Norusis(1986b)), estima parâmetros de modelos log-lineares hierárquicos para as frequências esperadas nas caselas, pelo método de MV, empregando o algoritmo de Ajuste Proporcional Iterativo. O segundo é mais geral, incluindo o ajuste de alguns modelos não hierárquicos e também de modelos log-lineares na formulação logitos adjacentes (ver Seção 2.2) e de modelos log-lineares específicos para variáveis ordinais (ver Agresti(1984, p. 253-254)). Esse procedimento produz estimativas de parâmetros pelo método de MV, utilizando o algoritmo de Newton-Raphson.

GENCAT - (Landis et al.(1976)) é um programa para computadores de grande porte, projetado para análise de dados categorizados pelo método de MQG. O programa ajusta modelos de regressão linear a funções do vetor de proporções observadas nas caselas. Essas funções podem ser geradas empregando-se transformações sucessivas - linear, logarítmica, exponencial e adição de vetor de constantes - ao vetor das proporções observadas e devem

ser escolhidas de tal modo que sua matriz de covariância seja não singular. Nesse sentido o programa é bastante geral e o ajuste de modelos log-lineares constitui um caso especial de aplicação.

Outras alternativas são: ECTA (Fay & Goodman (1975)), FREQ (Haberman (1979)), GLIM (Baker & Nelder (1978)), MULTIQUAL (Bock & Yates (1973)). Agresti (1984, p. 251-260) indica como empregar os programas citados na análise de dados categorizados envolvendo variáveis ordinais.

Como parte deste trabalho desenvolvemos um programa (**LOGLIN.AJU**) para ajuste de modelos estruturais log-lineares (hierárquicos ou não) pelos métodos de MV e MQG através do módulo CM do pacote SOC. Esse programa utiliza a formulação de logitos múltiplos e fornece estimativas de MQG e de MV (através do algoritmo de Newton-Raphson) para os parâmetros e respectivos desvios padrões. O programa produz ainda frequências, proporções e logitos múltiplos ajustados pelos dois procedimentos. Também foi desenvolvida uma subrotina (**LOGLIN.HIP**) para testar hipóteses sobre parâmetros por meio de estatísticas de Wald. Os comandos relativos aos dois programas estão apresentados no Anexo A, juntamente com as informações necessárias para sua utilização.

Na próxima seção alguns modelos apresentados no Capítulo 2, são ajustados através dos programas LOGLIN.AJU, GENCAT e BMDP/4F (os dois últimos nas versões para equipamento de grande porte disponíveis no Centro de Computação Eletrônica da Universidade de São Paulo).

4.2 - Exemplos ilustrativos de ajuste de modelos estruturais log-lineares

Nesta seção apresentamos resultados para alguns modelos estudados no Capítulo 2. Para efeito de comparação, esses resultados foram obtidos sob as duas metodologias descritas (MV e MQG) através dos programas computacionais BMDP/4F, GENCAT e LOGLIN.AJU/HIP.

Exemplo 4.2.1

Consideramos os dados da Tabela 1.6, envolvendo 3 variáveis dicotômicas: A (Sobrevivência da planta), B (Época de plantio) e C (Comprimento do corte) e selecionamos os modelos log-lineares hierárquicos (2.1.9), (2.2.6) e (2.3.5).

O modelo (2.1.9) pode ser diretamente ajustado pelo BMDP/4F pois trata-se de um modelo log-linear hierárquico associado a uma distribuição Multinomial. Os comandos do BMDP/4F para o ajuste estão apresentados em B.1 (Anexo B) e as estimativas e valores das estatísticas de ajuste obtidos da saída do programa constam da Tabela 4.2.1. Os modelos (2.2.6) e (2.3.5) não podem ser diretamente ajustados pelo BMDP/4F pois tem como modelo probabilístico associado uma distribuição Produto de Multinomiais. No entanto, as estimativas de MV para os parâmetros desses modelos puderam ser obtidas a partir das estimativas correspondentes ao modelo (2.1.9), dadas as relações existentes entre os parâmetros desses modelos, apresentadas nas Tabelas 4.2.2 e 4.2.3 (com base nos resultados da Seção 2.5).

Utilizando o programa GENCAT, foram diretamente calculadas as estimativas por MQG relativas ao modelo hierárquico (2.2.6); os resultados estão na Tabela 4.2.2. Note que a função $F(\mathbf{p}) = \mathbf{A} \log \mathbf{p}$ associada a esse modelo possui matriz de covariância $\mathbf{V}_F$ não singular; esta característica é própria dos modelos expressos na formulação generalizada (2.2.2), como é o caso do modelo (2.2.6). No Anexo B (B.2) apresentamos a listagem

com os dados de entrada do programa GENCAT para o ajuste do modelo (2.2.6). Para a obtenção das estimativas de MQG sob o modelo (2.1.9), é necessário utilizar uma formulação alternativa para a qual a matriz $\mathbf{V}_F$ é não singular. Nesse sentido, consideramos logitos múltiplos da forma:

$$L_{ijk} = \log \left(\frac{\pi_{ijk}}{\pi_{222}} \right) \quad i, j, k = 1, 2 \quad \text{com} \quad ijk \neq 222$$

e assim expressamos o modelo (2.1.9) na formulação generalizada (2.2.2). O vetor de funções $\mathbf{F}(\boldsymbol{\pi})$ é dado por $\mathbf{F}(\boldsymbol{\pi}) = \mathbf{A} \log \boldsymbol{\pi}$ com $\mathbf{A} = [\mathbf{I}_7, -\mathbf{1}_7]$; a matriz de especificação para o novo modelo é:

$$\mathbf{X}_L = \mathbf{A} \mathbf{X} = \begin{pmatrix} 2 & 2 & 2 & 0 & 0 & 0 \\ 0 & 2 & 2 & -2 & -2 & 0 \\ 2 & 0 & 2 & -2 & 0 & -2 \\ 0 & 0 & 2 & 0 & -2 & -2 \\ 2 & 2 & 0 & 0 & -2 & -2 \\ 0 & 2 & 0 & -2 & 0 & -2 \\ 2 & 0 & 0 & -2 & -2 & 0 \end{pmatrix} \quad (4.2.1)$$

onde $\mathbf{X}$ é a matriz de especificação correspondente ao modelo (2.1.9) e o vetor de parâmetros é:

$$\boldsymbol{\beta} = (u_1^A \ u_1^B \ u_1^C \ u_{11}^{AB} \ u_{11}^{AC} \ u_{11}^{BC})' .$$

As estimativas dos parâmetros por MQG para o modelo (2.3.5), apresentadas na Tabela 4.2.3, puderam ser calculadas tanto a partir das estimativas de MQG obtidas para o modelo (2.1.9) quanto a partir das estimativas correspondentes ao modelo (2.2.6), como indicado nas Seções 2.3 e 2.5 .

As estimativas de MV e de MQG para os parâmetros do modelo estrutural log-linear na formulação logitos múltiplos (2.2.6) também foram obtidas a partir do programa LOGLIN.AJU. Para a obtenção das estimativas de MV e MQG relativas ao modelo (2.1.9), utilizamos sua formulação em termos de logitos múltiplos (modelo (4.2.1)) e quanto ao modelo (2.3.5), obtivemos as estimativas com base na relação entre seus parâmetros e os parâmetros do modelo (2.2.6).

Nas Tabelas 4.2.1, 4.2.2 e 4.2.3, também apresentamos os valores das estatísticas de testes de ajuste desses modelos. Em termos comparativos, verificamos que os resultados obtidos sob os dois métodos de estimação são bastante semelhantes.

Tabela 4.2.1 - Estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões(dp) e valores das estatísticas de ajuste para o modelo (2.1.9)

parâmetro	M V		M Q G	
	estimativa	dp estimado	estimativa	dp estimado
u	-2,197	—	-2,196	—
u_1^A	-0,254	0,036	-0,253	0,037
u_1^B	0,081	0,035	0,080	0,036
u_1^C	0,056	0,037	0,055	0,034
u_{11}^{AB}	0,357	0,034	0,355	0,037
u_{11}^{AC}	0,254	0,036	0,253	0,037
u_{11}^{BC}	-0,081	0,035	-0,080	0,036
Ajuste do modelo ^(a)	$Q_P = 2,27$ $Q_L = 2,29$	$(p = 0,132)$ $(p = 0,130)$	$Q_W = 2,26$	$(p = 0,132)$

(a): graus de liberdade = 1

Tabela 4.2.2 - Estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões(dp) e valores das estatísticas de ajuste para o modelo (2.2.6)

parâmetro	M V		M Q G	
	estimativa	dp estimado	estimativa	dp estimado
w	-0,508	0,072	-0,506	0,073
w_1^B	0,714	0,068	0,711	0,074
w_1^C	0,508	0,072	0,506	0,073
Ajuste do modelo ^(a)	$Q_P = 2,27$ $Q_L = 2,29$	$(p = 0,132)$ $(p = 0,130)$	$Q_W = 2,26$	$(p = 0,132)$

(a): graus de liberdade = 1 $w = 2 u_1^A$; $w_1^B = 2 u_{11}^{AB}$; $w_1^C = 2 u_{11}^{AC}$.

Tabela 4.2.3 - Estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões(dp) e valores das estatísticas de ajuste para o modelo (2.3.5)

parâmetro	M V		M Q G	
	estimativa	dp estimado	estimativa	dp estimado
$v_{A(1)}$	-0,254	0,036	-0,253	0,037
$v_{A(1)}^{B(1)}$	0,357	0,034	0,355	0,037
$v_{A(1)}^{C(1)}$	0,254	0,036	0,253	0,037
Ajuste do modelo ^(a)	$Q_P = 2,27$ $Q_L = 2,29$	$(p = 0,132)$ $(p = 0,130)$	$Q_W = 2,26$	$(p = 0,132)$

(a): graus de liberdade = 1 $v_{A(1)} = u_1^A$; $v_{A(1)}^{B(1)} = u_{11}^{AB}$; $v_{A(1)}^{C(1)} = u_{11}^{AC}$. Também

$$v_{A(1)} = w / 2 ; v_{A(1)}^{B(1)} = w_1^B / 2 ; v_{A(1)}^{C(1)} = w_1^C / 2 .$$

Quando um modelo é considerado adequado aos dados a tabela com frequências ajustadas ($\hat{\mathbf{m}}$) pelo modelo pode ser de interesse. Os programas BMDP/4F e LOGLIN.AJU fornecem essa tabela na saída padrão. O GENCAT não o faz diretamente; contudo, lembrando que $\hat{\mathbf{m}}$ é função de $\hat{\boldsymbol{\pi}}$, vetor de proporções ajustadas pelo modelo, podemos obter $\hat{\mathbf{m}}$ a partir da formulação (2.3.7). Assim podemos escrever:

$$\hat{\mathbf{m}} = N \hat{\boldsymbol{\pi}} = N \mathbf{D}_{\hat{\boldsymbol{\eta}}}^{-1} \exp(\mathbf{X} \mathbf{b})$$

onde N é o total geral da tabela, $\hat{\boldsymbol{\eta}} = [\mathbf{1}_r, \mathbf{1}_r' \otimes \mathbf{I}_s] \exp(\mathbf{X} \mathbf{b})$, a matriz $\mathbf{X}$ corresponde à formulação do modelo em termos de $\log \boldsymbol{\pi}$ e $\mathbf{b}$ é a estimativa de $\boldsymbol{\beta}$. O programa permite obter esses valores através das transformações mencionadas na Seção 4.1 aplicadas a $\mathbf{b}$.

Como exemplo, na Tabela 4.2.4 apresentamos as frequências ajustadas relativas ao modelo (2.1.9) segundo os dois métodos de estimação; apresentamos também as frequências observadas e desvios padronizados. Os resultados obtidos através do GENCAT necessitaram dois processamentos. No primeiro, a partir do vetor $\mathbf{b}$ relativo a (4.2.1) (formulação do modelo (2.1.9) em termos de logitos múltiplos descrita anteriormente), obteve-se o vetor

$\hat{\eta}$ e em seguida a matriz $D_{\hat{\eta}}^{-1}$. No segundo processamento, calcularam-se as frequências ajustadas $\hat{m}$ a partir de b , N , $D_{\hat{\eta}}^{-1}$ e X , matriz de especificação do modelo em termos de $\log \pi$, dada em (2.1.9). No Anexo B (B.2 e B.4) apresentamos os comandos do GENCAT utilizados nesses dois processamentos.

Tabela 4.2.4 - Frequências observadas e frequências ajustadas ($\hat{m}$) e desvios padronizados (d) para o modelo (2.1.9)

níveis das variáveis			Frequência observada	MV		MQG	
A	B	C		$\hat{m}$	d	$\hat{m}$	d
1	1	1	156	161,1	-0,4	160,9	-0,4
2	1	1	84	78,9	0,6	79,1	0,6
1	2	1	84	78,9	0,6	79,1	0,6
2	2	1	156	161,1	-0,4	160,9	-0,4
1	1	2	107	101,9	0,5	102,0	0,5
2	1	2	133	138,1	-0,4	137,9	-0,4
1	2	2	31	36,1	-0,8	36,4	-0,9
2	2	2	209	203,9	0,4	203,8	0,4

$$d = \frac{(O - \hat{m})}{\sqrt{\hat{m}}}, \text{ onde } O \text{ é a frequência observada}$$

Exemplo 4.2.2

Para os dados da Tabela 1.5, admitimos o modelo probabilístico Produto de Multinomiais com $s = 4$ subpopulações (diferentes cirurgias) e $r = 3$ categorias de resposta (diferentes severidades de uma sequela) e consideramos os modelos log-lineares (2.2.8) e (2.2.10).

Os dois modelos não puderam ser ajustados pelo BMDP/4F pois são modelos log-lineares não hierárquicos. As estimativas de MQG puderam ser obtidas facilmente usando o GENCAT; ambos os modelos estão expressos na forma generalizada (2.2.2), com a matriz A definindo logitos múltiplos o que implica que a matriz de covariância V_F é não singular. Dada a formulação em termos de logitos múltiplos, esses modelos foram prontamente

ajustados também pelo programa LOGLIN.AJU, que produziu as estimativas de MV além de estimativas de MQG. Nas Tabelas 4.2.5 e 4.2.6 apresentamos as estimativas de MV e de MQG e os valores das estatísticas de ajuste. Os comandos relativos aos vários programas executados constam de B.4 e B.5 no Anexo B.

Tabela 4.2.5 - Estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões(dp) e valores das estatísticas de ajuste para o modelo (2.2.8)

parâmetro	M V		M Q G	
	estimativa	dp estimado	estimativa	dp estimado
u_1	2,165	0,399	2,165	0,399
v_1	1,386	0,423	1,386	0,423
α_2	-0,510	0,501	-0,510	0,501
β_2	-0,816	0,547	-0,816	0,547
α_3	-0,589	0,510	-0,589	0,510
β_3	-0,182	0,536	-0,182	0,536
α_4	-0,967	0,491	-0,967	0,491
β_4	-0,521	0,517	-0,521	0,517
Ajuste do modelo ^(a)	—		—	

(a): modelo saturado

Tabela 4.2.6 - Estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões(dp) e valores das estatísticas de ajuste para o modelo (2.2.10)

parâmetro	M V		M Q G	
	estimativa	dp estimado	estimativa	dp estimado
u_1	2,082	0,333	2,081	0,341
v_1	1,241	0,230	1,244	0,236
β_2	-0,080	0,219	-0,105	0,215
β_3	-0,318	0,210	-0,323	0,215
β_4	-0,457	0,208	-0,456	0,212
Ajuste do modelo ^(a)	$Q_P = 4,40$	$(p = 0,222)$	$Q_W = 4,31$	$(p = 0,230)$
	$Q_L = 4,40$	$(p = 0,221)$		

(a): graus de liberdade = 3

Exemplo 4.2.3

Considerando agora, os dados da Tabela 1.8, que envolve duas variáveis respostas dicotômicas (Ocorrência de Infarto (I) e Sequela Miocárdica (S)) e duas variáveis explicativas (Idade (A) com dois níveis e Combinação de Localidade e Raça (L) com tres níveis) mostraremos como ajustar o modelo (2.4.5) ou (2.4.6) através dos recursos computacionais existentes nos programas BMDP/4F, GENCAT e LOGLIN.AJU .

Inicialmente observemos que o modelo probabilístico associado é uma distribuição Produto de Multimoniais com $s = 6$ subpopulações e $r = 4$ perfis de respostas e que o modelo (2.4.5) está na forma (2.4.1), com $\mathbf{A}^* = [\mathbf{I}_6 \otimes (\begin{smallmatrix} 1 & -1 & -1 & 1 \end{smallmatrix})]$ definindo $\mathbf{F} = \mathbf{A}^* \log \boldsymbol{\pi}$ (vetor cujos elementos são logaritmos de razões de produtos cruzados) e a matriz de covariância associada ($\mathbf{V}_F$) é não singular. Esse modelo é diretamente ajustado pelo GENCAT e não o é pelos outros dois programas computacionais considerados.

Para utilizar a subrotina LOGLIN.AJU devemos reescrever o modelo (2.4.5) na formulação (2.2.2), com $\mathbf{A}$ definindo logitos múltiplos. Conforme indicado na Seção 2.4, esse modelo log-linear generalizado pode ser escrito como um modelo log-linear na formulação (2.2.2) tomando, por exemplo,

$$\mathbf{A} = \left[\mathbf{I}_6 \otimes \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \right]$$

$$\mathbf{Z} = \left[\mathbf{I}_6 \otimes \begin{pmatrix} 0 & -1 \\ -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \right]$$

$$\begin{aligned}
 \mathbf{X}_L &= \mathbf{A} \left[\mathbf{Z}, \mathbf{A}^{*'} (\mathbf{A}^* \mathbf{A}^{*'})^{-1} \mathbf{X}_L^* \right] \\
 &= \left\{ \left[\mathbf{I}_6 \otimes \begin{pmatrix} 0 & -2 \\ -1 & -1 \\ 1 & -1 \end{pmatrix} \right], \begin{pmatrix} 0 & 0 & 0 & 0 \\ -0,5 & -0,5 & -0,5 & 0 \\ -0,5 & -0,5 & -0,5 & 0 \\ 0 & 0 & 0 & 0 \\ -0,5 & 0,5 & -0,5 & 0 \\ -0,5 & 0,5 & -0,5 & 0 \\ 0 & 0 & 0 & 0 \\ -0,5 & -0,5 & 0 & -0,5 \\ -0,5 & -0,5 & 0 & -0,5 \\ 0 & 0 & 0 & 0 \\ -0,5 & 0,5 & 0 & -0,5 \\ -0,5 & 0,5 & 0 & -0,5 \\ 0 & 0 & 0 & 0 \\ -0,5 & -0,5 & 0,5 & 0,5 \\ -0,5 & -0,5 & 0,5 & 0,5 \\ 0 & 0 & 0 & 0 \\ -0,5 & 0,5 & 0,5 & 0,5 \\ -0,5 & 0,5 & 0,5 & 0,5 \end{pmatrix} \right\} \quad (4.2.2)
 \end{aligned}$$

e

$$\boldsymbol{\beta} = (\boldsymbol{\alpha}' \boldsymbol{\beta}^{*'})'$$

onde $\boldsymbol{\beta}^{*'} = (\mu \ \alpha_1 \ \beta_1 \ \beta_2)'$ são os parâmetros correspondentes ao modelo (2.4.5) especificados na Seção 2.4 (p. 54) e $\boldsymbol{\alpha}'$ são parâmetros que não tem interesse interpretativo.

A utilização do BMDP/4F exige como modelo probabilístico a distribuição *Multinomial* além de um modelo log-linear *hierárquico* na formulação (2.3.1), ou mais especificamente na formulação (2.1.2). Pelo exposto na Seção 2.5, um modelo log-linear hierárquico, formulado sob o modelo probabilístico Multinomial, que permite obter as estimativas para o modelo log-linear hierárquico (2.4.5) ou (2.4.6), expresso na forma polinomial, é dado por:

$$\begin{aligned}
 \log \pi_{ijkl} &= u + u_i^I + u_j^S + u_k^A + u_l^L + u_{ij}^{IS} + u_{ik}^{IA} + u_{il}^{IL} + u_{jk}^{SA} + u_{jl}^{SL} + u_{kl}^{AL} \\
 &\quad + u_{ijk}^{ISA} + u_{ijl}^{ISL} + u_{ikl}^{IAL} + u_{jkl}^{SAL}
 \end{aligned}$$

com restrições $\sum_i u_i^I = \sum_j u_j^S = \dots = \sum_j u_{jkl}^{SAL} = \sum_k u_{jkl}^{SAL} = \sum_l u_{jkl}^{SAL} = 0$ e restrição natural $\mathbf{1}_{24}' \boldsymbol{\pi} = 1$. Na forma matricial (2.1.2) esse modelo é dado por:

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ -1 & 1 & -1 & -1 & 1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ -1 & -1 & 1 & -1 & -1 & 1 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ -1 & -1 & 1 & 1 & 1 & -1 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 \\ -1 & -1 & 1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 \\ -1 & -1 & 1 & 1 & 1 & -1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix} \quad (4.2.3)$$

e

$$\beta = (u_1^I \ u_1^S \ u_1^A \ u_1^L \ u_2^L \ u_{11}^{IS} \ u_{11}^{IA} \ u_{11}^{IL} \ u_{12}^{IL} \ u_{11}^{SA} \ u_{11}^{SL} \ u_{12}^{SL} \\ u_{11}^{AL} \ u_{12}^{AL} \ u_{111}^{ISA} \ u_{111}^{ISL} \ u_{112}^{ISL} \ u_{111}^{IAL} \ u_{112}^{IAL} \ u_{111}^{SAL} \ u_{112}^{SAL})'$$

No BMDP/4F é usada a notação de classe geradora para definir os modelos log-lineares hierárquicos. As classes geradoras, correspondem aos parâmetros relativos às interações de ordem mais alta. Como pode ser visto em B.6 (Anexo B), este modelo é especificado pelos termos geradores $ISA \ ISL \ IAL \ SAL$. A partir desse modelo e das restrições associadas obtemos:

$$\begin{aligned}
\log \theta_{kl} &= \log \left(\frac{\pi_{11kl} \pi_{22kl}}{\pi_{12kl} \pi_{21kl}} \right) \\
&= \log \pi_{11kl} - \log \pi_{21kl} - \log \pi_{12kl} + \log \pi_{22kl} \\
&= 4 u_{11}^{IS} + 4 u_{11k}^{ISA} + 4 u_{11l}^{ISL}
\end{aligned}$$

com $\sum_k u_{11k}^{ISA} = \sum_l u_{11l}^{ISL} = 0$.

Em analogia com o modelo (2.4.6) temos que a relação entre os parâmetros desses modelos é:

$$\mu = 4 u_{11}^{IS} \quad ; \quad \alpha_k = 4 u_{11k}^{ISA} \quad \text{e} \quad \beta_l = 4 u_{11l}^{ISL} .$$

Na Tabela 4.2.7 apresentamos as estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões e valores das estatísticas de ajuste.

Tabela 4.2.7 - Estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões(dp) e valores das estatísticas de ajuste para o modelo (2.4.5)

parâmetro	M V		M Q G	
	estimativa	dp estimado	estimativa	dp estimado
μ	-1,047	0,207	-1,042	0,209
α_1	0,669	0,195	0,666	0,197
β_1	0,022	0,260	0,020	0,261
β_2	0,744	0,277	0,737	0,280
Ajuste do modelo ^(a)	$Q_P = 2,05$	$(p = 0,358)$	$Q_W = 2,05$	$(p = 0,359)$
	$Q_L = 2,06$	$(p = 0,357)$		

(a): graus de liberdade = 2

Calcularam-se também as estimativas de MV e MQG das razões de produtos cruzados e respectivos desvios padrões a partir das correspondentes estimativas dos parâmetros e

matriz de covariância estimada (Tabela 4.2.8). Para o cálculo dos desvios padrões utilizamos a seguinte expressão obtida a partir de Koch et al. (1985, p. 84–86):

$$dp(\hat{\theta}) = \sqrt{\mathbf{V}_{\hat{\theta}}} = \sqrt{\mathbf{H}(\mathbf{b}^*) \mathbf{V}_{\mathbf{b}^*} \mathbf{H}(\mathbf{b}^*)'}$$

onde $\mathbf{H}(\mathbf{b}^*) = \frac{\partial \mathbf{F}(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} \big|_{\boldsymbol{\beta}^* = \mathbf{b}^*}$. No caso $\hat{\boldsymbol{\theta}} = \exp(\mathbf{X}_L^* \mathbf{b}^*)$ e $\mathbf{H}(\mathbf{b}^*) = \text{DIAG}(\hat{\boldsymbol{\theta}}) \mathbf{X}_L^*$ com $\mathbf{X}_L^*$ matriz de especificação do modelo (2.4.5). Os comandos dos programas BMDP/4F, GENCAT e LOGLIN.AJU são apresentados em B.6, B.7 e B.8 (Anexo B).

Tabela 4.2.8 -Valores observados e ajustados pelo modelo (2.4.5) das razões de produtos cruzados e respectivos desvios padrões (). Estimativas de MV e MQG

Idade	Localidade e Raça	valores observados	estimativas de MV	estimativas de MQG
1	1	0,508 (0,214)	0,700 (0,247)	0,701 (0,249)
2	1	0,233 (0,084)	0,184 (0,059)	0,185 (0,059)
1	2	2,101 (1,228)	1,442 (0,597)	1,435 (0,600)
2	2	0,329 (0,117)	0,378 (0,120)	0,379 (0,122)
1	3	0,408 (0,239)	0,319 (0,156)	0,322 (0,158)
2	3	0,055 (0,044)	0,084 (0,044)	0,085 (0,045)

Exemplo 4.2.4

Os dados da Tabela 4.2.9 são provenientes de uma pesquisa sobre prevalência de hipertensão arterial entre integrantes da força de trabalho de 10 ramos de atividades na grande São Paulo (para informações sobre a pesquisa ver Debert-Ribeiro et al. (1981)). Um dos objetivos dessa pesquisa estava relacionado com a busca de efeitos de atributos bio-sociais (Escolaridade, Idade e Paridade) no Sobrepeso (medido através do Índice Quetelet).

Pretendemos utilizar esses dados com o intuito de exemplificar uma possível estratégia de análise para uma tabela multidimensional envolvendo uma variável resposta politômica. Não temos a intenção de fazer uma análise exaustiva dos dados; para isso seria importante levar em consideração os aspectos epidemiológicos do problema, o que não tem interesse no contexto deste trabalho.

Tabela 4.2.9 - Distribuição de 1450 mulheres segundo Escolaridade, Idade, Paridade e Sobrepeso

Escolaridade	Idade	Paridade	Sobrepeso			Total
			1	2	3	
Alta	< 35 anos	Sem filhos	27	95	242	364
		Com filhos	8	24	36	68
	≥ 35 anos	Sem filhos	5	13	10	28
		Com filhos	8	17	18	43
Intermediária	< 35 anos	Sem filhos	31	119	319	469
		Com filhos	13	23	17	53
	≥ 35 anos	Sem filhos	2	6	7	15
		Com filhos	6	13	5	24
Baixa	< 35 anos	Sem filhos	9	71	93	173
		Com filhos	21	35	23	79
	≥ 35 anos	Sem filhos	11	7	10	28
		Com filhos	54	29	23	106
Total			195	452	803	1450

Consideramos Escolaridade (E), Idade (I) e Paridade (P) como variáveis explicativas e Sobrepeso (S) como variável resposta. Escolaridade é indexada por i com $i = 1$ indicando escolaridade alta (12 anos ou mais), $i = 2$ indicando escolaridade intermediária (5 a 11 anos) e $i = 3$ indicando escolaridade baixa (0 a 4 anos); Idade é indexada por j com $j = 1$ indicando < 35anos e $j = 2$ indicando ≥ 35anos; Paridade é indexada por k com $k = 1$ indicando sem filhos e $k = 2$ indicando com filhos e Sobrepeso é indexada por

l com $l = 1$ indicando a categoria com alto sobrepeso, $l = 2$ indicando a categoria com sobrepeso intermediário e $l = 3$ indicando a categoria com baixo sobrepeso. A combinação dos níveis das variáveis explicativas definem 12 subpopulações sendo os totais marginais correspondentes tratados como fixos. Denotemos por n_{ijkl} o número de mulheres classificadas na l -ésima categoria de resposta correspondente à subpopulação (i, j, k) e seja π_{ijkl} a probabilidade de uma mulher selecionada aleatoriamente da subpopulação (i, j, k) ser classificada na l -ésima categoria de resposta.

Admitindo-se ainda que as subpopulações assim definidas sejam independentes, um modelo probabilístico apropriado é a distribuição Produto de Multinomiais cuja função de probabilidades é dada por (2.2.1) com $s = 12$ subpopulações, $r = 3$ categorias de resposta e vetor de probabilidades $\pi = (\pi'_{111} \pi'_{112} \pi'_{121} \pi'_{122} \cdots \pi'_{322})'$ com $\pi'_{ijk} = (\pi_{ijk1} \pi_{ijk2} \pi_{ijk3})$, $0 < \pi_{ijkl} < 1$ e $\sum_l \pi_{ijkl} = 1$ para $i, l = 1, 2, 3$, $j, k = 1, 2$.

Consideremos como função de interesse o vetor de logitos múltiplos $\log \pi^*$, onde $\pi^{*'} = (\pi^{*'}_{111} \pi^{*'}_{112} \pi^{*'}_{121} \pi^{*'}_{122} \cdots \pi^{*'}_{322})'$ com $\pi^{*'}_{ijk} = \left(\frac{\pi_{ijk1}}{\pi_{ijk3}}, \frac{\pi_{ijk2}}{\pi_{ijk3}} \right)$. Esses logitos podem ser obtidos empregando-se $A \log \pi$ com $A = \left[I_{12} \otimes \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \right]$.

Na busca de um modelo simples que possa ser considerado adequado aos dados, tentamos utilizar a informação de ordinalidade da variável resposta. Agresti (1983)⁴ sugere atribuir valores aos níveis da variável ordinal que reflitam as distâncias entre eles o que proporciona uma interpretação para o modelo relacionada com razões de produtos cruzados ("o logaritmo da razão de produtos cruzados é um valor proporcional à distância entre os níveis da variável ordinal"). Na ausência de tal informação, o autor indica a escolha de valores arbitrários igualmente espaçados. Um modelo empregando valores igualmente espaçados foi testado. Os valores das estatísticas de Wald, $Q_W = 19,45$ ($gl = 11$ e $p = 0,053$), de Pearson $Q_P = 20,36$ ($gl = 11$ e $p = 0,041$) e da Razão de Verossimilhança, $Q_L = 19,85$ ($gl = 11$ e $p = 0,047$) indicam falta de ajuste.

Dado que esse modelo não se ajusta, uma alternativa é considerar o problema sob um ponto de vista multivariado tentando obter modelos que expliquem a variabilidade dos logitos múltiplos simultaneamente. Com esse objetivo partimos de um modelo log-linear saturado escrito na forma (2.2.2) com

$$\mathbf{X}_L = \mathbf{I}_{24} \quad (4.2.4)$$

Os parâmetros correspondem aos próprios logitos definidos acima. Essa parametrização conhecida como *modelo de médias das caselas* é extremamente útil para verificar a significância estatística de diversas fontes de variação através de hipóteses lineares da forma $H : \mathbf{C}\boldsymbol{\beta} = 0$ utilizando as estatísticas de Wald dadas em (3.1.12) e (3.2.15). Por exemplo, para a hipótese de não interação entre as tres variáveis explicativas relativamente aos dois logitos uma possível matriz $\mathbf{C}$ é :

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \end{pmatrix} ;$$

analogamente, o efeito da variável Idade pode ser testado com

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \end{pmatrix} .$$

Os cálculos correspondentes podem ser facilmente realizados através das subrotinas LOGLIN.AJU e LOGLIN.HIP. A Tabela 4.2.10, apresenta resultados para diversas hipóteses de interesse; podemos verificar que os efeitos de interação EIP e IP não são significativos para os níveis de significância usuais ($\alpha < 0,10$, por exemplo), sugerindo um modelo log-linear hierárquico definido pelos termos de interação de primeira ordem EI e EP .

Tabela 4.2.10 - Estatísticas de Wald* e níveis descritivos para hipóteses envolvendo parâmetros do modelo (4.2.4)

fonte de variação	g.l.	Q_W	nível descritivo
E	4	13,65	0,009
I	2	21,80	0,000
P	2	23,33	0,000
EI	4	12,55	0,014
EP	4	8,67	0,070
IP	2	3,69	0,158
EIP	4	2,14	0,710

Este modelo também pode ser expresso na formulação (2.2.5) envolvendo restrições com

$$W = \begin{pmatrix} 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 \end{pmatrix}$$

Note que as linhas dessa matriz definem os parâmetros do modelo considerados nulos; por exemplo, as duas primeiras linhas definem a interação IP e as demais linhas a interação EIP . O ajuste desse modelo pode ser verificado através do teste de hipótese $H_0 : W\beta = 0$. O resultado obtido através do programa LOGLIN.HIP indica $Q_W = 8,09$ ($gl = 6$, $p = 0,232$) sugerindo que o modelo é adequado para explicar a variabilidade dos dados.

Na forma matricial (2.2.2) esse modelo é definido por:

$$X_L = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \quad (4.2.5)$$

e

$$\beta = (u \quad v \quad u_1^E \quad v_1^E \quad u_2^E \quad v_2^E \quad u_1^I \quad v_1^I \quad u_1^P \quad v_1^P \quad u_{11}^{EI} \quad v_{11}^{EI} \quad u_{21}^{EI} \quad v_{21}^{EI} \quad u_{11}^{EP} \quad v_{11}^{EP} \quad u_{21}^{EP} \quad v_{21}^{EP})'$$

onde u é a média dos $\log \frac{\pi_{ijk1}}{\pi_{ijk3}}$ e v é a média dos $\log \frac{\pi_{ijk2}}{\pi_{ijk3}}$; os parâmetros denotados por $\underline{u}$'s envolvendo índices correspondentes a uma única variável (Escolaridade, Idade ou Paridade) são efeitos principais dessas variáveis sobre $\log \frac{\pi_{ijk1}}{\pi_{ijk3}}$; analogamente aqueles denotados por $\underline{v}$'s são efeitos principais dessas variáveis sobre $\log \frac{\pi_{ijk2}}{\pi_{ijk3}}$; os demais parâmetros são efeitos de interação entre Escolaridade e Idade ou Escolaridade e Paridade sobre os mesmos logitos. Os valores das estatísticas de ajuste ($Q_W = 8,09$ ($gl = 6$,

$p = 0,232$), $Q_P = 8,30$ ($gl = 6$, $p = 0,217$) e $Q_L = 8,12$ ($gl = 6$, $p = 0,230$) não contradizem o modelo. Note que a estatística Q_W é idêntica àquela obtida para o modelo na formulação (2.2.5) (ver p. 98).

Para estudar a estrutura dessas interações, utilizamos uma outra parametrização, equivalente à do modelo (4.2.5), que facilita o estudo dos efeitos de Idade e Paridade nos diversos níveis de Escolaridade. Na forma (2.2.2) esse modelo é dado por :

$$X_L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (4.2.6)$$

e

$$\beta = (\mu_1 \ \mu_2 \ \alpha_1 \ \beta_1 \ \alpha_2 \ \beta_2 \ \alpha_3 \ \beta_3 \ \alpha_4 \ \beta_4 \ \alpha_5 \ \beta_5 \ \alpha_6 \ \beta_6 \ \alpha_7 \ \beta_7 \ \alpha_8 \ \beta_8)'$$

onde $\mu_1 = \log \frac{\pi_{1111}}{\pi_{1113}}$ e $\mu_2 = \log \frac{\pi_{1112}}{\pi_{1113}}$, logitos múltiplos correspondentes à Escolaridade *alta*, Idade *< 35 anos* e Paridade *sem filhos*, são valores de referência e $\{\alpha_h\}$ e $\{\beta_h\}$ representam variações nesses valores de referência relativas a alterações de Escolaridade,

Idade ou Paridade. Mais especificamente,

- α_1 e β_1 são variações relativas à μ_1 e μ_2 devido a Escolaridade *intermediária*;
- α_2 e β_2 são variações relativas à μ_1 e μ_2 devido a Escolaridade *baixa*;
- α_3 e β_3 são variações relativas à μ_1 e μ_2 devido a Idade ≥ 35 anos nas subpopulações com Escolaridade *alta*;
- α_4 e β_4 são variações relativas à μ_1 e μ_2 devido a Idade ≥ 35 anos nas subpopulações com Escolaridade *intermediária*;
- α_5 e β_5 são variações relativas à μ_1 e μ_2 devido a Idade ≥ 35 anos nas subpopulações com Escolaridade *baixa*;
- α_6 e β_6 são variações relativas à μ_1 e μ_2 devido a Paridade *com filhos* nas subpopulações com Escolaridade *alta*;
- α_7 e β_7 são variações relativas à μ_1 e μ_2 devido a Paridade *com filhos* nas subpopulações com Escolaridade *intermediária*;
- α_8 e β_8 são variações relativas à μ_1 e μ_2 devido a Paridade *com filhos* nas subpopulações com Escolaridade *baixa*.

Estimativas dos parâmetros desse modelo e de seus erros padrões estão indicadas na Tabela 4.2.11. Obviamente as estatísticas de ajuste são idênticas àquelas correspondentes ao modelo (2.4.5).

Em seguida, comparamos as variações relativas a Idade e Paridade entre os diversos níveis de Escolaridade. As hipóteses testadas, as estatísticas de Wald correspondentes bem como os níveis descritivos associados estão indicados na Tabela 4.2.12. A partir desses resultados foi testado o ajuste de um modelo (modelo (2.4.7)) em que as variações devidas a Idade ≥ 35 anos para as subpopulações com Escolaridade *alta* ou *intermediária* são admitidas iguais ($\alpha_3 = \alpha_4$, $\beta_3 = \beta_4$). As igualdades $\alpha_6 = \alpha_8$, $\beta_6 = \beta_8$ e $\alpha_7 = \alpha_8$, $\beta_7 = \beta_8$ detectadas não foram incorporadas no modelo em função da incongruência oriunda do método de comparações múltiplas adotado. (Note que $\alpha_6 = \alpha_7$, $\beta_6 = \beta_7$ é rejeitada.)

Tabela 4.2.11 - Estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões(dp) e valores das estatísticas de ajuste para o modelo (2.4.6)

parâmetro	M V		M Q G	
	estimativa	dp estimado	estimativa	dp estimado
μ_1	-2,158	0,195	-2,158	0,196
μ_2	-0,904	0,118	-0,904	0,118
α_1	-0,156	0,268	-0,154	0,270
β_1	-0,081	0,159	-0,081	0,159
α_2	0,203	0,329	0,276	0,351
β_2	0,636	0,194	0,636	0,193
α_3	1,088	0,416	1,074	0,410
β_3	0,770	0,305	0,759	0,306
α_4	0,699	0,541	0,712	0,525
β_4	0,773	0,416	0,772	0,415
α_5	1,414	0,324	1,398	0,318
β_5	-0,090	0,300	-0,090	0,309
α_6	0,496	0,377	0,500	0,370
β_6	0,343	0,257	0,346	0,255
α_7	1,978	0,388	1,972	0,383
β_7	1,262	0,307	1,261	0,309
α_8	1,567	0,334	1,508	0,334
β_8	0,638	0,264	0,634	0,272
Ajuste do modelo ^(a)	$Q_P = 8,30$	$(p = 0,217)$	$Q_W = 8,09$	$(p = 0,232)$
	$Q_L = 8,12$	$(p = 0,230)$		

(a): graus de liberdade = 6

Tabela 4.2.12 - Estatísticas de Wald e níveis descritivos para hipóteses envolvendo parâmetros do modelo (2.4.6)

hipótese nula	g.l.	M V		M Q G	
		Q_W	nível descritivo	Q_W	nível descritivo
$\alpha_3 = \alpha_4 = \alpha_5$; $\beta_3 = \beta_4 = \beta_5$	4	10,64	0,031	10,51	0,033
$\alpha_3 = \alpha_4$; $\beta_3 = \beta_4$	2	0,40	0,817	0,38	0,826
$\alpha_3 = \alpha_5$; $\beta_3 = \beta_5$	2	6,49	0,039	6,46	0,040
$\alpha_4 = \alpha_5$; $\beta_4 = \beta_5$	2	7,75	0,021	7,66	0,022
$\alpha_6 = \alpha_7 = \alpha_8$; $\beta_6 = \beta_7 = \beta_8$	4	10,66	0,031	10,38	0,035
$\alpha_6 = \alpha_7$; $\beta_6 = \beta_7$	2	9,48	0,009	9,46	0,009
$\alpha_6 = \alpha_8$; $\beta_6 = \beta_8$	2	4,52	0,104	4,10	0,129
$\alpha_7 = \alpha_8$; $\beta_7 = \beta_8$	2	2,41	0,300	2,41	0,300

Os resultados relativos às estatísticas de ajuste para o modelo (2.4.7) apresentados na Tabela 4.2.13 indicam que esse modelo com 16 parâmetros parece ser aceitável ($Q_W = 8,47$ ($gl = 8$, $p = 0,389$), $Q_P = 8,69$ ($gl = 8$, $p = 0,369$), $Q_L = 8,52$ ($gl = 8$, $p = 0,384$)). Na mesma tabela, verificamos também que os parâmetros β_5 , α_6 e β_6 são não significativos. Testadas ainda a hipótese simultânea de igualdade entre α_7 e α_8 e entre β_7 e β_8 e a hipótese de igualdade entre α_3^* e β_3^* (variações devido a idade nas subpopulações com Escolaridade *alta* ou *intermediária*) verificamos que tais hipóteses não são rejeitadas para os níveis de significância usuais (Tabela 4.2.14).

Tabela 4.2.13 - Estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões(dp) e valores das estatísticas de ajuste para o modelo (2.4.7)

parâmetro	M V		M Q G	
	estimativa	dp estimado	estimativa	dp estimado
μ_1	-2,139	0,191	-2,137	0,193
μ_2	-0,906	0,117	-0,906	0,118
α_1	-0,185	0,263	-0,186	0,264
β_1	-0,079	0,157	-0,077	0,157
α_2	0,185	0,327	0,256	0,350
β_2	0,638	0,193	0,639	0,193
α_3^*	0,931	0,328	0,930	0,322
β_3^*	0,781	0,246	0,770	0,246
α_5	1,414	0,324	1,398	0,318
β_5	-0,090	0,300	-0,090	0,309
α_6	0,555	0,362	0,548	0,359
β_6	0,336	0,251	0,343	0,249
α_7	1,915	0,375	1,920	0,371
β_7	1,269	0,298	1,269	0,289
α_8	1,567	0,334	1,508	0,334
β_8	0,638	0,264	0,634	0,272
Ajuste do modelo ^(a)	$Q_P = 8,69$ $Q_L = 8,52$	$(p = 0,369)$ $(p = 0,384)$	$Q_W = 8,47$	$(p = 0,389)$

(a): graus de liberdade = 8

Tabela 4.2.14 - Estatísticas de Wald e níveis descritivos para hipóteses envolvendo parâmetros do modelo (2.4.7)

hipótese nula	g.l.	M V		M Q G	
		Q_W	nível descritivo	Q_W	nível descritivo
$\alpha_3^* = \alpha_5 ; \beta_3^* = \beta_5$	2	10,31	0,006	10,12	0,006
$\alpha_3^* = \beta_3^*$	1	0,22	0,642	0,26	0,610
$\alpha_6 = \alpha_7 ; \beta_6 = \beta_7$	2	9,62	0,008	9,78	0,007
$\alpha_6 = \alpha_8 ; \beta_6 = \beta_8$	2	4,23	0,121	3,83	0,147
$\alpha_7 = \alpha_8 ; \beta_7 = \beta_8$	2	2,52	0,283	2,49	0,287

Posteriormente, ajustamos um modelo incorporando esses resultados. Na forma matricial (2.2.2) esse modelo é definido por :

$$X_L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.4.8)$$

e

$$\beta = (\mu_1 \ \mu_2 \ \alpha_1 \ \beta_1 \ \alpha_2 \ \beta_2 \ \gamma^I \ \lambda^I \ \alpha^P \ \beta^P)'$$

onde:

- γ^I variação devido a Idade ≥ 35 anos nas subpopulações com Escolaridade *alta* ou *intermediária* para os dois logitos,
- λ^I variação devido a Idade ≥ 35 anos nas subpopulações com Escolaridade *baixa* para o logito correspondente a alto sobrepeso versus baixo sobrepeso,
- α^P variação devido a Paridade *com filhos* nas subpopulações com Escolaridade *intermediária* ou *baixa* para o logito correspondente a alto sobrepeso versus baixo sobrepeso,
- β^P variação devido a Paridade *com filhos* nas subpopulações com Escolaridade *intermediária* ou *baixa* para o logito correspondente a intermediário sobrepeso versus baixo sobrepeso.

Na Tabela 4.2.15 apresentamos as estimativas dos parâmetros desse modelo e de seus erros padrões e também os valores das estatísticas de ajuste.

Tabela 4.2.15 - Estimativas de MV e de MQG dos parâmetros, respectivos desvios padrões(dp) e valores das estatísticas de ajuste para o modelo (2.4.8)

parâmetro	M V		M Q G	
	estimativa	dp estimado	estimativa	dp estimado
μ_1	-1,997	0,159	-1,982	0,162
μ_2	-0,864	0,106	-0,863	0,106
α_1	-0,309	0,230	-0,316	0,232
β_1	-0,082	0,144	-0,084	0,144
α_2	-0,012	0,289	0,030	0,305
β_2	0,490	0,176	0,492	0,175
γ^I	0,979	0,216	0,987	0,217
λ^I	1,458	0,275	1,436	0,270
α^P	1,704	0,244	1,662	0,245
β^P	0,866	0,189	0,852	0,189
Ajuste do modelo ^(a)	$Q_P = 15,28$	$(p = 0,359)$	$Q_W = 14,98$	$(p = 0,379)$
	$Q_L = 14,92$	$(p = 0,384)$		

(a): graus de liberdade = 14

Calcularam-se também as estimativas de MV e de MQG para as razões entre probabilidade de Sobrepeso alto e probabilidade de Sobrepeso baixo e para as razões entre probabilidade de Sobrepeso intermediário e probabilidade de Sobrepeso baixo. Esses valores estão indicados na Tabela 4.2.16.

Tabela 4.2.16 - Valores observados e ajustados pelo modelo (2.4.8) por MV e MQG para as razões entre probabilidades.

Escol.	Idade	Paridade	Sobrepeso					
			Alto vs Baixo			Interm. vs Baixo		
			OBS.	MV	MQG	OBS.	MV	MQG
Alta	< 35 anos	Sem filhos	0,11	0,14	0,14	0,39	0,42	0,42
		Com filhos	0,22	0,14	0,14	0,67	0,42	0,42
	≥ 35 anos	Sem filhos	0,50	0,36	0,37	1,30	1,12	1,13
		Com filhos	0,44	0,36	0,37	0,94	1,12	1,13
Interm.	< 35 anos	Sem filhos	0,10	0,10	0,10	0,37	0,39	0,39
		Com filhos	0,76	0,55	0,53	1,35	0,92	0,91
	≥ 35 anos	Sem filhos	0,29	0,27	0,27	0,86	1,03	1,04
		Com filhos	1,20	1,46	1,42	2,60	2,46	2,44
Baixa	< 35 anos	Sem filhos	0,10	0,13	0,14	0,76	0,69	0,69
		Com filhos	0,91	0,74	0,75	1,52	1,64	1,62
	≥ 35 anos	Sem filhos	1,10	0,58	0,60	0,70	0,69	0,69
		Com filhos	2,35	3,17	3,14	1,26	1,64	1,62

Note que na Escolaridade alta apenas a Idade parece influenciar a chance de um maior Sobrepeso; para as subpopulações com Idade ≥ 35 anos as razões entre as probabilidades de Sobrepeso alto versus baixo e Sobrepeso intermediário versus baixo são cerca de 2,7 vezes mais elevadas que àquelas relativas à Idade < 35 anos. Na Escolaridade intermediária, além de um mesmo efeito de Idade, também ocorre um efeito de Paridade; no entanto, esse efeito é mais acentuado para as razões correspondentes a Sobrepeso alto versus baixo onde os valores relativos a Paridade com filhos são cerca de 5,5 vezes maiores que os relativos

a Paridade sem filhos. Nas subpopulações com Escolaridade baixa, também ocorre efeito dessas duas variáveis apenas para as razões entre Sobrepeso alto e baixo; nesse caso, a Idade contribui com um aumento de cerca de 4,5 e a Paridade em torno de 5,2. As razões entre probabilidade de Sobrepeso intermediário e probabilidade de Sobrepeso baixo nas subpopulações com Escolaridade baixa são afetadas apenas pela Paridade; aquelas correspondentes a paridade com filhos apresentam valores 2,4 vezes maiores.

Ressaltamos que a estratégia de análise desenvolvida para a seleção de um modelo não é única. Alternativamente, por exemplo, propor modelos dirigidos para o estudo da influência das variáveis explicativas em cada logito separadamente pode ser uma estratégia interessante e talvez adequada do ponto de vista epidemiológico.

Outro ponto importante é a verificação do ajuste de um modelo particular. Para concluir que o modelo é adequado não devemos nos basear apenas nas estatísticas globais de ajuste (ver Bishop et al. (1985, cap.4) por exemplo). Devemos examinar o ajuste para cada casela, através de análise de resíduos e análise detalhada das observações. Com tal procedimento é possível detectar observações influentes ou "outliers". Diversas técnicas de diagnóstico disponíveis para modelos lineares generalizados podem ser empregadas em modelos log-lineares.

Referências Bibliográficas

- AGRESTI, A. (1984) *Analysis of Ordinal Categorical Data*. New York, John Wiley and Sons 287 p.
- ALTHAM, P. M. E. (1970a) The measurement of association of rows and columns for an $r \times s$ contingency table. *J. Roy. Statist. Soc., Ser. B*, **32**: 63–73.
- ALTHAM, P. M. E. (1970b) The measurement of association in a contingency table: three extensions of the cross-ratios and metric methods. *J. Roy. Statist. Soc., Ser. B*, **32**: 395–407.
- ANDERSEN, A. H. (1974) Multi-dimensional contingency tables. *Scand. J. Statist.*, **1**: 115–127.
- BAKER, R. J. & NELDER, J. A. (1978) *The GLIM System. Release 3. Generalised Linear Interactive Modelling Manual*. Oxford, N.A.G.
- BARTLETT, M. S. (1935) Contingency table interactions. *J. R. Statist.*

- BENÍCIO, M. H. D'A. (1983) *Fatores de baixo peso ao nascer em recém-nascidos vivos - Município de São Paulo, 1978*. São Paulo. 137 p. Tese (Doutoramento) - Faculdade de Medicina da USP.
- BHAPKAR, V. P. (1966) A note on the equivalence of two test criteria for hypotheses in categorical data. *J. Amer. Statist. Assoc.*, **61**: 228-235.
- BHAPKAR, V. P. & KOCH, G. G. (1968a) Hypotheses of "no interaction" in multidimensional contingency tables. *Technometrics*, **10**: 107-123.
- BHAPKAR, V. P. & KOCH, G. G. (1968b) On the hypotheses of "no interaction" in multidimensional contingency tables. *Biometrics*, **24**: 567-594.
- BIRCH, M. W. (1963) Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc., Ser. B*, **25**: 220-233.
- BISHOP, Y. M. M., FIENBERG, S. E.; HOLLAND, P. W. (1975) *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MIT Press 557 p.
- BRESLOW, N. E. & DAY, N. E. (1980) *Statistical Methods in Cancer Research: The analysis of case-control studies*. (IARC Scientific Publications, n. 32). v.1 338 p.
- COX, D. R. (1970) *The analysis of Binary Data*. London, Methuen. 142 p.
- DEBERT-RIBEIRO, M. B. et al. (1981) Hypertension and economic activities in São Paulo, Brazil. *Hypertension*, supp. II, **3** (6): 233-237.
- DEMING, W. E. & STEPHAN, F. F. (1940) On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.*, **11**: 427-444.
- DIXON, J. W. ed. (1985) *BMDP Statistical Software*, 1985 edition. Berkeley, University of California Press. 733p.

- EVERITT, B. S. (1977) *The Analysis of Contingency Tables*. London, Chapman & Hall. 128 p.
- FAY, R. & GOODMAN, L. A. (1975) *ECTA Program: Description for Users*. Chicago, Department of Statistics, University of Chicago.
- FIENBERG, S. E. (1980) *The Analysis of Cross-Classification Categorical Data*. 2.ed. Cambridge, MIT Press. 198 p.
- GOODMAN, L. A. (1968) The analysis of cross-classified data: independence, quasi-independence, and interaction in contingency tables with or without missing cells. *J. Amer. Statist. Assoc.*, **63**: 1091-1131.
- GOODMAN, L. A. (1970) The multivariate analysis of qualitative data: interactions among multiple classifications. *J. Amer. Statist. Assoc.*, **65**: 226-256.
- GOODMAN, L. A. (1971) Partitioning of chi square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. *J. Amer. Statist. Assoc.*, **66**: 339-344.
- GOODMAN, L. A. (1978) *Analysing Qualitative Categorical Data: Log-Linear Analysis and Latent Structure Analysis*. Cambridge, Abt. Associates.
- GOODMAN, L. A. (1979) Simple models for the analysis of association in cross-classifications having ordered categories in contingency tables with or without missing cells. *J. Amer. Statist. Assoc.*, **74**: 537-552.
- GOODMAN, L. A. & KRUSKAL, W. H. (1954) Measures of association for cross-classifications. *J. Amer. Statist. Assoc.*, **49**: 732-764.

- GOODMAN, L. A. & KRUSKAL, W. H. (1959) Measures of association for cross-classifications, II: further discussion and references. *J. Amer. Statist. Assoc.*, **54**: 123-163.
- GOODMAN, L. A. & KRUSKAL, W. H. (1963) Measures of association for cross-classifications, III: approximate sampling theory. *J. Amer. Statist. Assoc.*, **58**: 310-364.
- GOODMAN, L. A. & KRUSKAL, W. H. (1972) Measures of association for cross-classifications, IV: simplification of asymptotic variances. *J. Amer. Statist. Assoc.*, **67**: 415-421.
- GRIZZLE, J. E., STARMER, C. F.; KOCH, G. G. (1969) The analysis of categorical data by linear models. *Biometrics*, **25**: 489-504.
- HABER, M. (1985) Maximum likelihood methods for linear and log-linear models in categorical data. *Computational Statistics & Data Analysis*, **3**: 1-10.
- HABERMAN, S. J. (1974) *The Analysis of Frequency Data*. Chicago, University of Chicago Press. 419 p. (Statistical Research Monographs, 4).
- HABERMAN, S. J. (1978) *The Analysis of Qualitative Data: Introductory Topics*. New York, Academic Press. v.1
- IMREY, P. B. et al. (1981) Categorical data analysis: some reflections on the log linear model and logistic regression. Part I: Historical and methodological overview. *Int. Statist. Rev.*, **49**: 265-283.
- IMREY, P. B. et al. (1982) Categorical data analysis: some reflections on the log linear model and logistic regression. Part II: Data analysis. *Int. Statist. Rev.*, **50**: ~~305~~ 35-63.

- KOCH, G. G. et al. (1985) *Analysis of Categorical Data*. Montréal, Les Presses de l'Université de Montréal. 290 p.
- KOCH, G. G., SINGER, J. M. & AMARA, I. A. (1985) A two stage procedure for the analysis of ordinal categorical data. In: *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences - The Greenberg Volume*, P. K. Sen, editor, North Holland, Amsterdam.
- LANDIS, J. R. et al. (1976) A computer program for the generalized chi-square analysis of categorical data using least squares. *Comput. Programs Biomed.*, 6: 196-231.
- MONTGOMERY, D. C. & PECK, E. A. (1982) *Introduction to linear regression analysis*. New York, John Wiley & Sons. 504 p.
- NORUSIS, M. J. ed. (1986a) *SPSS^X Advanced Statistics Guide*. 2.ed. Chicago, McGraw-Hill, SPSS Inc. 505 p.
- NORUSIS, M. J. ed. (1986b) *SPSS/PC⁺ Advanced Statistics for the IBM PC/XT/AT*. Chicago, McGraw-Hill, SPSS Inc.
- PAULINO, C. D. M. (1988) Análise de dados categorizados incompletos: fundamentos, métodos e aplicações. São Paulo. 254 p. Tese (Doutoramento) - Instituto de Matemática e Estatística da USP.
- PLACKETT, R. L. (1974) *The Analysis of Categorical Data*. London, Griffin. 157 p.
- RAO, C. R. (1973) *Linear Statistical Inference and its Applications*. 2.ed. New York, John Wiley & Sons.
- SAS: *User's Guide: Statistics*. Cary, NC: SAS Institute Inc., 1985. (Versão 5) 956 p.
- SEARLE, S. R. (1982) *Matrix Algebra useful for Statistics*. New York, John Wiley & Sons. 438 p.

- SEBER, G. A. F. (1977) *Linear Regression Analysis*. New York, John Wiley & Sons. 465 p.
- SEMENYA, K. et al. (1983) Linear models methods for some rank function analysis of ordinal categorical data. *Commun. Statist. Ser. A*, **12**: 1277-1298.
- SINGER, J. M. (1987) Alguns procedimentos para análise de dados categorizados ordinais. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, **15**, Poços de Caldas, 1987. *Atas*. Rio de Janeiro, IMPA. p. 217-231.
- SOC Software científico elaborado pelo Núcleo tecnológico de Informática para Agropecuária da EMBRAPA (NTIA/EMBRAPA) - Caixa Postal 5.010 - CEP 13.081 - Campinas, SP.
- UPTON, G. J. G. (1978) *The Analysis of Cross-Tabulated Data*. New York, John Wiley & Sons. 148 p.
- WALD, A. (1943) Tests of statistical hypotheses concerning general parameters when the number of observations is large. *Trans. Am. Math. Soc.*, **54**: 426-482.
- WILLIAMS, O. D. & GRIZZLE, J. E. (1971) *Analysis of categorical data with more than one response variable by linear models*. Chapel Hill, NC. Department of Biostatistics. University of North Carolina at Chapel Hill (Institute of Statistics Mimeo Series n. 715).
- WINER, B. J. (1971) *Statistical principles in experimental design*. 2.ed. New York, McGraw-Hill. 907 p.

Anexo A

LOGLIN.AJU e LOGLIN.HIP

LOGLIN.AJU e LOGLIN.HIP são programas criados a partir do módulo CM do pacote científico SOC.

O programa LOGLIN.AJU ajusta modelos log-lineares na formulação logitos múltiplos a tabelas de contingência para as quais a distribuição Produto de Multinomiais é adequada. Note que o ajuste para situações em que a distribuição Multinomial é adequada, pode ser obtido como um caso particular. Mais especificamente, esse programa ajusta modelos da forma:

$$F(\pi) = A \log \pi = X_L \beta$$

onde π ($sr \times 1$) é o vetor de probabilidades nas caselas, $A = \{ I_s \otimes [I_{(r-1)}, -1_{(r-1)}] \}$, X_L [$s(r-1) \times t$] é a matriz de especificação do modelo e β ($t \times 1$), $t < s(r-1)$, é o vetor de parâmetros a ser estimado.

As estimativas são obtidas por Mínimos Quadrados Generalizados (MQG) e por Máxima Verossimilhança (MV); as estimativas de MV são calculadas através do algoritmo de Newton-Raphson.

O programa LOGLIN.HIP testa hipóteses do tipo $H_0 : C\beta = 0$ através da estatística de Wald da forma (3.2.13). Este programa só pode ser utilizado após a execução do programa LOGLIN.AJU pois são utilizadas as estimativas de MQG e MV para o vetor de parâmetros e correspondente matriz de covariância calculadas por esse programa.

A.1 - Entrada de Dados

Para a execução do LOGLIN.AJU são necessários os seguintes dados:

- s número de subpopulações
- r número de perfis de respostas
- MN tabela de contingência ($s \times r$)
- XL matriz de especificação do modelo linear para os logitos [$s(r-1) \times t$] ou
- X matriz de especificação do modelo linear para $\log \pi$ ($sr \times t$)

- ENTX = 0 se o usuário especificar XL
- ENTX = 1 se o usuário especificar X
- ENTB = 0 para que as estimativas de MQG sejam usadas como estimativas iniciais no processo iterativo de Newton-Raphson
- ENTB = 1 se o usuário quiser especificar as estimativas iniciais para o processo iterativo de Newton-Raphson; nesse caso, indicar **b**: vetor de estimativas iniciais dos parâmetros
- COVP = 0 para que não seja impressa a matriz de covariância para as proporções observadas
- COVP = 1 para que seja impressa a matriz de covariância para proporções observadas
- COVBO = 0 para que não seja impressa a matriz de covariância estimada para os parâmetros por MQG
- COVBO = 1 para que seja impressa a matriz de covariância estimada para os parâmetros por MQG
- COVB = 0 para que não seja impressa a matriz de covariância estimada para os parâmetros por MV
- COVB = 1 para que seja impressa a matriz de covariância estimada para os parâmetros por MV
- COV = 0 para que não sejam impressas as matrizes de covariâncias estimadas para proporções ajustadas por MQG e por MV
- COV = 1 para que as matrizes de covariâncias estimadas para proporções ajustadas por MQG e por MV sejam impressas

Para a execução de LOGLIN.HIP é necessário especificar C , matriz de contrastes correspondente a uma hipótese específica $H_0 : C\beta = 0$.

A.2 - Dados de Saída

As saídas que o programa fornece são:

MN	tabela de contingência
MP	matriz de proporções observadas
XL	matriz de especificação do modelo
ESTMQG	estimativas de MQG para os parâmetros (β) , respectivos desvios padrões e valores dessas estimativas em unidades de desvios padrões
QW	estatística de Wald para testar o ajuste do modelo por MQG
EST0	estimativa inicial de β
CONV	número de iterações e critério de convergência
ESTMV	estimativas de MV para os parâmetros (β) , respectivos desvios padrões e valores dessas estimativas em unidades de desvios padrões
QP	estatística de Pearson para testar o ajuste do modelo por MV
QL	estatística da Razão de Verossimilhança para testar o ajuste do modelo por MV
LOGI	logitos múltiplos observados e ajustados por MQG e MV e respectivos desvios padrões
PROP	proporções observadas e ajustadas por MQG e MV e respectivos desvios padrões

Opcionalmente,

VP	estimativa da matriz de covariância para as proporções observadas
Vb0	estimativa da matriz de covariância para os parâmetros por MQG
HINV	estimativa da matriz de covariância para os parâmetros por MV
VPII0	estimativa da matriz de covariância para as proporções ajustadas pelo método de MQG
VPII	estimativa da matriz de covariância para as proporções ajustadas pelo método de MV

No caso do programa LOGLIN.HIP as saídas são:

C	matriz que especifica o contraste a ser testado
Cb0	estimativa do contraste por MQG
QWCO	valor da estatística de Wald para o teste do contraste utilizando as estimativas de MQG
Cb	estimativa do contraste por MV
QWC	valor da estatística de Wald para o teste do contraste utilizando as estimativas de MV

A.3 - Exemplos

Exemplo 1 - Os resultados para o modelo (2.1.9) apresentados na Seção 4.2, Tabela 4.2.1, podem ser obtidos através do programa:

```

ANOTE "          ***** OUTUBRO/88 ***** ";
ANOTE " " ;
ANOTE "          DADOS DE BARTLETT ";
ANOTE " " ;
ANOTE "          Tabela 1.6 - modelo 2.1.9 (AB AC BC) ";
ANOTE " ";
s = 1 ;
r = 8 ;
MN = { 156  84  84 156 107 133  31 209 } ;
X = {
      1      1      1      1      1      1 ,
      (-1)    1      1      (-1) (-1)    1 ,
      1      (-1)    1      (-1)    1      (-1),
      (-1) (-1)    1      1      (-1) (-1),
      1      1      (-1)    1      (-1) (-1),
      (-1)    1      (-1) (-1)    1      (-1),
      1      (-1) (-1) (-1) (-1)    1 ,
      (-1) (-1) (-1)    1      1      1 } ;
ENTX = 1 ;
ENTB = 0 ;
COVP = 0 ;
COVBO = 0 ;
COVB = 0 ;
COV = 0 ;
IMPRIME X $ 5:0 ;
ANOTE " " ;
EXEC "A:LOGLIN.AJU";
FIM;

```

Exemplo 2 - Os resultados para o ajuste do modelo (2.2.8) apresentados na Seção 4.2, Tabela 4.2.5, podem ser obtidos através do programa a seguir. Observe que após o comando << EXEC "A:LOGLIN.AJU"; >> são definidos contrastes C que são testados com execuções sucessivas do programa LOGLIN.HIP.

```

ANOTE "          ***** OUTUBRO/88          ***** ";
ANOTE " ";
ANOTE "          CIRURGIAS X SEVERIDADE DA SEQUELA ";
ANOTE " ";
ANOTE "          MODELO 2.2.8 ";
ANOTE " ";
s = 4;
r = 3;
MN = { 61 28 7, 68 23 13, 58 40 12, 53 38 16 };
XL = { 1 0 1 0 1 0 1 0,
        0 1 0 1 0 1 0 1,
        0 0 1 0 0 0 0 0,
        0 0 0 1 0 0 0 0,
        0 0 0 0 1 0 0 0,
        0 0 0 0 0 1 0 0,
        0 0 0 0 0 0 1 0,
        0 0 0 0 0 0 0 1 }';
ENTX = 0;
ENTB = 0;
COVP = 1;
COVBO = 1;
COVB = 1;
COV = 0;
EXEC "A:LOGLIN.AJU";
C = { 0 0 1 0 0 (-1) 0 0 };
EXEC "A:LOGLIN.HIP";
C = { 0 0 0 1 0 0 (-1) 0 };
EXEC "A:LOGLIN.HIP";
C = { 0 0 0 0 1 0 0 (-1) };
EXEC "A:LOGLIN.HIP";
C = { 0 0 1 0 0 0 (-1) 0 0,
        0 0 0 1 0 0 0 (-1) 0,
        0 0 0 0 1 0 0 0 (-1) };
EXEC "A:LOGLIN.HIP";
FIM;

```

A.4 - Listagens dos programas LOGLIN.AJU e LOGLIN.HIP

```

/*          *****   PROGRAMA   LOGLIN.AJU   *****          */
/*
/*          Versao 1.0 -   Setembro/88          */
/* Este programa ajusta modelos log-lineares na formulacao logitos
/* multiplos via MQG e MV, pelo metodo iterativo de Newton-Raphson,
/* considerando a distribuicao Produto de Multinomiais */
/* ----- */
/*          /*   MATRIZES AUXILIARES UTILIZADAS NOS CALCULOS   */          */
/* ----- */
    Ur = VDIAG(IDENT(r));          /* vetor unitario (rx1) */
    U = VDIAG(IDENT(r-1));          /* vetor unitario ((r-1)x1) */
    BD1r = (IDENT(s)@(Ur * Ur'));  /* matriz bloco diagonal com
                                   blocos 1rr (sr x sr) */
    BD1 = (IDENT(s)@(U * U'));     /* matriz bloco diagonal com blocos
                                   1(r-1)x(r-1) (dim. s(r-1) x s(r-1)) */
    J = {IDENT(r-1)(-U)};          /* matriz [I(r-1), -1(r-1)]
                                   (dim. (r-1) x r) */
/* ----- */
/*          ADICAO DE 0.5 A TODAS AS CASELAS PARA CALCULO DAS ESTIMATIVAS
/*          DE MQG QUANDO ALGUM Nij = 0 .          */
/* ----- */
    SE (MIN(MN) == 0) {   Us = VDIAG(IDENT(s)) ;
                          MN = MN + ( 0.5 * (Us @ (Ur')) ) ;
    }
/* ----- */
/*          /*   MATRIZES UTILIZADAS NOS CALCULOS   */          */
/* ----- */
    T = LSOMA(MN);          /* vetor com totais marginais (srx1) */
    TI = INV(DIAG(T));       /* inversa da matriz diag. T (sxs) */
    BDTI = ( TI@(Ur * Ur') ); /* matriz bloco diagonal com blocos
                               (1/ni).1rr (sr x sr) */
/* ----- */
/*          CALCULO DAS PROPORCOES E LOGITOS OBSERVADOS E RESPECTIVAS
/*          MATRIZES DE COVARIANCIAS ESTIMADAS          */
/* ----- */
    MP = TI * MN ;          /* matriz P (srxr) */
    N = (MN[1,])';          /* vetor N1 (rx1) */
    P = (MP[1,])';          /* vetor P1 (rx1) */
/* Vetor de frequencias e de proporcoes observadas - N e P (srx1) */
    i=2;
    ENQUANTO ( i <= s ) {
        N = { N, (MN[i,])' };
        P = { P, (MP[i,])' };
        i = i+1;
    }

```



```

VP = (DIAG(P) - P*P')*!BDTI;          /* matriz de cov.assint.estim.
                                         (dim. sr x sr) */
/* VETOR DE LOGITOS F E RESPECTIVA MATRIZ DE COV.ASSINT. VF */
A = (IDENT(s) @ J);          /* matriz A (s(r-1) x sr) */
F = A*LOG(P);                 /* vetor de logitos F (s(r-1)x1) */
H = A/(DIAG(P));
VF = H*(VP)*H';              /* matriz de cov.assint.estim.para
                             os logitos F (s(r-1)x s(r-1)) */
dpF = (VDIAG(VF))#.5;        /* vetor com d.p. para F */
/* ----- */
/* CALCULO DA MATRIZ XL A PARTIR DA MATRIZ X */
SE (ENTX == 1) XL = (A*X);
/* ----- */
/* ESTIMACAO PELO METODO DE MINIMOS QUADRADOS GENERALIZADOS (MQG) */
/* ----- */
/* ESTIMATIVAS DOS PARAMETROS */
Vb0 = INV(XL'*(VF\ XL));      /* matriz de cov.
                             assint. de b0 (txt) */
dpb0 = (VDIAG(Vb0))#.5;      /* desvio padrao de b0 (tx1) */
b0 = (Vb0*(XL')*(VF\ F));    /* vetor b0 (tx1) */
/* ESTATISTICA DE AJUSTE */
GL = s*(r-1) - NLIN(b0);
SE (GL > 0)
    Q = (F - XL*b0)'*(VF\ (F - XL*b0));
CC
    Q = 0;                    /* estatistica Q-Wald */
/* VALORES PREDITOS */
FF0 = XL*b0;                 /* logitos ajustados */
VFF0 = XL*Vb0*(XL');         /* matriz de cov.assint. */
dpFF0 = VDIAG(VFF0)#.5;      /* d.p. para os logitos
                             ajustados */
NIO = (VDIAG(IDENT(s*(r-1))) + (BD1 * EXP(XL*b0)));
PIO = DIAG(NIO)\ EXP(XL*b0);
PIRO = (1 - CSOMA(PIO[1:(r-1),]));
PIIO = { PIO[1:(r-1),], PIRO }; /* vetor PI1 (rx1) */
i = 1;
ENQUANTO (i<s) {
    g1 = (i*(r-1) + 1);
    g2 = ((i+1)*(r-1));
    OIRO = 1 - CSOMA(PIO[g1:g2,]);
    OII0 = { PIO[g1:g2,], OIRO };
    PII0 = { PIIO, OII0 };
    i = i+1;
}
VPIIO = (DIAG(PIIO) - PII0*PIIO')*!BDTI; /* vetor PII0 (srx1) */
/* matriz de cov.
assint.(srxsr) */
/* FREQUENCIAS AJUSTADAS */
MO = ((T @ Ur)*!PIIO);

```

```

/* ----- */
/*      ESTIMACAO POR MAXIMA VEROSSIMILHANCA (MV) PELO METODO
      ITERATIVO DE NEWTON-RAPHSON      */
/* ----- */

/* inicializacao do processo iterativo */
SE (ENTB == 1) ina = b;
CC
  ina = b0 ;
  b = ina ;                      /* estimativa inicial de 'BETA' */
  delta = VDIAG(IDENT(NLIN(b))) ;
  P1 = (IDENT(s) @ {IDENT(r-1) (0*U) }) * P ;
k=0;
ENQUANTO ((MAX(ABS(delta)) > 0.0001) && (k<=25))
/* criterio de parada */
{
  k = k+1;
  NI = (VDIAG(IDENT(s*(r-1))) + (BD1 * EXP(XL*b))) ;
  PI = DIAG(NI) \ EXP(XL*b) ;
  G = XL'*(DIAG(T)@(IDENT(r-1)))*(P1 - PI) ; /* vetor GRADIENTE */
  IVF = (DIAG(T) @ IDENT(r-1))*((DIAG(PI) - PI*PI')*!BD1);
  HINV = INV((XL')*IVF*XL); /* matriz HESSIANA */
  delta = HINV * G ;
  b = b + delta;
}

Vb = HINV; /* matriz de cov.assint. de b (txt) */
dp = (VDIAG(Vb))#.5 ; /* desvio padrao de b (tx1) */

/* VALORES PREDITOS */
FF = XL*b; /* logitos ajustados */
VFF = XL*HINV*(XL'); /* matriz de cov.assint. */
dpFF = VDIAG(VFF)#.5; /* d.p. para os logitos ajust.*/
PIR = (1 - CSOMA(PI[1:(r-1),]));
PII = { PI[1:(r-1),], PIR }; /* vetor PI1 (rx1) */

i = 1;
ENQUANTO (i<s) {
  g1 = (i*(r-1) + 1);
  g2 = ((i+1)*(r-1));
  OIR = 1 - CSOMA(PI[g1:g2,]);
  OII = { PI[g1:g2,], OIR };
  PII = { PII, OII };
  i = i+1;
}
VPII = (DIAG(PII) - PII*PII')*!BDTI; /* vetor PII (srx1) */
/* matriz de cov.assint.
(srxsr) */

/* FREQUENCIAS AJUSTADAS */
M = ((T @ Ur)*!PII);

```

```

/* ----- */
/*          ESTATISTICAS DE AJUSTE          */
/* ----- */
SE (GL > 0) {
    QQP = CSOMA(((N - M)# 2)/!M) ;
    QQL = 2*N'*LOG(N/!M) ;
}

CC
{ QQP = 0 ;
  QQL = 0 ;
}

/* ----- */
/*          IMPRESSAO DOS RESULTADOS          */
/* ----- */
ANOTE " ";
ANOTE " ";
ANOTE "      MODELOS LOG-LINEARES - LOGITOS MULTIPLOS      ";
ANOTE " ";
ANOTE " ";
ANOTE " ";
ANOTE "  TABELA DE CONTINGENCIA ";
ANOTE " ";
IMPRIME MN 10:1 ;
ANOTE " ";
ANOTE "  PROPORCOES OBSERVADAS ";
ANOTE " ";
IMPRIME MP 10:5 ;
ANOTE " ";
SE (COVP == 1) {
ANOTE "  MATRIZ DE COVARIANCIA ESTIMADA ";
IMPRIME VP $ 10:5 ;
ANOTE " ";
ANOTE " ";
}

ANOTE " ";
ANOTE " ";
ANOTE "  MATRIZ DE ESPECIFICACAO DO MODELO ";
ANOTE " ";
IMPRIME XL $ 5:2 (" ") ;
ANOTE " ";
ANOTE " ";
ANOTE " ";
ANOTE "  ESTIMATIVAS DOS PARAMETROS ATRAVES DO METODO MQG " ;
ANOTE " ";
ANOTE " ";
ANOTE "  PARAMETROS - ESTIM.(b0),DESVIO PADRAO (dp0) E z0 = (b0/dpb0)";
ANOTE " ";
z0 = (b0 /! dpb0);
ESTMQG = b0 dpb0 z0 ;
IMPRIME ESTMQG 10:5 (" b0 , dp(b0), z0 ") ;
ANOTE " ";

```

```

SE ( COVBO == 1 ) {
  ANOTE "MATRIZ DE COV. (Vb0)";
  IMPRIME Vb0 10:5 ;
  ANOTE " ";
}

ANOTE " ";
ANOTE " ";
ANOTE "  ESTATISTICA DE AJUSTE ";
ANOTE " ";
ANOTE " ";
SE ( GL > 0 ) {
  QW = { GL Q XPROB(Q,GL) } ;
  IMPRIME QW $ 15:4 (" GL, WALD, P-DESCRITIVO") ;
}

CC {
  ANOTE "  **  MODELO SATURADO  -  AJUSTE PERFEITO  **  ";
}

ANOTE " ";
ANOTE " ";
ANOTE " ";
ANOTE " ";
ANOTE "  ESTIMATIVAS DOS PARAMETROS ATRAVES DO METODO DE MAXIMA ";
ANOTE " ";
ANOTE "  VEROSSIMILHANCA VIA ALGORITMO DE NEWTON-RAPHSON ";
ANOTE " ";
ANOTE " ";
ANOTE "  ESTIMATIVAS INICIAIS";
ESTO = (ina') ;
IMPRIME ESTO $ 10:5 ;
ANOTE " ";
ANOTE "  CONVERGENCIA";
CONV = { k MAX(ABS(delta)) } ;
IMPRIME CONV $ 15:5 (" No.Iteracao, Criterio") ;
ANOTE " ";
ANOTE " ";
ANOTE "  PARAMETROS - ESTIMATIVA (b), DESVIO PADRAO (dp) E z = (b/dp)";
ANOTE " ";
z = (b /! dp);
ESTMV = b dp z ;
IMPRIME ESTMV 10:5 (" b , dp(b) , z  ") ;
ANOTE " ";
SE (COVB == 1) {
  ANOTE "  MATRIZ DE COV. (Vb)";
  IMPRIME Vb 10:5 ;
  ANOTE " ";
}

ANOTE " ";
ANOTE " ";
ANOTE "  ESTATISTICAS DE AJUSTE ";
ANOTE " ";
ANOTE " ";

```



```

SE ( GL > 0 ) {
    QP = { GL QQP XPROB(QQP,GL) } ;
    IMPRIME QP $ 15:4 (" GL, PEARSON, P-DESCRITIVO") ;
    QL = { GL QQL XPROB(QQL,GL) } ;
    IMPRIME QL $ 15:4 (" GL, RAZAO VEROSS., P-DESCRITIVO");
}
CC {
    ANOTE "    **  MODELO SATURADO  -  AJUSTE PERFEITO  ** ";
}
ANOTE " ";
ANOTE " ";
ANOTE " ";
ANOTE " ";
ANOTE "    LOGITOS OBSERVADOS E AJUSTADOS POR  MQG  E  MV  ";
ANOTE " ";
LOGI = { F dpF FFO dpFFO FF dpFF } ;
ANOTE "    OBSERVADO          AJUSTE POR MQG          AJUSTE POR MV ";
IMPRIME LOGI 10:4 (" logito, dp , logito, dp , logito, dp ") ;
ANOTE " ";
ANOTE " ";
ANOTE "    PROPORCOES OBSERVADAS E AJUSTADAS POR  MQG  E  MV ";
ANOTE " ";
PROP = { P PIIO PII } ;
ANOTE "    OBSERVADO          MQG          MV ";
IMPRIME PROP 12:5 (" P , PIIO , PII ") ;
ANOTE " ";
ANOTE " ";
ANOTE "    FREQUENCIAS OBSERVADAS E AJUSTADAS POR  MQG  E  MV ";
ANOTE " ";
CONT = { N MO M } ;
ANOTE "    OBSERVADO          MQG          MV ";
IMPRIME CONT 12:5 (" N , MO, M ") ;
ANOTE " ";
ANOTE " ";
ANOTE " ";
ANOTE " ";
SE (COV == 1) {
    ANOTE "    MATRIZ DE COV.ESTIM. PARA PROPORCOES AJUSTADAS VIA MQG";
    IMPRIME VPIIO $ 10:5 ;
    ANOTE " ";
    ANOTE " ";
    ANOTE "    MATRIZ DE COV.ESTIM. PARA PROPORCOES AJUSTADAS VIA MV";
    IMPRIME VPPII $ 10:5 ;
}
ANOTE " ";
ANOTE " ";
ANOTE "    ***  PROGRAMA    LOGLIN.AJU    TERMINADO    *** ";
/* ----- */
/*          FIM DO PROGRAMA LOGLIN.AJU          */
/* ----- */

```

```

/*          *****   PROGRAMA   LOGLIN.HIP   *****          */
/*
/*          Versao 1.0 - setembro/88          */
/*
/*  Testa hipoteses da forma  $H_0: C*BETA = 0$   atraves da estatistica
/*      de Wald considerando as estimativas de MQG e de MV obtidas
/*      com a execucao do programa LOGLIN.AJU  */
/* ----- */
ANOTE " ";
ANOTE " ";
SE (GL>0)
{
ANOTE "  ESTATISTICA GENERALIZADA DE WALD PARA A HIPOTESE
               $H_0: C*BETA = 0$  ";
ANOTE " ";
ANOTE "  MATRIZ DE CONTRASTES ";
ANOTE " ";
              IMPRIME C $ 5:2 (" ") ;
              ANOTE " ";ANOTE " ";
ANOTE "  RESULTADO DO TESTE UTILIZANDO AS ESTIMATIVAS DE MQG ";
ANOTE " ";
Cb0 = C*b0;
Qb0 = (Cb0)'*((C*Vb0*C')\ (Cb0));
              IMPRIME Cb0  10:5 ;
              ANOTE " ";
QWCO = { NLIN(C) Qb0 XPROB(Qb0,NLIN(C)) };
              IMPRIME QWCO $ 15:4 (" GL , Q-WALD , P-DESCRITIVO ");
              ANOTE " ";
              ANOTE " ";
ANOTE "  RESULTADO DO TESTE UTILIZANDO AS ESTIMATIVAS DE MV ";
ANOTE " ";
Cb = C*b;
Qb = (Cb)'*((C*HINV*C')\ (Cb));
              IMPRIME Cb  10:5 ;
              ANOTE " ";
QWC = { NLIN(C) Qb XPROB(Qb,NLIN(C)) };
              IMPRIME QWC $ 15:4 (" GL , Q-WALD , P-DESCRITIVO ");
              ANOTE " ";
              ANOTE " ";
ANOTE " ----- ";
ANOTE " ";
}
CC
{
ANOTE "  ESTATISTICA GENERALIZADA DE WALD PARA A HIPOTESE
               $H_0: C*BETA = 0$  ";
ANOTE " ";
ANOTE "  MATRIZ DE CONTRASTES ";
ANOTE " ";
              IMPRIME C $ 5:2 (" ") ;

```

```

      ANOTE " "; ANOTE " ";
ANOTE "  MODELO SATURADO -- ESTIM. DE MV = ESTIM. DE MQG ";
ANOTE " ";
Cb0 = C*b0;
Qb0 = (Cb0)'*((C*Vb0*C')\ (Cb0));
      IMPRIME Cb0  10:5 ;
      ANOTE " ";
QWCO = { NLIN(C) Qb0 XPROB(Qb0,NLIN(C)) };
      IMPRIME QWCO $ 15:4 (" GL , Q-WALD , NIV.DESCRIT. ");
      ANOTE " ";
      ANOTE " ";
ANOTE " ----- ";
ANOTE "  ";
    }
/* ----- */
      /*      FIM DO PROGRAMA  LOGLIN.HIP      */
/* ----- */

```

Anexo B

Listagens de comandos de programas utilizados nos exemplos

B.1 - Comandos do programa BMDP/4F para o ajuste do modelo (2.1.9)

```

-----
/PROBLEM          TITLE IS 'CRESCIMENTO DE PLANTAS SUBMETIDAS A
                   DIVERSOS TRATAMENTOS'.
/INPUT            VARIABLES ARE 3.
                   FORMAT IS '(2F4.0)'.
                   TABLE IS 2,2,2.
/TABLE            INDICES ARE A,B,C.
/VARIABLES        NAMES ARE A,B,C.
/FIT              MODEL IS AB,AC,BC.
/PRINT            EXPECTED.CHISQ.LRCHI.LAMB.BETA.VAR.
/END
156  84
 84 156
107 133
 31 209
-----

```

B.2 - Comandos do programa GENCAT para o ajuste do modelo (2.2.6)

```

-----
               1         2         3                               8
12345678901234567890123456789012345678901234567890 .... 0

    5    1    1                                CRESCIMENTO DE PLANTAS SUBMETIDAS A
    4    2                                DIVERSOS TRATAMENTOS
156  84                                (2F4.0)
 84 156
107 133
 31 209

    2
    1    2    4    1    1                (2F3.0)
1 -1
    7    1    3                (4F3.0)      MODELO (2.2.6)
1  1  1  1
1 -1  1 -1
1  1 -1 -1
-----

```

Obs.: As duas primeiras linhas servem de orientação pois os comandos do GENCAT ocupam posições definidas.

B.3 - Comandos do programa GENCAT para obtenção das frequências ajustadas pelo modelo (2.1.9)

Processamento 1

```

      1          2          3          8
12345678901234567890123456789012345678901234567890 .... 0

      5      1      1          CRESCIMENTO DE PLANTAS SUBMETIDAS A
                                DIVERSOS TRATAMENTOS
                                (8F4.0)
      1      8
156  84  84 156 107 133  31 209
      2          1
      1      1      7          1      (8F3.0)
      1  0  0  0  0  0  0 -1
      0  1  0  0  0  0  0 -1
      0  0  1  0  0  0  0 -1
      0  0  0  1  0  0  0 -1
      0  0  0  0  1  0  0 -1
      0  0  0  0  0  1  0 -1
      0  0  0  0  0  0  1 -1
      7      1      6          1  (7F3.0)  MODELO (2.1.9) -
                                FORM. LOGITOS

      2  0  2  0  2  0  2
      2  2  0  0  2  2  0
      2  2  2  2  0  0  0
      0 -2 -2  0  0 -2 -2
      0 -2  0 -2 -2  0 -2
      0  0 -2 -2 -2 -2  0
      6
      1      1      8
      1  1  1  1  1  1
     -1  1  1 -1 -1  1
      1 -1  1 -1  1 -1
     -1 -1  1  1 -1 -1
      1  1 -1  1 -1 -1
     -1  1 -1 -1  1 -1
      1 -1 -1 -1 -1  1
     -1 -1 -1  1  1  1
      3          1
      1      1      1          1
     1  1  1  1  1  1  1  1      (8F2.0)  (resulta exp(X.b))
                                           (resulta NI)

```

(logitos múltiplos)

(matriz XL)

(matriz X)

Processamento 2

.
.
.
(comandos do Processamento 1 até matriz XL)

```

      1      2      3      8
12345678901234567890123456789012345678901234567890 .... 0

      6
      1      1      8
      1  1  1  1  1  1
     -1  1  1 -1 -1  1
      1 -1  1 -1  1 -1
     -1 -1  1  1 -1 -1
      1  1 -1  1 -1 -1
     -1  1 -1 -1  1 -1
      1 -1 -1 -1 -1  1
     -1 -1 -1  1  1  1

      3      1
      1      2      8      1      1
0.11120749 (F11.8)
      1      2      8      1      (F4.0)
960
                                     (matriz X)
                                     (resulta exp(X.b))
                                     (resulta )
                                     (resulta m )

```

B.4 - Comandos do programa LOGLIN.AJU para o ajuste dos modelos (2.2.8) e (2.2.10)

```

-----
ANOTE "      ****   OUTUBRO/88   **** ";
ANOTE " ";
ANOTE "      EXEMPLO: CIRURGIAS X SEVERIDADE DA SEQUELA ";
ANOTE " ";
ANOTE "      MODELO 2.2.8 ";
ANOTE " ";
ANOTE " ";
s = 4;
r = 3;
/* no. de linhas (subpopulacoes) */
/* no. de colunas (categorias de respostas) */
/* Tabela de contingencia - dados por subpopulacao (sxr) */
MN = { 61 28 7, 68 23 13, 58 40 12, 53 38 16 };
XL = { 1 0 1 0 1 0 1 0,
      0 1 0 1 0 1 0 1,
      0 0 1 0 0 0 0 0,
      0 0 0 1 0 0 0 0,
      0 0 0 0 1 0 0 0,
      0 0 0 0 0 1 0 0,
      0 0 0 0 0 0 1 0,
      0 0 0 0 0 0 0 1 }';
/*      matriz XL      */

```

```

ENTX = 0;
ENTB = 0;
COVP = 1;
COVBO = 1;
COVB = 1;
COV = 0;
EXEC "A:LOGLIN.AJU";

LIMPE;

ANOTE " ";
ANOTE "      MODELO 2.2.10 ";
ANOTE " ";
ANOTE " ";

s = 4;
r = 3;
MN = { 61 28 7, 68 23 13, 58 40 12, 53 38 16 };
XL = { 1 0 1 0 1 0 1 0,
        0 1 0 1 0 1 0 1,
        0 0 2 1 0 0 0 0,
        0 0 0 0 2 1 0 0,
        0 0 0 0 0 0 2 1 }';

ENTX = 0;
ENTB = 0;
COVP = 1;
COVBO = 1;
COVB = 1;
COV = 0;
EXEC "A:LOGLIN.AJU";

FIM;                      /*      FIM      */

```

B.5 - Comandos do programa GENCAT para o ajuste dos modelos (2.2.8) e (2.2.10)

```

          1          2          3          8
12345678901234567890123456789012345678901234567890 .... 0

      5      1      1          TABELA 1.5-CIRURGIAS VS SEV.SEQUELA
      4      3          (3F4.0)
61  28  7
68  23  13
58  40  12
53  38  16

      2          1
      1      2      8      2      1      (3F3.0)
1  0 -1
0  1 -1

```


	7		1		8				(8F3.0)	MODELO 2.2.8
1	0	1	0	1	0	1	0			
0	1	0	1	0	1	0	1			
0	0	1	0	0	0	0	0			
0	0	0	1	0	0	0	0			
0	0	0	0	1	0	0	0			
0	0	0	0	0	1	0	0			
0	0	0	0	0	0	1	0			
0	0	0	0	0	0	0	1			

	7		1		5				(8F3.0)	MODELO 2.2.10
1	0	1	0	1	0	1	0			
0	1	0	1	0	1	0	1			
0	0	2	1	0	0	0	0			
0	0	0	0	2	1	0	0			
0	0	0	0	0	0	2	1			

B.6 - Comandos do programa BMDP/4F para o ajuste do modelo (2.4.5)

```

/PROBLEM          TITLE IS 'TABELA 1.8 - DADOS SOBRE ARTERIOSCLEROSE'.
/INPUT            VARIABLES ARE 4.
                  FORMAT IS FREE.
                  TABLE IS 2,2,2,3.
/VARIABLES        NAMES ARE INFARTO, SEQMIO, AGE, LOCRACA.
/CATEGORY         CODES(1) ARE 1,2.
                  NAMES(1) ARE NAO, SIM.
                  CODES(2) ARE 1,2.
                  NAMES(2) ARE NAO, SIM.
                  CODES(3) ARE 1,2.
                  NAMES(3) ARE '35-54', '55-69'.
                  CODES(4) ARE 1,2,3.
                  NAMES(4) ARE 'NEW B', OSLO, 'NEW N'.
/TABLE            INDICES ARE INFARTO, SEQMIO, AGE, LOCRACA.
/FIT              MODEL IS   ISA, ISL, IAL, SAL .
/PRINT            EXPECTED.STANDARDIZED.CHISQ.LRCHI.LAMB.BETA.VAR.
/END

```

19	34	22	20
21	60	39	26
13	11	9	16
15	38	66	55
14	15	16	7
4	17	17	4

B.7 - Comandos do programa GENCAT para o ajuste do modelo (2.4.5)

```

-----
      1      2      3      8
12345678901234567890123456789012345678901234567890 .... 0

      5      1      1      1      TAB. 1.8-DADOS SOBRE ARTERIOSCLEROSE
      6      4      (6F3.0)
19 34 22 20
21 60 39 26
13 11 9 16
15 38 66 55
14 15 16 7
4 17 17 4
      2      1
      1      2      6      1      (4F3.0)
1 -1 -1 1
      3
      2      1
      7      1      4      1 (6F3.0)      MODELO (2.4.5)
1 1 1 1 1 1
1 -1 1 -1 1 -1
1 1 0 0 -1 -1
0 0 1 1 -1 -1

      6      CALC. PRODUTOS CRUZADOS AJUSTADOS
      1      1      6      (4F3.0)
1 1 1 0
1 -1 1 0
1 1 0 1
1 -1 0 1
1 1 -1 -1
1 -1 -1 -1
      3
-----

```

B.8 - Comandos do programa LOGLIN.AJU para o ajuste do modelo (2.4.5)

```

-----
ANOTE " ";
ANOTE " *** ESTUDO SOBRE ARTERIOSCLEROSE ***";
ANOTE " ";
s = 6 ; /* no. subpopulacoes */
r = 4 ; /* perfis de resposta */
MN = { 19 34 22 20, 21 60 39 26,
        13 11 9 16, 15 38 66 55,
        14 15 16 7, 4 17 17 4 }; /* tabela de contingencia */
ANOTE " ";
B = IDENT(6) @ { 1 (-1) (-1) 1 } ;

```

```

Y = { 1 1 1 0,
      1 (-1) 1 0,
      1 1 0 1,
      1 (-1) 0 1,
      1 1 (-1) (-1),
      1 (-1) (-1) (-1) };

ANOTE " " ;
IMPRIME B $ 6:1 ;
ANOTE " " ;
IMPRIME Y $ 6:1 ;

A = IDENT(6) @ { 1 0 0 (-1),
                  0 1 0 (-1),
                  0 0 1 (-1) };

T = A*B'*(inv(B*B'))*Y;
Z = IDENT(6) @ { 0 (-1), (-1) 0, 1 0, 0 1 };
m = A*Z;

ANOTE " " ;
IMPRIME Z $ 6:1 ;
ANOTE " " ;
ANOTE " XL = { m T } " ;
ANOTE " " ;
ANOTE " onde m = A*Z e T = A*B'*(inv(B*B'))*Y " ;
ANOTE " " ;
IMPRIME m $ 6:2 ;
IMPRIME T $ 6:2 ;

XL = { m T } ;
ENTX = 0;
ENTB = 0;
COV = 0;
COVP = 1;
COVBO = 1;
COVB = 1;
COV = 1;
EXEC "LOGLIN.AJU";

ANOTE " " ;
ANOTE " VALORES OBSERVADOS PARA OS PRODUTOS CRUZADOS(OR) " ;
ANOTE " " ;

A = (IDENT(s) @ { 1 (-1) (-1) 1 });
ORP = EXP(A*LOG(P)) ;
HP = ((DIAG(ORP)) * A * (INV(DIAG(P)))) ;
VORP = HP * VP * (HP)';
SORP = VDIAG(VORP) # 0.5 ;
OBS = { ORP SORP } ;
IMPRIME OBS 10:5 (" OR , DP(OR)");
ANOTE " " ;

```

```

/* CALCULO DOS ODDS AJUST.A PARTIR DE BETA E V(BETA) */
ANOTE " VALORES AJUSTADOS PELO MODELO PARA OS PRODUTOS CRUZADOS(OR) ";
ANOTE " ";
bb0 = b0[13:16,];
Vbb0 = Vb0[13:16, 13:16];
bb = b[13:16,];
Vbb = Vb[13:16, 13:16];
A = { 1 1 1 0,
      1 (-1) 1 0,
      1 1 0 1,
      1 (-1) 0 1,
      1 1 (-1) (-1),
      1 (-1) (-1) (-1) } ;
ANOTE " ";
ORb0 = EXP(A*bb0) ;
Hb0 = (DIAG(ORb0) * A) ;
VORb0 = Hb0 * Vbb0 * (Hb0)';
SORb0 = VDIAG(VORb0) # 0.5 ;
MQGb0 = { ORb0 SORb0 } ;
IMPRIME MQGb0 10:5 (" ORb0 , DP(ORb0)");
ANOTE " ";
ORb = EXP(A*bb) ;
Hb = (DIAG(ORb) * A) ;
VORb = Hb * Vbb * (Hb)';
SORb = VDIAG(VORb) # 0.5 ;
MVb = { ORb SORb } ;
IMPRIME MVb 10:5 (" ORb , DP(ORb)");
FIM;

```