

ANÁLISE DE TABELAS DE CONTINGÊNCIA TRIDIMENSIONAIS: USO DA INFORMAÇÃO MÍNIMA DE DISCRIMINAÇÃO

LUSBI ANTONIO HERRERA SANDOVAL

Orientador: Prof. Dr. ANTONIO F. IEMMA

Tese apresentada à Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo, para obtenção do título de Doutor em Agronomia. Área de Concentração: Estatística e Experimentação Agronômica.

PIRACICABA
Estado de São Paulo - Brasil
Junho - 1987

A Deus,

A meu pai (*in memoriam*),

A minha mãe Carmen Sandoval de Herrera e a meus irmãos,

A minha sogra,

A minha querida esposa Mariela,

A meus filhos Daniel e Andr es,

DEDICO.

AGRADECIMENTOS

- Ao Prof. Dr. Antonio F. Iemma, pela valiosa orientação durante o curso e na realização deste trabalho, assim como também pela sua amizade.
- Aos Professores do curso de Pós-Graduação em Estatística e Experimentação Agronômica da ESALQ/USP pelos ensinamentos recebidos, especialmente aos professores Dr. Humberto de Campos, Dr. Décio Barbin e Dr. Cássio R. de Melo Godoi.
- Aos colegas das disciplinas de Matemática da Faculdade de Agronomia da Universidade Central da Venezuela pelo apoio e oportunidade de realização do curso.
- Aos colegas do curso de Pós-Graduação pelo convívio amigável, especialmente Rubens Braga Jr., João Gil de Luna, Antonio C. de Oliveira, Geraldo de Souza Barbosa, Edmilson de Araújo Silva e Ana B. Costa Czermainski.
- Ao Centro de Informática na Agricultura (CIAGRI) pela cooperação, especialmente ao Analista de Sistemas Jr. Gabriel Adrián Sarriés e ao Operador de Sistemas Jr. Marcelo Z. Silva pela indispensável contribuição na elaboração do programa computacional ANDACAT.
- Aos funcionários do Departamento de Matemática e Estatística da ESALQ/USP, pela amizade e cooperação.
- À Sra. Marcia B. Correia e ao Sr. Jorge Luiz Diorio, pelo excelente trabalho datilográfico.

SUMÁRIO

	Página
RESUMO	viii
SUMMARY	x
1. INTRODUÇÃO	01
2. REVISÃO BIBLIOGRÁFICA	04
3. DESENVOLVIMENTO TEÓRICO	23
3.1. Introdução	23
3.2. Distribuição multinomial	25
3.2.1. Estimacão dos parâmetros de uma dis- tribuição multinomial	28
3.2.2. Distribuição dos \hat{p}_i	30
3.2.3. Resumo dos resultados obtidos no item 3.2.	32
3.3. Tabelas de contingência $R \times S \times T$	33
3.3.1. Introdução	33
3.3.2. Modelo log-linear	35
3.3.3. Modelos hierárquicos	39
3.3.4. Hipóteses de interesse	40
3.3.5. Estimativas de máxima verossimilhan- ça	41
3.3.6. Resumo dos resultados obtidos no item 3.3.	44
3.4. Teoria da informacão	46
3.4.1. Introdução	46
3.4.1.1. Informacão esperada	48
3.4.1.2. Alguns comentários	51
3.4.2. Informacão de discriminacão	54
3.4.2.1. Definicao de informacão ...	55
3.4.2.2. Propriedades da informacão.	59
3.4.2.3. Informacão e suficiencia ..	60
3.4.3. Informacão mínima de discriminacão $I(*;2)$	61

	Página
3.4.4. Estatística da informação	68
3.4.4.1. Estimativa de $I(*:2)$	68
3.4.4.2. Distribuição de $\hat{I}(*:2)$	70
3.4.4.3. Distribuição conjugada da distribuição multinomial ...	72
3.5. Formulação e teste de hipóteses de interesse em tabelas de contingência $R \times S \times T$, consideran <u>do</u> do fixas as frequências marginais observadas	76
3.5.1. Introdução	76
3.5.2. Representação log-linear	77
3.5.3. Estimativa da frequência esperada na (i,j,k) -ésima casela, através da in- formação mínima de discriminação	80
3.5.4. Independência mútua completa	82
3.5.4.1. Representação log-linear ...	82
3.5.4.2. Estimativas das frequências esperadas	89
3.5.4.2.1. Processo iterati <u>v</u> vo	89
3.5.4.2.2. Computação das es <u>t</u> timativas das fre <u>q</u> uências esperadas.	93
3.5.4.3. Estatística da informação mí <u>n</u> ima de discriminação	96
3.5.4.4. Estimativa dos parâmetros do modelo log-linear	103
3.5.4.5. Estimativa da matriz de co- variâncias das estimativas dos parâmetros	105
3.5.5. Independência parcial	106
3.5.5.1. Representação log-linear ...	106
3.5.5.2. Estimativas das frequências esperadas	108
3.5.5.2.1. Processo iterati <u>v</u> vo	108

	Página
3.5.5.2.2. Computação das estimativas das frequências esperadas.	111
3.5.5.3. Estatística da informação mínima de discriminação	113
3.5.5.4. Estimativa dos parâmetros do modelo log-linear	116
3.5.5.5. Estimativa da matriz de covariâncias das estimativas dos parâmetros	116
3.5.6. Independência condicional	116
3.5.6.1. Representação log-linear ...	117
3.5.6.2. Estimativas das frequências esperadas	118
3.5.6.2.1. Processo iterativo	118
3.5.6.2.2. Computação das estimativas das frequências esperadas.	118
3.5.6.3. Estatística da informação mínima de discriminação	119
3.5.6.4. Estimativa dos parâmetros do modelo log-linear	120
3.5.6.5. Estimativa da matriz de covariâncias das estimativas dos parâmetros	121
3.5.7. Inexistência de interação de segunda ordem	121
3.5.7.1. Representação log-linear ...	122
3.5.7.2. Estimativas das frequências esperadas	123
3.5.7.2.1. Processo iterativo	123
3.5.7.2.2. Computação das estimativas das frequências esperadas.	123

3.5.7.3. Estatística da informação mínima de discriminação	124
3.5.7.4. Estimativa dos parâmetros do modelo log-linear	125
3.5.7.5. Estimativa da matriz de covariâncias das estimativas dos parâmetros	126
3.6. Análise da informação	126
3.6.1. $I(p:\pi) = I(p:p^*) + I(p^*:\pi)$	126
3.6.2. Quadros de análise de informação	129
3.6.3. Modelos hierárquicos em tabelas de contingência $R \times S \times T$	134
4. EXEMPLO ILUSTRATIVO	136
5. CONCLUSÕES	147
6. REFERÊNCIAS BIBLIOGRÁFICAS	150
APÊNDICE 1 - DEMONSTRAÇÃO DA REGRA DE BAYES	156
APÊNDICE 2 - PROGRAMA COMPUTACIONAL ANDACAT	158

ANÁLISE DE TABELAS DE CONTINGÊNCIA TRIDIMENSIONAIS:
USO DA INFORMAÇÃO MÍNIMA DE DISCRIMINAÇÃO

Autor: LUSBI ANTONIO HERRERA SANDOVAL

Orientador: Prof. Dr.: ANTONIO FRANCISCO IEMMA

RESUMO

O problema considerado foi o de estimar as frequências esperadas m_{ijk} de uma tabela de contingência $R \times S \times T$ quando algumas frequências marginais são consideradas fixas. Admite-se que os dados são provenientes de uma única população multinomial.

O princípio da informação mínima de discriminação é usado para gerar hipóteses de interesse concernentes a interações e efeitos na análise de uma tabela de contingência tridimensional. O método baseado na informação mínima de discriminação conduz a modelos log-lineares e permite calcular as estimativas das frequências esperadas em cada casela sob várias hipóteses de interesse, assim como também testar essas hipóteses.

Para as condições estabelecidas, isto é, quando impõem-se restrições nas frequências marginais observadas, a x^2 -tabela fornece estimativas de máxima verossimilhança e a estatística da informação mínima de discriminação é igual a duas vezes o logaritmo natural da razão de verossimilhança com sinal menos.

A estatística da informação mínima de discriminação, sob uma determinada hipótese, pode ser analisada em

componentes que por sua vez são estatísticas de informação mínima de discriminação sob sub-hipóteses, dado que essas estatísticas são aditivas, assim como também os graus de liberdade associados. Uma tabela de análise de informação pode ser construída onde a tabela apresenta uma análise aditiva completa da tabela de contingência e não apenas uma parte da análise.

Apresenta-se um exemplo ilustrativo e, para uma tabela de contingência de três entradas, são dadas tabelas de análise de informação. Foi elaborado um programa computacional em Pascal.

ON THE ANALYSIS OF THREE-DIMENSIONAL CONTINGENCY TABLES:
 APPLYING THE PRINCIPLE OF MINIMUM DISCRIMINATION INFORMATION.

Author: LUSBI ANTONIO HERRERA SANDOVAL
 Adviser: Prof. Dr. ANTONIO FRANCISCO IEMMA

SUMMARY

In its simplest formulation the problem considered is to estimate the cell frequencies m_{ijk} of an $R \times S \times T$ contingency table for which certain marginals are considered as fixed. We shall look upon a contingency table as a single sample from a multinomial population.

The principle of minimum discrimination information is used to generate hypotheses of interest concerning various interactions and effects in the analysis of three-dimensional contingency tables. The minimum discrimination information approach leads to loglinear models and enables one to find estimates of cell frequencies under various hypotheses and to test these hypotheses.

For the procedures discussed herein with observed marginals as the given constraints, the x^* -table provides maximum likelihood estimates and the minimum discrimination information statistic turn out to be equal to minus twice the natural logarithm of a likelihood ratio.

The minimum discrimination information statistic under an hypothesis can be analysed into components under sub hypotheses, since the minimum discrimination information statistics are additive, as are the associated degrees of

freedom. An analysis of information table may be constructed wherein the table presents an additive analysis of the complete contingency table, rather than just a especial segment of the analysis.

An illustrative example and analysis of information tables are given for a three-way contingency table. A Pascal program to aid in the computation has been prepared.

1. INTRODUÇÃO

Consideremos, dentre outras, duas maneiras usuais de coletar dados estatísticos. Uma maneira é registrar a medida de interesse para cada indivíduo. Uma outra maneira é classificar os indivíduos como pertencentes a diferentes categorias de acordo a possuírem ou não, um ou vários atributos.

Se os indivíduos de uma população podem ser descritos como pertencentes a uma das i categorias do atributo R , a uma das j categorias do atributo S e a uma das k categorias ou níveis do atributo T , então tem-se uma distribuição de freqüências dos indivíduos em $R \times S \times T$ classes. Ao tomar-se uma amostra de tais indivíduos o resultado será uma tabela de classificação cruzada de três entradas.

A informação que se deseja obter de uma tabela de contingência é geralmente de natureza preliminar. Pode ser de interesse conhecer a inter-relação ou associação entre uma classificação e uma ou mais das outras classificações. Se em um experimento concluiu-se que um determinado atributo desejável está associado à presença de outra característica, pode-se orientar uma futu-

ra pesquisa no sentido de se obterem indivíduos com essa característica.

A estrutura de uma tabela de contingência tridimensional, geralmente é avaliada mediante a formulação das seguintes hipóteses: (a) independência mutua completa das três variáveis; (b) independência parcial de uma variável das outras duas conjuntamente; (c) independência condicional de duas variáveis em cada nível da terceira e d) ausência de interação de segunda ordem ou ausência de interação entre as três variáveis. Para testar essas hipóteses, as estatísticas usuais tem sido a χ^2_p de Pearson, a χ^2_n modificada de Neyman, o teste da razão de verossimilhança e o método de análise proposto por GRIZZLE, STARMER & KOCH (1969), conhecido como método GSK, que considera o uso de modelos lineares na análise de tabelas de contingência. Também, quando as observações têm distribuição de Poisson, pode-se utilizar a metodologia dos modelos lineares generalizados propostos por NELDER (1974).

Um procedimento alternativo pode ser obtido através da aplicação do princípio da informação mínima de discriminação, formulado por KULLBACK (1959). Ocorre que, apesar de ser altamente eficiente, esse procedimento tem sido pouco divulgado nos livros-textos elementares, dado o excesso de formalismo e rigor com que é apresentado, em geral, pelos autores, exigindo do leitor um conhecimento elevado de estatística matemática. Nesse contexto, o tema praticamente inexiste em periódicos inerentes às áreas aplicadas. Isto sem dúvida restringe drástica-

mente sua divulgação e por conseguinte suas aplicações. Assim, por exemplo, encontra-se em IRELAND & KULLBACK (1968), um estudo formal sobre o uso da informação mínima de discriminação para tabelas de dupla entrada quando as probabilidades marginais são conhecidas.

A revisão bibliográfica não registrou estudos formais para tabelas com mais de duas entradas. Além da verdade, KU *et alii* (1971), KULLBACK & FISHER (1973) e KU & KULLBACK (1974), apresentam apenas resultados para tabelas tridimensionais com freqüências marginais fixas.

Pretende-se aqui, como uma aplicação da informação mínima de discriminação, apresentar um estudo formal sobre a análise de tabelas de contingência completas de três entradas, com dados provenientes de uma única população multinomial e com algumas freqüências marginais consideradas como fixas, visando atingir profissionais de áreas aplicadas em geral.

Para tanto, o estudo da teoria da informação é abordada de modo tão didático quanto possível. Os principais teoremas são demonstrados em linguagem acessível a profissionais da pesquisa agropecuária, visando otimizar a assimilação. Apresentam-se exemplos ilustrativos, além de um programa computacional auto explicativo.

2. REVISÃO DE LITERATURA

BISHOP *et alii* (1977), KULLBACK (1959), FIENBERG (1978), dentre outros, concordam ao afirmar que a análise de dados discretos multivariados, especialmente na forma de tabelas de contingência tem ocupado um lugar de destaque na literatura estatística, a partir dos trabalhos de Karl Pearson em 1904 e de R.A.Fisher em 1925.

KULLBACK (1959), observa que Fisher em 1925, em seu trabalho sobre Teoria da Estimacão Estatística, definiu uma medida da quantidade de informacão que é proporcionada pelos dados acerca de um parâmetro desconhecido. O autor considera que essa medida é o primeiro uso da informacão em estatística matemática.

WALD (1945), define um teste seqüencial de hipóteses estatísticas como sendo aquele em que em qualquer etapa do experimento, seguindo uma regra estabelecida, pode-se tomar uma das seguintes decisões: (i) aceitar a hipótese nula que está sendo testada; (ii) rejeitar a hipótese nula (aceitar a hipótese alternativa) ou (iii) continuar o experimento fazendo observações adicionais.

Considerando a primeira observacão ou ensaio, uma das três decisões mencionadas acima é tomada.

Ao tomar-se a primeira ou a segunda decisão, o processo é terminado. Se pelo contrário, toma-se a terceira decisão, então realiza-se um segundo ensaio. Sobre a base das duas primeiras observações uma das três decisões é tomada e se a terceira é a escolhida, então realiza-se um terceiro ensaio e assim sucessivamente. O processo continua até que se tome uma das duas primeiras decisões. Seja \underline{n} o número de observações necessárias para que o processo termine. Então \underline{n} é uma variável aleatória pois o valor de \underline{n} depende do resultado das observações. Suponha-se que antes de tirar a amostra exista uma probabilidade conhecida que H_0 seja verdadeira. Denotemos esta probabilidade por g^0 . Então a probabilidade *a priori* que H_1 seja verdadeira é $g^1 = 1 - g^0$. Depois que tivermos efetuado várias observações, teremos obtido alguma informação adicional, a qual afetará a probabilidade que H_i ($i = 0; 1$) seja verdadeira. Sejam g_m^0 e g_m^1 as probabilidades *a posteriori* que H_0 e H_1 sejam verdadeiras, respectivamente, depois que \underline{m} observações foram feitas, então é conhecido pelo teorema de Bayes que: (ver Apêndice 1)

$$g_m^0 = \frac{g^0 p_m^0}{g^0 p_m^0 + g^1 p_m^1} \quad \text{e} \quad g_m^1 = \frac{g^1 p_m^1}{g^0 p_m^0 + g^1 p_m^1}, \quad \text{onde}$$

$p_m^i = p_m^i(x_1, x_2, \dots, x_m)$, denota a f.d.p. no espaço amostral m -dimensional sob as hipóteses H_i ($i = 0; 1$). Este trabalho, segundo KULLBACK (1959), pode ser considerado como a maior contribuição às aplicações estatísticas da teoria da informação.

BARTLETT (1935), estudou a hipótese de não-interação entre os três fatores numa tabela de contingência $2 \times 2 \times 2$ e estabeleceu que a hipótese de não-interação indica que a associação entre duas variáveis é a mesma para os dois níveis da terceira variável. Essa associação é medida pelo logaritmo da razão dos produtos cruzados, isto é, não há interação de segunda ordem se:

$$\ell_n \frac{P_{111} P_{221}}{P_{121} P_{211}} = \ell_n \frac{P_{112} P_{222}}{P_{122} P_{212}}, \quad \underline{i} \text{ e } \underline{j} \text{ para } \underline{k}$$

$$\ell_n \frac{P_{111} P_{212}}{P_{112} P_{211}} = \ell_n \frac{P_{121} P_{222}}{P_{122} P_{221}}, \quad \underline{i} \text{ e } \underline{k} \text{ para } \underline{j}$$

$$\ell_n \frac{P_{111} P_{122}}{P_{112} P_{121}} = \ell_n \frac{P_{211} P_{222}}{P_{212} P_{221}}, \quad \underline{j} \text{ e } \underline{k} \text{ para } \underline{i}$$

ROY & KASTENBAUM (1956), numa situação em que as observações independentes são freqüências de uma tabela de contingência multidimensional e onde apenas o número total de observações é considerado fixo, justificaram e generalizaram o teste de hipóteses de não interação. Mostraram que os resultados de BARTLETT (1935), são casos particulares dos resultados obtidos por eles e destacaram as seguintes diferenças: (i) os resultados de BARTLETT (1935) estão referidos a situações em que os totais marginais de dupla entrada são considerados fixos e, (ii) no trabalho de BARTLETT (1935) não aparece qualquer indicação do mecanismo matemático que está por trás de sua

formulação de hipóteses de não interação. Para uma tabela geral $R \times S \times T$, estabeleceram a seguinte formulação para a hipótese de "não-interação":

$$\frac{p_{rst} p_{ijjt}}{p_{ist} p_{rjt}} = \frac{p_{rsk} p_{ijk}}{p_{isk} p_{rjk}}, \quad \text{para} \quad \begin{cases} k = 1, 2, \dots, (t-1) \\ j = 1, 2, \dots, (s-1) \\ i = 1, 2, \dots, (r-1) \end{cases}$$

observa-se então, que têm-se $(t-1) (s-1) (r-1)$ restrições sobre os p_{ijk} ; $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ e $k = 1, 2, \dots, t$.

KULLBACK (1959), apresenta um estudo geral sobre teoria da informação e sua aplicação nos testes de hipóteses estatísticas. Afirma que a teoria da informação fornece uma unificação de resultados estatísticos conhecidos e conduz à obtenção de novos resultados. Este livro é de fundamental importância para o nosso trabalho.

O teste da razão de verossimilhança é usado por DARROCH (1962) para testar a hipótese de não interação de segunda ordem numa tabela tridimensional e afirma que se numa tabela desse tipo temos um conjunto completo e consistente de restrições de segunda ordem, isto é, todos os totais das marginais, então a hipótese de não interação de segunda ordem conduz a um único conjunto de probabilidades para as diferentes caselas.

GOOD (1963), utiliza o princípio de máxima entropia para gerar hipóteses nulas e o aplica principalmente em tabelas de contingência multidimensionais. De-

monstra também que se um conjunto completo de restrições de ordem r são dadas, então a hipótese de não interação de ordem r e interações de ordem superior, leva a obter um único conjunto de probabilidades para as caselas se as restrições são consistentes. Um tipo de dualidade entre máxima entropia e máxima verossimilhança é provada.

É importante mencionar o trabalho de BIRCH (1963), que obteve estimativas de máxima verossimilhança para as frequências esperadas numa tabela de contingência tridimensional. O autor demonstra que sob vários esquemas amostrais obtêm-se as mesmas estimativas de máxima verossimilhança, além de ressaltar as seguintes propriedades: (i) as equações de máxima verossimilhança têm solução única quando a verossimilhança é maximizada; (ii) as estimativas de máxima verossimilhança são as mesmas sob vários esquemas amostrais e; (iii) essas estimativas são estatísticas suficientes.

GOODMAN (1964), apresenta um teste baseado na análise das frequências observadas para testar a hipótese H_0 : a interação entre os três fatores é zero, numa tabela de contingência $2 \times 2 \times K$. Quando H_0 é verdadeira, o teste estatístico tem uma distribuição assintótica de χ^2 com $K-1$ graus de liberdade e o autor introduz um método para particionar esse teste estatístico em $K-1$ componentes assintoticamente independentes, cada um com um grau de liberdade. Nesse trabalho é desenvolvido um teste aproximado para H_0 em tabelas de contingência $2 \times J \times K$ e $l \times J \times K$. Temos que salientar que o autor apresenta

métodos que fazem possível estimar a magnitude da interação entre os três fatores, obter intervalos de confiança para essa interação e, ainda, testar a hipótese que essa interação é igual a algum valor especificado.

BHAPKAR (1966), demonstrou que para testar hipóteses lineares do tipo $H_0 : F_k(\underline{p}) = \phi$, $k = 1, 2, \dots, t$, onde as linhas de F_k são \underline{t} funções independentes das \underline{p} 's e $\underline{p}' = [p_{11}, p_{21}, \dots, p_{(r-1)1}, \dots, p_{1s}, p_{2s}, \dots, p_{(r-1)s}]$ em dados categorizados provenientes de \underline{s} populações multinomiais com \underline{r} categorias, a estatística X_1^2 de Neyman é equivalente (isto é, idêntica algébricamente) à estatística de Wald. Mostra também que para testar hipóteses não lineares essas estatísticas continuam sendo equivalentes se aplicamos a técnica de linearização de Neyman. Em ambos os casos a estatística X_1^2 deve ser definida, isto é, todas as freqüências das diferentes categorias são positivas.

BHAPKAR & KOCH (1968), consideram o relacionado com a formulação de hipóteses de não interação para diferentes situações experimentais que produzem dados na forma de tabelas de contingência multidimensionais. Enfatizam a distinção entre "fatores", os quais denotam uma classificação descrevendo algum aspecto da sub-população à que o indivíduo pertence e "respostas", as quais denotam uma classificação descrevendo alguma característica observável que possa acontecer com ele durante o experimento. Consideram ainda que "fatores" têm as marginais totais fixas e as "respostas" têm as marginais totais a-

leatorias. No caso de tabelas de contingência tridimensionais, têm-se os seguintes modelos: (i) três respostas, nenhum fator. Os aspectos considerados são análogos aos problemas de independência e correlação na análise multivariada; (ii) duas respostas, um fator. Neste caso estaremos interessados não apenas na associação entre as respostas mas também no efeito do fator sobre elas; (iii) uma resposta, dois fatores. Os aspectos considerados são análogos aos problemas da análise da variância. Os correspondentes testes estatísticos estão baseados nos critérios de Wald.

GOODMAN (1968), apresenta vários métodos adequados à análise de tabelas de contingência não truncadas (aqueias nas quais nenhuma das caselas é omitida) que podem também ser aplicados à análise de tabelas de contingência truncadas (aquelas nas quais as frequências de algumas das caselas são omitidas da análise ou porque foram perdidas, ou por serem vazias, ou por algum tipo de restrição). Numa tabela de contingência $R \times C$ é considerado um subconjunto S de caselas e as linhas e colunas são definidas como "quase-independentes" se as proporções p_{ij} podem ser escritas como $p_{ij} = a_i b_j$, para as caselas $(i, j) \in S$, onde a_i e b_j são constantes positivas. O autor destaca que esse conceito é muito útil para analisar tabelas nas quais algumas caselas são omitidas, pois permite uma análise mais detalhada da associação entre os níveis das variáveis categorizadas R e C .

De forma similar, ao caso de tabelas não

truncadas, quando numa tabela de contingência as linhas e as colunas não são quase-independentes, é de interesse estimar o tipo e a magnitude da interação ou associação presentes na tabela. GOODMAN (1968) discute a estimação das interações presentes nesses tipos de tabela e apresenta três métodos de análise para as tabelas truncadas: (i) um método baseado na partição da tabela truncada em apropriadas subtabelas; (ii) um método baseado na estimação das frequências esperadas calculadas sob a hipótese de quase-independência, e (iii) um método baseado na estimação de certas interações.

IRELAND & KULLBACK (1968), estimaram as probabilidades p_{ij} das caselas de uma tabela de contingência $R \times C$ quando as probabilidades marginais $p_{i.}$ e $p_{.j}$ são conhecidas e fixas, minimizando a informação de discriminação $I(p : \pi) = \sum_{i,j} p_{ij} \ln(p_{ij}/\pi_{ij})$, onde π_{ij} são as proporções observadas em cada casela.

Eles apontaram $p_{ij}^* = a_i b_j \pi_{ij}$; $p_{i.} = a_i \sum_j b_j \pi_{ij}$ e $p_{.j} = b_j \sum_i a_i \pi_{ij}$, como estimadores das probabilidades, onde os a_i e os b_j são determinados sujeitos às restrições das probabilidades marginais. Em lugar de achar os valores correspondentes aos a_i e aos b_j , consideram um processo iterativo para obter os valores de p_{ij}^* . Os autores usaram os resultados de Neyman (1949)¹ e demonstraram que os estimadores são BAN (*Best asymptotically normal*) e que o procedimento iterativo é convergente.

¹ NEYMAN, J. Contribution to the theory of the χ^2 test. Berkeley Symp. 239-273, 1949.

Usando as noções básicas da teoria da informação, KU & KULLBACK (1968) apresentam um estudo unificado da análise de tabelas de contingência multidimensionais e propuzeram o princípio da informação mínima de discriminação para gerar hipóteses de interesse quando certas marginais são consideradas como fixas. Para cada conjunto de marginais fixadas, é obtido um único conjunto de estimativas das probabilidades das caselas $\{p^*\}$, minimizando a informação de discriminação. Esse conjunto $\{p^*\}$ corresponde às estimativas das probabilidades das caselas representando não interação e podem ser expressas numa forma linear na escala logaritmica. Por exemplo,

$$\ln p_{ij}^* = \text{const.} + \ln a_i + \ln b_j, \text{ onde}$$

a_i e b_j são funções das probabilidades das caselas das correspondentes tabelas de uma entrada com as marginais fixas. A diferença entre as probabilidades estimadas dos dados (proporções observadas) e $\{p^*\}$ é por esse motivo uma medida da interação. A analogia entre a análise de informação e a análise de variância é ressaltada. Para reduzir as dimensões de uma tabela e também a complexidade da análise, a hipótese de não interação é interpretada como equivalente a "as tabelas com marginais fixas são suficientes e contêm toda a informação de uma tabela completa".

GRIZZLE *et alii* (1969), assumindo que há n_i , $i = 1, 2, \dots, s$, amostras de s distribuições mul-

tinomiais, cada uma com r categorias de respostas, definem μ funções das probabilidades verdadeiras de cada casela ($\pi_{ij} : i = 1, 2, \dots, s; j = 1, 2, \dots, r; \sum_j \pi_{ij} = 1$). Sob certas condições descrevem um procedimento geral não iterativo para ajustar essas funções a um modelo linear, para testar o grau de ajustamento do modelo e para testar hipóteses a respeito dos parâmetros do modelo linear. Nesse trabalho é considerada a análise de tabelas de contingência como uma subclasse de problemas que podem ser abordados através de regressão ponderada. Os autores acham que essa unificação é meritória devido à simplicidade com que os modelos e hipóteses podem ser formulados e testados.

PLACKETT (1969), considera que os métodos amostrais para arranjar dados em tabelas de contingência tridimensionais conduzem a uma distribuição de probabilidade que pode ser obtida supondo que as freqüências de cada casela são variáveis Poisson independentes, as quais são observadas sujeitas a certas restrições das freqüências marginais ou totais. Apresenta um resumo das idéias de BARTLETT (1935), ROY & KASTENBAUM (1956), DARROCH (1962), BIRCH (1963), GOOD (1963), GOODMAN (1964, 1968) e KU & KULLBACK (1968) sobre associação, interação, estimação e teste de hipóteses e diz que quando as freqüências esperadas são conhecidas, elas podem ser comparadas com as freqüências observadas, calculando-se

$$A^2 = 2 \sum_{i,j,k} n_{ijk} \ln (n_{ijk}/N\hat{p}_{ijk}) \quad \text{ou}$$

$$B^2 = \sum_{i,j,k} (n_{ijk} - N\hat{p}_{ijk})^2 / N\hat{p}_{ijk}.$$

Nesse caso, sob H_0 : não interação de segunda ordem, ambas as estatísticas convergem em distribuição para $\chi^2(v)$, onde $v = (r - 1)(s - 1)(t - 1)$, porém, para um grupo de hipóteses hierárquicas, os valores de A^2 são aditivos e fornecem uma exata partição de χ^2 , mas os valores de B^2 não são aditivos, exceto assintoticamente.

Para tabelas de contingência \underline{m} -dimensionais, GOODMAN (1970) discute tanto a estimação direta das interações multiplicativas entre as \underline{m} variáveis, como o teste indireto de hipóteses relativas a essas interações. O autor considera hipóteses hierárquicas que podem ser expressas em termos de um ou mais dos seguintes conceitos: (a) o conceito usual de independência e equiprobabilidade; (b) conceitos descrevendo propriedades condicionais relativas a um subconjunto das \underline{m} variáveis, dado o nível de algumas das variáveis restantes; (c) conceitos relacionados à usual logit-análise; e (d) conceitos relativos a um modelo log-linear.

FIENBERG (1970), descreve o procedimento iterativo de ajustamento proporcional (PIAP) proposto inicialmente por Deming & Stephan (1940)¹ para estimar as probabilidades de cada casela de uma tabela de contingência $R \times C$, baseadas nas observações e quando as marginais totais $p_{i.} = \sum_j p_{ij}$, $i = 1, 2, \dots, r$; e $p_{.j} = \sum_i p_{ij}$,

¹ DEMING, W.E. & STEPHAN, F.F. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* Baltimore, 11:427-444, 1940.

$j = 1, 2, \dots, c$, são conhecidas e fixas. Deming e Stephan propuseram o PIAP como uma forma de obter estimativas que minimizam

$$\sum_i \sum_j (n_{ij} - Np_{ij})^2 / n_{ij},$$

sujeito às restrições das marginais totais e assumindo $n_{ij} > 0$, mas não conseguiram provar a convergência. FIENBERG (1970), também apresenta um procedimento geométrico para demonstrar a convergência do processo iterativo numa tabela de contingência $R \times C$ e termina o seu trabalho com uma discussão sobre a extensão dos resultados obtidos a tabelas de contingência multidimensionais e a tabelas com alguns zeros nas caselas.

Suponha que \underline{X} é uma variável aleatória com função de distribuição de probabilidade $F(x)$, desconhecida. Antes de realizar um experimento é feita uma conjectura inicial e determina-se que a função de distribuição é $F_0(x)$. Depois de realizado o experimento é calculado o valor

$$\hat{a} = n^{-1} \sum_k T(X_k), \quad \text{onde}$$

T é uma função conhecida e X_1, X_2, \dots, X_k são os valores de \underline{X} em n ensaios independentes. De que forma pode o experimentador modificar a sua suposição acerca de $F(x)$, levando em conta a nova informação dada por esse número? CAMPBELL (1970), considera que se pode modificar essa conjectura através do uso de uma extensão do princípio de Gauss para a obtenção da distribuição de probabilidades

normal e/ou através do método da informação mínima de discriminação e demonstra que esses métodos conduzem à mesma afirmação acerca de $F(x)$.

KU *et alii* (1971) descrevem a estimação através do princípio de informação mínima de discriminação e é usado para obter estimativas para testes de hipóteses referentes a interações e efeitos na análise de tabelas de contingência multidimensionais, quando algumas marginais são consideradas fixas. Eles destacam que quando se passa de uma tabela de contingência de dupla entrada à uma tabela multidimensional, aparecem novos problemas, dentre os quais destacam: (i) formulação de hipóteses de não interação; (ii) computação das frequências esperadas e (iii) a interpretação de resultados.

Um método passo a passo (*step-by-step*) é apresentado por GOODMAN (1971) para particionar um certo tipo de hipóteses H acerca de uma tabela de contingência multidimensional. No caso de tabelas tridimensionais, uma hipótese H pode ser particionada em (a) uma hipótese H' correspondente a uma tabela marginal de dupla entrada ignorando uma das três dimensões da tabela e (b) uma hipótese H'' sobre independência, independência condicional ou equiprobabilidade condicional na tabela tridimensional. É também proposto um procedimento passo a passo para calcular estimativas de máxima verossimilhança para as frequências esperadas numa tabela de contingência multidimensional sob H . Aplicando os métodos propostos nesse trabalho numa tabela tridimensional, pode-se facilitar a

análise de seis hipóteses hierárquicas acerca dos correspondentes modelos log-lineares.

Um procedimento iterativo é proposto por GOKHALE (1971), para obter estimativas de máxima verossimilhança das freqüências esperadas para uma classe de modelos log-lineares em experimentos multinomiais. A convergência do método também é demonstrada, porém o número de ciclos necessários para consegui-la é elevado.

GOKHALE (1972) apresenta um processo iterativo para obter estimativas de máxima verossimilhança das freqüências esperadas das caselas e dos parâmetros de um modelo log-linear num experimento multinomial, com a diferença do proposto por GOKHALE (1971), que os coeficientes dos parâmetros do modelo log-linear podem ser diferentes de zero ou um. GRIZZLE & WILLIAMS (1972), descrevem métodos de ajustamento de modelos log-lineares para dados multinomiais dispostos em tabelas de contingência multidimensionais. São desenvolvidos métodos estatísticos para testar independência numa tabela de contingência multidimensional, os quais são baseados na correspondência ou similaridade entre a análise de experimentos fatoriais e testes de independências marginais. Modelos lineares da análise de regressão através dos mínimos quadrados ponderados são usados para calcular estimativas de probabilidades.

KULLBACK & FISHER (1973), examinaram uma particular tabela de contingência $9 \times 2 \times 2$, na qual a variável idade tem nove categorias e a hipótese de não

interação de segunda ordem foi rejeitada. Eles consideraram razoável supor que a presença da interação de segunda ordem deve estar relacionada a um efeito de idade e então propuseram uma partição dessa hipótese. O estudo foi baseado no princípio da informação mínima de discriminação, a representação dos modelos log-lineares associados e na análise de tabelas de informação. As estimativas das frequências esperadas foram obtidas usando o algoritmo iterativo de ajustamento de marginais de Deming-Stephan.

Um procedimento geral para achar a distribuição discreta que minimiza a informação de discriminação com respeito a uma distribuição de probabilidade dada e sujeita a um conjunto de restrições lineares sobre as probabilidades é apresentada por GOKHALE (1973).

Considere s variáveis aleatórias discretas independentes X_i , as quais podem tomar d_i valores $(x_{i1}, x_{i2}, \dots, x_{id_i}) = \underline{x}'_i$, com probabilidades $\underline{p}'_i = (p_{i1}, p_{i2}, \dots, p_{id_i})$; $i = 1, 2, \dots, s$. Seja $\underline{p}' = (\underline{p}'_1, \underline{p}'_2, \dots, \underline{p}'_s)$ um vetor de dimensão $d = \sum_i d_i$ e \underline{P} um conjunto de todos os vetores \underline{p} . Para um dado $\underline{\pi}' = (\underline{\pi}'_1, \underline{\pi}'_2, \dots, \underline{\pi}'_s) \in P$ e $\underline{p} \in P$, a função de informação generalizada é dada por:

$$I(\underline{p} : \underline{\pi}) = \sum_i W_i \sum_{k=1}^{d_i} p_{ik} \ln(p_{ik}/\pi_{ik}), \text{ onde}$$

as constantes W_i são conhecidas e tais que $\sum_i W_i = 1$; $0 < W_i < 1$ e $\pi_{ik} > 0$. Suponha que as probabilidades p_{ik} estão sujeitas a r restrições lineares, além da restrição normal que $\sum_k p_{ik} = 1$. Isso pode ser expresso como

$B \underline{p} = \underline{\theta}$, onde \underline{B} é uma matriz de dimensão $(s + r) \times d$.

No trabalho de GOKHALE (1973), é considerado o problema da escolha de um $\underline{p}_0 \in P$ tal que minimiza $l(\underline{p}; \underline{\pi})$ sujeito à restrição $B \underline{p} = \underline{\theta}$. O \underline{p}_0 é obtido por um processo iterativo que não requer inversão de matrizes, usando um algoritmo que é uma extensão do algoritmo de Darroch e Ratcliff.

KU & KULLBACK (1974), ilustram com dois exemplos a estimação e teste de parâmetros no modelo log-linear o qual é para tabelas de contingência o correlativo do modelo linear no caso de variáveis contínuas. Em ambos os casos, tem-se que propor e testar se o modelo é adequado e depois melhorá-lo se é necessário; adicionando ou excluindo parâmetros. Se o modelo é adequado, procede-se a estimar e testar os parâmetros. Porém, tem as seguintes diferenças:

1. O princípio usado na estimação.
2. Nos modelos log-lineares, em lugar de primeiro computar as estimativas dos parâmetros, computam-se através de um processo iterativo as estimativas das frequências de cada casela e depois obtêm-se as estimativas dos parâmetros.
3. Nos modelos de análise da variância, geralmente nosso interesse está centrado nos efeitos principais e nas interações de ordem mais baixa, enquanto que na análise de tabelas de contingência multidimensionais algumas interações de alta ordem são também importantes em nosso estudo.

Modelos log-lineares para tabelas de contingência são formulados por NELDER (1974) como um caso especial de modelos generalizados com uma componente aditiva sistemática \underline{Y} , erros com distribuição Poisson e uma função de ligação exponencial entre os valores esperados das observações e a componente sistemática \underline{Y} .

Se as estimativas de máxima verossimilhança das freqüências das caselas são funções explícitas de estatísticas suficientes os modelos log-lineares são denominados ou classificados como diretos. Em caso contrário são classificados como indiretos. Para modelos saturados (portanto, diretos), GOODMAN (1970) e BISHOP *et alii* (1975), usaram o método δ para calcular as variâncias assintóticas dos $\hat{\mu}$ termos dos modelos log-lineares. LEE (1977), faz uma extensão para os modelos log-lineares hierárquicos diretos e não saturados.

NUNES MAIA (1977), faz um estudo resumido baseado no livro de KULLBACK (1959) no referente a tabelas de contingência completas com duas e três entradas e dá alguns exemplos de aplicação da informação estatística.

GOKHALE & KULLBACK (1978), fizeram uma revisão sobre a aplicação da informação mínima de discriminação (IMD), em análise de dados categorizados na forma de perguntas e respostas e acreditam que a aproximação IMD tem, principalmente, as seguintes vantagens:

1. Fornece um tratamento unificado para dados categorizados não necessariamente arranjados como

uma tabela de contingência ou para tabelas de contingência de qualquer dimensão,

2. Estão disponíveis programas computacionais para realizar as análises necessárias, e

3. A aproximação IMD é razoável no sentido que as estimativas IMD são, em geral, estimativas BAN e para um grande tipo de análise são estimativas de máxima verossimilhança.

CANTON & FESTA (1978), fizeram um levantamento cronológico dos principais artigos envolvendo aplicações de dados categorizados usando modelos lineares e descrevem o artigo de GRIZZLE *et alii* (1969) que, segundo eles, originou essa série de trabalhos.

Além da listagem cronológica dos artigos é apresentado um breve resumo dos objetivos inerentes a cada um dos artigos.

PERES (1978), comparando o uso prático da estatística " χ^2 de Pearson" e a estatística obtida através da IMD, diz que o teste da Teoria da Informação apresenta duas vantagens:

1. O cálculo do valor observado da estatística é mais fácil e,

2. A estatística da IMD pode ser particionada em parcelas que permitem explorar a rejeição da hipótese nula, fornecendo uma interpretação completa dos resultados.

O autor destaca que, embora o teste da teo

ria da informação apresenta algumas vantagens em relação ao χ^2 de Pearson, ele não é divulgado nos livros textos elementares, da mesma forma que o χ^2 de Pearson, devido ao fato de que o teste da Teoria da Informação é sempre apresentado nos livros e artigos com um formalismo rigoroso, exigindo do leitor um conhecimento bastante grande de estatística matemática.

CANTON (1980), apresenta de maneira sucinta uma metodologia geral para análise de dados categorizados baseado no trabalho de GRIZZLE *et alii* (1969) e afirma que esse enfoque tem a vantagem de permitir ao analista maior flexibilidade na escolha de modelos e testes de hipóteses.

LIMA (1986), apresenta o modelo log-linear como um caso particular do modelo linear generalizado e dá vários exemplos de sua utilização através de GLIM (*Generalized Linear Interactive Modelling*).

3. DESENVOLVIMENTO TEÓRICO

3.1. Introdução

O desenvolvimento da metodologia baseia-se em um experimento onde os indivíduos de uma população podem ser classificados como pertencentes a uma das i categorias da variável R , a uma das j categorias da variável S e a uma das k categorias ou níveis da variável ou atributo T . Ao tomar uma amostra de tamanho N de tais indivíduos; o resultado será uma tabela de contingência tridimensional, onde as frequências observadas em cada célula são consideradas como provenientes de uma única população multinomial.

Considere a Tabela 1 de contingência tridimensional, para efeito de se estabelecer a notação.

Tabela 1. Distribuição de frequências observadas.

	j = 1			j = 2			j = s		
	k=1	k=2	...k=t	k=1	k=2	...k=t	k=1	k=2	...k=t
i = 1	x_{111}	x_{112}	$\dots x_{11t}$	x_{121}	x_{122}	$\dots x_{12t}$	x_{1s1}	x_{1s2}	$\dots x_{1st}$
i = 2	x_{211}	x_{212}	$\dots x_{21t}$	x_{221}	x_{222}	$\dots x_{22t}$	x_{2s1}	x_{2s2}	$\dots x_{2st}$
.....										
.....										
.....										
i = r	x_{r11}	x_{r12}	$\dots x_{r1t}$	x_{r21}	x_{r22}	$\dots x_{r2t}$	x_{rs1}	x_{rs2}	$\dots x_{rst}$

Tem-se que x_{ijk} é o número de indivíduos da amostra pertencentes à casela com as (i, j, k) -ésimas categorias e cumpre-se que

$$\sum_i x_{ijk} = x_{.jk}; \quad \sum_j x_{ijk} = x_{i.k}; \quad \sum_k x_{ijk} = x_{ij.};$$

$$\sum_{i,j} x_{ijk} = \sum_j x_{.jk} = \sum_i x_{i.k} = x_{..k};$$

$$\sum_{i,k} x_{ijk} = \sum_k x_{.jk} = \sum_i x_{ij.} = x_{.j.};$$

$$\sum_{j,k} x_{ijk} = \sum_j x_{ij.} = \sum_k x_{i.k} = x_{i..} \quad e$$

$$\sum_{i,j,k} x_{ijk} = \sum_{i,j} x_{ij.} = \sum_{i,k} x_{i.k} = \sum_{j,k} x_{.jk} = \sum_i x_{i..} =$$

$$= \sum_j x_{.j.} = \sum_k x_{..k} = x_{...} = N.$$

A probabilidade de um indivíduo ser classificado ou pertencer à (i, j, k) -ésima casela é p_{ijk} , com a restrição que $p_{ijk} > 0$ e que $\sum_{i,j,k} p_{ijk} = 1$. Isto é, neste trabalho considera-se a análise e interpretação de dados categorizados provenientes de uma única população multinomial, dispostos em uma tabela de contingência completa tridimensional $R \times S \times T$. Assume-se que os indivíduos são amostrados independentemente ao acaso, assim como também que a probabilidade de ser classificado na (i, j, k) -ésima casela, permanece constante de indivíduo para indivíduo.

Para as probabilidades p_{ijk} , definem-se relações similares àquelas das freqüências observadas, isto é, $\sum_i p_{ijk} = p_{.jk}$; $\sum_{jk} p_{ijk} = p_{i..}$; $\sum_{i,j,k} p_{ijk} =$

$$\sum_{ij} p_{ij.} = \sum_i p_{i..} = 1; \text{ etc.}$$

3.2. Distribuição multinomial

Assume-se que um indivíduo da população sob estudo pode ser classificado como pertencente a uma e só uma das $R \times S \times T$ categorias com probabilidade p_{ijk} , de tal forma que $\sum_{i,j,k} p_{ijk} = 1$. Isto é, essas categorias são mutuamente exclusivas e exaustivas (FELLER, 1968).

Uma distribuição multinomial é especificada por um vetor $\underline{p}' = (p_{111}, p_{112}, \dots, p_{11t}, p_{121}, \dots, p_{12t}, \dots, p_{1st}, \dots, p_{rst})$. Se N ensaios independentes produzem x_{111} eventos do tipo 111, x_{112} eventos do tipo 112, e assim sucessivamente até produzir x_{rst} eventos do tipo rst, então tem-se que

$$P(X_{111} = x_{111}, X_{112} = x_{112}, \dots, X_{rst} = x_{rst}) =$$

$$= \frac{N!}{x_{111}! x_{112}! \dots x_{rst}!} p_{111}^{x_{111}} p_{112}^{x_{112}} \dots p_{rst}^{x_{rst}}, \quad \text{onde}$$

$$x_{rst} = N - x_{111} - x_{112} - \dots - x_{rs(t-1)} \text{ e } p_{rst} = 1 - p_{111} -$$

$p_{112} - \dots - p_{rs(t-1)}$. Com o intuito de simplificar a notação, as caselas serão enumeradas em ordem lexicográfica

a partir da casela 111 = 1, até a casela rst = ℓ . Considerando-se, por exemplo, uma tabela de contingência 2 x 2 x 3, isto é, R = 2; S = 2 e T = 3, as freqüências das caselas enumeradas em ordem lexicográfica, são:

$$x_{111} = x_1; x_{112} = x_2; x_{113} = x_3; x_{121} = x_4; x_{122} = x_5;$$

$$x_{123} = x_6; x_{211} = x_7; x_{212} = x_8; x_{213} = x_9; x_{221} = x_{10};$$

$$x_{222} = x_{11}; x_{223} = x_{12}.$$

A função densidade de probabilidade pode ser expressa como sendo:

$$P(X_1 = x_1, \dots, X_{\ell-1} = x_{\ell-1}, X_{\ell} = x_{\ell}) = \frac{N!}{x_1! \dots x_{\ell-1}! x_{\ell}!} p_1^{x_1} \dots p_{\ell-1}^{x_{\ell-1}}$$

$$\text{onde } x_{\ell} = N - x_1 - x_2 - \dots - x_{\ell-1} \text{ e}$$

$$p_{\ell} = 1 - p_1 - p_2 - \dots - p_{\ell-1}.$$

A função geradora de momentos da distribuição multinomial (HOGG & CRAIG, 1978; MOOD *et alii*, 1974) é

$$M(t_1, t_2, \dots, t_{\ell-1}) = (p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_{\ell-1} e^{t_{\ell-1}} + p_{\ell})^N$$

$$\text{Sabe-se que } E(X_i) = \left. \frac{\delta M}{\delta t_i} \right|_{t_i = 0, \forall i} \quad i=1, 2, \dots, \ell-1. \quad \therefore$$

$$\frac{\delta M}{\delta t_i} = N (p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_{\ell-1} e^{t_{\ell-1}} + p_{\ell})^{N-1} p_i e^{t_i} \therefore$$

$$E(X_i) = N p_i \quad (1)$$

$$E(X_i^2) = \frac{\delta^2 M}{\delta t_i^2} \Big|_{t_i=0, \forall i}, \quad i=1, 2, \dots, \ell-1. \therefore$$

$$\frac{\delta^2 M}{\delta t_i^2} = N(N-1) (p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_{\ell-1} e^{t_{\ell-1}} + p_{\ell})^{N-2} (p_i e^{t_i})^2 + N (p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_{\ell-1} e^{t_{\ell-1}} + p_{\ell})^{N-1} p_i e^{t_i} \therefore$$

$$E(X_i^2) = N(N-1) p_i^2 + N p_i = N^2 p_i^2 - N p_i^2 + N p_i, \quad \text{logo}$$

$$V(X_i) = E(X_i^2) - [E(X_i)]^2 = N^2 p_i^2 - N p_i^2 + N p_i - N^2 p_i^2 \therefore$$

$$V(X_i) = N p_i (1 - p_i) \quad (2)$$

$$E(X_i X_j) = \frac{\delta^2 M}{\delta t_i \delta t_j} \Big|_{t_i=0; t_j=0, \forall i, j} \therefore$$

$$\frac{\delta^2 M}{\delta t_i \delta t_j} = N(N-1) p_i e^{t_i} (p_1 e^{t_1} + p_2 e^{t_2} + \dots + p_{\ell-1} e^{t_{\ell-1}} + p_{\ell})^{N-2} p_j e^{t_j}$$

$$E(X_i X_j) = N(N-1) p_i p_j$$

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i) E(X_j), \text{ logo}$$

$$\text{Cov}(X_i, X_j) = N(N-1)p_i p_j - Np_i Np_j \dots$$

$$\text{Cov}(X_i, X_j) = -N p_i p_j. \text{ Tem-se então que se} \quad (3)$$

$$\underline{X}' = (X_1, X_2, \dots, X_{\ell-1}), \text{ então}$$

$$V(\underline{X}) = N \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & -p_1 p_3 & \dots & -p_1 p_{\ell-1} \\ & p_2(1-p_2) & -p_2 p_3 & \dots & -p_2 p_{\ell-1} \\ & & p_3(1-p_3) & \dots & -p_3 p_{\ell-1} \\ \dots & \dots & \dots & \dots & \dots \\ & & & & p_{\ell-1}(1-p_{\ell-1}) \end{bmatrix} \quad (4)$$

Geralmente, as probabilidades populacionais não são conhecidas e têm que ser estimadas. Para isso, pode-se aplicar o método de estimação da máxima verossimilhança. Os estimadores obtidos através desse método têm ótimas propriedades relacionadas com as ótimas propriedades das grandes amostras (MOOD *et alii*, 1974).

3.2.1. Estimação das Probabilidades (parâmetros) de uma Distribuição Multinomial

Ao tomar uma amostra de N indivíduos de

uma população com distribuição multinomial, tem-se a função de verossimilhança

$$L(\underline{x}; \underline{p}) = \frac{N!}{x_1! x_2! \cdots x_{\ell-1}! x_{\ell}!} p_1^{x_1} p_2^{x_2} \cdots p_{\ell-1}^{x_{\ell-1}} p_{\ell}^{x_{\ell}} \quad (5)$$

e sabe-se que o ponto de máximo de $L(\underline{x}, \underline{p})$ coincide com o ponto de máximo de $\ln L(\underline{x}, \underline{p})$.

$$\ln L(\underline{x}, \underline{p}) = \ln \frac{N!}{\prod_{i=1}^{\ell} x_i!} + \sum_{i=1}^{\ell} x_i \ln p_i.$$

Maximizando essa função, sujeito à restrição que $\sum_i p_i = 1$; $p_i > 0$, $i = 1, 2, \dots, \ell$ e usando o método dos multiplicadores de Lagrange, tem-se que achar o máximo da função

$$L(\underline{p}, \lambda) = \ln \frac{N!}{\prod_{i=1}^{\ell} x_i!} + \sum_{i=1}^{\ell} x_i \ln p_i + \lambda \left(\sum_{i=1}^{\ell} p_i - 1 \right) \dots$$

$$\frac{\partial L(\underline{p}, \lambda)}{\partial p_i} = \frac{x_i}{p_i} + \lambda \quad \text{e} \quad \frac{\partial L(\underline{p}, \lambda)}{\partial \lambda} = \sum_{i=1}^{\ell} p_i - 1. \quad \text{Portanto,}$$

$$\frac{x_i}{\hat{p}_i} + \hat{\lambda} = 0 \quad \text{e} \quad \sum_{i=1}^{\ell} \hat{p}_i - 1 = 0 \quad \dots$$

$$x_i + \hat{\lambda} \hat{p}_i = 0 \quad \text{e} \quad \sum_{i=1}^{\ell} \hat{p}_i = 1, \quad \text{logo}$$

$$\sum_{i=1}^{\ell} x_i + \hat{\lambda} \sum_{i=1}^{\ell} \hat{p}_i = 0, \quad \text{mas} \quad \sum_{i=1}^{\ell} x_i = N \quad \text{e} \quad \sum_{i=1}^{\ell} \hat{p}_i = 1,$$

então $N + \hat{\lambda} = 0 \Rightarrow \hat{\lambda} = -N$. Logo

$$\frac{x_i}{\hat{p}_i} - N = 0 \Rightarrow \hat{p}_i = \frac{x_i}{N}, \quad i = 1, 2, \dots, \ell$$

$\hat{p}_i = x_i/N$ é uma estimativa de p_i e o estimador de máxima verossimilhança de p_i é dado por

$$\hat{p}_i = \frac{X_i}{N} \quad (6)$$

3.2.2. Distribuição dos \hat{p}_i

$$E(\hat{p}_i) = E\left(\frac{X_i}{N}\right) = \frac{1}{N} E(X_i) = \frac{1}{N} N p_i = p_i$$

$$V(\hat{p}_i) = V\left(\frac{X_i}{N}\right) = \frac{1}{N^2} V(X_i) = \frac{1}{N^2} N p_i (1 - p_i) \dots$$

$$V(\hat{p}_i) = \frac{p_i (1 - p_i)}{N}$$

$$\text{Cov}(\hat{p}_i, \hat{p}_j) = \text{Cov}\left(\frac{X_i}{N}, \frac{X_j}{N}\right) = \frac{1}{N^2} \text{Cov}(X_i, X_j) =$$

$$\frac{1}{N^2} (-N p_i p_j) \dots \text{Cov}(\hat{p}_i, \hat{p}_j) = -\frac{p_i p_j}{N}$$

Da função geradora de momentos da distribuição multinomial, tem-se que a função geradora de momentos da distribuição marginal de X_i , é:

$$M(0, 0, \dots, t_i, 0, \dots, 0) = (p_1 + p_2 + \dots + p_i e^{t_i} + p_{i+1} + \dots + p_\ell)^N$$

$$M(0, 0, \dots, t_i, 0, \dots, 0) = [p_i e^{t_i} + (1 - p_i)]^N,$$

que é a função geradora de momentos de uma distribuição binomial. Isto é, $X_i \sim B(N, p_i)$,

$$f(x_i) = \frac{N!}{x_i! (N - x_i)!} p_i^{x_i} (1 - p_i)^{N - x_i}, \quad x_i = 0, 1, \dots, N$$

De forma análoga, verifica-se que a distribuição marginal das variáveis X_i e X_j , é trinomial, isto é,

$$f(x_i, x_j) = \frac{N!}{x_i! x_j! (N - x_i - x_j)!} p_i^{x_i} p_j^{x_j} (1 - p_i - p_j)^{N - x_i - x_j}$$

De (6) tem-se que o estimador de máxima verossimilhança de p_i é a estatística

$$\hat{p}_i = \frac{X_i}{N} \quad \therefore \quad X_i = N \hat{p}_i, \quad \text{logo}$$

$$g(\hat{p}_i) = \frac{N!}{(N \hat{p}_i)! (N - N \hat{p}_i)!} p_i^{N \hat{p}_i} (1 - p_i)^{N - N \hat{p}_i};$$

$$\hat{p}_i = 0, \frac{1}{N}, \dots, 1 \quad (7)$$

3.2.3. Resumo dos Resultados Obtidos no Ítem 3.2.

Considerando de novo a tabela de contingência tridimensional $R \times S \times T$, tem-se que:

(i) x_{ijk} é a freqüência observada numa amostra de tamanho N do número de indivíduos pertencentes à (i, j, k) -ésima casela, com a restrição que $\sum_{i,j,k} x_{ijk} = N$

(ii) p_{ijk} , é a probabilidade de um indivíduo ser classificado como pertencente à (i, j, k) -ésima casela, com a restrição que $p_{ijk} > 0$ e $\sum_{i,j,k} p_{ijk} = 1$

(iii) admite-se que as variáveis aleatórias X_{ijk} , $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ e $k = 1, 2, \dots, t$, estão distribuídas multinomialmente com parâmetros $\underline{p} = (p_{111}, p_{112}, \dots, p_{11t}, p_{121}, \dots, p_{12t}, \dots, p_{rst}, \dots, p_{rst})$, portanto $E(X_{ijk}) = N p_{ijk}$,

$$V(X_{ijk}) = N p_{ijk} (1 - p_{ijk}) \text{ e}$$

$$\text{Cov}(X_{ijk}, X_{lmn}) = -N p_{ijk} p_{lmn}; i \neq l, j \neq m \text{ e } k \neq n$$

(iv) o estimador de máxima verossimilhança de p_{ijk} é $\hat{p}_{ijk} = X_{ijk}/N$, que é uma variável aleatória distribuída com uma função densidade de probabilidade dada pela expressão (7), e $E(\hat{p}_{ijk}) = p_{ijk}$

$$V(\hat{p}_{ijk}) = (1/N) p_{ijk} (1 - p_{ijk}) \text{ e}$$

$$\text{Cov}(\hat{p}_{ijk}, \hat{p}_{lmn}) = -\frac{1}{N} p_{ijk} p_{lmn}; i \neq l, j \neq m \text{ e } k \neq n$$

(v) se m_{ijk} é a freqüência esperada do número de indivíduos pertencentes à (i, j, k) -ésima casela, numa amostra de tamanho N , então $m_{ijk} = E(X_{ijk}) = N p_{ijk}$, portanto, a estimativa de máxima verossimilhança da fre-

freqüência esperada na (i, j, k) -ésima casela é $\hat{m}_{ijk} = E(\widehat{X}_{ijk}) = N \cdot \hat{p}_{ijk} = N \hat{p}_{ijk} = x_{ijk}$.

Evidentemente, cumpre-se também que

$$\sum_{i,j,k} \hat{m}_{ijk} = \sum_{i,j,k} N \hat{p}_{ijk} = N \sum_{i,j,k} \hat{p}_{ijk} = N \sum_{i,j,k} \frac{x_{ijk}}{N}$$

$$\sum_{i,j,k} \hat{m}_{ijk} = \sum_{i,j,k} x_{ijk} = N,$$

$$\hat{m}_{ij.} = \sum_k \hat{m}_{ijk} = \sum_k N \hat{p}_{ijk} = N \sum_k \frac{x_{ijk}}{N} = \sum_k x_{ijk} = x_{ij.},$$

$$\hat{m}_{i..} = \sum_{j,k} \hat{m}_{ijk} = N \sum_{j,k} \hat{p}_{ijk} = \sum_{j,k} x_{ijk} = x_{i..}, \quad \text{e assim}$$

sucessivamente.

Tem-se então que a freqüência observada x_{ijk} é a estimativa de máxima verossimilhança da freqüência esperada m_{ijk} ; $x_{ij.}$ é a estimativa de máxima verossimilhança de $m_{ij.}$; e assim sucessivamente. Tem que se destacar que ao impor outras restrições sobre as probabilidades marginais unidimensionais ou bidimensionais, obtêm-se outras estimativas de máxima verossimilhança para as freqüências esperadas m_{ijk} .

3.3. Tabelas de Contingência R x S x T

3.3.1. Introdução

Quando se observam três variáveis categorizadas simultaneamente, tem-se uma tabela de contingência tridimensional, onde cada variável corresponde a uma dimensão da tabela. Essas tabelas apresentam problemas

especiais de análise e interpretação e segundo FIENBERG (1978), esses problemas têm ocupado um lugar de destaque na literatura estatística desde o primeiro artigo de BARTLETT (1935), relacionado com teste de hipóteses numa tabela de contingência $2 \times 2 \times 2$.

Uma tabela de contingência está relacionada basicamente com categorias qualitativas e dados de contagem. Segundo KU *et alii* (1971), a informação que se deseja obter de uma tabela de contingência é de natureza preliminar ou exploratória. Deseja-se inferir, com base numa amostra, se essas características qualitativas estão associadas uma com outra ou se elas são independentes.

Na análise de variância univariada a presença ou existência de interação indica que os efeitos de algum fator sobre a variável resposta depende do nível de algum outro fator. Em tabelas de contingência o termo interação tem sido usado para indicar a dependência ou uma adequada medida de associação entre várias respostas dentro dos níveis ou categorias de outras variáveis respostas ou dentro das categorias de outras variáveis consideradas como fatores, de acordo com a distinção feita por BHAPKAR & KOCH (1968).

A relação entre o conceito de independência ou associação e interação foi expressado por BARTLETT (1935), ao afirmar que o teste de independência numa tabela 2×2 pode ser considerado como testar a significância da interação entre as duas classificações. Ele definiu a hipótese de "não interação de segunda ordem" numa

tabela 2 x 2 x 2, como implicando que

$$\frac{P_{111} P_{221}}{P_{121} P_{211}} = \frac{P_{112} P_{222}}{P_{122} P_{212}},$$

que, pela sua vez, estabelece a igualdade de associação entre as classificações \underline{R} e \underline{S} nas categorias T_1 e T_2 ; entre as classificações \underline{S} e \underline{T} nas categorias R_1 e R_2 e entre as classificações \underline{R} e \underline{T} nas categorias S_1 e S_2 .

Toda a informação que o pesquisador espera obter de um experimento está contida nas observações arranjadas na tabela de contingência. Na análise de uma tabela de contingência o objetivo é poder expressar a tabela observada através de um número de parâmetros que dependam de algumas ou de todas as freqüências marginais. Por exemplo, se a interação de primeira ordem é não significativa, então admite-se que toda a informação da tabela está contida nas freqüências marginais de primeira ordem ou freqüências marginais unidimensionais, no sentido que conhecendo essas marginais, a tabela completa pode ser construída dentro do limite do erro experimental.

3.3.2. Modelo log-linear

Na análise de tabelas de contingência, admite-se um modelo linear na escala logaritmica para as probabilidades das respectivas caselas. BISHOP *et alii* (1977), por analogia com os modelos de análise da variân-

cia (ANOVA), propõem o seguinte modelo log-linear:

$$\log p_{ijk} = \mu + \mu_1(i) + \mu_2(j) + \mu_3(k) + \mu_{12}(ij) + \mu_{13}(ik) + \mu_{23}(jk) + \mu_{123}(ijk), \quad (1)$$

com $i = 1, 2, \dots, r$, $j = 1, 2, \dots, s$ e $k = 1, 2, \dots, t$.

$$\text{Impondo as restrições } \sum_i \mu_1(i) = \sum_j \mu_2(j) =$$

$$\sum_k \mu_3(k) = \sum_i \mu_{12}(ij) = \sum_j \mu_{12}(ij) = \sum_{i,k} \mu_{13}(ik) = \sum_{j,k} \mu_{23}(jk) =$$

$$\sum_{i,j,k} \mu_{123}(ijk) = 0, \text{ tem-se}$$

$$\sum_{i,j,k} \log p_{ijk} = IJK \mu \quad \therefore \mu = \frac{\sum_{i,j,k} \log p_{ijk}}{IJK}, \text{ isto é,}$$

μ é a média dos logaritmos das probabilidades.

$$\sum_{j,k} \log p_{ijk} = JK \mu + JK \mu_1(i) \quad \therefore \mu + \mu_1(i) = \frac{\sum_{j,k} \log p_{ijk}}{JK} \quad (2)$$

isto é, $\mu + \mu_1(i)$ é a média dos logaritmos das probabilidades no nível i da primeira variável, ou seja que

$$\mu_1(i) = \frac{\sum_{j,k} \log p_{ijk}}{JK} - \mu,$$

representam desvios com relação à média μ .

Chamando $\log p_{ijk} = \ell_{ijk}$, tem-se que

$$\sum_{j k} \log p_{ijk} = \sum_{i j k} \ell_{ijk} = \ell_{...}; \quad \sum_{j k} \log p_{ijk} =$$

$\sum_{j k} \ell_{ijk} = \ell_{i...}$, e assim sucessivamente, obtem-se:

$$\mu = \frac{\ell_{...}}{IJK}, \text{ média geral}$$

$$\mu_1(i) = \frac{\ell_{i..}}{JK} - \mu, \text{ efeito principal do nível } i \text{ da variável } \underline{R},$$

$$\mu_2(j) = \frac{\ell_{.j.}}{IK} - \mu, \text{ efeito principal do nível } j \text{ da variável } \underline{S},$$

$$\mu_3(k) = \frac{\ell_{..k}}{IJ} - \mu, \text{ efeito principal do nível } k \text{ da variável } \underline{T}.$$

As interações de primeira ordem são

$$\mu_{12}(ij) = \frac{\ell_{ij.}}{K} - [\mu + \mu_1(i) + \mu_2(j)],$$

$$\mu_{13}(ik) = \frac{\ell_{i.k}}{J} - [\mu + \mu_1(i) + \mu_3(k)] \text{ e}$$

$$\mu_{23}(jk) = \frac{\ell_{.jk}}{I} - [\mu + \mu_2(j) + \mu_3(k)].$$

A interação de segunda ordem é

$$\begin{aligned} \mu_{123}(ijk) = & \ell_{ijk} - [\mu + \mu_1(i) + \mu_2(j) + \mu_3(k) + \dots + \\ & + \mu_{23}(jk)]. \end{aligned}$$

Observa-se que os efeitos principais são funções das somas marginais dos logaritmos das probabili-

dades em uma determinada categoria da variável, mas essas somas não correspondem às somas marginais $p_{i..}$, $p_{.j.}$ e $p_{..k}$ na escala original. Observação análoga pode-se fazer com respeito às interações de primeira e de segunda ordem.

As restrições impostas no modelo (1), reduzem o número de parâmetros independentes representados em cada termo. Por exemplo, o valor de $\mu_1(i)$ é diferente para cada uma das i categorias da variável R , mas a restrição imposta reduz o número de parâmetros independentes a $i - 1$. Pode-se estabelecer então o seguinte quadro:

Quadro 1. Graus de liberdade associados a cada termo do modelo log-linear (1).

Termo	Parâmetros independentes = Graus de liberdade
μ	1
$\mu_1(i)$	$(I - 1)$
$\mu_2(j)$	$(J - 1)$
$\mu_3(k)$	$(K - 1)$
$\mu_{12}(ij)$	$(I - 1)(J - 1)$
$\mu_{13}(ik)$	$(I - 1)(K - 1)$
$\mu_{23}(jk)$	$(J - 1)(K - 1)$
$\mu_{123}(ijk)$	$(I - 1)(J - 1)(K - 1)$
TOTAL	IJK

Observa-se que o modelo log-linear (1) tem

um número de parâmetros independentes igual ao número de caselas, ou seja que é um modelo saturado.

3.3.3. Modelos hierárquicos

Em geral, tem-se que se v é o número de dimensões de uma tabela de contingência, então o número de μ -termos no modelo log-linear saturado é igual a 2^v . No caso de tabelas de contingência tridimensionais verifica-se que o modelo log-linear saturado contém $2^3 = 8$ μ -termos. Cada μ -termo inclui um número de parâmetros independentes mostrados no Quadro 1.

Considere dois μ -termos. Um com c sub-índices e outro com b sub-índices, onde $c > b$. Diz-se que os termos são relativos se os c sub-índices contêm entre eles todos os b sub-índices. Por exemplo, $\mu_{12}(ij)$ é um relativo de ordem superior tanto de $\mu_1(i)$ como de $\mu_2(j)$ e $\mu_{123}(ijk)$ é um relativo de ordem superior de todos os μ -termos. Um conjunto de modelos é definido como hierárquico no caso em que quando um μ -termo é igual a zero, então todos os μ -termos relativos de ordem superior também são iguais a zero. Inversamente, se um μ -termo é diferente de zero, então os seus μ -termos de ordem inferiores devem estar presentes no modelo log-linear. Por exemplo, se $\mu_{12}(ij) = 0$, então $\mu_{123}(ijk) = 0$ e se $\mu_{12}(ij) \neq 0$, os termos $\mu_1(i)$ e $\mu_2(j)$ devem estar presentes no modelo. Neste trabalho apenas serão considerados modelos log-lineares hierárquicos.

3.3.4. Hipóteses de interesse

Numa tabela de contingência tridimensional são de interesse as seguintes hipóteses:

(i) Independência mútua completa entre as três variáveis.

$$H_0 : p_{ijk} = p_{i..} p_{.j.} p_{..k}, \quad \begin{cases} i = 1, 2, \dots, r \\ j = 1, 2, \dots, s \\ k = 1, 2, \dots, t \end{cases}$$

(ii) Independência parcial. Quando uma variável é independente das diferentes combinações das outras duas variáveis conjuntamente.

$$H_0 : p_{ijk} = p_{i..} p_{.jk}, \quad \begin{cases} i = 1, 2, \dots, r \\ j = 1, 2, \dots, s \\ k = 1, 2, \dots, t \end{cases}$$

A probabilidade correspondente a uma particular combinação das variáveis S e T, é dada pela probabilidade total marginal $p_{.jk}$. A variável R é completamente independente de S e T, mas, as variáveis S e T estão associadas. Há três hipóteses deste tipo.

(iii) Independência condicional. Quando duas variáveis são independentes, dentro de cada categoria ou nível da terceira variável.

$$H_0 \quad p_{ijk} = \frac{p_{i.k} p_{.jk}}{p_{..k}}, \quad \begin{cases} i = 1, 2, \dots, r \\ j = 1, 2, \dots, s \\ k = 1, 2, \dots, t \end{cases}$$

As variáveis \underline{R} e \underline{S} são independentes para cada nível da variável \underline{T} , mas cada uma delas está associada com a variável \underline{T} , ou seja que as variáveis \underline{R} e \underline{S} são independentes condicionalmente dado um nível da variável \underline{T} . Há três hipóteses deste tipo.

(iv) Não interação de segunda ordem. Quando alguma medida de associação entre qualquer par de variáveis é constante nos níveis da terceira variável. ROY & KASTENBAUM (1956), generalizaram os critérios de BARLETT (1935) e estabeleceram que a hipótese de não interação de segunda ordem, implica

$$H_0 : \frac{p_{rst} p_{ijjt}}{p_{ist} p_{rjt}} = \frac{p_{rsk} p_{ijjk}}{p_{isk} p_{rjk}}, \text{ para } \begin{cases} k = 1, 2, \dots, (t-1) \\ j = 1, 2, \dots, (s-1) \\ i = 1, 2, \dots, (r-1) \end{cases}$$

Isso indica que há $(t-1)(s-1)(r-1)$ restrições sobre os p_{ijk} e que contrastes entre logaritmos das probabilidades das caselas para cada sub-tabela $2 \times 2 \times 2$ são todos iguais a zero.

3.3.5. Estimativas de máxima verossimilhança

Considerando a hipótese (i) de independência mútua completa, tem-se que, sob H_0 , nenhuma das configurações bidimensionais apresentam interação, isto é, $\mu_{12}(ij) = \mu_{13}(ik) = \mu_{23}(jk) = 0$. Portanto, o modelo a ser testado é $\log p_{ijk} = \mu + \mu_1(i) + \mu_2(j) + \mu_3(k) \quad \therefore$

$$p_{ijk} = e^{\mu + \mu_1(i) + \mu_2(j) + \mu_3(k)} \quad \text{Sabe-se que}$$

$$p_{i..} = \sum_{j,k} p_{ijk} = e^{\mu + \mu_1(i)} \sum_{j,k} e^{\mu_2(j) + \mu_3(k)} \quad (1),$$

$$p_{.j.} = \sum_{i,k} p_{ijk} = e^{\mu + \mu_2(j)} \sum_{i,k} e^{\mu_1(i) + \mu_3(k)} \quad (2),$$

$$p_{..k} = \sum_{i,j} p_{ijk} = e^{\mu + \mu_3(k)} \sum_{i,j} e^{\mu_1(i) + \mu_2(j)} \quad (3) \text{ e}$$

$$p_{...} = \sum_{i,j,k} p_{ijk} = e^{\mu} \sum_{i,j,k} e^{\mu_1(i) + \mu_2(j) + \mu_3(k)} = 1 \quad (4).$$

Verifica-se então que $p_{ijk} = (1) (2) (3) = p_{i..} p_{.j.} p_{..k}$.
Portanto, no modelo de independência mútua completa a hipótese a ser testada é $H_0 : p_{ijk} = p_{i..} p_{.j.} p_{..k} \quad (5)$.

No item 3.2., demonstrou-se que

$$m_{ijk} = E(X_{ijk}) = N p_{ijk} \therefore p_{ijk} = \frac{m_{ijk}}{N}.$$

BIRCH (1963), demonstrou que sob H_0 , as estimativas de máxima verossimilhança das marginais unidimensionais são as marginais unidimensionais das frequências observadas. Logo, (5), pode ser expressa como sendo

$$H_0 : \frac{m_{ijk}}{N} = \frac{m_{i..}}{N} \frac{m_{.j.}}{N} \frac{m_{..k}}{N} \quad \therefore$$

$$H_0 : m_{ijk} = \frac{m_{i..} m_{.j.} m_{..k}}{N^2}.$$

As estimativas de máxima verossimilhança para m_{ijk} , neste modelo, são

$$\hat{m}_{ijk} = \frac{x_{i..} \cdot x_{.j.} \cdot x_{..k}}{N^2} \quad (6)$$

É conveniente ter presente que sob H_0 , cum prem-se as seguintes relações:

$$\hat{m}_{i..} = \sum_{j,k} \hat{m}_{ijk} = \sum_{j,k} \frac{x_{i..} \cdot x_{.j.} \cdot x_{..k}}{N^2} = x_{i..},$$

$$\hat{m}_{.j.} = \sum_{i,k} \hat{m}_{ijk} = \sum_{i,k} \frac{x_{i..} \cdot x_{.j.} \cdot x_{..k}}{N^2} = x_{.j.}, \quad e$$

$$\hat{m}_{..k} = x_{..k}.$$

Isto é, a tabela de contingência construída com as freqüências estimadas \hat{m}_{ijk} , sob H_0 , cumpre com a restrição de que as freqüências marginais unidimensionais são iguais às freqüências marginais unidimensionais da tabela de contingência das freqüências observadas.

Observa-se que em geral se verifica que $\hat{m}_{ij.} \neq x_{ij.}$; $\hat{m}_{i.k} \neq x_{i.k}$ e $\hat{m}_{.jk} \neq x_{.jk}$. Por exemplo,

$$\hat{m}_{.jk} = \sum_i \hat{m}_{ijk} = \frac{x_{.j.} \cdot x_{..k}}{N^2} \sum_i x_{i..} \dots \quad \hat{m}_{.jk} = \frac{x_{.j.} \cdot x_{..k}}{N}.$$

Essa última expressão não necessariamente é igual a $x_{.jk}$.

Procedimento análogo pode ser seguido para se obter as estimativas de máxima verossimilhança das freqüências esperadas sob as hipóteses (ii) e (iii). Porém, no caso da hipótese (iv) não pode-se obter diretamente u-

ma estimativa para m_{ijk} em função das marginais bidimensionais. Uma alternativa para obtê-las, é usar um processo iterativo. Esse assunto será discutido posteriormente.

3.3.6. Resumo dos resultados obtidos no ítem 3.3.

Para a análise de dados categorizados, tem-se usado o método da máxima verossimilhança para estimar os parâmetros do modelo log-linear admitido como sendo:

$$\log p_{ijk} = +\mu_1(i) + \mu_2(j) + \mu_3(k) + \mu_{12}(ij) + \mu_{13}(ik) + \mu_{23}(jk) +$$

$$\mu_{123}(ijk), \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, s$$

$$\text{e } k = 1, 2, \dots, t, \quad \text{com } p_{ijk} > 0 \text{ e } \sum_{ijk} p_{ijk} = 1.$$

Pode-se associar a cada hipótese de interesse um modelo log-linear não saturado, isto é, um modelo com menos parâmetros que o número total de caselas da tabela de contingência.

Tem-se, então, o seguinte quadro:

Quadro 2. Modelos associados a hipóteses de interesse em tabelas de contingência tridimensionais.

Modelo	Hipótese testada
(i) $\mu + \mu_1(i) + \mu_2(j) + \mu_3(k)$	Independência mútua completa
(ii) $\mu + \mu_1(i) + \mu_2(j) + \mu_3(k) + \mu_{23}(jk)$	Independência parcial
Há três hipóteses deste tipo	
(iii) $\mu + \mu_1(i) + \mu_2(j) + \mu_3(k) + \mu_{23}(jk) + \mu_{13}(ik)$	Independência condicional
Há três hipóteses deste tipo	
(iv) $\mu + \mu_1(i) + \mu_2(j) + \mu_3(k) + \mu_{23}(jk) + \mu_{13}(ik) + \mu_{12}(ij)$	Não interação de segunda ordem

Cada uma dessas hipóteses leva implicitamente algumas restrições sobre as probabilidades marginais unidimensionais e/ou bidimensionais.

KULLBACK (1959), KU & KULLBACK (1968) e KU *et alii* (1971), principalmente, têm usado o princípio da informação mínima de discriminação para estimar frequências esperadas em amostras de populações distribuídas multinomialmente, assim como também a sua aplicação em tabelas de contingência. Um dos aspectos interessantes desse método é que não se pressupõe *a priori* um determinado modelo para testar uma hipótese. O modelo log-linear será obtido a partir das restrições impostas às frequências unidimensionais e/ou bidimensionais da tabela de contingência observada.

Através do método da informação mínima de discriminação será abordado nos próximos itens o relacionado com a formulação, interpretação e testes de hipóteses de interesse em tabelas de contingência tridimensionais, numa linguagem acessível aos pesquisadores da área agrônômica, visando a sua posterior aplicação.

3.4. Teoria da Informação

3.4.1. Introdução

Estudar-se-ã a teoria da informação como um ramo da teoria matemática das probabilidades e da estatística. A informação contida numa mensagem acerca da ocorrência de um evento está desprovida de qualquer aspecto emocional da mensagem. O conceito será dado apenas em função da probabilidade p de que um evento E possa ocorrer antes da chegada da mensagem fidedigna e confiável que o evento E realmente ocorreu. O conteúdo de informação $h(p)$ de uma mensagem é uma função decrescente da probabilidade p , isto é, quanto menor for a probabilidade de um evento antes da mensagem que ele ocorreu, maior será o conteúdo de informação dessa mensagem.

Uma função decrescente e que apresenta algumas outras vantagens é:

$$h(p) = \log \frac{1}{p} ; \quad 0 < p \leq 1 ; \quad 0 \leq h(p) < \infty \quad (1)$$

Essa é uma função monótona decrescente no intervalo $(0, 1]$, como pode ser ilustrado esquematicamente na Figura 1.

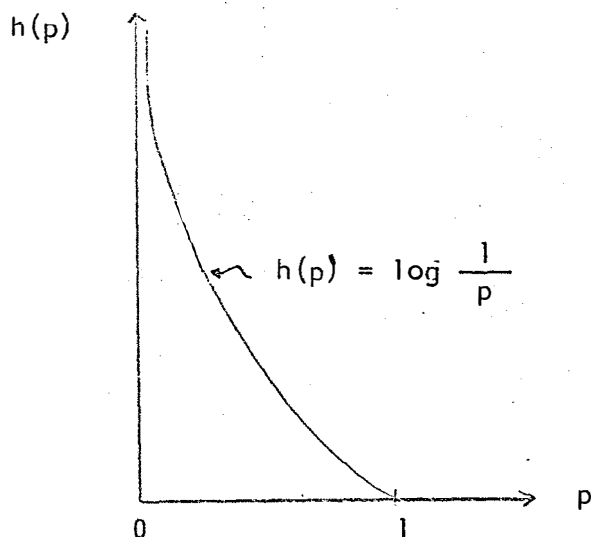


Figura 1. Conteúdo de informação $h(p)$ definida na equação (1).

Segundo THEIL (1967), a definição logaritmica da informação é o resultado de alguns axiomas:

(i) Deseja-se definir a informação como uma função dependente unicamente da probabilidade p .

(ii) Deseja-se que $h(p)$ seja uma função contínua no intervalo $0 < p \leq 1$.

(iii) Admite-se infinita surpresa quando é recebida uma mensagem afirmando que um evento \underline{E} ocorreu quando ele tinha probabilidade zero de acontecer; e admite-se zero surpresa quando é recebida uma mensagem afirmando que um evento \underline{E} ocorreu quando ele tinha probabilidade 1 de acontecer; isto é, o conteúdo de informação $h(p)$, nesses casos é $h(0) = \infty$ e $h(1) = 0$.

(iv) Baseado em (iii), $h(p)$ é uma função monotona decrescente. Em símbolos:

$$h(p_1) > h(p_2) \iff p_1 < p_2$$

(v) Deseja-se que o conteúdo de informação seja aditivo no caso de eventos independentes, isto é, $h(p_1 p_2) = h(p_1) + h(p_2)$.

É fácil verificar que a definição logarítmica da informação, ou seja, $h(p) = \log 1/p$, satisfaz a esses cinco axiomas. Por exemplo, o axioma (v) estabelece que $h(p_1 p_2) = h(p_1) + h(p_2)$, $0 < p_1, p_2 \leq 1$. Se dois eventos E_1 e E_2 são eventos independentes com probabilidades p_1 e p_2 , respectivamente, então $P(E_1 \cap E_2) = P(E_1 E_2) = p_1 p_2$; logo

$$h(p_1 p_2) = \log \frac{1}{p_1 p_2} = \log \frac{1}{p_1} + \log \frac{1}{p_2} = h(p_1) + h(p_2).$$

Conclui-se que $h(p)$ como definida em (1) é apropriada e que ela é única quando é referida a algum número positivo como a base logarítmica. Neste trabalho considera-se o número e como a base logarítmica, isto é, os logaritmos naturais.

3.4.1.1. Informação esperada

Seja um conjunto de n eventos E_1, E_2, \dots, E_n , mutuamente exclusivos e exaustivos (FELLER, 1968), com probabilidades p_1, p_2, \dots, p_n , respectivamente tal que $p_i > 0$, $\sum_{i=1}^n p_i = 1$; $i = 1, 2, \dots, n$. Quando é recebida uma mensagem de que E_i ocorreu, então o conteúdo de informação da mensagem é $h(p_i) = \ln 1/p_i$. Antes de receber a mensagem não se conhece a quantia da informação

por quanto pode ser qualquer dos números $h(p_1), h(p_2), \dots, h(p_n)$, porém, pode-se calcular o valor da informação esperada antes de que a mensagem tenha chegado. Como E_i tem probabilidade p_i de ocorrer então a mensagem indicando que E_i ocorreu, será recebida com a mesma probabilidade, de tal forma que o conteúdo de informação será $h(p_i)$ com probabilidade p_i e o valor esperado de informação é

$$H(p) = \sum_{i=1}^n p_i h(p_i) = \sum_{i=1}^n p_i \ln \frac{1}{p_i} = - \sum_{i=1}^n p_i \ln p_i.$$

Isto é, o valor esperado de informação é a Esperança dos logaritmos das probabilidades com sinal negativo.

$H(p)$ não pode ser negativa dado que é uma média ponderada de valores individuais de informação $h(p_i)$ que são não negativos e cujos pesos são positivos. Observe que se $p_i = 0$, o produto $p_i \ln p_i$ é da forma $0 \times (-\infty)$, que não está definido. Admite-se então que $p_i \ln p_i = 0$, se $p_i = 0$.

Dado que $H(p) \geq 0$, tem-se que o valor mínimo de $H(p)$ é zero, valor que é atingido quando $p_i = 1$, para algum i e $p_j = 0$, para todo $j \neq i$. Isto é, nenhuma informação deve ser esperada de uma mensagem sobre a realização de um entre n eventos, quando é conhecido com antecedência que algum E_i tem probabilidade 1. Para achar o maior valor que $H(p)$ pode tomar, maximiza-se essa função sujeita à restrição que $\sum p_i = 1$. Tem-se então que maximizar a função

$$Z(p_i, \lambda) = - \sum_{i=1}^n p_i \ln p_i + \lambda \left(\sum_{i=1}^n p_i - 1 \right), \text{ logo}$$

$$\frac{\partial Z}{\partial p_i} = -1 - \ln \hat{p}_i + \lambda = 0, \quad i = 1, 2, \dots, n \quad (1)$$

$$\frac{\partial Z}{\partial \lambda} = \sum_{i=1}^n \hat{p}_i - 1 = 0 \quad \therefore \quad \sum_{i=1}^n \hat{p}_i = 1 \quad (2)$$

$$\text{Em (1)} \quad \ln \hat{p}_i = \lambda - 1 \quad \therefore \quad \hat{p}_i = e^{\lambda - 1} \quad \therefore$$

$$\sum_{i=1}^n \hat{p}_i = n e^{\lambda - 1} \quad \therefore \quad e^{\lambda - 1} = \frac{1}{n} \quad \therefore$$

$$\lambda - 1 = \ln \frac{1}{n} \quad \therefore \quad \lambda = 1 - \ln n \quad (3)$$

Substituindo-se (3) em (1), tem-se que

$$-1 - \ln \hat{p}_i + 1 - \ln n = 0 \quad \therefore$$

$$\ln \hat{p}_i = \ln \frac{1}{n} \quad \therefore \quad \hat{p}_i = \frac{1}{n} \quad (4)$$

Isto é, a informação esperada assume seu máximo valor quando todos os eventos têm a mesma probabilidade. Esse valor é

$$H(p) = - \sum_{i=1}^n \frac{1}{n} \ln \frac{1}{n} = \ln n, \text{ logo}$$

$$0 \leq H(p) \leq \ln n \quad (5)$$

3.4.1.2. Alguns comentários

O conteúdo de informação de uma mensagem fidedigna, afirmando que um evento E_i ocorreu, foi definido como

$$h(p_i) = \ln \frac{1}{p_i} = \ln 1 - \ln p_i = - \ln p_i \quad (1)$$

Evidentemente, depois que o evento ocorreu, a probabilidade do evento E_i é 1. A diferença $\ln 1 - \ln p_i$, é definida como a informação recebida com a mensagem, que, em geral é definida como, $I = \ln p'_i - \ln p_i$, (2) isto é, $I = \ln p'_i/p_i$, onde p'_i é a probabilidade de E_i *a posteriori* da mensagem e p_i é a probabilidade *a priori* da mensagem. Observe que em (1) a informação recebida com a mensagem é igual ao conteúdo da mensagem. A equação (2) pode ser expressa como

$$I = \ln \frac{p'_i}{p_i} = h(p_i) - h(p'_i) \quad (3)$$

Seja por exemplo um experimento que consiste em jogar um dado e seja o evento $E_1 = \{\text{"obter um número par"}\}$.

Neste caso, o espaço amostral $\Omega = \{1, 2, 3, 4, 5, 6\}$; $E_1 = \{2, 4, 6\}$; $\bar{E}_1 = E_2 = \{1, 3, 5\}$; $E_1 \cup E_2 = \Omega$; $p(E_1) = 1/2$; $p(E_2) = 1/2$; $p(E_1) + p(E_2) = 1$.

Considere que depois de realizado o experimento é recebida a seguinte mensagem:

Caso 1: "Foi obtido um número maior ou igual a 4"

Tem-se então que $p(E_1) = p_1 = 1/2 \implies h(p_1) =$

$$= \ln 2, \text{ e } p'(E_1) = p'_1 = 2/3 \implies h(p'_1) = \ln 3/2 \text{ e}$$

$$I = h(p_1) - h(p'_1) = \ln 2 - \ln 3/2 > 0.$$

Houve ganho de informação depois da mensagem.

Caso 2: "Foi obtido um número menor ou igual a 2"

$$\text{Então, } p_1 = 1/2 \implies h(p_1) = \ln 2, \text{ e}$$

$$p'_1 = 1/2 \implies h(p'_1) = \ln 2, \text{ logo}$$

$$I = h(p_1) - h(p'_1) = \ln 2 - \ln 2 = 0.$$

Não houve ganho de informação depois da mensagem.

Caso 3: "Foi obtido um número primo"

$$\text{Então, } p_1 = 1/2 \implies h(p_1) = \ln 2 \text{ e}$$

$$p'_1 = 1/4 \implies h(p'_1) = \ln 4, \text{ logo}$$

$$I = h(p_1) - h(p'_1) = \ln 2 - \ln 4 < 0.$$

Houve perda de informação depois da mensagem.

Observe que as mensagens M_i , $i = 1, 2, 3$, podem ser consideradas como eventos, as probabilidades *a priori* podem ser consideradas como probabilidades incondicionais e as probabilidades *a posteriori* podem ser consideradas como probabilidades condicionais. Tem-se então que:

Caso 1: $E_1 = \{2, 4, 6\}$; $M_1 = \{4, 5, 6\}$,

$$E_1 \cap M_1 = \{4, 6\}; \quad P(E_1) = \frac{1}{2}; \quad P(M_1) = \frac{1}{2} \quad e$$

$$P(E_1 \cap M_1) = \frac{1}{3} \neq P(E_1) P(M_1) = \frac{1}{4}. \quad \text{Sabe-se que}$$

$$P(E_1 | M_1) = \frac{P(E_1 \cap M_1)}{P(M_1)} = \frac{2}{3} > P(E_1) = \frac{1}{2}$$

Caso 2: $E_1 = \{2, 4, 6\}$; $M_2 = \{1, 2\}$;

$$E_1 \cap M_2 = \{2\}; \quad P(E_1) = \frac{1}{2}; \quad P(M_2) = \frac{1}{3} \quad e$$

$$P(E_1 \cap M_2) = \frac{1}{6} = P(E_1) P(M_2). \quad \text{Sabe-se que}$$

$$P(E_1 | M_2) = \frac{P(E_1 \cap M_2)}{P(M_2)} = \frac{1}{2} = P(E_1)$$

Caso 3: $E_1 = \{2, 4, 6\}$; $M_3 = \{1, 2, 3, 5\}$;

$$E_1 \cap M_3 = \{2\}; \quad P(E_1) = \frac{1}{2}; \quad P(M_3) = \frac{2}{3};$$

$$P(E_1 \cap M_3) = \frac{1}{6} \neq P(E_1) P(M_3) = \frac{1}{3}. \quad \text{Sabe-se que}$$

$$P(E_1 | M_3) = \frac{P(E_1 \cap M_3)}{P(M_3)} = \frac{1}{4} < P(E_1) = \frac{1}{2}.$$

Tem-se então que a informação sobre um evento E recebida de uma mensagem pode ser positiva, nula ou negativa. No caso em que a probabilidade *a priori* ou incondicional é igual à probabilidade *a posteriori* ou con

dicional, não se obtém nenhuma informação adicional sobre o evento \underline{E} .

No apêndice 1 é demonstrada uma propriedade interessante da probabilidade condicional conhecida como Regra de Bayes, que será de muita utilidade nos próximos itens.

Na estatística experimental em geral e na experimentação agronômica em particular, quando se faz observações de realizações de uma ou várias variáveis aleatórias está-se procurando informação. Isto é, através de um conjunto particular de observações deseja-se inferir sobre algumas características ou parâmetros da população amostrada.

Considere novamente a expressão (3), $I = \ln(p'_{ij}/p_j)$. No caso de uma tabela de contingência tridimensional, p'_{ij} pode ser a frequência relativa observada na casela \underline{i} , já que é *a posteriori*. p_j pode ser considerada como a probabilidade esperada na casela \underline{i} , sob uma determinada hipótese H_0 , já que é uma probabilidade *a priori*.

3.4.2. Informação de discriminação

No ítem 3.4.1. apresentou-se uma abordagem intuitiva e axiomática do conceito de conteúdo de informação de uma determinada mensagem. Agora esses conceitos vão ser formalizados procurando a sua aplicação na experimentação agronômica.

3.4.2.1. Definição de informação

Considere os espaços de probabilidades (Ω, β, P_1) e (Ω, β, P_2) , isto é, tem-se um conjunto de elementos $x \in \Omega$, um campo de Borel β , que é uma coleção de subconjuntos do espaço amostral Ω e P_1 e P_2 são duas medidas de probabilidade para todos os conjuntos ou eventos do campo de Borel β . Os elementos de Ω podem ser as possíveis amostras de tamanho N de uma população multinomial, β pode ser um campo de Borel do espaço euclidiano n -dimensional e P_1 e P_2 podem definir as probabilidades das diferentes amostras para diferentes valores dos parâmetros da população. Admite-se que a medida de probabilidade P_1 é absolutamente contínua com respeito a P_2 , isto é, $P_1(E) = 0$ para todo conjunto mensurável $E \in \beta$ para o qual $P_2(E) = 0$.

Seja \underline{X} uma variável aleatória com função de massa de probabilidade $f_1(x)$ quando \underline{X} assume valores em uma população 1 e $f_2(x)$ quando \underline{X} assume valores em uma população 2. Considere uma única população constituída pelos elementos das populações 1 e 2, e seja \underline{x} o valor observado da variável \underline{X} selecionado ao acaso nessa nova população.

Se H_i , $i = 1; 2$, é a hipótese que \underline{X} é da população estatística com medida de probabilidade P_i , então pela regra de Bayes sobre probabilidade condicional (ver Apêndice 1), tem-se que

$$P(H_i | x) = \frac{P(X = x | H_i) P(H_i)}{P(H_1) P(X = x | H_1) + P(H_2) P(X = x | H_2)}, \quad i = 1; 2 \quad (1)$$

onde $P(X = x | H_i) = f_i(x)$, $i = 1; 2$, logo

$$P(H_i | x) = \frac{f_i(x) P(H_i)}{P(H_1) f_1(x) + P(H_2) f_2(x)}, \quad i = 1; 2 \quad (2)$$

$$\text{Portanto } \frac{P(H_1 | x)}{P(H_2 | x)} = \frac{f_1(x) P(H_1)}{f_2(x) P(H_2)}$$

$$\frac{f_1(x)}{f_2(x)} = \frac{P(H_1 | x) P(H_2)}{P(H_2 | x) P(H_1)} \quad (3)$$

onde $P(H_i)$, $i = 1; 2$ é a probabilidade *a priori* de H_i e $P(H_i | x)$ é a probabilidade *a posteriori* de H_i ou a probabilidade condicional de H_i dado $X = x$. Tomando-se logaritmos neperianos a ambos os membros de (3), tem-se

$$\ln \frac{f_1(x)}{f_2(x)} = \ln \frac{P(H_1 | x)}{P(H_2 | x)} - \ln \frac{P(H_1)}{P(H_2)} \quad (4)$$

O segundo membro da equação (4) é uma medida da diferença entre os logaritmos dos *odds* (*chances*) em favor de H_1 depois da observação $X = x$ e antes dessa mesma observação. Essa diferença pode ser positiva, negativa ou zero.

O logaritmo da razão de verossimilhança $\ln f_1(x)/f_2(x)$, é definido como a informação contida em $X = x$ para discriminar em favor de H_1 contra H_2 e é representada por

$$I(1 : 2; x) = \ln \frac{f_1(x)}{f_2(x)} \quad (5)$$

Exemplo: Seja \underline{X} uma variável aleatória dis-
tribuída trinomialmente na população i com função de mas-
sa de probabilidade

$$f_i (X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{N!}{x_1! x_2! x_3!} p_{1i}^{x_1} p_{2i}^{x_2} p_{3i}^{x_3},$$

$$i = 1; 2; p_{1i} + p_{2i} + p_{3i} = 1 \quad \text{e} \quad x_1 + x_2 + x_3 = N,$$

$$\text{então } I(1 : 2; \underline{x}) = \ln \frac{f_1(\underline{x})}{f_2(\underline{x})} = \ln \frac{p_{11}^{x_1} p_{21}^{x_2} p_{31}^{x_3}}{p_{12}^{x_1} p_{22}^{x_2} p_{32}^{x_3}} \dots$$

$$I(1 : 2; \underline{x}) = x_1 \ln \frac{p_{11}}{p_{12}} + x_2 \ln \frac{p_{21}}{p_{22}} + x_3 \ln \frac{p_{31}}{p_{32}}$$

A medida de informação $I(1 : 2; \underline{x})$ definida em (5) depende de cada observação \underline{x} . Ao considerar todos os valores que a variável aleatória \underline{X} pode assumir, define-se informação média por observação para discriminar em favor de H_1 versus H_2 dado $x \in \Omega \in \beta$, por P_1 , como

$$I(1 : 2; \Omega) = I(1 : 2) = \int f_1(\underline{x}) \ln \frac{f_1(\underline{x})}{f_2(\underline{x})} d\underline{x} \quad (6)$$

quando \underline{X} é uma variável contínua e

$$I(1 : 2) = \sum_{i=1}^N P_1(X = x_i) \ln \frac{P_1(X = x_i)}{P_2(X = x_i)}, \quad i = 1, 2, \dots, N$$

quando \underline{X} é uma variável aleatória discreta. Obviamente \underline{X} pode ser um vetor de variáveis aleatórias.

No exemplo anterior, tem-se então que

$$l(1 : 2) = \sum_{x_1=0}^N \sum_{x_2=0}^{N-x_1} \frac{N!}{x_1! x_2! x_3!} p_{11}^{x_1} p_{21}^{x_2} p_{31}^{x_3} \ln \frac{p_{11}^{x_1} p_{21}^{x_2} p_{31}^{x_3}}{p_{12}^{x_1} p_{22}^{x_2} p_{32}^{x_3}},$$

onde $p_{1i} + p_{2i} + p_{3i} = 1$; $i = 1, 2$ e $x_3 = N - x_1 - x_2$, \therefore

$$l(1 : 2) = \sum_{x_1=0}^N \sum_{x_2=0}^{N-x_1} f_1(\underline{x}) \left[x_1 \ln \frac{p_{11}}{p_{12}} + x_2 \ln \frac{p_{21}}{p_{22}} + x_3 \ln \frac{p_{31}}{p_{32}} \right] \therefore$$

$$l(1 : 2) = \ln \frac{p_{11}}{p_{12}} \sum_{x_1=0}^N \sum_{x_2=0}^{N-x_1} x_1 f_1(\underline{x}) + \dots + \ln \frac{p_{31}}{p_{32}} \sum_{x_1=0}^N \sum_{x_2=0}^{N-x_1} x_3 f_1(\underline{x})$$

onde sabe-se que

$$\sum_{x_1=0}^N \sum_{x_2=0}^{N-x_1} x_j f_1(\underline{x}) = E_1(X_c) = N p_{c1}, \quad c = 1, 2, 3, \quad \text{logo}$$

$$l(1 : 2) = N \left[p_{11} \ln \frac{p_{11}}{p_{12}} + p_{21} \ln \frac{p_{21}}{p_{22}} + p_{31} \ln \frac{p_{31}}{p_{32}} \right] \therefore$$

$$l(1 : 2) = N \sum_{c=1}^3 p_{c1} \ln \frac{p_{c1}}{p_{c2}} \quad (7)$$

3.4.2.2. Propriedades da informação

No ítem 3.4.1. demonstraram-se algumas das seguintes propriedades:

(i) $I(1 : 2)$ é não negativa, isto é $I(1 : 2) \geq 0$, verificando-se a igualdade se e somente se $f_1(x) = f_2(x)$. Não há nenhuma informação de discriminação se as distribuições das observações são as mesmas sob ambas as hipóteses.

(ii) $I(1 : 2)$ é aditiva para eventos aleatórios independentes, isto é, para \underline{X} e \underline{Y} independentes sob H_i , $i = 1; 2$, tem-se que $I(1 : 2; X, Y) = I(1 : 2; X) + I(1 : 2; Y)$.

Aditividade da informação para eventos independentes é intuitivamente uma condição fundamental. Uma amostra de \underline{N} observações independentes da mesma população fornece \underline{N} vezes a informação média de uma única observação. Isto é, se X_1, X_2, \dots, X_n I.I.D. $f(x)$, então

$$I(1 : 2; X_1, X_2, \dots, X_n) = I(1 : 2; X_1) + I(1 : 2; X_2) + \dots + I(1 : 2; X_n) \quad \dots$$

$$I(1 : 2; \underline{X}) = N I(1 : 2; X_1) \quad (8)$$

(iii) Se $E_i \in \beta$, $i = 1, 2, \dots$; $E_i \cap E_j = \phi$, $i \neq j$ e $\Omega = \cup_i E_i$, isto é, quando se faz uma partição de Ω em eventos E_1, E_2, \dots , tal que sejam disjun-

tos dois a dois, tem-se que

$$I(1 : 2) \geq \sum P_1(E_i) \ln \frac{P_1(E_i)}{P_2(E_i)}, \quad \text{para } x \in E_i, i = 1, 2, \dots,$$

verificando-se a igualdade se e somente se (KULLBACK, 1959)

$$\frac{f_1(x)}{f_2(x)} = \frac{P_1(E_i)}{P_2(E_i)}, \quad \text{para } x \in E_i, \quad i = 1, 2, \dots, \text{ ou}$$

$$\frac{f_1(x)}{P_1(E_i)} = \frac{f_2(x)}{P_2(E_i)}, \quad \text{para } x \in E_i, \quad i = 1, 2, \dots \quad (9)$$

A expressão (9) implica que uma condição necessária e suficiente para que a informação não seja diminuída devido ao agrupamento de observações em eventos E_i é que as densidades condicionais de \underline{X} dado E_i sejam as mesmas sob ambas as hipóteses.

3.4.2.3. Informação e suficiência

Se a partição do espaço amostral na propriedade (iii) do ítem 3.4.2.2., é tal que a condição necessária e suficiente para a igualdade é satisfeita, isto é, que a densidade condicional de \underline{X} dado E_i seja a mesma sob ambas as hipóteses para todo E_i da partição, então diz-se que a partição de $\Omega = \cup_i E_i, i = 1, 2, \dots$, é uma partição suficiente para discriminação. KULLBACK (1959), baseado no fato de que uma estatística é uma partição do espaço amostral Ω , estabelece que uma estatística

ca é suficiente para discriminação se a condição necessária e suficiente (9) para a igualdade é satisfeita.

Seja $Y = T(x)$ uma estatística, tal que é possível fazer uma transformação mensurável do espaço de probabilidades (Ω, β, P_i) sobre o espaço de probabilidades (γ, σ, μ_i) , $i = 1; 2$. KULLBACK (1959) demonstrou o seguinte teorema:

Teorema: $I(1 : 2; \Omega) \geq I(1 : 2; \gamma)$, com a igualdade prevalecendo se e somente se

$$\frac{f_1(x)}{f_2(x)} = \frac{g_1(T(x))}{g_2(T(x))} \quad \text{ou equivalentemente} \quad \frac{f_1(x)}{g_1(y)} = \frac{f_2(x)}{g_2(y)} \quad (10),$$

onde $g_i(y)$ é a Esperança condicional de $f_i(x)$ dado $y = T(x)$, isto é, $g_i(y) = E[f_i(x) | Y]$, $i = 1; 2$. Uma estatística que satisfaz a condição (10) é chamada uma estatística suficiente para discriminação. O teorema anterior estabelece basicamente que quando se grupam ou se transformam observações através de uma estatística, em geral, implica uma perda de informação. Se a estatística é suficiente, não há nenhuma perda de informação.

3.4.3. Informação mínima de discriminação $I(* : 2)$

Viu-se que a informação para discriminar em favor de H_1 versus H_2 , por P_1 , é:

$$I(1 : 2) = \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx \quad (1)$$

Para aplicações posteriores, $f_2(x)$ será uma função de densidade associada com o conjunto de populações da hipótese nula e $f_1(x)$ é um elemento que pertence ao conjunto de densidades de probabilidades que está associado à hipótese alternativa.

Para uma $f_2(x)$ dada, procurar-se-á a função de densidade $f_1(x)$ que esteja mais próxima ou que tenha mais "semelhança" com a medida de probabilidade P_2 no sentido de minimizar a informação de discriminação $I(1 : 2)$. Como $I(1 : 2) \geq 0$, com a igualdade se e somente se $f_1(x) = f_2(x)$, é lógico que se deve impor algumas outras restrições adicionais sobre $f_1(x)$ para se obter uma outra medida de probabilidade diferente de $f_2(x)$.

Deseja-se então minimizar a informação de discriminação sujeita à restrição que $\int T(x) f_1(x) dx = \theta$, onde θ pode ser um parâmetro multidimensional populacional e $T(x)$ é uma estatística, que pela sua vez é um estimador imparcial de θ .

Usando multiplicadores de Lagrange, o problema consiste então em minimizar a função

$$Z = \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx - \tau \left[\int T(x) f_1(x) dx - \theta \right] + \lambda \left[\int f_1(x) dx - 1 \right] \quad (2)$$

KULLBACK (1959), demonstrou que o mínimo de (2) ocorre quando $\hat{f}_1(x) = f^*(x) = f_2(x) e^{-\tau T(x) - \lambda - 1}$ (3)

Integrando em ambos os membros de (3) tem-se que

$$\int f^*(x) dx = \int f_2(x) e^{\tau T(x)} e^{-\lambda-1} dx \quad \therefore$$

$$1 = e^{-\lambda-1} \int f_2(x) e^{\tau T(x)} dx \quad (4),$$

mas sabe-se que a função geradora de momentos de $T(x)$ é

$$E(e^{\tau T(x)}) = \int e^{\tau T(x)} f_2(x) dx = M_2(\tau) < \infty \quad (5),$$

$$\text{logo } 1 = e^{-\lambda-1} M_2(\tau) \quad (6)$$

$$\therefore e^{\lambda+1} = M_2(\tau) \quad (7) \text{ e}$$

$$\ln M_2(\tau) = \lambda + 1 \quad (8)$$

$$\text{Definindo } \int f^*(x) \ln \frac{f^*(x)}{f_2(x)} dx = I(* : 2),$$

e lembrando que, pela equação (3), $\frac{f^*(x)}{f_2(x)} = e^{\tau T(x) - \lambda - 1} \therefore$

$$\frac{f^*(x)}{f_2(x)} = e^{\tau T(x)} e^{-\lambda-1}, \text{ mas, por (6), sabe-se que } e^{-\lambda-1} =$$

$$\frac{1}{M_2(\tau)}, \text{ logo } \frac{f^*(x)}{f_2(x)} = \frac{e^{\tau T(x)}}{M_2(\tau)} \quad \therefore$$

$$\ln \frac{f^*(x)}{f_2(x)} = \tau T(x) - \ln M_2(\tau) \quad (9),$$

$$\text{logo, } I(* : 2) = \int f^*(x) [\tau T(x) - \ln M_2(\tau)] dx \quad \therefore$$

$$I(* : 2) = \int f^*(x) \tau T(x) dx - \ln M_2(\tau) \int f^*(x) dx \quad \therefore$$

$$I(* : 2) = \tau \int f^*(x) T(x) dx - \ln M_2(\tau) \dots$$

$$I(* : 2) = \tau \theta - \ln M_2(\tau) \quad (10),$$

onde $\theta = \int T(x) f^*(x) dx = \int T(x) f_2(x) e^{\tau T(x) - \lambda - 1} dx \dots$

$$\theta = \int \frac{T(x) f_2(x) e^{\tau T(x)}}{e^{\lambda+1}} dx, \text{ mas por (7),}$$

$$\theta = \int \frac{T(x) f_2(x) e^{\tau T(x)}}{M_2(\tau)} dx \dots$$

$$\theta = \frac{1}{M_2(\tau)} \int T(x) f_2(x) e^{\tau T(x)} dx, \quad (11),$$

mas por (5) $M_2(\tau) = \int f_2(x) e^{\tau T(x)} dx \dots$

$$\frac{\partial M_2(\tau)}{\partial \tau} = \int \frac{\partial [f_2(x) e^{\tau T(x)}]}{\partial \tau} dx \dots$$

$$\frac{\partial M_2(\tau)}{\partial \tau} = \int T(x) f_2(x) e^{\tau T(x)} dx \quad (12)$$

substituindo-se (12) em (11), tem-se que

$$\theta = \frac{1}{M_2(\tau)} \frac{\partial M_2(\tau)}{\partial \tau} \dots \theta = \frac{\frac{\partial M_2(\tau)}{\partial \tau}}{M_2(\tau)} = \frac{\partial \ln M_2(\tau)}{\partial \tau} \quad (13)$$

Com base nos resultados anteriores, enuncia-se o seguinte teorema (KULLBACK, 1959): "Se $f_1(x)$ e uma dada $f_2(x)$ são funções de densidade de probabilidade

do espaço mensurável (Ω, β) e $Y = T(x)$ é uma estatística tal que $\int T(x) f_1(x) dx = \theta$ e $M_2(\tau) = \int f_2(x) e^{\tau T(x)} dx$, existe para τ em algum intervalo, então cumpre-se que

$$I(1 : 2) \geq I(* : 2) = \theta\tau - \ln M_2(\tau);$$

$\theta = \frac{\partial}{\partial \tau} \ln M_2(\tau)$, com a igualdade se e somente se

$$f_1(x) = f^*(x) = \frac{e^{\tau T(x)} f_2(x)}{M_2(\tau)} \quad (14)$$

Posteriormente se estudará a relação de θ com valores amostrais observados e as implicações no teste de hipóteses estatísticas. Os valores amostrais serão usados para determinar a "semelhança" entre a amostra, como um possível elemento do conjunto de populações da hipótese alternativa, e a população da hipótese nula definida por $f_2(x)$, através de uma estimativa da informação mínima de discriminação.

Exemplo:

Seja $f_2(x_1, x_2) = \frac{N!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$, uma distribuição

trinomial para $N = 3$; $p_1 = 0.3$; $p_2 = 0.5$ e $T(x) = \underline{x}$.

Sabe-se que $M_2(\underline{\tau}) = (p_1 e^{\tau_1} + p_2 e^{\tau_2} + p_3)^3$. Logo

$$f^*(x_1, x_2) = \frac{e^{\underline{\tau}' \underline{x}} f_2(x_1, x_2)}{M_2(\underline{\tau})} = \frac{e^{\tau_1 x_1 + \tau_2 x_2} f_2(x_1, x_2)}{(p_1 e^{\tau_1} + p_2 e^{\tau_2} + p_3)^3} \dots$$

$$f^*(x_1, x_2) = \frac{3!}{x_1! x_2! x_3!} \frac{(p_1 e^{\tau_1})^{x_1} (p_2 e^{\tau_2})^{x_2} p_3^{x_3}}{(p_1 e^{\tau_1} + p_2 e^{\tau_2} + p_3)^3},$$

mas $x_1 + x_2 + x_3 = 3$, logo

$$f^*(x_1, x_2) = \frac{3!}{x_1! x_2! x_3!} (p_1^*)^{x_1} (p_2^*)^{x_2} (p_3^*)^{x_3}, \text{ onde}$$

$$p_1^* = \frac{p_1 e^{\tau_1}}{p_1 e^{\tau_1} + p_2 e^{\tau_2} + p_3}; \quad p_2^* = \frac{p_2 e^{\tau_2}}{p_1 e^{\tau_1} + p_2 e^{\tau_2} + p_3} \quad e$$

$$p_3^* = \frac{p_3}{p_1 e^{\tau_1} + p_2 e^{\tau_2} + p_3}.$$

Observe que $f^*(x_1, x_2)$ também é uma função de densidade trinomial.

Por (13), tem-se que $\underline{\theta} = \frac{\partial}{\partial \underline{\tau}} \ln M_2(\underline{\tau})$, logo

$$\underline{\theta} = \frac{\partial}{\partial \underline{\tau}} \ln (p_1 e^{\tau_1} + p_2 e^{\tau_2} + p_3)^3 \quad \dots$$

$$\theta_1 = \frac{3 p_1 e^{\tau_1}}{p_1 e^{\tau_1} + p_2 e^{\tau_2} + p_3} = 3 p_1^* \quad e$$

$$\theta_2 = \frac{3 p_2 e^{\tau_2}}{p_1 e^{\tau_1} + p_2 e^{\tau_2} + p_3} = 3 p_2^*.$$

Suponha que houvesse interesse em procurar

$$\theta_1 = E_1(T(X_1)) = E_1(X_1) = 1 \quad e \quad \theta_2 = E_1(T(X_2)) = E_1(X_2) = 1,$$

então

$$p_1^* = \frac{1}{3} \quad e \quad p_2^* = \frac{1}{3}.$$

Entre todas as funções densidades de pro-

babilidades $f_1(x_1, x_2)$, tal que $E_1(X_1) = 1$ e $E_1(X_2) = 1$, verifica-se que quando

$$f_1^*(x_1, x_2) = f^*(x_1, x_2) = \frac{3!}{x_1! x_2! x_3!} \left(\frac{1}{3}\right)^{x_1} \left(\frac{1}{3}\right)^{x_2} \left(\frac{1}{3}\right)^{x_3},$$

obtem-se o menor valor para $I(1 : 2)$. Neste exemplo, tem-se que

$$I(* : 2) = \sum_{x_1=0}^3 \sum_{x_2=0}^{3-x_1} f^*(x_1, x_2) \ln \frac{f^*(x_1, x_2)}{f_2(x_1, x_2)}, \quad \text{onde}$$

x_1	x_2	x_3	$f^*(x_1, x_2)$	$f_2(x_1, x_2)$	$f^*(x_1, x_2) \ln \frac{f^*(x_1, x_2)}{f_2(x_1, x_2)}$
0	0	3	0,0370	0,008	0,0567
0	1	2	0,1111	0,060	0,0684
0	2	1	0,1111	0,150	-0,0334
0	3	0	0,0370	0,125	-0,0450
1	0	2	0,1111	0,036	0,1252
1	1	1	0,2224	0,130	0,0468
1	2	0	0,1111	0,225	-0,0784
2	0	1	0,1111	0,054	0,0802
2	1	0	0,1111	0,135	-0,0216
3	0	0	0,0370	0,027	0,0117
			1,0000	1,000	0,21048

logo, $I(* : 2) = 0,21048$.

3.4.4. Estatística da informação3.4.4.1. Estimativa de $I(* : 2)$

No ítem anterior viu-se que a informação mínima de discriminação $I(* : 2)$ é o valor que minimiza a expressão

$$I(* : 2) = \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx, \text{ para uma dada } f_2(x) \text{ e to-}$$

da $f_1(x)$, tal que $\theta = \int T(x) f_1(x) dx$. Achou-se que o valor mínimo $I(* : 2) = \theta\tau - \ln M_2(\tau)$, é atingido quando

$$f_1(x) = f^*(x) = \frac{e^{\tau T(x)} f_2(x)}{M_2(\tau)} \quad (1),$$

$$M_2(\tau) = \int e^{\tau T(x)} f_2(x) dx \quad (2) \text{ e}$$

$$\theta = \frac{d}{d\tau} \ln M_2(\tau) \quad (3).$$

A função $f^*(x)$, dada pela expressão (1) é chamada distribuição conjugada de $f_2(x)$. Da expressão (3), é fácil verificar que τ é função de θ e se não há nenhum valor de τ de modo que $\theta = \frac{d}{d\tau} \ln M_2(\tau)$, a informação mínima de discriminação é zero. Pode-se escrever então que $I(* : 2) = \theta\tau(\theta) - \ln M_2(\tau(\theta))$ (4), onde $\tau(\theta)$ é o valor de τ para o qual

$$\frac{d \ln M_2(\tau)}{d\tau} = \theta.$$

Seja O_N uma amostra de N observações inde-

pendentes com função de densidade de probabilidade $f_2(x)$. Pode-se então estimar $I(* : 2)$ usando-se o valor observado de $T(x)$ na amostra 0_N , como uma estimativa de θ , isto é, $T(x) = \hat{\theta}(x)$ e também para se obter uma estimativa de τ , isto é, $\hat{\tau}(x) = \tau(\hat{\theta}(x))$. Então

$$T(x) = \hat{\theta}(x) = \left[\frac{d}{d\tau} \ln M_2(\tau) \right]_{\tau = \hat{\tau}(x) = \tau(\hat{\theta}(x))} \quad (5)$$

A estimativa de $I(* : 2)$ e, então

$$\hat{I}(* : 2; 0_N) = \hat{\theta}(x) \tau(\hat{\theta}(x)) - \ln M_2[\tau(\hat{\theta}(x))] \quad (6) \text{ ou}$$

$$\hat{I}(* : 2; 0_N) = \hat{\theta} \tau(\hat{\theta}) - \ln M_2(\tau(\hat{\theta})) \quad (7)$$

Definição: $\hat{I}(* : 2; 0_N)$ é a informação mínima de discriminação entre a população com função de densidade $f^*(x)$, dada por (1), com o valor do parâmetro θ sendo o mesmo valor de $\hat{\theta}$ da amostra, e a população com função de densidade $f_2(x)$. Devido a que $\hat{I}(* : 2; 0_N) \geq 0$, com a igualdade se e somente se $\hat{\theta}$ é igual ao valor do parâmetro da população com densidade de probabilidade $f_2(x)$, tem-se que, quanto maior for o valor de $\hat{I}(* : 2; 0_N)$, maior será a divergência entre a amostra e $f_2(x)$.

A estatística da informação $\hat{I}(* : 2; 0_N)$, dada pelas expressões (6) ou (7), foi calculada considerando apenas uma função de densidade de probabilidade $f_2(x)$ dada. Porém, em muitos problemas de aplicação $f_2(x)$ representa um elemento de um conjunto \underline{H} de funções de

densidade de probabilidade. Seja $\hat{I}(* : H)$, o valor mínimo de $\hat{I}(* : 2; 0_N)$, para toda $f_2(x) \in H$, isto é:

$$\hat{I}(* : H) = \min_{f_2 \in H} \hat{I}(* : 2; 0_N). \quad (3)$$

O valor de $\hat{I}(* : H)$ é uma medida da divergência entre a amostra e aquele membro da família de populações H , que tem mais "semelhança" com a amostra. Evidentemente que se o valor de $\hat{\theta}$ na amostra é igual ao parâmetro θ para algum dos membros de H , então de fato $\hat{I}(* : H) = 0$, isto é, a amostra não produz nenhuma informação para discriminar contra H .

A estimativa $\hat{I}(* : H)$ é definida como a estatística da informação mínima de discriminação (EIMD).

3.4.4.2. Distribuição de $\hat{I}(* : 2)$

Da equação (1) do item anterior, tem-se que

$$f^*(x) = \frac{e^{\tau T(x)} f_2(x)}{M_2(\tau)}, \quad \text{é uma função do parâmetro } \tau,$$

portanto, o estimador de máxima verossimilhança de τ é $\hat{\tau}(x)$. De fato,

$$\ln f^*(x) = \tau T(x) + \ln f_2(x) - \ln M_2(\tau) \quad \therefore \quad (1),$$

$$\frac{d}{d\tau} \ln f^*(x) = T(x) - \frac{d}{d\tau} \ln M_2(\tau). \quad \text{Igualando a zero,}$$

tem-se que $T(x) = \frac{d}{d\tau} \ln M_2(\tau)$ e o valor para o qual

$\frac{d}{d\tau} \ln f^*(x) = 0$ é único e dado por

$$T(x) = \left[\frac{d}{d\tau} \ln M_2(\tau) \right]_{\tau = \hat{\tau}(x)}$$

Substituindo-se o valor de $\hat{\tau}(x)$, na expressão (1), tem-se

$$\ln \max_{\tau} f^*(x) = \hat{\tau}(x) T(x) + \ln f_2(x) - \ln M_2(\hat{\tau}(x)) \quad \therefore$$

$$\ln \frac{\max_{\tau} f^*(x)}{f_2(x)} = \hat{\tau}(x) T(x) - \ln M_2(\hat{\tau}(x)) \quad (2),$$

então por (6), (7) e (8) do item 3.4.4.1., tem-se que

$$\ln \frac{\max_{\tau} f^*(x)}{f_2(x)} = \hat{T}(* : 2; 0_N) \quad (3) \text{ e}$$

$$\hat{T}(* : H) = \min_{f_2 \in H} \left[\ln \frac{\max_{\tau} f^*(x)}{f_2(x)} \right] \quad (4)$$

Obviamente (4) será mínima quando $f_2(x)$ for máxima, isto é,

$$\hat{T}(* : H) = \ln \frac{\max_{\tau} f^*(x)}{\max_{f_2 \in H} f_2(x)} \quad (5)$$

Admite-se que as populações de H são membros de uma família de funções exponenciais que pela sua vez é um subconjunto das funções exponenciais geradas por $f^*(x)$, (KULLBACK, 1959). Denotando o espaço dos parâmetros τ por Ω e um subconjunto de valores de τ correspon-

dentess a \underline{H} , por \underline{W} , tem-se que

$$\max_{f_2 \in H} f_2(x) = \max_{\tau \in W} f_2(x) = \max_{\tau \in W} f^*(x), \text{ logo (5) fica:}$$

$$\hat{T}(* : H) = \ln \frac{\max_{\tau \in \Omega} f^*(x)}{\max_{\tau \in W} f^*(x)} = - \ln \lambda \quad (6),$$

onde λ é a razão de verossimilhança de Neyman-Pearson, isto é;

$$\lambda = \frac{\max_{\tau \in W} f^*(x)}{\max_{\tau \in \Omega} f^*(x)}, \text{ logo } 2 \hat{T}(* : H) = - 2 \ln \lambda \quad (7)$$

WILKS (1938), demonstrou o seguinte teorema:

Teorema: Seja uma população com uma variável \underline{X} distribuída com a função de densidade $f(x; \theta_1, \theta_2, \dots, \theta_h)$, tal que existam estimadores ótimos para os θ 's. Seja \underline{H} a hipótese simples $H : \theta_i = \theta_{0i}; i = m + 1, m + 2, \dots, h$. Sob \underline{H} , a distribuição de $- 2 \ln \lambda$ é distribuída como um χ^2 com $(h-m)$ graus de liberdade, exceto para termos de ordem $n^{-1/2}$, onde

$$\lambda = \frac{P_W(0_N)}{P_\Omega(0_N)}$$

3.4.4.3. Distribuição conjugada da distribuição multinomial.

Seja $p(x) = p(x_1, x_2, \dots, x_c) =$

$$\frac{N!}{x_1! x_2! \dots x_c!} p_1^{x_1} p_2^{x_2} \dots p_c^{x_c} \quad (1),$$

onde $p_i > 0, i = 1, 2, \dots, c; \sum_{i=1}^c p_i = 1$ e $\sum_{i=1}^c x_i = N$.

Baseando-se nos resultados obtidos no ítem 3.4.3., pode-se afirmar que

$$l(p^*(x) : p(x)) = \sum_{x_1+x_2+\dots+x_c=N} p^*(x) \ln \frac{p^*(x)}{p(x)}$$

toma um valor mínimo quando

$$p^*(x) = \frac{e^{\tau_1 x_1 + \tau_2 x_2 + \dots + \tau_c x_c} p(x)}{(p_1 e^{\tau_1} + p_2 e^{\tau_2} + \dots + p_c)^N} \quad \text{ou}$$

$$p^*(x) = \frac{N!}{x_1! x_2! \dots x_c!} (p_1^*)^{x_1} (p_2^*)^{x_2} \dots (p_c^*)^{x_c} \quad (2),$$

$$\text{onde } p_i^* = \frac{p_i e^{\tau_i}}{(p_1 e^{\tau_1} + p_2 e^{\tau_2} + \dots + p_c)} \quad (3)$$

(ver exemplo do ítem 3.4.3., na página 65).

Pelas restrições conhecidas, tem-se que $E^*(X_i) = \theta_i = N p_i^*$, os τ 's são parâmetros reais e

$$\theta_i = \frac{\partial}{\partial \tau_i} \ln (p_1 e^{\tau_1} + p_2 e^{\tau_2} + \dots + p_c)^N.$$

A distribuição $p^*(x)$ é a distribuição conjugada da distribuição multinomial $p(x)$. Além disso, tem-se que:

$$\theta_i = N p_i^* = \frac{N p_i e^{\tau_i}}{(p_1 e^{\tau_1} + p_2 e^{\tau_2} + \dots + p_c)}, \quad \dots \quad (4)$$

$$e^{\tau_i} = \frac{\theta_i}{N p_i} (p_1 e^{\tau_1} + p_2 e^{\tau_2} + \dots + p_c) = \frac{\theta_i}{N p_i} K, \quad K > 0,$$

$$\tau_i = \ln \frac{\theta_i}{N p_i} + \ln K \quad (5),$$

$$I(* : 2; 0_N) = I(p^*(x) : p(x)) = \sum_{x_1+x_2+\dots+x_c=N} p^*(x) \ln \frac{p^*(x)}{p(x)} \quad (6),$$

$$\text{mas, } \ln \frac{p^*(x)}{p(x)} = \ln \frac{e^{\tau_1 x_1 + \tau_2 x_2 + \dots + \tau_c x_c} p(x)}{(p_1 e^{\tau_1} + p_2 e^{\tau_2} + \dots + p_c e^{\tau_c})^N \cdot p(x)}$$

$$\ln \frac{p^*(x)}{p(x)} = \tau_1 x_1 + \tau_2 x_2 + \dots + \tau_c x_c - N \ln K,$$

mas por (5) e lembrando que $E(X_i) = N p_i$, tem-se que

$$\ln \frac{p^*(x)}{p(x)} = \tau_1 \theta_1 + \tau_2 \theta_2 + \dots + \tau_c \theta_c - N \ln K \quad (7)$$

substituindo-se (7) em (6), obtem-se:

$$I(* : 2; 0_N) = \sum_{x_1+x_2+\dots+x_c=N} p^*(x) (\tau_1 \theta_1 + \tau_2 \theta_2 + \dots + \tau_c \theta_c - N \ln K) \quad \dots$$

$$I(* : 2; 0_N) = \tau_1 \theta_1 + \tau_2 \theta_2 + \dots + \tau_c \theta_c - N \ln K, \quad \text{mas}$$

$$\text{por (5)} \quad I(* : 2; 0_N) = \theta_1 \ln \frac{\theta_1}{N p_1} + \theta_1 \ln K + \theta_2 \ln \frac{\theta_2}{N p_2}$$

$$+ \theta_2 \ln K + \dots + \theta_c \ln \frac{\theta_c}{N p_c} + \theta_c \ln K - N \ln K, \quad \dots$$

$$I(* : 2; 0_N) = \theta_1 \ln \frac{\theta_1}{N p_1} + \theta_2 \ln \frac{\theta_2}{N p_2} + \dots + \theta_c \ln \frac{\theta_c}{N p_c} +$$

$$(\theta_1 + \theta_2 + \dots + \theta_c) \ln K - N \ln K; \quad \text{mas por (4),}$$

tem-se que $\sum_{i=1}^c \theta_i = \sum_{i=1}^c N p_i^* = N$, logo

$$l(*:2; 0_N) = \theta_1 \ln \frac{\theta_1}{N p_1} + \theta_2 \ln \frac{\theta_2}{N p_2} + \dots + \theta_c \ln \frac{\theta_c}{N p_c} \quad (8)$$

$$l(*:2; 0_N) = \sum_{i=1}^c \theta_i \ln \frac{\theta_i}{N p_i}, \quad i = 1, 2, \dots, c \quad (9)$$

Para aplicações a testes de hipóteses sobre populações multinomiais, a distribuição $p(x)$ em (1) representa a hipótese nula, enquanto que a distribuição conjugada $p^*(x)$, dada por (2) representa a hipótese alternativa. Na distribuição conjugada $p^*(x)$, os parâmetros serão considerados como sendo iguais às estimativas imparciais observadas na amostra, isto é,

$$\hat{\theta}_i = N \hat{p}_i^* = x_i, \quad i = 1, 2, \dots, c \quad (10)$$

Da equação (5) tem-se que $\tau_i = \ln \frac{\theta_i}{N p_i} + \ln K$. Devido

ao fato de que o valor de K é arbitrário e que $l(*:2; 0_N)$ não depende do valor de $K = p_1 e^{\tau_1} + p_2 e^{\tau_2} + \dots + p_c$, então K pode ser considerado como sendo igual a 1, portanto

$$\tau_i = \ln \frac{\theta_i}{N p_i}, \quad i = 1, 2, \dots, c \quad (11) \text{ e}$$

a sua estimativa é $\hat{\tau}_i = \ln \frac{\hat{\theta}_i}{N p_i} = \frac{x_i}{N p_i}$, $i = 1, 2, \dots, c$.

Substituindo-se este valor em (8), obtem-se a estatística da informação mínima de discriminação (EIMD),

$$\begin{aligned}
 T(*:2; 0_N) &= T(\hat{p}^*(x) : p(x)) = x_1 \ln \frac{x_1}{Np_1} + x_2 \ln \frac{x_2}{Np_2} + \\
 &x_3 \ln \frac{x_3}{Np_3} + \dots + x_c \ln \frac{x_c}{Np_c} \dots \\
 T(*:2; 0_N) &= \sum_{i=1}^c x_i \ln \frac{x_i}{Np_i} \quad (12)
 \end{aligned}$$

De acordo com o estabelecido no ítem 3.4.4.2.,

$$2 T(*:2; 0_N) \sim \chi^2(c-1) \quad (13)$$

Conclui-se então, que pode-se testar a hipótese nula H_2 , *versus* a hipótese alternativa H_1 , através de $2 T(p^*(x) : p(x))$, rejeitando H_2 para valores grandes da EIMD. Admite-se que $x_i \ln x_i = 0$, se $x_i = 0$.

3.5. Formulacão e Teste de Hipóteses de Interesse em Tabelas de Contingência R x S x T, Considerando Fixas as Freqüências Marginais Observadas

3.5.1. Introdução

Nos ítems 3.3.3. e 3.3.4., foram apresentadas as hipóteses de interesse numa tabela de contingência R x S x T e foi obtida a estimativa de máxima verossimilhança para a (i, j, k)-ésima casela, sob a hipótese de independência mútua completa, a partir do modelo log-linear $\ln p_{ijk} = \mu + \mu_1(i) + \mu_2(j) + \mu_3(k)$.

Neste caso, foi admitido *a priori* um determinado modelo log-linear para as probabilidades das

$R \times S \times T$ caselas da tabela de contingência. Neste trabalho será usado o princípio da informação mínima de discriminação, que, dentre outras, tem a vantagem de não pressupor *a priori* um determinado modelo para testar uma hipótese. O modelo log-linear será obtido a partir de restrições impostas nas frequências marginais observadas. As restrições impostas, especificam que a tabela de contingência estimada tenha o mesmo conjunto de marginais iguais às marginais observadas. GOKHALE & KULLBACK (1978), consideram que as restrições impostas especificam uma tabela de contingência estimada tal que funções lineares de caselas estimadas tenham o mesmo valor que as mesmas funções lineares de caselas observadas.

3.5.2. Representação log-linear

Neste ítem será demonstrado um dos resultados apresentados por GOKHALE & KULLBACK (1978).

Seja uma tabela de contingência $R \times S \times T$, isto é, contendo $R \times S \times T$ caselas. Seja Ω , o conjunto formado pelas $R \times S \times T$ caselas e seja w uma casela (i, j, k) qualquer. Isto é, $\Omega = \{w : w = (i, j, k), i = 1, 2, \dots, R; j = 1, 2, \dots, S, \text{ e } k = 1, 2, \dots, T\}$. Considere-se duas distribuições de probabilidade ou duas tabelas de contingência definidas no conjunto de caselas Ω . Sejam essas duas distribuições denotadas por $p(w)$ e $\pi(w)$, tal que $\sum_{\Omega} p(w) = \sum_{\Omega} \pi(w) = 1$.

No ítem 3.4.2.1. foi definida a informação

de discriminação em favor de $p(w)$ versus $\pi(w)$, como sendo

$$I(p : \pi) = \sum_{\Omega} p(w) \ln \frac{p(w)}{\pi(w)} \quad (1)$$

Visando aplicações posteriores, $p(w)$ é um elemento da família \underline{P} de distribuições que satisfazem restrições impostas. $\pi(w)$ é uma distribuição arbitrária e a sua escolha será feita de acordo com o problema estudado.

Como se viu no ítem 3.4.3., procura-se o valor $p(w)$ que minimiza (1) entre os elementos da família \underline{P} de distribuições que satisfazem algumas restrições linearmente independentes. Usando notação matricial, as restrições podem ser expressas como:

$$\underline{C} \underline{p} = \underline{\theta}, \quad (2), \text{ onde}$$

\underline{C} é uma matriz $(r + 1) \times \Omega$, \underline{p} é o valor das probabilidades das $R \times S \times T$ caselas, isto é, \underline{p} tem dimensões $\Omega \times 1$ e $\underline{\theta}$ é um vetor de restrições de ordem $(r + 1) \times 1$, e o *rank* de \underline{C} é $(r + 1) \leq \Omega$, isto é, as $(r + 1)$ linhas de \underline{C} são linearmente independentes. Denotando os elementos da matriz \underline{C} por $C_i(w)$, $i = 0, 1, 2, \dots, r$ e $w = 1, 2, \dots, \Omega$, tem-se então que (2) pode ser expresso como sendo

$$\underline{C} \underline{p} = \underline{\theta} \implies \sum_{\Omega} C_i(w) p(w) = \theta_i, \quad i = 0, 1, 2, \dots, r.$$

Ilustrando com um caso genérico, tem-se que:

$$c \cdot p = \theta \implies \begin{matrix} c_0(1) & c_0(2) & c_0(3) & \dots & c_0(w) \\ c_1(1) & c_1(2) & c_1(3) & \dots & c_1(w) \\ c_2(1) & c_2(2) & c_2(3) & \dots & c_2(w) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_r(1) & c_r(2) & c_r(3) & \dots & c_r(w) \end{matrix} \begin{matrix} p(1) \\ p(2) \\ p(3) \\ \vdots \\ p(w) \end{matrix} = \begin{matrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_r \end{matrix} \quad (3)$$

Por exemplo, $c_0(1) p(1) + c_0(2) p(2) + c_0(3) p(3) + \dots + c_0(w) p(w) = \theta_0 \quad \therefore$

$$\sum_{\Omega} c_0(w) p(w) = \theta_0 \quad (4)$$

Para expressar a restrição natural que $\sum_{\Omega} p(w) = 1$, toma-se $c_0(w)=1$, para todo w e $\theta_0 = 1$.

Minimizando (1) com as restrições (3), tem-se que

$$Z = \sum_{\Omega} p(w) \ln \frac{p(w)}{\pi(w)} = \tau'_0 (\sum_{\Omega} p(w) - 1) - \tau_1 (\sum_{\Omega} c_1(w) p(w) - \theta_1) - \tau_2 (\sum_{\Omega} c_2(w) p(w) - \theta_2) - \dots - \tau_r (\sum_{\Omega} c_r(w) p(w) - \theta_r)$$

$$\text{e } \frac{\partial Z}{\partial p(w)} = \ln \frac{p(w)}{\pi(w)} + 1 - \tau'_0 - \tau_1 c_1(w) - \tau_2 c_2(w) - \dots - \tau_r c_r(w).$$

$$\tau'_0 c_0(w).$$

Fazendo $\tau'_0 = \tau_0 + 1$, tem-se que

$$\frac{\partial Z}{\partial p(w)} = \ln \frac{p(w)}{\pi(w)} - \tau_0 - \tau_1 c_1(w) - \tau_2 c_2(w) - \dots - \tau_r c_r(w).$$

Igualando a zero, obtem-se

$$\ln \frac{p^*(w)}{\pi(w)} = \tau_0 + \tau_1 C_1(w) + \tau_2 C_2(w) + \dots + \tau_r C_r(w) \quad (4)$$

$$\text{ou } p^*(w) = \left[e^{\tau_0 + \tau_1 C_1(w) + \tau_2 C_2(w) + \dots + \tau_r C_r(w)} \right] \pi(w)$$

$w = 1, 2, \dots, \Omega$, onde os τ 's serão determinados tal que $cp^* = \underline{0}$. A representação (4) é também conhecida como modelo log-linear. Observa-se que $p^*(w)$ gera uma família de distribuições exponenciais, que como se sabe, têm propriedades estatísticas desejáveis (HOGG & KRAIG, 1978; WILKS, 1962, dentre outros). Os τ 's são os parâmetros exponenciais ou parâmetros naturais da família exponencial e representam parâmetros de efeitos principais e de interações.

3.5.3. Estimativa da freqüência esperada na (i, j, k) - é-sima casela através da informação mínima de discriminação

Uma vez que uma hipótese é selecionada, o próximo passo é estimar as freqüências das caselas sob a hipótese nula, usando as freqüências marginais. No ítem 3.5.2., viu-se que a estimativa pelo princípio da informação mínima de discriminação (IMD) é $p^*(w)$ tal que

$$I(p^* : \pi) = \sum_{\Omega} p^*(w) \ln \frac{p^*(w)}{\pi(w)} = \min I(p : \pi), p, p^* \in P.$$

No caso particular de tabelas de contingência tridimensionais, será estimada $p(ijk)$ minimizando a informação de discriminação

$$I(p : \pi) = \sum_i \sum_j \sum_k p(ijk) \ln \frac{p(ijk)}{\pi(ijk)} \quad (1)$$

$i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ e $k = 1, 2, \dots, t$,

impondo a restrição que a tabela estimada tenha as marginais de interesse iguais às marginais da tabela observada e $\pi(ijk)$ será tomada como a distribuição uniforme, isto

$$\text{é, } \pi(ijk) = \frac{1}{R \times S \times T}.$$

Para qualquer tabela de contingência tridimensional $R \times S \times T$, podem-se considerar três tabelas associadas da mesma dimensão (KU *et alii*, 1971):

1) A π -tabela $\{\pi(ijk)\}$, $\sum_i \sum_j \sum_k \pi(ijk) = 1$. A π -tabela será admitida como especificada pela distribuição uniforme, isto é,

$$\pi(ijk) = \frac{1}{R \times S \times T}, \text{ para toda } i, j, k.$$

2) A família das p -tabelas denotadas por $\{p(ijk)\}$, $\sum_i \sum_j \sum_k p(ijk) = 1$. Esta tabela satisfaz certas condições de interesse, geralmente uma especificação das marginais, por exemplo, as marginais unidimensionais $P(i..)$, $P(.j.)$ e $P(..k)$, $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ e $k = 1, 2, \dots, t$.

3) A p^* -tabela $\{p^*(ijk)\}$, $\sum_{i,j,k} p^*(ijk) = 1$. A p^* -tabela é a estimativa de $p(ijk)$ através do método da informação mínima de discriminação. É o elemento da família das p -tabelas que tem mais "semelhança" com a π -tabela, isto é, a p^* -tabela minimiza a informação de discriminação

$$I(p : \pi) = \sum_{i,j,k} p(ijk) \ln \frac{p(ijk)}{\pi(ijk)}, \text{ sujeita às restrições}$$

impostas.

3.5.4. Independência mútua completa

3.5.4.1. Representação log-linear

Admite-se uma tabela de contingência conhecida $\pi(ijk)$, $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ e $k = 1, 2, \dots, t$; $\sum_i \sum_j \sum_k \pi(ijk) = 1$ e é requerida uma tabela de contingência $p(ijk)$ tal que as probabilidades marginais $p(i..)$, $p(.j.)$ e $p(..k)$, sejam fixas e iguais às probabilidades observadas numa tabela de contingência $R \times S \times T$, de tal forma que

$$I(p : \pi) = \sum_i \sum_j \sum_k p(ijk) \ln \frac{p(ijk)}{\pi(ijk)} \text{ seja um mínimo para}$$

toda $p(ijk)$ com as probabilidades marginais dadas.

Para aplicar o teorema de KULLBACK (1959), que foi enunciado no ítem 3.4.3., página 64, necessitam-se estatísticas $T_{\alpha..}(ijk)$, $\alpha = 1, 2, \dots, r$;

$$T_{\cdot\beta\cdot}(ijk), \quad \beta = 1, 2, \dots, s \quad e$$

$$T_{\dots\gamma}(ijk), \quad \gamma = 1, 2, \dots, t, \quad \text{tal que}$$

$$\left. \begin{aligned} \sum_i \sum_j \sum_k T_{\alpha\cdot\cdot}(ijk) P(ijk) &= P(i\cdot\cdot), \\ \sum_i \sum_j \sum_k T_{\cdot\beta\cdot}(ijk) P(ijk) &= P(\cdot j \cdot), \\ \sum_i \sum_j \sum_k T_{\dots\gamma}(ijk) P(ijk) &= P(\cdot\cdot k) \end{aligned} \right\} e \quad (1)$$

$$\text{Definindo } T_{\alpha\cdot\cdot}(ijk) = \begin{cases} 1, & \text{se } \alpha = i \\ 0, & \text{se } \alpha \neq i \end{cases} \quad (2)$$

$$T_{\cdot\beta\cdot}(ijk) = \begin{cases} 1, & \text{se } \beta = j \\ 0, & \text{se } \beta \neq j \end{cases} \quad (3)$$

$$e \quad T_{\dots\gamma}(ijk) = \begin{cases} 1, & \text{se } \gamma = k \\ 0, & \text{se } \gamma \neq k \end{cases} \quad (4)$$

Observe que (2), (3) e (4) satisfazem (1), por exemplo, para $\alpha = 1$, tem-se

$$\sum_i \sum_j \sum_k T_{1\cdot\cdot}(ijk) P_{ijk} =$$

$$T_{1\cdot\cdot}(111)P(111) + T_{1\cdot\cdot}(112)P(112) + \dots + T_{1\cdot\cdot}(11t)P(11t)^+$$

$$T_{1\cdot\cdot}(121)P(121) + T_{1\cdot\cdot}(122)P(122) + \dots + T_{1\cdot\cdot}(12t)P(12t)$$

+ ... +

$$\begin{aligned}
& T_{1..(1s1)}P(1s1) + T_{1..(1s2)}P(1s2) + \dots + T_{1..(1st)}P(1st) + \\
& T_{1..(211)}P(211) + T_{1..(212)}P(212) + \dots + T_{1..(21t)}P(21t) + \\
& T_{1..(221)}P(221) + T_{1..(222)}P(222) + \dots + T_{1..(22t)}P(22t) + \\
& + \dots + \\
& T_{1..(2s1)}P(2s1) + T_{1..(2s2)}P(2s2) + \dots + T_{1..(2st)}P(2st) + \\
& + \dots + \\
& T_{1..(r11)}P(r11) + T_{1..(r12)}P(r12) + \dots + T_{1..(r1t)}P(r1t) + \\
& T_{1..(r21)}P(r21) + T_{1..(r22)}P(r22) + \dots + T_{1..(r2t)}P(r2t) + \\
& + \dots + \\
& T_{1..(rs1)}P(rs1) + T_{1..(rs2)}P(rs2) + \dots + T_{1..(rst)}P(rst) +
\end{aligned}$$

Observa-se que a partir de $i = 2$, $\alpha = 1 \neq i = 2$, portanto $T_{1..} = 0$, logo

$$\sum_j \sum_k T_{1..(ijk)} P(ijk) = \sum_j \sum_k P(ijk) = P_{1..}$$

Da equação (14) do ítem 3.4.3., página 65, tem-se que

$$p^*(ijk) = \frac{e^{\tau(ijk)} \tau(ijk)}{M_2(\tau(ijk))} P_2(ijk) \quad (5), \text{ logo}$$

$$p^*(ijk) = \frac{e^{\sum \tau_{\alpha..} T_{\alpha..}(ijk) + \sum \tau_{\beta.} T_{\beta.}(ijk) + \sum \tau_{\gamma} T_{\gamma}(ijk)}}{M(\tau_{\alpha..}, \tau_{\beta.}, \tau_{\gamma})} \cdot \pi(ijk)$$

seja $M = M(\tau_{\alpha..}, \tau_{\beta.}, \tau_{\gamma})$, logo

$$M = \sum_i \sum_j \sum_k e^{\sum \tau_{\alpha..} T_{\alpha..}(ijk) + \sum \tau_{\beta.} T_{\beta.}(ijk) + \sum \tau_{\gamma} T_{\gamma}(ijk)} \cdot \pi(ijk) \dots$$

$$M = \sum_i \sum_j \sum_k e^{\tau_{i..} + \tau_{.j.} + \tau_{..k}} \cdot \pi(ijk), \text{ logo}$$

$$p^*(ijk) = \frac{e^{\tau_{i..} + \tau_{.j.} + \tau_{..k}} \cdot \pi(ijk)}{\sum_i \sum_j \sum_k e^{\tau_{i..} + \tau_{.j.} + \tau_{..k}} \cdot \pi(ijk)}$$

Pela equação (13), ítem 3.4.3., página 64, têm-se que

$$\underline{\theta} = \frac{\partial \ln M_2(\underline{\tau})}{\partial \tau} = \frac{M_2'(\underline{\tau})}{M_2(\underline{\tau})}$$

onde

$$\ln M_2(\tau) = \ln M = \ln \sum_i \sum_j \sum_k e^{\tau_{i..} + \tau_{.j.} + \tau_{..k}} \cdot \pi(ijk), \text{ isto é,}$$

$$\ln M = \ln \sum_i e^{\tau_{i..}} \sum_j e^{\tau_{.j.}} \sum_k e^{\tau_{..k}} \cdot \pi(ijk) \quad (6)$$

Observe que (6) pode ser expresso como

$$\ln M = \ln (e^{\tau_{1..}} + e^{\tau_{2..}} + \dots + e^{\tau_{r..}}) \sum_j e^{\tau_{.j.}} \sum_k e^{\tau_{..k}} \pi(ijk)$$

(7),

logo, tem-se:

$$p_{i..} = \frac{\partial \ln M}{\partial \tau_{i..}} = \frac{e^{\tau_{i..}} \sum_j \sum_k e^{\tau_{.j.} + \tau_{..k}} \cdot \pi(ijk),}{M}$$

$i = 1, 2, \dots, r$ (8)

$$p_{.j.} = \frac{\partial \ln M}{\partial \tau_{.j.}} = \frac{e^{\tau_{.j.}} \sum_i \sum_k e^{\tau_{i..} + \tau_{..k}} \cdot \pi(ijk),}{M}$$

$j = 1, 2, \dots, s$ (9)

$$p_{..k} = \frac{\partial \ln M}{\partial \tau_{..k}} = \frac{e^{\tau_{..k}} \sum_i \sum_j e^{\tau_{i..} + \tau_{.j.}} \cdot \pi(ijk),}{M}$$

$k = 1, 2, \dots, t$ (10)

Por (10), do ítem 3.4.3., página 64, sabe-se que o valor mínimo de $I(p : \Pi)$ é $I(p^* : \Pi) = \tau\theta - \ln M$, donde

$$I(p^* : \Pi) = [\tau_{i..} \ \tau_{.j.} \ \tau_{..k}] \begin{bmatrix} p_{i..} \\ p_{.j.} \\ p_{..k} \end{bmatrix} - \ln M \quad (11),$$

$$\text{logo } I(p^* : \Pi) = \sum_i \tau_{i..} p_{i..} + \sum_j \tau_{.j.} p_{.j.} + \sum_k \tau_{..k} p_{..k} - \ln M \quad (12)$$

Para usar a mesma notação de KU & KULLBACK

(1963), KU *et alii* (1971) e KU & KULLBACK (1974), sejam

$$e^{\tau_{i..}} = a_i; e^{\tau_{.j.}} = b_j; e^{\tau_{..k}} = c_k \text{ e considerando-se}$$

$$\text{que } \sum_i \sum_j \sum_k a_i b_j c_k \pi(ijk) = 1, \text{ obtem-se:}$$

$$p^*(ijk) = a_i b_j c_k \pi(ijk) \quad (13),$$

$$p_{i..} = p_{i..}^* = a_i \sum_j \sum_k b_j c_k \pi(ijk) \quad (14),$$

$$p_{.j.} = p_{.j.}^* = b_j \sum_i \sum_k a_i c_k \pi(ijk) \quad (15),$$

$$p_{..k} = p_{..k}^* = c_k \sum_i \sum_j a_i b_j \pi(ijk) \quad (16) \text{ e}$$

$$l(p^* : \pi) = \sum_i p_{i..} \ln a_i + \sum_j p_{.j.} \ln b_j + \sum_k p_{..k} \ln c_k \quad (17)$$

Observações importantes:

(i) Por (13), $p^*(ijk) = a_i b_j c_k \pi(ijk)$. Observe que é um modelo multiplicativo nos parâmetros que pode ser transformado em linear na escala logarítmica, isto é,

$$\ln p^*(ijk) = \ln \pi(ijk) + \ln a_i + \ln b_j + \ln c_k \quad (18)$$

Esse é o modelo log-linear obtido sob a hipótese de independência mútua completa, tendo surgido naturalmente, sem que fosse assumido *a priori*, como no ítem 3.3.4., página 42.

(ii) A equação (18), pode ser expressa como sendo $\ln p^*(ijk) = \ln \pi(ijk) + \tau_{i..} + \tau_{.j.} + \tau_{..k}$, onde os τ 's são funções das probabilidades marginais unidi-

mensionais. Lembre que $\theta = \frac{d \ln M(\tau)}{d\tau}$.

Por (18), $\ln \frac{p^*(ijk)}{\pi(ijk)} = \ln a_i + \ln b_j + \ln c_k$.

Em lugar de determinar a_i , b_j e c_k , tal que $p^*(ijk)$ cumpra com as restrições marginais unidimensionais, será estimado $p^*(ijk)$ e a partir desses valores serão calculados os valores correspondentes a a_i , b_j e c_k . As equações (13), (14), (15) e (16), motivaram um procedimento iterativo usado por DEMING & STEPHAN (1940)¹, citado por IRELAND & KULLBACK (1968), para obter as estimativas de $p^*(ijk)$ sob H_0 . Este procedimento iterativo será considerado no próximo ítem.

(iii) Para facilitar a notação, as demonstrações foram realizadas sem mencionar explicitamente que tanto $p_{i..}$, $p_{.j.}$ e $p_{..k}$ são estimativas de máxima verossimilhança dos parâmetros populacionais. Isto é, na verdade, deve-se considerar todo o procedimento tomando-se $\hat{p}_{i..} = \frac{x_{i..}}{N}$; $\hat{p}_{.j.} = \frac{x_{.j.}}{N}$; $\hat{p}_{..k} = \frac{x_{..k}}{N}$ em lugar de

$p_{i..}$, $p_{.j.}$ e $p_{..k}$, respectivamente. Isto deve-se, a que a restrição imposta foi de que a tabela de contingência $p^*(ijk)$ tivesse as mesmas probabilidades observadas ou estimadas numa tabela de contingência $R \times S \times T$. Lembre que no ítem 3.4.4.1., página 68, foi estabelecido que em caso de uma amostra de N observações independentes,

¹ DEMING, W.E. & STEPHAN, F.F. On a least squares adjustment of a sampled frequency table when the expected marginals totals are known. *Ann. Math. Statist.*, Baltimore, 11:427-44. 1940

pode-se usar o valor observado de $T(x)$ como uma estimativa de θ , isto é, $T(x) = \hat{\theta}(x)$. Observe na equação (11), que $\underline{\theta}' = [p_{i..} \ p_{.j.} \ p_{..k}]$, logo $\hat{\underline{\theta}}' = [\hat{p}_{i..} \ \hat{p}_{.j.} \ \hat{p}_{..k}]$.

3.5.4.2. Estimativas das freqüências esperadas

3.5.4.2.1. Processo iterativo

Das equações (13), (14), (15) e (16), do ítem 3.5.4.1., têm-se o seguinte sistema de equações:

$$\left. \begin{aligned} p^*(ijk) &= a_i b_j c_k \pi(ijk) \\ p_{i..} &= a_i \sum_j \sum_k b_j c_k \pi(ijk) \\ p_{.j.} &= b_j \sum_i \sum_k a_i c_k \pi(ijk) \\ p_{..k} &= c_k \sum_i \sum_j a_i b_j \pi(ijk) \end{aligned} \right\} \quad (1)$$

A solução iterativa do sistema de equações (1), consiste em obter $p^*(ijk)$ em lugar de resolver para a_i , b_j e c_k . Este processo iterativo será realizado ajustando em forma sistemática as probabilidades marginais unidimensionais. Considere os seguintes passos:

$$(i) \quad p_{i..} = a_i^{(1)} \sum_j \sum_k b_j^{(1)} c_k^{(1)} \pi(ijk) \implies p_{i..}^{(1)} = a_i^{(1)} b_j^{(1)} c_k^{(1)} \pi(ijk)$$

$$(ii) \quad p_{.j.} = b_j^{(2)} \sum_i \sum_k a_i^{(1)} c_k^{(1)} \pi_{(ijk)} \implies p_{(ijk)}^{(2)} =$$

$$a_i^{(1)} b_j^{(2)} c_k^{(1)} \pi_{(ijk)}$$

$$(iii) \quad p_{..k} = c_k^{(2)} \sum_i \sum_j a_i^{(1)} b_j^{(2)} \pi_{(ijk)} \implies p_{(ijk)}^{(3)} =$$

$$a_i^{(1)} b_j^{(2)} c_k^{(2)} \pi_{(ijk)}$$

$$(iv) \quad p_{i..} = a_i^{(2)} \sum_j \sum_k b_j^{(2)} c_k^{(2)} \pi_{(ijk)} \implies p_{(ijk)}^{(4)} =$$

$$a_i^{(2)} b_j^{(2)} c_k^{(2)} \pi_{(ijk)},$$

assim sucessivamente. Seja inicialmente $b_j^{(1)} = c_k^{(1)} = 1$, logo, tem-se que

$$(i) \quad p_{i..} = a_i^{(1)} \sum_j \sum_k \pi_{(ijk)} = a_i^{(1)} \pi_{i..} \dots a_i^{(1)} = \frac{p_{i..}}{\pi_{i..}},$$

$$\text{então } p_{(ijk)}^{(1)} = a_i^{(1)} \pi_{(ijk)} = \frac{p_{i..}}{\pi_{i..}} \pi_{(ijk)}$$

$$(ii) \quad p_{.j.} = b_j^{(2)} \sum_i \sum_k a_i^{(1)} \pi_{(ijk)} = b_j^{(2)} \sum_i a_i^{(1)} \pi_{ij.} \dots$$

$$b_j^{(2)} = \frac{p_{.j.}}{\sum_i a_i^{(1)} \pi_{ij.}}, \text{ mas } \sum_i \sum_k p_{(ijk)}^{(1)} = p_{.j.} =$$

$$\sum_i a_i^{(1)} \pi_{ij.}, \text{ portanto } b_j^{(2)} = \frac{p_{.j.}}{p_{.j.}}, \text{ então}$$

$$p_{(ijk)}^{(2)} = a_i^{(1)} b_j^{(2)} \pi_{(ijk)} \dots$$

$$p_{(ijk)}^{(2)} = b_j^{(2)} a_i^{(1)} \pi_{(ijk)} = \frac{p_{.j.}}{p_{.j.}} p_{(ijk)}^{(1)}$$

$$(iii) \quad p_{..k} = c_k^{(2)} \sum_i \sum_j a_i^{(1)} b_j^{(2)} \pi_{(ijk)} \dots$$

$$c_k^{(2)} = \frac{p_{..k}}{\sum_i \sum_j a_i^{(1)} b_j^{(2)} \pi_{(ijk)}}, \text{ mas}$$

$$\sum_i \sum_j p_{(ijk)}^{(2)} = p_{..k}^{(2)} = \sum_i \sum_j a_i^{(1)} b_j^{(2)} \pi_{(ijk)}, \text{ portanto}$$

$$c_k^{(2)} = \frac{p_{..k}}{p_{..k}^{(2)}}, \text{ então } p_{(ijk)}^{(3)} = a_i^{(1)} b_j^{(2)} c_k^{(2)} \pi_{(ijk)} \dots \quad (2)$$

$$p_{(ijk)}^{(3)} = c_k^{(2)} a_i^{(1)} b_j^{(2)} \pi_{(ijk)} = \frac{p_{..k}}{p_{..k}^{(2)}} p_{(ijk)}^{(2)}$$

$$(iv) \quad p_{i..} = a_i^{(2)} \sum_j \sum_k b_j^{(2)} c_k^{(2)} \pi_{(ijk)} \dots$$

$$a_i^{(2)} = \frac{p_{i..}}{\sum_j \sum_k b_j^{(2)} c_k^{(2)} \pi_{(ijk)}} \quad (3),$$

$$\text{mas, por (2), } \frac{p_{(ijk)}^{(3)}}{a_i^{(1)}} = b_j^{(2)} c_k^{(2)} \pi_{(ijk)}, \text{ logo}$$

$$\sum_j \sum_k \frac{p_{(ijk)}^{(3)}}{a_i^{(1)}} = \sum_j \sum_k b_j^{(2)} c_k^{(2)} \pi_{(ijk)} = \frac{p_{i..}^{(3)}}{a_i^{(1)}} \quad (4)$$

$$\text{logo } a_i^{(2)} = \frac{p_{i..} a_i^{(1)}}{p_{i..}^{(3)}}. \text{ Então}$$

$$p_{(ijk)}^{(4)} = a_i^{(2)} b_j^{(2)} c_k^{(2)} \pi_{(ijk)} = \frac{p_{i..} a_i^{(1)}}{p_{i..}^{(3)}} \cdot \frac{p_{(ijk)}^{(3)}}{a_i^{(1)}} \dots$$

$$p_{(ijk)}^{(4)} = \frac{p_{i..}^{(3)}}{p_{i..}^{(3)}} p_{(ijk)}^{(3)}.$$

Os valores da p*-tabela devem ser computados por um pro-

cesso iterativo de ajustamento da π -tabela, até satisfazer as restrições marginais dadas, levando-se em conta uma tolerância previamente estabelecida.

Em geral, para a hipótese de independência mutua completa, o processo iterativo pode ser expresso como:

$$\left. \begin{aligned} p^{(3n+1)}(ijk) &= \frac{p_{i..}}{p_{i..}^{(3n)}} p^{(3n)}(ijk) \\ p^{(3n+2)}(ijk) &= \frac{p_{.j.}}{p_{.j.}^{(3n+1)}} p^{(3n+1)}(ijk) \\ p^{(3n+3)}(ijk) &= \frac{p_{..k}}{p_{..k}^{(3n+2)}} p^{(3n+2)}(ijk), \quad n = 0, 1, 2, \dots \end{aligned} \right\} (5)$$

Esse processo iterativo é convergente. Seja, por exemplo,

$$\begin{aligned} I(p^* : p^{(3n+2)}) &= \sum_i \sum_j \sum_k p^*_{ijk} \ln \frac{p^*_{ijk}}{p_{ijk}^{(3n+2)}} \dots \\ I(p^* : p^{(3n+2)}) &= \sum_i \sum_j \sum_k p^*_{ijk} \ln \frac{p^*_{ijk}}{\frac{p_{.j.} p_{ijk}}{p_{.j.}^{(3n+1)}}} = \\ &= \sum_i \sum_j \sum_k p^*_{ijk} \ln \frac{p^*_{ijk}}{p_{ijk}^{(3n+1)}} - \sum_i \sum_j \sum_k p^*_{ijk} \ln \frac{p_{.j.}}{p_{.j.}^{(3n+1)}} = \\ &= \sum_i \sum_j \sum_k p^*_{ijk} \ln \frac{p^*_{ijk}}{p_{ijk}^{(3n+1)}} - \sum_i \sum_j p^*_{ij.} \ln \frac{p_{.j.}}{p_{.j.}^{(3n+1)}} \end{aligned}$$

Como esses valores de informação são maiores ou iguais a

$$\text{zero, então } \sum_i \sum_j \sum_k p^*_{ijk} \ln \frac{p^*_{ijk}}{(3n+2)} \leq \sum_i \sum_j \sum_k p^*_{ijk} \ln \frac{p^*_{ijk}}{(3n+1)}$$

Com a igualdade se e somente se

$$\sum_i \sum_j p^*_{ij.} \ln \frac{p_{.j.}}{(3n+1)} = 0, \text{ o que é verdade, se e somente se}$$

$$p_{.j.}^{(3n+1)} = p_{.j.}, \quad j = 1, 2, \dots, s$$

Operações similares podem ser feitas para demonstrar que

$$p_{i..}^{(3n+3)} = p_{i..}, \quad i = 1, 2, \dots, r \quad \text{e}$$

$$p_{..k}^{(3n+2)} = p_{..k}, \quad k = 1, 2, \dots, t, \quad n = 0, 1, 2, \dots$$

3.5.4.2.2. Computação das estimativas das frequências esperadas

No processo iterativo expresso pelas equações (5) do ítem anterior é recomendável considerar-se uma distribuição $\pi(ijk)$ na qual se cumpram ou verifiquem as restrições impostas sobre as marginais unidimensionais.

Uma distribuição com essas características é a distribuição uniforme, isto é, $\pi(ijk) = \frac{1}{R \times S \times T}$, para todo $i =$

$1, 2, \dots, r; j = 1, 2, \dots, s$ e $k = 1, 2, \dots, t$.

Quando $n = 0$, obtem-se

$$P_{(ijk)}^{(1)} = \frac{P_{i..}}{P_{i..}^{(0)}} P_{(ijk)}^{(0)} \quad (1)$$

$$P_{(ijk)}^{(2)} = \frac{P_{.j.}}{P_{.j.}^{(1)}} P_{(ijk)}^{(1)} \quad (2) \text{ e}$$

$$P_{(ijk)}^{(3)} = \frac{P_{..k}}{P_{..k}^{(2)}} P_{(ijk)}^{(2)} \quad (3),$$

onde $P_{ijk}^{(0)} = \pi(ijk) = \frac{1}{R \times S \times T}$, logo

$$P_{i..}^{(0)} = \sum_j \sum_k \pi(ijk) = \frac{1}{R}, \text{ logo}$$

$$P_{(ijk)}^{(1)} = \frac{P_{i..}}{\frac{1}{R}} \frac{1}{R \times S \times T} = \frac{P_{i..}}{S \times T} \dots$$

$$P_{.j.}^{(1)} = \sum_i \sum_k P_{(ijk)}^{(1)} = \sum_i \sum_k \frac{P_{i..}}{S \times T} = \frac{T}{S \times T} = \frac{1}{S}, \text{ logo}$$

$$P_{(ijk)}^{(2)} = \frac{P_{.j.}}{\frac{1}{S}} \frac{P_{i..}}{S \times T} = \frac{P_{.j.} \cdot P_{i..}}{T} \dots$$

$$P_{..k}^{(2)} = \sum_i \sum_j P_{(ijk)}^{(2)} = \sum_i \sum_j \frac{P_{.j.} \cdot P_{i..}}{T} = \frac{1}{T}, \text{ logo}$$

$$P_{(ijk)}^{(3)} = \frac{P_{..k}}{\frac{1}{T}} \frac{P_{.j.} \cdot P_{i..}}{T} \dots$$

$$p_{(ijk)}^{(3)} = p_{i..} p_{.j.} p_{..k} \quad \text{logo}$$

$$p_{i..}^{(3)} = \sum_j \sum_k p_{(ijk)}^{(3)} = \sum_j \sum_k p_{i..} p_{.j.} p_{..k} = p_{i..}, \quad \text{logo}$$

$$p_{(ijk)}^{(4)} = \frac{p_{i..}}{p_{i..}} p_{i..} p_{.j.} p_{..k} = p_{i..} p_{.j.} p_{..k},$$

Observa-se que $p_{(ijk)}^{(4)} = p_{(ijk)}^{(3)}$. Além disso

$$p_{.j.}^{(3)} = \sum_i \sum_k p_{(ijk)}^{(3)} = p_{.j.} \quad e$$

$$p_{..k}^{(3)} = \sum_i \sum_j p_{(ijk)}^{(3)} = p_{..k}, \quad \text{isto é, verifica-se que se}$$

$$p_{(ijk)}^{(3)} = p_{i..} p_{.j.} p_{..k} \tag{4}$$

então, a tabela de contingência construída dessa forma cumpre com a restrição imposta de que as marginais unidimensionais sejam iguais às marginais unidimensionais da tabela observada.

Lembrando-se da observação (iii) do item 3.5.4.1., página 88, tem-se que

$$p^*(ijk) = \hat{p}_{i..} \hat{p}_{.j.} \hat{p}_{..k}, \quad \text{onde} \quad \hat{p}_{i..} = \frac{x_{i..}}{N}, \quad \hat{p}_{.j.} = \frac{x_{.j.}}{N},$$

$$\hat{p}_{..k} = \frac{x_{..k}}{N} \quad \text{e que} \quad Np^*(ijk) = x^*_{ijk}, \quad \text{então}$$

$$x^*(ijk) = \frac{x_{i..} x_{.j.} x_{..k}}{N^2} \tag{5}$$

Observa-se que $x^*(ijk)$ dada pela expressão (5), que é a estimativa obtida através do método da informação mínima de discriminação, coincide com a estimativa de máxima verossimilhança de m_{ijk} obtida no ítem 3.3.4. e dada pela expressão (6) da página 43.

3.5.4.3. Estatística da informação mínima de discriminação

Testar a hipótese $H_0 : p_{ijk} = p_{i..} p_{.j.} p_{..k}$ é equivalente a testar

$$H_0 : x(ijk) = \frac{x_{i..} x_{.j.} x_{..k}}{N^2} .$$

A estatística da informação mínima de discriminação

$$2 N I(\hat{p} : p^*) = 2 I(x : x^*) = 2 \sum_i \sum_j \sum_k x_{ijk} \ln \frac{x_{ijk}}{x^*_{ijk}}, \text{ sob}$$

H_0 , é distribuída assintoticamente como χ^2 com um número de graus de liberdade calculados com base no teorema de Wilks enunciado no ítem 3.4.4.2., página 72.

No ítem 3.5.2., demonstrou-se que ao minimizar a informação de discriminação

$$I(p : \pi) = \sum_{\Omega} p(w) \ln \frac{p(w)}{\pi(w)}, \text{ sujeita às restrições } C_p = \underline{\theta},$$

obtem-se o modelo log-linear

$$\ln \frac{p^*(w)}{\pi(w)} = \tau_0 + \tau_1 C_1(w) + \tau_2 C_2(w) + \dots + \tau_r C_r(w),$$

onde os τ 's são determinados tal que $C \underline{p}^* = \underline{\theta}$, onde a matriz C é de ordem $(r+1) \times \Omega$, \underline{p} é um vetor de probabilidades $\Omega \times 1$ e $\underline{\theta}$ é vetor com os valores das restrições de ordem $(r+1) \times 1$ e o *rank* de \underline{C} é $(r+1) \leq \Omega$, isto é, de *rank* linha completo. Tinha-se visto que as restrições linearmente independentes $C \underline{p} = \underline{\theta}$ são equivalentes a $\sum_{\Omega} C_i(w) p(w) = \theta_i$, $i = 0, 1, 2, \dots, r$, onde $C_i(w)$, $i = 0, 1, 2, \dots, r$; $w = 1, 2, \dots, \Omega$, são os elementos da i -ésima linha e da w -ésima coluna da matriz \underline{C} .

Em termos de freqüências estimadas x^*_{ijk} , as quais serão denotadas por $x^*(w)$, tem-se que

$$N p^*(w) = x^*(w) \therefore p^*(w) = \frac{x^*(w)}{N} \text{ e}$$

$$\frac{x^*(w)}{N \pi(w)} = e^{\tau_0 + \tau_1 C_1(w) + \tau_2 C_2(w) + \dots + \tau_r C_r(w)} \quad (1), \dots$$

$$x^*(w) = \left[e^{\tau_0 + \tau_1 C_1(w) + \tau_2 C_2(w) + \dots + \tau_r C_r(w)} \right] N \pi(w)$$

Observe que $\tau_0 + \tau_1 C_1(w) + \tau_2 C_2(w) + \dots + \tau_r C_r(w)$, para cada w , pode ser expressado em forma matricial do seguinte modo:

$$\begin{bmatrix} 1 & C_1(w) & C_2(w) & \dots & C_r(w) \\ 1 & C_1(w) & C_2(w) & \dots & C_r(w) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & C_1(w) & C_2(w) & \dots & C_r(w) \end{bmatrix}_{\Omega \times (r+1)} \begin{bmatrix} \tau_0 \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_r \end{bmatrix}_{(r+1) \times 1} = T \underline{\tau},$$

onde é fácil verificar que $T = C'$. O sistema (1) pode ser então escrito na forma matricial como sendo

$$\ln(\underline{x}^*/N\pi) = T \underline{\tau} \quad (2),$$

onde $\ln(\underline{x}^*/N\pi)$ é um vetor de dimensão $\Omega \times 1$ das estimativas das frequências esperadas divididas pelas frequências da tabela especificada por $\pi(w)$; T , pela similaridade com a representação dos modelos lineares, é a matriz de delineamento de ordem $\Omega \times (r+1)$ de $rank(r+1) \leq \Omega$, isto é, de $rank$ coluna completo. Denotando as colunas linearmente independentes de \underline{T} por $T_i(w)$, $i = 0, 1, 2, \dots, r$ e $w = 1, 2, \dots, \Omega$, tem-se, que a condição imposta para que as estimativas $x^*(w)$ cumpram com a restrição de que algum conjunto de marginais são iguais às correspondentes marginais da tabela observada, pode ser escrita como

$$C \underline{p}^* = \underline{\theta} \quad \therefore C N \underline{p}^* = N \underline{\theta} \quad \therefore C \underline{x}^* = N \underline{\theta} \quad \therefore C \underline{x} = N \underline{\theta} \quad \therefore$$

$$C \underline{x}^* = C \underline{x} \quad \therefore T' \underline{x}^* = T' \underline{x}.$$

As colunas de \underline{T} , que implicam uma restrição marginal, podem ser consideradas como funções indicadoras das respectivas marginais, isto é, $T_i(w)$ estará constituída por $\underline{1}$ e $\underline{0}$ para qualquer casela w , dependendo se a casela faz parte ou não da respectiva frequência marginal. Por exemplo, seja

$$\Omega = \{(1, 1, 1), (1, 1, 2), (1, 2, 1), (1, 2, 2), (2, 1, 1), (2, 1, 2), (2, 2, 1), (2, 2, 2)\}, \text{ e considere a seguinte}$$

restrição: $x^*_{i..} = x_{i..}$, $i = 1, 2$. Neste caso, tem-se que

$$T \underline{\tau} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \tau_0 \\ \tau_1 \end{bmatrix} = [I_0(w) \ I_1(w)] \begin{bmatrix} \tau_0 \\ \tau_1 \end{bmatrix}$$

$T_0(w) = 1$, para todo w , de forma de satisfazer a restrição natural

$$\sum_{\Omega} x^*(w) = \sum_{\Omega} x(w) = N, \text{ isto é, } \sum_{\Omega} T_0(w) x^*(w) = \sum_{\Omega} T_0(w)$$

$x(w) = N$. Pela restrição imposta, cumpre-se também que

$$\sum_{\Omega} T_1(w) x^*(w) = \sum_{\Omega} T_1(w) x(w) = x_{1..} \quad (3)$$

A equação (1), pode ser expressa como

$$\frac{x^*(w)}{N \pi(w)} = e^{\tau_0 + \tau_1 T_1(w) + \tau_2 T_2(w) + \dots + \tau_r T_r(w)} \quad (4)$$

e em concordância com (2) $\ln(x^*/N \pi) = T \underline{\tau}$

O parâmetro normalizador τ_0 sera denotado por $\underline{\mu}$ e os parâmetros $\tau_1, \tau_2, \dots, \tau_r$ representarão os efeitos principais e as interações. Tem-se então o modelo log-linear:

$$\ln \frac{x^*(w)}{N \pi(w)} = L + \tau_1 T_1(w) + \tau_2 T_2(w) + \dots + \tau_r T_r(w),$$

$$w = 1, 2, \dots, \Omega$$

Da equação (4), tem-se que

$$\sum_{\Omega} x^*(w) = N = \sum_{\Omega} e^{L + \tau_1 T_1(w) + \tau_2 T_2(w) + \dots + \tau_r T_r(w)} N \pi(w) \dots$$

$$1 = e^L \sum_{\Omega} e^{\tau_1 T_1(w) + \tau_2 T_2(w) + \dots + \tau_r T_r(w)} \pi(w) \dots$$

$$\ln 1 = 0 = L + \ln \sum_{\Omega} e^{\tau_1 T_1(w) + \tau_2 T_2(w) + \dots + \tau_r T_r(w)} \pi(w),$$

$$\text{mas } \sum_{\Omega} e^{\tau_1 T_1(w) + \tau_2 T_2(w) + \dots + \tau_r T_r(w)} \pi(w) = M(\tau_1, \tau_2, \dots, \tau_r) \dots$$

$$L = - \ln M(\tau_1, \tau_2, \dots, \tau_r) \quad (5)$$

Lembre que $N \pi(w)$ é uma constante, já que como foi estabelecido anteriormente, $\pi(w)$ denota a distribuição uniforme. Desse modo $N \pi(w)$ pode ser absorvido por L e o modelo log-linear que se adotará é o seguinte:

$$\ln x^*(w) = \tau_0 + \tau_1 T_1(w) + \tau_2 T_2(w) + \dots + \tau_r T_r(w) \quad (6),$$

em forma matricial

$$\ln (x^*) = T \underline{\tau} \quad (7) \text{ ou}$$

$$\ln \widehat{E(X)} = T \underline{\tau} \quad (8)$$

onde a estimativa da $E[X_{ijk}]$ é obtida pelo método da informação mínima de discriminação e, sob H_0 , é expressa pela equação (5) do ítem 3.5.4.2.2., página 95.

Por analogia com o modelo (1), do Quadro 2, do ítem 3.3.5., página 44, o modelo log-linear sob H_0 , é

$$\ln x^*(ijk) = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} \quad (9).$$

Como as restrições impostas são:

$$\sum_i \sum_j \sum_k x_{ijk} = \sum_i \sum_j \sum_k x^*_{ijk} = N,$$

$$\sum_j \sum_k x_{ijk} = \sum_j \sum_k x^*_{ijk} = x_{i..},$$

$$\sum_i \sum_k x_{ijk} = \sum_i \sum_k x^*_{ijk} = x_{.j.} \quad e$$

$$\sum_i \sum_j x_{ijk} = \sum_i \sum_j x^*_{ijk} = x_{..k}, \quad \text{então o número de parâmetros no modelo log-linear (9) será:}$$

$$\tau_0 = 1,$$

$$\tau_1^{(1)}, \tau_1^{(2)}, \tau_1^{(3)}, \dots, \tau_1^{(i-1)} = R - 1,$$

$$\tau_2^{(1)}, \tau_2^{(2)}, \tau_2^{(3)}, \dots, \tau_2^{(j-1)} = S - 1 \quad e$$

$$\tau_3^{(1)}, \tau_3^{(2)}, \tau_3^{(3)}, \dots, \tau_3^{(k-1)} = T - 1.$$

Ao se testar a hipótese

$$H_0 : x_{ijk} = \frac{x_{i..} \cdot x_{.j.} \cdot x_{..k}}{N^2},$$

estã-se testando o modelo log-linear

$$\ln x_{ijk}^* = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)}, \text{ que tem } 1 + (R - 1)$$

+ (S - 1) + (T - 1) parâmetros. O modelo completo ou mo-

$$\text{delo saturado } \ln x_{ijk} = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} +$$

$$\tau_{12}^{(ij)} + \tau_{13}^{(ik)} + \tau_{23}^{(jk)} + \tau_{123}^{(ijk)}, \text{ tem}$$

$$1 + (R - 1) + (S - 1) + (T - 1) + (R - 1)(S - 1) + (R - 1)$$

$$(T - 1) + (S - 1)(T - 1) + (R - 1)(S - 1)(T - 1) \text{ parâ-}$$

metros. Então, a estatística

$$2 \uparrow (x : x^*) = 2 \sum_i \sum_j \sum_k x_{ijk} \ln \frac{x_{ijk}}{x_{ijk}^*}, \text{ é distribuída como}$$

$$\chi^2 \text{ com } (R - 1)(S - 1) + (R - 1)(T - 1) + (S - 1)(T - 1)$$

$$+ (R - 1)(S - 1)(T - 1) = RST - R - S - T + 2 \text{ graus de}$$

liberdade. Em geral, FIENBERG (1978), dá a seguinte fórmula para determinar os graus de liberdade:

$$G.L. = n^\circ \text{ de caselas} - n^\circ \text{ de parâmetros ajustados.}$$

Têm-se então, que para testar a hipótese nula $H_0 : P_{ijk} = P_{i..} P_{.j.} P_{..k}$, que é equivalente a

testar $H_0 : x_{ijk} = \frac{x_{i..} \cdot x_{.j.} \cdot x_{..k}}{N^2}$, usa-se a estatística

da informação mínima de discriminação

$2 \hat{I}(x : x^*) = 2 \sum_i \sum_j \sum_k x_{ijk} \ln \frac{x_{ijk}}{x^*_{ijk}}$, que é distribuída

assintoticamente como um χ^2 com $R \times S \times T - R - S - T + 2$, graus de liberdade, rejeitando H_0 se

$2 \hat{I}(x : x^*) > \chi^2_{(RST - R - S - T + 2)}$, isto é, se $\chi^2_{\text{calc}} > \chi^2_{\text{tab}}$, com $RST - R - S - T + 2$ graus de liberdade e um nível de significância α escolhido, geralmente $\alpha = 0.05$ ou $\alpha = 0.01$.

3.5.4.4. Estimativa dos parâmetros do modelo log-linear

No ítem anterior, demonstrou-se que o modelo log-linear, sob

$H_0 : x_{ijk} = \frac{x_{i..} \cdot x_{.j.} \cdot x_{..k}}{N^2}$, é $\ln x_{ijk} = \tau_0 + \tau_1^{(i)} +$

$\tau_2^{(j)} + \tau_3^{(k)}$, com $(RST - 2)$ parâmetros. Em forma matricial

$$\ln \widehat{E[X]} = T \underline{\tau} \quad (1)$$

No modelo (1), primeiro serão estimadas as frequências esperadas nas $R \times S \times T$ caselas aplicando o princípio da informação mínima de discriminação e a partir dessas estimativas serão calculadas as estimativas dos parâmetros τ de maneira simples e direta. Seja por exemplo uma tabela

de contingência $3 \times 2 \times 2$, que tem 12 caselas. A representação log-linear, de acordo com o modelo (1) é:

$$\begin{array}{c}
 \text{casela} \\
 \begin{bmatrix} 111 \\ 112 \\ 121 \\ 122 \\ 211 \\ 212 \\ 221 \\ 222 \\ 311 \\ 312 \\ 321 \\ 322 \end{bmatrix}
 \end{array}
 =
 \begin{array}{c}
 \tau_0 \quad \tau_1^{(1)} \quad \tau_1^{(2)} \quad \tau_2^{(1)} \quad \tau_3^{(1)} \\
 \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}
 \end{array}
 \begin{array}{c}
 \tau_0 \\
 \tau_1^{(1)} \\
 \tau_1^{(2)} \\
 \tau_2^{(1)} \\
 \tau_3^{(1)}
 \end{array}$$

Tem-se então que, por exemplo;

$$\tau_0 = \ln x^*_{322}, \quad (2),$$

$$\ln x^*_{321} = \tau_0 + \tau_3^{(1)} = \ln x^*_{322} + \tau_3^{(1)} \quad \therefore$$

$$\tau_3^{(1)} = \ln \frac{x^*_{321}}{x^*_{322}} \quad (3),$$

$$\ln x^*_{312} = \tau_0 + \tau_2^{(1)} = \ln x^*_{322} + \tau_2^{(1)} \quad \therefore$$

$$\tau_2^{(1)} = \ln \frac{x^*_{312}}{x^*_{322}} \quad (4),$$

e assim sucessivamente.

3.5.4.5. Estimativa da matriz de covariância das estimativas dos parâmetros

KU & KULLBACK (1974), estabelecem que a estimativa assintótica da matriz de variâncias e covariâncias dos τ 's para x^* calculada sob H_0 é obtida da seguinte maneira:

(i) Compute $S = T'DT$, onde T é a matriz de delineamento de ordem $\Omega \times (r + 1)$, para a representação log-linear de x^* e D é uma matriz diagonal cujos elementos são os valores estimados x^* , na mesma ordem em que foram colocadas as caselas na representação log-linear. D é uma matriz $\Omega \times \Omega$.

(ii) Particione a matriz S , como sendo

$$S = \begin{bmatrix} S_{11} & | & S_{12} \\ \hline & & \\ S_{21} & | & S_{22} \end{bmatrix}, \text{ onde } S_{11} \text{ é } 1 \times 1.$$

(iii) A matriz de variâncias e covariâncias dos τ 's é

$S_{22.1}^{-1} = (S_{22} - S_{21} S_{11}^{-1} S_{12})^{-1}$. Para efeitos computacionais, uma vez obtida a matriz S , é fácil obter $S_{22.1}^{-1}$, já que

$$S^{-1} = \begin{bmatrix} S_{11} & | & S_{12} \\ \hline & & \\ S_{21} & | & S_{22} \end{bmatrix}^{-1} = \begin{bmatrix} S_{11}^{-1} & | & M \\ \hline M' & | & S_{22.1}^{-1} \end{bmatrix}, \text{ onde}$$

$$M = -S_{11}^{-1} S_{12} S_{22.1}^{-1} \quad e$$

$S_{22.1}^{-1}$ é de ordem $(R + S + T - 3) \times (R + S + T - 3)$.

3.5.5. Independência parcial

Numa tabela de contingência tridimensional, ao testar-se a hipótese de independência parcial, esta-se testando se uma variável é independente das outras duas variáveis conjuntamente. Têm-se então três hipóteses deste tipo:

(i) \underline{R} é independente de \underline{S} e \underline{T} conjuntamente

(ii) \underline{S} é independente de \underline{R} e \underline{T} conjuntamente

(iii) \underline{T} é independente de \underline{R} e \underline{S} conjuntamente.

Serão apresentados os resultados correspondentes à hipótese (i), isto é, \underline{R} é independente de \underline{S} e \underline{T} conjuntamente. Por analogia, é fácil se obter os resultados correspondentes às hipóteses (ii) e (iii).

3.5.5.1. Representação log-linear

Seja uma tabela de contingência conhecida $\pi(ijk)$, $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ e $k = 1, 2, \dots, t$; $\sum_{i,j,k} \pi(ijk) = 1$ e é requerida uma tabela de

contingência $p(ijk)$ tal que as probabilidades marginais $p_{i..}$ e $p_{.jk}$, sejam fixas e iguais às correspondentes probabilidades observadas numa tabela de contingência $R \times S \times T$, de tal forma que

$$I(p : \pi) = \sum_i \sum_j \sum_k p(ijk) \ln \frac{p(ijk)}{\pi(ijk)}, \text{ seja um m\u00ednimo para}$$

toda $p(ijk)$ com as probabilidades marginais dadas.

Para aplicar o teorema de KULLBACK (1959), que foi enunciado no item 3.4.3., necessitam-se estat\u00edsticas $T_{\alpha..}(ijk)$, $\alpha = 1, 2, \dots, r$, $T_{.\beta\gamma}(ijk)$, $\beta = 1, 2, \dots, s$ e $\gamma = 1, 2, \dots, t$, tais

$$\text{que } \sum_i \sum_j \sum_k T_{\alpha..}(ijk) p(ijk) = p_{i..} \text{ e}$$

$$\sum_i \sum_j \sum_k T_{.\beta\gamma}(ijk) p(ijk) = p_{.jk}$$

Por procedimento an\u00e1logo ao efetuado no item 3.5.4.1., tem-se que

$$p^*(ijk) = a_i b_{jk} \pi(ijk) \quad (1),$$

$$p_{i..} = p^*_{i..} = a_i \sum_j \sum_k b_{jk} \pi(ijk) \quad (2) \text{ e}$$

$$p_{.jk} = p^*_{.jk} = b_{jk} \sum_i a_i \pi(ijk) \quad (3) \text{ e}$$

$$I(p^* : \pi) = \sum_i p_{i..} \ln a_i + \sum_j \sum_k \ln b_{jk} \quad (4)$$

De (1), tem-se que
$$\ln \frac{p^*(ijk)}{\pi(ijk)} = \ln a_i + \ln b_{jk} \quad (5)$$

A equação (5) representa o modelo log-linear obtido sob a hipótese de independência parcial da variável categorizada \underline{R} das variáveis \underline{S} e \underline{T} conjuntamente.

A equação (5) pode ser expressa como

$$\ln p^*(ijk) = \ln \pi(ijk) + \tau_{i..} + \tau_{.jk} \quad (6),$$

onde os τ 's são funções das probabilidades marginais unidimensionais da variável \underline{R} e das probabilidades marginais bidimensionais das variáveis \underline{S} e \underline{T} .

3.5.5.2. Estimativas das freqüências esperadas

3.5.5.2.1. Processo iterativo

Das equações (1), (2) e (3) do ítem 3.5.5.1., têm-se o seguinte sistema de equações:

$$\left. \begin{aligned} p^*(ijk) &= a_i b_{jk} \pi(ijk) \\ p_{i..} &= a_i \sum_j \sum_k b_{jk} \pi(ijk) \\ p_{.jk} &= b_{jk} \sum_i a_i \pi(ijk) \end{aligned} \right\} \quad (1)$$

A solução iterativa do sistema de equações (1), consiste em obter-se $p^*(ijk)$ em lugar de resolver para a_i e b_{jk} . Este processo iterativo será realizado ajustando-se em forma sistemática, as probabilidades mar-

ginais unidimensionais da variável \underline{R} e as probabilidades marginais bidimensionais das variáveis \underline{S} e \underline{T} . Considere os seguintes passos:

$$(i) \quad p_{i..} = a_i^{(1)} \sum_j \sum_k b_{jk}^{(1)} \pi(ijk) \implies p_{(ijk)}^{(1)} =$$

$$a_i^{(1)} b_{jk}^{(1)} \pi(ijk),$$

$$(ii) \quad p_{.jk} = b_{jk}^{(2)} \sum_i a_i^{(1)} \pi(ijk) \implies p_{ijk}^{(2)} =$$

$$a_i^{(1)} b_{jk}^{(2)} \pi(ijk),$$

$$(iii) \quad p_{i..} = a_i^{(2)} \sum_j \sum_k b_{jk}^{(2)} \pi(ijk) \implies p_{(ijk)}^{(3)} =$$

$$a_i^{(2)} b_{jk}^{(2)} \pi(ijk),$$

$$(iv) \quad p_{.jk} = b_{jk}^{(3)} \sum_i a_i^{(2)} \pi(ijk) \implies p_{(ijk)}^{(4)} =$$

$$a_i^{(2)} b_{jk}^{(3)} \pi(ijk), \quad \text{e assim sucessivamente. Seja ini-$$

cialmente $b_{jk}^{(1)} = 1$, logo, tem-se que

$$(i) \quad p_{i..} = a_i^{(1)} \sum_j \sum_k \pi(ijk) = a_i^{(1)} \pi_{i..} \dots a_i^{(1)} = \frac{p_{i..}}{\pi_{i..}},$$

$$\text{então } p_{(ijk)}^{(1)} = a_i^{(1)} \pi(ijk) = \frac{p_{i..}}{\pi_{i..}} \pi(ijk)$$

$$(ii) \quad p_{.jk} = b_{jk}^{(2)} \sum_i a_i^{(1)} \pi(ijk) \therefore b_{jk}^{(2)} = \frac{p_{.jk}}{\sum_i a_i^{(1)} \pi(ijk)},$$

$$\text{mas } \sum_i p_{(ijk)}^{(1)} = \sum_i a_i^{(1)} \pi_{(ijk)} = p_{.jk}^{(1)}, \text{ logo}$$

$$b_{jk}^{(2)} = \frac{p_{.jk}}{p_{.jk}^{(1)}} \text{ e } p_{(ijk)}^{(2)} = b_{jk}^{(2)} a_i^{(1)} \pi_{(ijk)} = \frac{p_{.jk}}{p_{.jk}^{(1)}} p_{(ijk)}^{(1)}$$

$$(iii) \quad p_{i..} = a_i^{(2)} \sum_{jk} b_{jk}^{(2)} \pi_{(ijk)} \dots$$

$$a_i^{(2)} = \frac{p_{i..}}{\sum_{jk} b_{jk}^{(2)} \pi_{(ijk)}}, \text{ mas}$$

$$\sum_{jk} p_{(ijk)}^{(2)} = p_{i..}^{(2)} = a_i^{(1)} \sum_{jk} b_{jk}^{(2)} \pi_{(ijk)}, \text{ logo}$$

$$a_{(i)}^{(2)} = \frac{p_{i..} a_i^{(1)}}{p_{i..}^{(2)}}. \text{ Então}$$

$$p_{(ijk)}^{(3)} = a_i^{(2)} b_{jk}^{(2)} \pi_{(ijk)} = \frac{p_{i..} a_i^{(1)}}{p_{i..}^{(2)}} \frac{p_{(ijk)}^{(2)}}{a_i^{(1)}} \dots$$

$$p_{(ijk)}^{(3)} = \frac{p_{i..}}{p_{i..}^{(2)}} p_{(ijk)}^{(2)}$$

(iv) Analogamente, obtem-se

$$p_{ijk}^{(4)} = \frac{p_{.jk}}{p_{.jk}^{(3)}} p_{(ijk)}^{(3)}$$

Os valores da p^* -tabela devem ser computados por um processo iterativo de ajustamento da π -tabela, até satisfazer as restrições marginais dadas, levando-se em conta uma tolerância previamente estabelecida. Por exemplo, po-

de-se fixar que a diferença entre a marginal observada e a marginal computada iterativamente não exceda de 0,1 ou 0.01.

Em geral, para a hipótese de independência parcial entre a variável R e as variáveis S e T , conjuntamente, o processo iterativo pode ser expresso como:

$$P_{(ijk)}^{(2n+1)} = \frac{P_{i..}}{P_{i..}^{(2n)}} P_{(ijk)}^{(2n)},$$

$$P_{(ijk)}^{(2n+2)} = \frac{P_{.jk}}{P_{.jk}^{(2n+1)}} P_{(ijk)}^{(2n+1)}, \quad n = 0, 1, 2, \dots \quad (2)$$

Esse processo iterativo é convergente e tem solução única (IRELAND & KULLBACK, 1968).

3.5.5.2.2. Computação das estimativas das frequências esperadas

No processo iterativo expresso pelas equações (2), do ítem anterior é recomendável considerar-se uma distribuição $\pi_{(ijk)}$ na que se cumpram ou verifiquem as restrições impostas sobre as marginais. Uma distribuição com essas características é a distribuição uniforme, isto é, $\pi_{(ijk)} = \frac{1}{R \times S \times T}$, para toda (i, j, k) -

ésima casela, $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ e $k = 1, 2, \dots, t$. Quando $n = 0$, obtêm-se

$$p_{ijk}^{(1)} = \frac{p_{i..}}{p_{i..}^{(0)}} p_{(ijk)}^{(0)} \quad \text{e} \quad p_{ijk}^{(2)} = \frac{p_{.jk}}{p_{.jk}^{(1)}} p_{(ijk)}^{(1)}, \quad \text{onde}$$

$$p_{ijk}^{(0)} = \pi(ijk) = \frac{1}{R \times S \times T}, \quad \text{logo} \quad p_{i..}^{(0)} = \frac{1}{R} \quad \therefore$$

$$p_{(ijk)}^{(1)} = \frac{p_{i..}}{\frac{1}{R}} \frac{1}{R \times S \times T} = \frac{p_{i..}}{S \times T} \quad \therefore$$

$$p_{.jk}^{(1)} = \sum_i p_{(ijk)}^{(1)} = \sum_i \frac{p_{i..}}{S \times T} = \frac{1}{S \times T}, \quad \text{logo}$$

$$p_{ijk}^{(2)} = \frac{p_{.jk}}{\frac{1}{S \times T}} \frac{p_{i..}}{S \times T} = p_{.jk} p_{i..}$$

$$p_{i..}^{(2)} = \sum_j \sum_k p_{ijk}^{(2)} = \sum_j \sum_k p_{.jk} p_{i..} = p_{i..}, \quad \text{logo}$$

$$p_{(ijk)}^{(3)} = \frac{p_{i..}}{p_{i..}} p_{i..} p_{.jk} = p_{i..} p_{.jk}.$$

Observe que $p_{(ijk)}^{(3)} = p_{(ijk)}^{(2)}$, além disso

$$p_{.jk}^{(3)} = \sum_i p_{(ijk)}^{(3)} = p_{.jk}, \quad \text{então, a tabela de contingência}$$

construída dessa forma, cumpre com as restrições impostas, isto é, as marginais da variável R e das variáveis S e T, conjuntamente, são iguais tanto na tabela de contingência observada como na tabela de contingência estimada. Tem-se então que $p^*(ijk) = p_{i..} p_{.jk}$. Aqui é válida também a observação (iii) do ítem 3.5.4.1., página 88. Logo,

$$p^*(ijk) = \hat{p}_{i..} \hat{p}_{.jk}, \text{ onde } \hat{p}_{i..} = \frac{x_{i..}}{N}, \hat{p}_{.jk} = \frac{x_{.jk}}{N} \text{ e}$$

$$N p^*(ijk) = x^*_{ijk}. \text{ Portanto, } x^*(ijk) = \frac{x_{i..} x_{.jk}}{N} \quad (1)$$

A expressão (1) estima a frequência esperada na (i, j, k)-ésima casela, sob H_0 , através do método da informação mínima de discriminação e coincide com a estimativa obtida pelo método da máxima verossimilhança.

3.5.5.3. Estatística da informação mínima de discriminação

Testar a hipótese $H_0 : p_{ijk} = p_{i..} p_{.jk}$ é equivalente a testar

$$H_0 : x_{ijk} = \frac{x_{i..} x_{.jk}}{N}. \text{ A estatística da informação mínima}$$

de discriminação, $2 N I(\hat{p} : p^*) = 2 I(x : x^*) =$

$$2 \sum_{i,j,k} x_{ijk} \ln \frac{x_{ijk}}{x^*_{ijk}} \quad (1),$$

sob H_0 , é distribuída assintoticamente como um χ^2 com um número de graus de liberdade calculado com base no teorema de Wilks, enunciado no ítem 3.4.4.2., página 72.

Em forma similar ao procedimento seguido ítem 3.5.4.3., página 96, e para não repetir os argumentos usados apresenta-se o seguinte resultado: o modelo lo-linear, sob H_0 , é:

$$\ln x_{(ijk)} = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} + \tau_{23}^{(jk)} \quad (2)$$

É importante destacar que as restrições impostas na tabela estimada são $x_{i..}^* = x_{i..}$, $i = 1, 2, \dots, r$ e

$$x_{.jk}^* = x_{.jk}, \quad j = 1, 2, \dots, s \quad \text{e} \quad k = 1, 2, \dots, t.$$

Demonstrou-se que sob H_0 , as estimativas das frequências

$$\text{esperadas são} \quad x_{(ijk)}^* = \frac{x_{i..} \cdot x_{.jk}}{N}.$$

Ao considerar-se fixas as marginais bidimensionais das variáveis \underline{S} e \underline{T} , obviamente também consideram-se fixas as marginais unidimensionais da variável \underline{S} e da variável \underline{T} , isto é,

$$x_{.j.} = \sum_k x_{.jk} \quad \text{e} \quad x_{..k} = \sum_j x_{.jk}.$$

Por outro lado, no ítem 3.3.2., página 39, sobre modelos hierárquicos, foi estabelecido que se o parâmetro $\tau_{23}^{(jk)}$ está presente no modelo então os termos $\tau_2^{(j)}$ e $\tau_3^{(k)}$ também devem estar.

Como as restrições impostas são:

$$\sum_i \sum_j \sum_k x_{ijk}^* = \sum_i \sum_j \sum_k x_{ijk} = N,$$

$$\sum_j \sum_k x_{ijk}^* = \sum_j \sum_k x_{ijk} = x_{i..} \quad \text{e}$$

$$\sum_i x_{ijk}^* = \sum_i x_{ijk} = x_{.jk}, \quad \text{então o número de parâmetros no}$$

modelo log-linear (2) é dado por: $\tau_0 = 1$,

$$\tau_1^{(1)}, \tau_1^{(2)}, \dots, \tau_1^{(i-1)} = R - 1,$$

$$\tau_2^{(1)}, \tau_2^{(2)}, \dots, \tau_2^{(j-1)} = S - 1,$$

$$\tau_3^{(1)}, \tau_3^{(2)}, \dots, \tau_3^{(k-1)} = T - 1,$$

$$\tau_{23}^{(11)}, \tau_{23}^{(12)}, \dots, \tau_{23}^{(1(K-1))}, \dots, \tau_{23}^{((j-1)(k-1))}$$

$$= (S - 1)(T - 1).$$

Logo, a estatística $2 \hat{T}(x : x^*)$, dada pela expressão (1) é distribuída como um χ^2 com $(R - 1)(S - 1) + (R - 1)(T - 1) + (R - 1)(S - 1)(T - 1) = (R - 1)(ST - 1)$ graus de liberdade.

Tem-se que então, que para testar a hipótese nula $H_0 : p_{ijk} = P_{i..} P_{.jk}$, que é equivalente a testar

$$H_0 : x_{ijk} = \frac{x_{i..} x_{.jk}}{N}, \text{ usa-se a estatística da informa-}$$

ção mínima de discriminação $2 \hat{T}(x : x^*)$, que é distribuída assintoticamente como um χ^2 com $(R - 1)(ST - 1)$, graus de liberdade, rejeitando-se H_0 , se $2 \hat{T}(x : x^*) > \chi^2((R - 1)(ST - 1))$, isto é, se $\chi^2_{\text{calc}} > \chi^2_{\text{tab}}$, com $(R - 1)(ST - 1)$ g. de l. ao nível de significância α desejado.

3.5.5.4. Estimativas dos parâmetros do modelo log-linear

Sob $H_0 : x_{ijk} = \frac{x_{i..} \cdot x_{.jk}}{N}$, o modelo log-

linear e $\ln x_{(ijk)} = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} + \tau_{23}^{(jk)}$

com $R + ST - 1$ parâmetros independentes. Em forma matricial $\ln E \widehat{[X]} = T \underline{\tau}$. (1)

No modelo (1), primeiro serão estimadas as frequências esperadas nas $R \times S \times T$ caselas usando o princípio da informação mínima de discriminação e a partir dessas estimativas serão calculadas as estimativas dos parâmetros τ de maneira fácil e direta, tal como foi mostrado no ítem 3.5.4.4., página 103.

3.5.5.5. Estimativa da matriz de covariância das estimativas dos parâmetros

Para se estimar a matriz de variâncias e covariâncias dos τ 's, para x^* calculada sob H_0 , é seguido o mesmo procedimento apresentado no ítem 3.5.4.5..

3.5.6. Independência condicional

Numa tabela de contingência tridimensional ao testar a hipótese de independência condicional, está-se testando a hipótese de que duas variáveis são independentes dentro de cada categoria ou nível da terceira

variável. Têm-se então três hipóteses deste tipo:

(i) \underline{R} e \underline{S} são independentes dentro de cada nível de \underline{T}

(ii) \underline{R} e \underline{T} são independentes dentro de cada nível de \underline{S} e

(iii) \underline{S} e \underline{T} são independentes dentro de cada nível de \underline{R} .

Serão apresentados os resultados correspondentes à hipótese (i). Por analogia, é fácil obter os resultados correspondentes às hipóteses (ii) e (iii).

3.5.6.1. Representação log-linear

Sem repetir os detalhes dos itens 3.5.4.1. e 3.5.5.1., obtem-se:

$$p^*(ijk) = a_{ik} b_{jk} \pi(ijk) \quad (1)$$

A equação (1) representa o modelo log-linear obtido sob a hipótese de independência condicional das variáveis categorizadas \underline{R} e \underline{S} , dentro de cada nível da variável \underline{T} . A equação (1) pode ser expressa como

$$\ln p^*(i \cdot k) = \ln \pi(\cdot \cdot k) + \tau_{i \cdot k} + \tau_{\cdot j k}, \quad \text{onde os } \tau \text{'s são}$$

funções das marginais bidimensionais das variáveis \underline{R} e \underline{T} , das variáveis \underline{S} e \underline{T} e das marginais unidimensionais da variável \underline{T} .

3.5.6.2. Estimativas das freqüências esperadas

3.5.6.2.1. Processo iterativo

Por um procedimento análogo considerado no ítem 3.5.5.2.1., para hipótese de independência condicional entre as variáveis R e S, dentro de cada nível da variável T, o processo iterativo para obter as estimativas das freqüências esperadas na (i, j, k)-ésima casela pode ser expresso como:

$$p^{(2n+1)}(ijk) = \frac{P(i.k)}{P(i.k)} p^{(2n)}(ijk),$$

$$p^{(2n+2)}(ijk) = \frac{P(.jk)}{P(.jk)} p^{(2n+1)}(ijk)$$

Esse processo é convergente e tem solução única (IRELAND & KULLBACK, 1968).

3.5.6.2.2. Computação das estimativas das freqüências esperadas

Por um procedimento similar ao seguido no ítem 3.5.5.2.2., obtem-se que:

$$p^*(ijk) = \frac{\hat{p}_{i.k} \hat{p}_{.jk}}{\hat{p}_{..k}}, \text{ onde } \hat{p}_{i.k} = \frac{x_{i.k}}{N}; \quad \hat{p}_{.jk} = \frac{x_{.jk}}{N};$$

$$\hat{p}_{..k} = \frac{x_{..k}}{N} \text{ e } N p^*(ijk) = x^*(ijk). \text{ Portanto}$$

$x^*(ijk) = \frac{x_{i.k} x_{.jk}}{x_{..k}}$, que é a estimativa da frequência esperada na (i, j, k) -ésima casela, sob H_0 , obtida através do método da informação mínima de discriminação e coincide com a estimativa obtida pelo método da máxima verossimilhança.

3.5.6.3. Estatística da informação mínima de discriminação

Testar a hipótese $H_0 : p_{ijk} = \frac{p_{.ik} p_{.jk}}{p_{..k}}$ é

equivalente a testar $H_0 : x_{ijk} = \frac{x_{i.k} x_{.jk}}{x_{..k}}$. A estatística

da informação mínima de discriminação

$$2 N \ln(\hat{p} : p^*) = 2 \ln(x : x^*) = 2 \sum_i \sum_j \sum_k x_{ijk} \ln \frac{x_{ijk}}{x^*_{ijk}} \quad (1),$$

sob H_0 , é distribuída assintoticamente como um χ^2 com um número de graus de liberdade calculado com base no teorema de Wilks.

De forma similar ao procedimento seguido no item 3.5.4.3., página 96, apresenta-se o seguinte resultado. O modelo log-linear, sob H_0 , é

$$\ln x_{ijk} = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} + \tau_{13}^{(ik)} + \tau_{23}^{(jk)} \quad (2)$$

O modelo log-linear (2), tem $1 + (R - 1) + (S - 1) +$

$(T - 1) + (R - 1)(T - 1) + (S - 1)(T - 1) = T(R + S - 1)$ parâmetros. Logo, a estatística $2 \hat{T}(x : x^*)$, dada pela expressão (1) é distribuída assintoticamente como um χ^2 com $(R-1)(S - 1) + (R - 1)(S - 1)(T - 1) = T(R - 1)(S - 1)$ graus de liberdade.

O valor da estatística $2 \hat{T}(x : x^*)$ é comparada com o valor tabelado de um $\chi^2(T(R - 1)(S - 1))$ ao nível de significância α e se $\chi^2_{\text{calc}} > \chi^2_{\text{tab}}$, rejeita-se H_0 .

3.5.6.4. Estimativa dos parâmetros do modelo log-linear

Os parâmetros do modelo log-linear

$$\ln x_{ijk} = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} + \tau_{13}^{(ik)} + \tau_{23}^{(jk)}, \text{ se}$$

rão estimados de maneira simples e direta de acordo como foi estabelecido no item 3.5.4.4..

3.5.6.5. Estimativa da matriz de covariâncias das estimativas dos parâmetros

Para se estimar a matriz de variâncias e covariâncias dos τ 's, para x^* calculada sob H_0 , segue-se o mesmo procedimento apresentado no item 3.5.4.5.

3.5.7. Inexistência de interação de segunda ordem

Numa tabela de contingência $2 \times 2 \times 2$, BARTEETT (1935), definiu a não interação de segunda ordem ou não interação entre as três variáveis se ocorrer a igualdade

$$\frac{P(111) P(221)}{P(121) P(211)} = \frac{P(112) P(222)}{P(122) P(212)} \quad (1)$$

Observe que (1), também pode ser escrito como

$$\frac{P(111) P(122)}{P(121) P(112)} = \frac{P(211) P(222)}{P(221) P(212)} \quad \text{ou}$$

$$\frac{P(111) P(212)}{P(112) P(211)} = \frac{P(121) P(222)}{P(122) P(221)}$$

ROY & KASTENBAUM (1956), obtiveram um conjunto de restrições de não interação de segunda ordem numa tabela de contingência $R \times S \times T$, que pode ser expresso como

$$\frac{P_{ijk} P_{rsk}}{P_{isk} P_{rjk}} = \frac{P_{rst} P_{ijt}}{P_{ist} P_{rjt}}, \quad \text{para} \quad \begin{array}{l} i = 1, 2, \dots, r-1 \\ j = 1, 2, \dots, s-1 \\ k = 1, 2, \dots, t-1 \end{array}$$

A hipótese ou modelo de não interação de segunda ordem é formulada e testada quando não é encontrado nenhum outro tipo de independência.

3.5.7.1. Representação log-linear

Quanto todas as marginais bidimensionais x_{ij} , $x_{i.k}$ e $x_{.jk}$ são consideradas como fixas, o que implica que as marginais unidimensionais também o são, usando-se um procedimento análogo ao apresentado no item 3.5.4.1., página 82, obtêm-se

$$p^*(ijk) = a_{ij} b_{jk} c_{ik} \pi(ijk) \quad (1)$$

$$p^*_{ij.} = p_{ij.} = a_{ij} \sum_k b_{jk} c_{ik} \pi(ijk) \quad (2)$$

$$p^*_{i.k} = p_{i.k} = c_{ik} \sum_j a_{ij} b_{jk} \pi(ijk) \quad (3)$$

$$p^*_{.jk} = p_{.jk} = b_{jk} \sum_i a_{ij} c_{ik} \pi(ijk) \quad (4)$$

A equação (1) representa o modelo log-linear obtido com as marginais bidimensionais consideradas fixas e pode-se expressar como

$$\ln p^*(ijk) = \ln \pi(ijk) + \tau_{ij.} + \tau_{i.k} + \tau_{.jk},$$

onde os τ 's são funções das marginais bidimensionais das variáveis R e S, R e T e S e T.

3.5.7.2. Estimativas das frequências esperadas

3.5.7.2.1. Processo iterativo

De forma similar à metodologia seguida no item 3.5.4.2.1., página 89, o processo iterativo para obter as estimativas das frequências esperadas na (i,j,k) -ésima casela, pode ser obtido de

$$P^{(3n+1)}(ijk) = \frac{P(ij.)^{(3n)}}{P(ij.)^{(3n)}} \cdot P^{(3n)}(ijk)$$

$$P^{(3n+2)}(ijk) = \frac{P(i.k)^{(3n+1)}}{P(i.k)^{(3n+1)}} \cdot P^{(3n+1)}(ijk)$$

$$P^{(3n+3)}(ijk) = \frac{P(.jk)^{(3n+2)}}{P(.jk)^{(3n+2)}} \cdot P^{(3n+2)}(ijk), \quad n = 0, 1, 2, \dots,$$

3.5.7.2.2. Computação das estimativas das frequências esperadas

Neste caso, ao finalizar o primeiro ciclo, não é possível determinar-se uma expressão direta para a estimativa de $p^*(ijk)$ em função das marginais bidimensionais. O processo iterativo pode ser posto em função das estimativas das frequências esperadas e das estimativas de máxima verossimilhança das probabilidades marginais, lembrando-se que:

$$Np^*(ijk) = x^*(ijk); \quad \hat{p}(ij.) = \frac{x_{ij.}}{N};$$

$$\hat{p}(i.k) = \frac{x_{i.k}}{N} \quad \text{e} \quad \hat{p}(.jk) = \frac{x_{.jk}}{N}. \quad \text{Logo,}$$

$$x_{(ijk)}^{(3n+1)} = \frac{x_{(ij.)}^{(3n)}}{x_{(ij.)}^{(3n)}} x_{(ijk)}^{(3n)},$$

$$x_{(ijk)}^{(3n+2)} = \frac{x_{(i.k)}^{(3n+1)}}{x_{(i.k)}^{(3n+1)}} x_{(ijk)}^{(3n+1)},$$

$$x_{(ijk)}^{(3n+3)} = \frac{x_{(.jk)}^{(3n+2)}}{x_{(.jk)}^{(3n+2)}} x_{(ijk)}^{(3n+2)}, \quad n = 0, 1, 2, \dots$$

O processo iterativo deve continuar até que as estimativas das frequências marginais bidimensionais sejam iguais às frequências marginais bidimensionais observadas, dentro de uma margem de erro que é estabelecido. Por exemplo, o processo iterativo deve continuar até que

$$|x_{ij.}^* - x_{ij.}| \leq 0,01$$

$$|x_{i.k}^* - x_{i.k}| \leq 0,01 \quad e$$

$$|x_{.jk}^* - x_{.jk}| \leq 0,01$$

3.5.7.3. Estatística da informação mínima de discriminação

No item 3.5.7.1, foi obtido que

$$p^*(ijk) = a_{ij} b_{jk} c_{ik} \pi(ijk).$$

Se $\pi(ijk) = \frac{1}{R \times S \times T} = \pi$, $p^*(ijk)$ satisfaz a definição de não interação de segunda ordem de Bartlett, isto é,

$$\frac{p^*(111) p^*(221)}{p^*(121) p^*(211)} = \frac{a_{11} b_{11} c_{11} \pi a_{22} b_{21} c_{21} \pi}{a_{12} b_{21} c_{11} \pi a_{21} b_{11} c_{21} \pi} = \frac{a_{11} a_{22}}{a_{12} a_{21}} \quad e$$

$$\frac{p^*(112) p^*(222)}{p^*(122) p^*(212)} = \frac{a_{11} a_{22}}{a_{12} a_{21}}, \text{ portanto}$$

$$\frac{p^*(111) p^*(221)}{p^*(121) p^*(211)} = \frac{p^*(112) p^*(222)}{p^*(122) p^*(212)}$$

A estatística da informação mínima de discriminação $2N\hat{T}(\hat{p}:p^*) = 2l(x:x^*) = 2 \sum_{i,j,k} x_{ijk} \ln \frac{x_{ijk}}{x_{ijk}^*}$ (1), sob H_0 , é distribuída assintoticamente com um número de graus de liberdade calculado em base ao teorema de Wilks.

De forma similar ao procedimento seguido no item 3.5.4.3., página 96, apresenta-se o seguinte resultado: O modelo linear sob H_0 , é

$$\ln x_{ijk} = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} + \tau_{12}^{(ij)} + \tau_{13}^{(ik)} + \tau_{23}^{(jk)} \quad (2)$$

O modelo log-linear (2) tem $1+(R-1) + (S-1) + (T-1) + (R-1)(S-1) + (R-1)(T-1) + (S-1)(T-1) = (R-1)(S+T-1) + ST$ parâmetros. Logo, a estatística da informação mínima de discriminação $2\hat{T}(x:x^*)$, dada pela equação (1) é distribuída como um χ^2 com $(R-1)(S-1)(T-1)$ graus de liberdade.

3.5.7.4. Estimativa dos parâmetros do modelo log-linear

Os parâmetros do modelo log-linear $\ln x_{ijk} = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} + \tau_{12}^{(ij)} + \tau_{13}^{(ik)} + \tau_{23}^{(jk)}$ são estimados de maneira fácil e direta de acordo como foi estabelecido no item 3.5.4.4.

3.5.7.5. Estimativa da matriz de covariâncias das estimativas dos parâmetros

Para se estimar a matriz de variâncias e covariâncias dos π 's, para x^* calculada sob H_0 , é seguido o mesmo procedimento apresentado no item 3.5.4.5.

3.6. Análise da informação

$$3.6.1. I(p:\pi) = I(p:p^*) + I(p^*:\pi)$$

Viu-se nos itens anteriores que uma vez que se decide testar uma hipótese H_0 , que, por sua vez implica testar um modelo log-linear, são estimadas as frequências esperadas em cada casela através do princípio da informação mínima de discriminação. Essas estimativas das frequências esperadas são obtidas por um processo iterativo convergente e de solução única. Posteriormente, a adequação do ajuste do modelo é testada usando a estatística da informação mínima de discriminação

$$2\hat{l}(x:x^*) = 2 \sum_i \sum_j \sum_k x_{(ijk)} \ln \frac{x_{(ijk)}}{x^*_{(ijk)}} \quad (1),$$

que, sob H_0 , é distribuída assintoticamente como um χ^2 com apropriados graus de liberdade. O valor dessa estatística é comparado com valores tabelados a um nível de significância α escolhido.

A análise de informação é baseada no seguinte teorema:

$$I(p:\pi) = I(p:p^*) + I(p^*:\pi) \quad (2),$$

se cumpre para p^* obtida pelo processo iterativo mostrado no item 3.5.4.2.1, página 89, onde as distribuições p^* e p têm as mesmas marginais especificadas".

O teorema será demonstrado para o caso em que todas as marginais bidimensionais são consideradas fixas, mas é válido para qualquer outra restrição sobre as marginais. Sabe-se que

$$\begin{aligned} \sum_k p_{ijk} &= \sum_k p_{ijk}^* = p_{ij.}, \\ \sum_j p_{ijk} &= \sum_j p_{ijk}^* = p_{i.k}, \\ \sum_i p_{ijk} &= \sum_i p_{ijk}^* = p_{.jk}. \end{aligned}$$

$$\text{Sabe-se também que } I(p^*:\pi) = \sum_{i,j,k} p_{ijk}^* \ln \frac{p_{ijk}^*}{\pi_{ijk}} \quad (3)$$

Da equação (1) do item 3.5.7.1, tem-se que

$$p_{ijk}^* = a_{ij} b_{jk} c_{ik} \pi_{ijk} \quad (4),$$

$$\text{Logo } I(p^*:\pi) = \sum_{ijk} p_{ijk}^* \ln p_{ijk}^* - \sum_{ijk} p_{ijk}^* \ln \pi_{ijk} =$$

$$\sum_{ijk} p_{ijk}^* \ln (a_{ij} b_{jk} c_{ik}) + \sum_{ijk} p_{ijk}^* \ln \pi_{ijk} -$$

$$\sum_{ijk} p_{ijk}^* \ln \pi_{ijk} \quad \therefore$$

$$I(p^*:\pi) = \sum_{ijk} p_{ijk}^* \ln a_{ij} + \sum_{ijk} p_{ijk}^* \ln b_{jk} + \sum_{ijk} p_{ijk}^* \ln c_{ik},$$

$$\text{mas } \sum_i \sum_j \sum_k p_{ijk}^* \ln a_{ij} = \sum_i \sum_j \ln a_{ij} \sum_k p_{ijk}^* = \sum_i \sum_j \ln a_{ij} \sum_k p_{ijk},$$

Logo $\sum_i \sum_j \sum_k p_{ijk}^* \ln a_{ij} = \sum_i \sum_j \ln a_{ij} p_{ij}$. Então

$$I(p^*:\pi) = \sum_{i,j} p_{ij} \ln a_{ij} + \sum_{j,k} p_{.jk} \ln b_{jk} + \sum_{i,k} p_{i.k} \ln c_{ik} \quad (5)$$

Por outro lado, $\sum_{i,j,k} p_{ijk} \ln \frac{p_{ijk}^*}{\pi_{ijk}} = \sum_{i,j,k} p_{ijk} \ln p_{ijk}^* - \sum_{i,j,k} p_{ijk} \ln \pi_{ijk} =$

$$= \sum_{i,j,k} p_{ijk} \ln (a_{ij} b_{jk} c_{ik} \pi_{ijk}) - \sum_{i,j,k} p_{ijk} \ln \pi_{ijk} \dots$$

$$\sum_{i,j,k} p_{ijk} \ln \frac{p_{ijk}^*}{\pi_{ijk}} = \sum_{i,j,k} p_{ijk}^* \ln \frac{p_{ijk}^*}{\pi_{ijk}} = I(p^*:\pi) \quad (6)$$

Sabe-se que $I(p:\pi) = \sum_{i,j,k} p_{ijk} \ln \frac{p_{ijk}}{\pi_{ijk}} \dots$

$$I(p:\pi) = \sum_{i,j,k} p_{ijk} \ln p_{ijk} - \sum_{i,j,k} p_{ijk} \ln \pi_{ijk} \quad (7)$$

Somando-se e subtraindo-se $\sum_{i,j,k} p_{ijk} \ln p_{ijk}^*$, tem-se que

$$I(p:\pi) = \sum_{i,j,k} p_{ijk} \ln p_{ijk} - \sum_{i,j,k} p_{ijk} \ln p_{ijk}^* +$$

$$\sum_{i,j,k} p_{ijk} \ln p_{ijk}^* - \sum_{i,j,k} p_{ijk} \ln \pi_{ijk} \dots$$

$$I(p:\pi) = \sum_{i,j,k} p_{ijk} \ln \frac{p_{ijk}}{p_{ijk}^*} + \sum_{i,j,k} p_{ijk} \ln \frac{p_{ijk}^*}{\pi_{ijk}},$$

mas $\sum_{i,j,k} p_{ijk} \ln \frac{p_{ijk}}{p_{ijk}^*} = I(p:p^*)$ e por (6)

$$\sum_{i,j,k} p_{ijk} \ln \frac{p_{ijk}^*}{\pi_{ijk}} = I(p^*:\pi); \text{ logo}$$

$$I(p:\pi) = I(p:p^*) + I(p^*:\pi) \text{ c.q.d.} \quad (8)$$

Evidentemente, cumpre-se também que

$$2I(p:\pi) = 2I(p:p^*) + 2I(p^*:\pi) \text{ e}$$

$$2I(x:n\pi) = 2I(x:x^*) + 2I(x^*:n\pi) \quad (9)$$

3.6.2. Quadros de análise de informação

A análise de informação está baseada na relação (9) para as estatísticas de informação mínima de discriminação. KU et alii (1971), estabelecem que se $Np_a^*(ijk) = x_a^*(ijk)$ é a estimativa obtida pelo método da informação mínima de discriminação correspondente a um conjunto H_a de marginais fixas e $Np_b^*(ijk) = x_b^*(ijk)$ é a estimativa obtida pelo método da informação mínima de discriminação correspondente a um conjunto H_b de marginais fixas, onde o conjunto H_a está explicitamente ou implicitamente contido no conjunto H_b , isto é, $H_a \subset H_b$, então

$$2l(x:n\pi) = 2l(x_a^*:n\pi) + 2l(x:x_a^*) \quad (1)$$

$$2l(x:n\pi) = 2l(x_b^*:n\pi) + 2l(x:x_b^*) \quad (2)$$

$$2l(x_b^*:n\pi) = 2l(x_a^*:n\pi) + 2l(x_b^*:x_a^*) \quad (3)$$

$$2l(x:x_a^*) = 2l(x_b^*:x_a^*) + 2l(x:x_b^*) \quad (4)$$

com a correspondente aditividade para os respectivos graus de liberdade associados. A equação (4) é de muita importância e será de muita utilidade em aplicações posteriores.

Da relação (4), tem-se que

$$2l(x:x_a^*) \geq 2l(x_b^*:x_a^*), \text{ com a igualdade se e somente se } x=x_b^* \text{ e}$$

$$2l(x:x_a^*) \geq 2l(x:x_b^*), \text{ com a igualdade se e somente se } x_a^* = x_b^*.$$

Medidas da forma $2l(x:x_a^*)$, isto é, a comparação de uma tabela de contingência observada com uma tabela

de contingência estimada são denominadas medidas de interação. Deve-se observar que $2l(x:x_a^*)$, testa a hipótese nula que os valores dos parâmetros τ 's que estão na representação log-linear da tabela de contingência observada, mas que não estão na representação log-linear da tabela de contingência estimada x_a^* , são todos iguais a zero. O número desses parâmetros é o número de graus de liberdade (ver final do item 3.5.4.3., página 103).

Medidas da forma $2l(x_b^*:x_a^*)$, isto é, a comparação de duas tabelas de contingência estimadas, são chamadas medidas de efeitos. Os efeitos das diferenças entre os conjuntos de restrições H_a e H_b , ou os efeitos dos parâmetros τ 's que estão em x_b^* mas não em x_a^* . Em forma similar, $2l(x_b^*:x_a^*)$, testa a hipótese nula que os valores dos parâmetros τ 's que estão na representação log-linear da tabela de contingência estimada x_b^* , mas que não estão na representação log-linear da tabela de contingência estimada x_a^* , são todos iguais a zero. O número desses parâmetros é o número de graus de liberdade.

Seja por exemplo,

$$H_a: x_{ijk} = \frac{x_{i..} \cdot x_{.j.} \cdot x_{..k}}{N^2} \quad (5)$$

$$H_b: x_{ijk} = \frac{x_{i..} \cdot x_{.jk}}{N} \quad (6)$$

A representação log-linear da tabela de contingência completa observada é:

$$\begin{aligned} \ln x_{ijk} = & \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} + \tau_{12}^{(ij)} + \tau_{13}^{(ik)} + \tau_{23}^{(jk)} + \\ & + \tau_{123}^{(ijk)} \end{aligned} \quad (7)$$

com $1 + (R-1) + (S-1) + \dots + (R-1)(S-1)(T-1) = R \times S \times T$ parâmetros.

A representação log-linear da tabela de contingência estimada sob H_a é:

$$\ln x_a^*(ijk) = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} \quad (8),$$

com $R + S + T - 2$ parâmetros.

A representação log-linear da tabela de contingência estimada sob H_b é:

$$\ln x_b^*(ijk) = \tau_0 + \tau_1^{(i)} + \tau_2^{(j)} + \tau_3^{(k)} + \tau_{23}^{(jk)} \quad (9),$$

com $ST + R - 1$ parâmetros.

De acordo com (4), têm-se que

$$2l(x:x_a^*) = 2l(x_b^*:x_a^*) + 2l(x:x_b^*), \quad \text{donde}$$

$$2l(x_b^*:x_a^*) = 2l(x:x_a^*) - 2l(x:x_b^*) \quad (10)$$

$2l(x:x_a^*)$, testa a hipótese nula de que os parâmetros $\tau_{12}^{(ij)}$, $\tau_{13}^{(ik)}$, $\tau_{23}^{(jk)}$ e $\tau_{123}^{(ijk)}$ são todos iguais a zero. Essa estatística tem distribuição assintótica como um χ^2 com

$$\begin{aligned} V_a &= (R-1)(S-1) + (R-1)(T-1) + (S-1)(T-1) + (R-1)(S-1)(T-1) = \\ &= RST - R - S - T + 2 \text{ graus de liberdade.} \end{aligned}$$

$2l(x_b^*:x_a^*)$, testa a hipótese que os $(S-1)(T-1)$ $\tau_{23}^{(jk)}$ parâmetros são todos iguais a zero, dado que os parâmetros $\tau_{12}^{(ij)}$, $\tau_{13}^{(ik)}$ e $\tau_{123}^{(ijk)}$ são iguais a zero. Essa estatística tem distribuição assintótica de qui-quadrado com $(S-1)(T-1)$ graus de liberdade.

$2I(x:x_b^*)$ testa a hipótese nula que os parâmetros $\tau_{12}^{(ij)}$, $\tau_{13}^{(ik)}$ e $\tau_{123}^{(ijk)}$ são todos iguais a zero. Essa estatística tem uma distribuição assintótica de q - quadrado com

$$V_b = (R-1)(S-1) + (R-1)(T-1) + (R-1)(S-1)(T-1) = (ST-1)(R-1)$$

graus de liberdade.

Quadro da análise de informação

Componentes	Informação	G.L.
Ha: Interação de primeira ordem	$2I(x:x_a^*)$	Va
Hb: Efeito de ST	$2I(x_b^*:x_a^*)$	Va-Vb
Interação	$2I(x:x_b^*)$	Vb

Observe que $2I(x:x_a^*) = 2I(x_b^*:x_a^*) + 2I(x:x_b^*)$ e que $V_a = (V_a - V_b) + V_b$.

A seguir serão apresentados dois quadros de análise da informação para tabelas de contingência $R \times S \times T$. O quadro 1, apresenta a análise de informação dos componentes das interações de primeira e segunda ordem. O quadro 2 apresenta uma análise dos efeitos das marginais bidimensionais.

Quadro 1. Análise de informação.

Componentes	Informação	G.L.
Uniformidade	$2I(x:N\pi)$	V
Efeitos das marginais unidimensionais	$2I(x_1^*:N\pi)$	V_1
Interação de primeira ordem	$2I(x:x_1^*)$	$V-V_1$
Efeitos das marginais bidimensionais	$2I(x_2^*:x_1^*)$	V_2
Interação de segunda ordem	$2I(x:x_2^*)$	$V-V_1-V_2=V_3$

Observações: $N\pi(ijk) = N/rxsxt,$

$$x_1^*(ijk) = (x_{i..} \cdot x_{.j.} \cdot x_{..k}) / N^2,$$

$$V = R \times S \times T - 1; \quad V_1 = (R-1) + (S-1) + (T-1);$$

$$V_2 = (R-1)(S-1) + (R-1)(T-1) + (S-1)(T-1) \text{ e}$$

$$V-V_1-V_2 = V_3 = (R-1)(S-1)(T-1); \quad V = V_1 + V_2 + V_3$$

Quadro 2. Análise de informação.

Componentes	Informação	G.L.
Interação de primeira ordem	$2I(x:x_1^*)$	$V-V_1$
(a) Efeito de RS	$2I(x_a^*:x_1^*)$	$(R-1)(S-1)$
Interação	$2I(x:x_a^*)$	$V-V_1-(R-1)(S-1)$
(b) Efeito de RT RS	$2I(x_b^*:x_a^*)$	$(R-1)(T-1)$
Interação	$2I(x:x_b^*)$	$V-V_1-(R-1)(S-1)-(R-1)(T-1)$
(c) Efeito de ST RS,RT	$2I(x_2^*:x_b^*)$	$(S-1)(T-1)$
Interação de segunda ordem	$2I(x:x_2^*)$	$V-V_1-V_2=V_3$

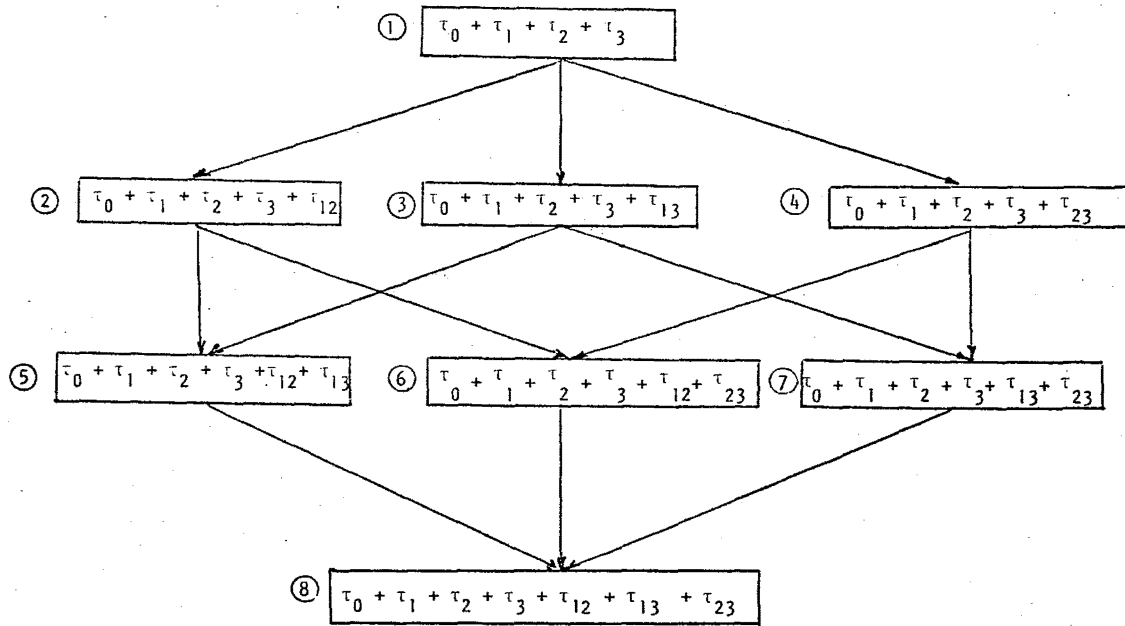
Observações: $x_a^* = \frac{x_{ij.} \cdot x_{..k}}{N}$

$$x_b^* = \frac{x_{ij.} \cdot x_{i.k}}{x_{i..}}$$

x_2^* , não pode ser expressa em termos das marginais bidimensionais. As suas estimativas são obtidas pelo processo iterativo discutido no item 3.5.7.2.1.

3.6.3. Modelos hierárquicos em tabelas de contingência $R \times S \times T$

Em geral, para tabelas de contingência tridimensionais há oito possíveis modelos hierárquicos que podem ser usados para ajustar os dados de uma tabela observada. Em forma esquemática, tem-se que:



Numa análise de tabelas de contingência tridimensionais têm-se então as seguintes sequências de modelos hierárquicos:

- (i) ① → ② → ⑤ → ⑧
- (ii) ① → ② → ⑥ → ⑧
- (iii) ① → ③ → ⑤ → ⑧
- (iv) ① → ③ → ⑦ → ⑧
- (v) ① → ④ → ⑥ → ⑧
- (vi) ① → ④ → ⑦ → ⑧

Observe que no quadro 2, apresenta-se uma análise de informação da sequência de modelos hierárquicos

① → ② → ⑤ → ⑧

Em muitos problemas de análise de dados arranjados numa tabela de contingência tridimensional, geralmente o pesquisador dispõe de um conhecimento prévio sobre as variáveis em estudo. Esse conhecimento prévio auxilia na escolha das frequências marginais de interesse e sugere a sequência na qual devem ser introduzidas na análise.

4. EXEMPLO ILUSTRATIVO

Para efeitos de ilustrar os conceitos teóricos desenvolvidos nos itens anteriores, suponha-se o seguinte conjunto de dados provenientes de uma única população multinomial, dispostos numa tabela de contingência completa de três entradas $R \times S \times T$, onde $R = 2$, $S = 2$ e $T = 2$.

TABELA 1. Frequências observadas.

	S_1		S_2	
	T_1	T_2	T_1	T_2
R_1	120	4	22	2
R_2	80	31	24	23

Para se analisarem esses dados será utilizado o programa auto explicativo ANDACAT (análise de dados categorizados), que foi escrito na linguagem Pascal no ambiente de desenvolvimento Turbo Pascal, em equipamentos da linha IBM-PC compatível (ver Apêndice 2).

Pretende-se estudar uma sequência de modelos hierárquicos e, posteriormente, baseados na construção de uma tabela de análise de informação, escolher o modelo log-linear que melhor ajuste a tabela de contingência observada.

Tem-se então que:

```

=====
| ENTRADA DE DADOS |
=====

```

Numero de niveis da variavel "R" = 2

Numero de niveis da variavel "S" = 2

Numero de niveis da variavel "T" = 2

Nome do arquivo: b:EXEMPLO

```

=====
| FREQUENCIAS MARGINAIS OBSERVADAS |
=====

```

X [1..] = 148.0000	X [.1.] = 235.0000	X [..1] = 246.0000
X [2..] = 158.0000	X [.2.] = 71.0000	X [..2] = 60.0000
X [11.] = 124.0000	X [1.1] = 142.0000	X [.11] = 200.0000
X [12.] = 24.0000	X [1.2] = 6.0000	X [.12] = 35.0000
X [21.] = 111.0000	X [2.1] = 104.0000	X [.21] = 46.0000
X [22.] = 47.0000	X [2.2] = 54.0000	X [.22] = 25.0000
X [...] = 306.0000		

O primeiro índice corresponde aos níveis da variável R, o segundo índice corresponde aos níveis da variável S e o terceiro índice corresponde aos níveis ou categorias da variável T.

Em seguida o programa calcula e fornece as estimativas das frequências esperadas sob o modelo de independência mútua completa das três variáveis, assim como também as frequências marginais unidimensionais sob esse modelo.

```

=====
| FREQUENCIAS ESTIMADAS E FREQUENCIAS MARGINAIS UNIDIMENSIONAIS |
|                               |
|           SOB O MODELO DE INDEPENDENCIA MUTUA COMPLETA           |
|                               |
=====

```

X(1) [1,1,1] = 91.3738	X(1) [2,1,1] = 97.5477
X(1) [1,1,2] = 22.2863	X(1) [2,1,2] = 23.7921
X(1) [1,2,1] = 27.6066	X(1) [2,2,1] = 29.4719
X(1) [1,2,2] = 6.7333	X(1) [2,2,2] = 7.1883

X [1..] = 148.0000	X [.1.] = 235.0000	X [..1] = 246.0000
X [2..] = 158.0000	X [.2.] = 71.0000	X [..2] = 60.0000

A seguir é calculada a estatística da informação mínima de discriminação $EIMD(1) = 2I[x:x(1)]$ e é pedido ao usuário que entre com um valor tabelado de qui-quadrado ao nível de significância desejado com a finalidade de se testar o modelo de independência mútua completa. Tem-se então que:

```

=====
| TESTE DO MODELO DE INDEPENDENCIA MUTUA COMPLETA |
|                               |
=====

```

Entre com qui-quadrado com 4 Graus de Liberdade = 9.49

$EIMD(1) = 65.1407$

Rejeita-se o Modelo de Independencia Mutua Completa

Continuando com a análise de uma sequência de modelos hierárquicos, se procederá a estudar o modelo de independência parcial. O "menu" para os diferentes modelos é:

```

=====
| MODELOS DE INDEPENDENCIA PARCIAL |
=====

```

2. R independente de S e T
3. S independente de R e T
4. T independente de R e S

Pressione nº da opção escolhida : 2

Suponha-se que neste exemplo, devido a um conhecimento prévio do pesquisador ou a alguma outra causa, o interessado deseje verificar se os níveis da variável R são homogêneos nos níveis das variáveis S e T. Sendo assim, escolheu-se a opção 2.

```

=====
| FREQUENCIAS ESTIMADAS E FREQUENCIAS MARGINAIS UNIDIMENSIONAIS E |
| BIDIMENSIONAIS SOB MODELO DE INDEPENDENCIA PARCIAL |
=====

```

X(a) [1,1,1] = 96.7320	X(a) [2,1,1] = 103.2680
X(a) [1,1,2] = 16.9281	X(a) [2,1,2] = 18.0719
X(a) [1,2,1] = 22.2484	X(a) [2,2,1] = 23.7516
X(a) [1,2,2] = 12.0915	X(a) [2,2,2] = 12.9085

X [1..] = 148.0000	X [.1.] = 235.0000	X [..1] = 246.0000
X [2..] = 158.0000	X [.2.] = 71.0000	X [..2] = 60.0000

X [11.] = 113.6601	X [1.1] = 118.9804	X [.11] = 200.0000
X [12.] = 34.3399	X [1.2] = 29.0196	X [.12] = 35.0000
X [21.] = 121.3399	X [2.1] = 127.0196	X [.21] = 46.0000
X [22.] = 36.6601	X [2.2] = 30.9804	X [.22] = 25.0000

X [...] = 306.0000

A seguir é calculada a estatística da informação mínima de discriminação $EIMD(a) = 2I[x:x(a)]$ e é pedido ao usuário que entre com um valor tabelado de qui-quadrado ao nível de significância desejado.

Tem-se então que

```

=====
| TESTE DO MODELO DE INDEPENDENCIA PARCIAL |
=====

```

Entre com qui-quadrado com 3 Graus de Liberdade = 7.81

$EIMD(a) = 52.1767$

Rejeita-se o Modelo de Independencia Parcial

O efeito de ter adicionado ST no modelo, é fornecido posteriormente na tabela de análise de informação. Como no "menu" anterior foi escolhida a opção 2, o "sub-menu" para os diferentes modelos de independência condicional é:

```

=====
| MODELOS DE INDEPENDENCIA CONDICIONAL |
=====

```

5. R e T são independentes em cada nível de S
6. R e S são independentes em cada nível de T.

Pressione nº da opção escolhida : 5

Suponha-se que o pesquisador estivesse interessado em verificar se as variáveis R e T são independentes em cada nível de S, isto é, em cada nível da variável S, o grau de associação entre as variáveis R e T é igual em termos esta

tísticos. Ao escolher a opção 5, obtêm-se:

```

=====
| FREQUENCIAS ESTIMADAS E FREQUENCIAS MARGINAIS UNIDIMENSIONAIS E |
| BIDIENSAIONAIS SOB MODELO DE INDEPENDENCIA CONDICIONAL |
=====

```

```

X(b) [1,1,1] = 105.5319      X(b) [2,1,1] = 94.4681
X(b) [1,1,2] = 18.4681      X(b) [2,1,2] = 16.5319
X(b) [1,2,1] = 15.5493      X(b) [2,2,1] = 30.4507
X(b) [1,2,2] = 8.4507       X(b) [2,2,2] = 16.5493

```

```

X [1..] = 148.0000  X [.1.] = 235.0000  X [..1] = 246.0000
X [2..] = 158.0000  X [.2.] = 71.0000   X [..2] = 60.0000

```

```

X [11.] = 124.0000      X [1.1] = 121.0812      X [.11] = 200.0000
X [12.] = 24.0000      X [1.2] = 26.9188      X [.12] = 35.0000
X [21.] = 111.0000     X [2.1] = 124.9188     X [.21] = 46.0000
X [22.] = 47.0000     X [2.2] = 33.0812     X [.22] = 25.0000

```

```

X [...] = 306.0000

```

Depois, como nos casos anteriores, tem-se que

```

=====
| TESTE DO MODELO DE INDEPENDENCIA CONDICIONAL |
=====

```

Entre com qui-quadrado com 2 Graus de Liberdade = 5.99

EIMD (b) = 44.1972

Rejeita-se o Modelo de Independencia Condicional

Continuando com a análise de uma sequência de modelos hierárquicos, se procederá a estudar o modelo onde a interação entre as três variáveis ou interação de segunda ordem, está ausente. O programa, neste exemplo, fornece o seguinte:


```

=====
| FREQUENCIAS ESTIMADAS SOB MODELO DE NAO |
| INTERACAO DE SEGUNDA ORDEM |
|=====

```

X(2) [1,1,1] =	119.8966	X(2) [2,1,1] =	80.1034
X(2) [1,1,2] =	4.1036	X(2) [2,1,2] =	30.8964
X(2) [1,2,1] =	22.1035	X(2) [2,2,1] =	23.8965
X(2) [1,2,2] =	1.8963	X(2) [2,2,2] =	23.1037

Numero de ciclos = 6

Neste modelo, as frequências marginais unidimensionais e bidimensionais estão ajustadas e, dentro dos limites da margem de erro admitido, são iguais às frequências marginais observadas.

```

=====
| TESTE DO MODELO DE NAO INTERACAO DE SEGUNDA ORDEM |
|=====

```

Entre com qui-quadrado com 1 Graus de Liberdade = 3.84

EIND(2) = 0.0102

Não se rejeita o Modelo de Não Interação de Segunda Ordem

Observe a grande semelhança entre as frequências observadas e as estimativas das frequências esperadas sob essa hipótese.

Em continuação é apresentado um quadro de análise de informação que contém de forma resumida, objetiva e aditiva uma análise completa da tabela de contingência estudada.

===== [ANÁLISE DE INFORMAÇÃO] =====		
COMPONENTES	INFORMAÇÃO	G.L.
R,S,T	$2I[X:X(1)] = 65.1407$	4
Efeito de ST	$2I[X(a):X(1)] = 12.9640$	1
R independente de ST	$2I[X:X(a)] = 52.1767$	3
Efeito de RS/ST	$2I[X(b):X(a)] = 7.9795$	1
Independencia Condicional de (R x T)/S	$2I[X:X(b)] = 44.1972$	2
Efeito de RT/ST,RS	$2I[X(2):X(b)] = 44.1871$	1
Interacao de segunda ordem	$2I[X:X(2)] = 0.0102$	1

Essa análise indica que o ajustamento desse conjunto de dados pode ser feito adequadamente usando como variáveis explanatórias as frequências marginais observadas das três tabelas $R \times S$, $R \times T$ e $S \times T$, isto é:

	S_1	S_2	
R_1	124	24	148
R_2	111	47	158
	235	71	306

	T_1	T_2	
R_1	142	6	148
R_2	104	54	158
	246	60	306

	T_1	T_2	
S_1	200	35	235
S_2	46	25	71
	246	60	306

A interpretação da hipótese de não interação de segunda ordem pode ser expressa como sendo:

- (i) a associação entre a variável R e a variável S é a mesma nos diferentes níveis da variável T;
- (ii) a associação entre as variáveis R e T é a mesma nos diferentes níveis da variável S e;
- (iii) a associação entre as variáveis S e T é a mesma nos diferentes níveis da variável R.

Tem-se então que o modelo escolhido é

$$\log x_{ijk} = \tau_0 + \tau_1(i) + \tau_2(j) + \tau_3(k) + \tau_{12}(ij) + \tau_{13}(ik) + \tau_{23}(jk),$$

onde $\tau_{1(2)} = \tau_{2(2)} = \tau_{3(2)} = 0$. Na forma matricial

$$\begin{bmatrix} \log X(2) & [1,1,1] \\ \log X(2) & [1,1,2] \\ \log X(2) & [1,2,1] \\ \log X(2) & [1,2,2] \\ \log X(2) & [2,1,1] \\ \log X(2) & [2,1,2] \\ \log X(2) & [2,2,1] \\ \log X(2) & [2,2,2] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \tau_0 \\ \tau_{1(1)} \\ \tau_{2(1)} \\ \tau_{3(1)} \\ \tau_{12(11)} \\ \tau_{13(11)} \\ \tau_{23(11)} \\ 0 \end{bmatrix}$$

Para se obterem as estimativas dos parâmetros do modelo selecionado, assim como também a estimativa da matriz de variâncias e covariâncias das estimativas dos parâmetros foi utilizado o programa ALGEMA de autoria do colega Carlos T. Santos Dias.

Tem-se que

$$\begin{aligned}\hat{\tau}_0 &= 3,1400 \\ \hat{\tau}_{1(1)} &= -2,5001 \\ \hat{\tau}_{2(1)} &= 0,2906 \\ \hat{\tau}_{3(1)} &= 0,0337\end{aligned}$$

$$\begin{aligned}\hat{\tau}_{12(11)} &= 0,4814 \\ \hat{\tau}_{13(11)} &= 2,4221 \\ \hat{\tau}_{23(11)} &= 0,9190\end{aligned}$$

e

$$V(\hat{\tau}) = \begin{bmatrix} 0.2295 & -0.0619 & -0.0076 & -0.0645 & -0.1775 & 0.0344 \\ -0.0619 & 0.0677 & 0.0055 & -0.0006 & 0.0002 & -0.0053 \\ -0.0076 & 0.0105 & 0.0775 & 0.0022 & -0.0045 & -0.0705 \\ -0.0645 & -0.0006 & 0.0022 & 0.0957 & -0.0117 & -0.0096 \\ -0.1775 & 0.0002 & -0.0045 & -0.0117 & 0.0045 & -0.0044 \\ 0.0344 & -0.0053 & -0.0705 & -0.0096 & -0.0044 & 0.1123 \end{bmatrix}$$

5. CONCLUSÕES

Face ao anteriormente exposto, concluiu-se que:

(i) A estrutura de uma tabela de contingência pode ser estudada com grande riqueza de detalhes e pode ser interpretada em termos das interrelações entre as três variáveis categorizadas.

(ii) o princípio da informação mínima de discriminação pode ser usado eficientemente para se obterem estimativas das frequências esperadas numa tabela de contingência $R \times S \times T$, sob várias hipóteses ou modelos de interesse. Para as condições deste trabalho, a x^* -tabela fornece estimativas de máxima verossimilhança.

(iii) as estimativas obtidas através do princípio da informação mínima de discriminação são formuladas como membros de uma família exponencial, que pode ser expressa como um modelo multiplicativo ou como um modelo linear aditivo na escala logaritmica. Esse fato é interessante no sentido que não é assumido um determinado modelo senão que ele é obtido segundo as restrições impostas sobre as frequências marginais observadas.

(iv) para cada conjunto de marginais fixas é obtido um único conjunto de estimativas de frequências esperadas através de um processo iterativo. Esse processo iterativo é convergente e para o caso das hipóteses de independência mútua completa, independência parcial e independência condicional são obtidas as estimativas logo ao final do primeiro ciclo.

(v) a adequacidade de ajuste de um modelo, sob qualquer hipótese nula, pode ser testado. A estatística da informação mínima de discriminação $2I(x:x^*)$ é distribuída assintoticamente como um χ^2 com apropriados graus de liberdade. A computação dessa estatística é feita de maneira simples e direta.

(vi) a estatística da informação mínima de discriminação, sob uma determinada hipótese, pode ser analisada em componentes que por sua vez são estatísticas de informação mínima de discriminação sob sub-hipóteses, dado que essas estatísticas são aditivas, assim como também os graus de liberdade associados. Essa propriedade é usada para construir tabelas de análise de informação que apresentam uma análise aditiva da tabela de contingência.

(vii) no modelo log-linear $\log x^* = T\tau$, as estimativas dos parâmetros que representam os efeitos principais e as interações é feita de uma maneira simples a partir das estimativas das frequências esperadas. Igualmente, pode ser obtida uma estimativa para a matriz de variâncias e covariâncias das estimativas dos parâmetros.

(viii) deve-se prosseguir o estudo de dados categorizados arranjados em tabelas de contingência tridimensionais visando, dentre outros, os seguintes aspectos:

- estudo de tabelas incompletas;
- estudo de tabelas de contingência quando as frequências marginais esperadas são conhecidas;
- estudo dos resíduos $x_{ijk} - x^*_{ijk}$, como uma ajuda na análise da tabela de contingência. De fato, no teste de ajustamento de um determinado modelo, se os valores computados de $2l(x:x^*)$ são significativos, pode haver uma grande discrepância entre os valores observados e os valores esperados;
- estender o estudo a tabelas de contingência com maiores dimensões;
- estudo do comportamento da estatística da informação mínima de discriminação em amostras pequenas.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BARTLETT, M.S. Contingency table interactions. J. Roy Statist. Soc., London, 2: 248-252, 1935.

BHAPKAR, V.P. A note on the equivalence of two test criteria for hypotheses in categorical data. J.A.S.A., Boston, 61: 228-235, 1966.

BHAPKAR, V.P. & KOCH, G.G. On the hypothesis of 'no interaction' in contingency tables. Biometrics, Raleigh, 16: 618-635, 1968.

BIRCH, M.W. Maximum likelihood in three-way contingency tables. J. Roy. Statist. Soc., London, 25: 220-233, 1963.

BISHOP, Y.M.M.; S.E. FIENBERG; P.W. HOLLAND. Discrete Multivariate Analysis: theory and practice. 3ª ed., The Mit Press, Cambridge. 557p, 1977.

CAMPBELL, L.L. Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities. Ann. Math. Statist., Baltimore, 41: 1011-1015, 1970.

CANTON, A.W.P. Análise de dados categorizados. 4º Simpósio Nac. de Prob. e Estat., Rio de Janeiro, 1980, 79p.

CANTON, A.W.P. & FESTA, M.N. Análise de dados categorizados usando modelos lineares e suas aplicações - uma revisão bibliográfica. Atas do 3º SINAPE, São Paulo, 182-189, 1979.

- CORDEIRO, G.M. Modelos lineares generalizados. VII Sim_pósio Nacional de Probabilidade e Estatística. Campinas. 286p, 1986.
- DARROCH, J.N. Interactions in multi-factor contingency tables. J. Roy. Statist. Soc. Ser. B, London, 24: 251-263, 1962.
- FELLER, W. An introduction to probability theory and its applications. 3^a ed., John Wiley & Sons, New York. 509p, 1968.
- FIENBERG, S.E. An iterative procedure for estimation in contingency tables. Ann. Math. Statist., Baltimore, 41: 907-917, 1970.
- FIENBERG, S.E. The analysis of cross-classified categorical data. 2^a ed., The Mit Press, Cambridge. 151p., 1978.
- FLEISS, J.L. Statistical methods for rates and proportions. John Wiley & Sons, New York. 223p, 1973.
- GOKHALE, D.V. An iterative procedure for analysing log-linear models. Biometrics, Raleigh, 27: 681-687, 1971.
- GOKHALE, D.V. Analysis of log-linear models. J. Roy. Statist. Soc., London, 34: 371-376, 1972.
- GOKHALE, D.V. Approximating discrete distributions with applications. JASA. Boston, 68: 1009-1012, 1973.
- GOKHALE, D.V. & KULLBACK, S. The minimum discrimination information approach in analysing categorical data. Commun. Statist. - Theor. Meth., New York, A7(10): 987-1005, 1978.
- GOOD, I.J. Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. Ann. Math. Statist. Baltimore, 34: 911-934, 1963.

- GOODMAN, L.A. Simple methods for analysing three-factor interaction in contingency tables. JASA. Boston, 59: 319-352, 1964.
- GOODMAN, L.A. The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries. JASA. Boston, 63: 1091-1131, 1968.
- GOODMAN, L.A. The multivariate analysis of qualitative data: interaction among multiple classifications. JASA. Boston, 65: 226-256, 1970.
- GODDMAN, L.A. Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. JASA. Boston, 66: 339-344, 1971.
- GRIZZLE, J.E. & WILLIMS, O.D. Log-linear models and tests of independence for contingency tables. Biometrics, Raleigh, 28: 137-156, 1972.
- GRIZZLE, J.E.; STARMER, C.F.; KOCH, G.G. Analysis of categorical data by linear models. Biometrics, Raleigh, 25: 489-504, 1969.
- HOGG, R. & A. CRAIG. Introduction to mathematical statistics. 4^a ed., MacMillan Publishing Co., Inc., New York. 438p, 1978.
- IRELAND, C.T. & KULLBACK, S. Contingency tables with given marginals. Biometrika, London, 55: 179-188, 1968.
- KU, H.H. & S. KULLBACK. Interaction in multidimensional contingency tables: An information theoretic Approach. Journal of Research of the National Bureau of Standards. Washington, 72: 159-199, 1968.

- KU, H.H. & KULLBACK, S. Log linear models in contingency tables analysis. The American Statistician, Washington, 28: 115-122, 1974.
- KU, H.H.; VARNER, R.N.; KULLBACK, S. On the analysis of multidimensional contingency tables. JASA, Boston, 66: 55-64, 1971.
- KULLBACK, S. Information theory and statistics. New York, Dover Publications, Inc., 1959.
- KULLBACK, S. & FISHER, M. Partitioning Second-order interaction in three-way contingency tables. J. Roy. Statist. Soc. London, 22: 172-184, 1973.
- LEE, S.K. On the asymptotic variances of \hat{u} terms in loglinear models of multidimensional contingency tables. JA SA. Boston, 72: 412-419, 1977.
- LIMA, M.A. Estudo dos modelos log-lineares através do GLIM. 7º Simp. Nac. de Prob. e Estat., Campinas, 1986. 11p.
- MOOD, A.M.; F.A. GRAYBILL & D.C. BOES. Introduction to the theory of statistics. 3ª ed., McGraw-Hill Koga - Kusha Ltd., Tokyo. 564p, 1974.
- NELDER, J.A. Log linear models for contingency tables: a generalization of classical least squares. J. Roy. Statist. Soc., London, 23: 323-329, 1974.
- NUNES MAIA, M.M.S. Análise de dados categóricos usando teoria da informação. São Paulo, 1977. (Mestrado - Instituto de Matemática e Estatística/USP).
- PERES, C.A. Análise de dados categóricos usando teoria da informação. Atas do 3º SINAPE, São Paulo, 190-195, 1979.

PLACKETT, R.L. Multidimensional contingency tables: a survey of models and methods. Bull. Int. Inst. Stat. London, 43: 133-142, 1969.

RAO, C.R. Linear statistical inference and its applications. John Wiley & Sons, Inc., New York, 522p, 1965.

ROY, S.N. & KASTENBAUM, M.A. On the hypothesis of no "interactions" in a multi-way contingency table. Ann. Math. Statist., Baltimore, 27: 749-757, 1956.

THEIL, H. Economics and information theory. Rand McNally e Company, Chicago. 486p, 1967.

WALD, A. Sequential tests of statistical hypotheses. Ann. Math. Statist. Baltimore, 16: 117-186, 1945.

WILKS, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Statist. Baltimore, 9: 60-62, 1938.

WILKS, S.S. Mathematical statistics. John Wiley & Sons, Inc., New York. 644p, 1962.

A P E N D I C E S

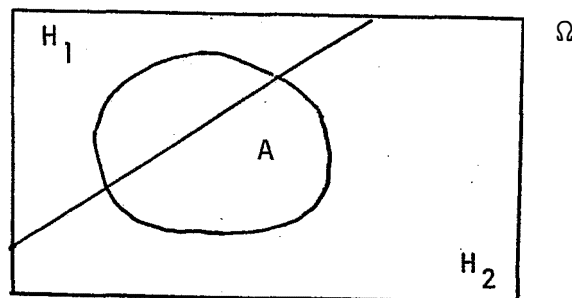
APÊNDICE 1 - PROBABILIDADE CONDICIONAL

Definição: Seja A um evento com probabilidade positiva. Para um evento qualquer A , tem-se que:

$$P(H|A) = \frac{P(HA)}{P(A)} \quad (1)$$

Segundo FELLER (1968), $P(H|A)$ assim definida e chamada probabilidade condicional de H sob a hipótese A ou para um dado A .

Sejam H_1 e H_2 dois eventos mutuamente exclusivos e exaustivos. Isto é, $H_1 \cup H_2 = \Omega$, onde Ω é o espaço amostral; $H_1 \cap H_2 = \emptyset$ e $P(H_1) + P(H_2) = 1$. Consequentemente, qualquer evento A pode ocorrer unicamente de maneira conjunta com algum H_i , $i = 1, 2$. Num gráfico tem-se que



Então $A \cap (H_1 \cup H_2) = (A \cap H_1) \cup (A \cap H_2) = AH_1 \cup AH_2 = A$, onde

$$AH_1 \cap AH_2 = \emptyset, \text{ portanto } P(A) = P(AH_1) + P(AH_2) \quad (2)$$

Pela definição (1), tem-se que $P(A|H_i) = \frac{P(AH_i)}{P(H_i)}$
isto é,

$$P(AH_i) = P(A|H_i) P(H_i), \quad i=1,2 \quad (3)$$

Substituindo (3) em (2), obtem-se

$$P(A) = P(A|H_1) P(H_1) + P(A|H_2) P(H_2) \quad (4)$$

e substituindo (4) em (1), obtêm-se

$$P(H_1|A) = \frac{P(H_1A)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2)} \quad (5)$$

e

$$P(H_2|A) = \frac{P(H_2A)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2)} \quad (6)$$

e em geral

$$P(H_i|A) = \frac{P(H_iA)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2)}, \quad i=1; 2 \quad (7)$$

ou

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2)}, \quad i=1; 2 \quad (8)$$

A equação (8) é conhecida como Regra de Bayes.


```

program ANDACAT;
{$R+}
type
  MatA = array[1..10,1..10,1..10] of real;
  MatB = array[1..10] of real;
  MatC = array[1..10,1..10] of real;
  str80 = string[80];
  strin255 = string[255];
var
  A,A1,Aa,Ab,A2           :MatA;
  B1,B2,B3,B1a0,B2a0,B3a0 :MatB;
  C1a1,C2a1,C3a1,C1,C2,C3,C1a,C2a,C3a :MatC;
  I,L,J,K,R,S,T          :integer;
  VI, VJ, VK,cont        :integer;
  Soma, SomaTotal,EIMD1,EIMDa,EIMDb,quadrado,GL1,GLa,GLb,GL2,EIMD2 : real;
  LinhaImp : array[1..10] of string[80];
  Arquivo   :text;
  Menuopt,Submenuopt :char;
  NomeDoArquivo :string[20];

procedure TeclEnter;
var
  Aux:char;
begin
  writeln;
  writeln('Tecl <ENTER> para continuar... ');
  writeln;
  write (#7);
  repeat
    read (kbd,Aux);
  until Aux = #13;
end;

procedure TelaInicial;
begin
  clrscr;
  gotoxy(23,5); writeln('ANALISE DE TABELAS DE CONTINGENCIA');
  gotoxy(22,7); writeln('TRIDIMENSIONAIS ATRAVES DO PRINCIPIO');
  gotoxy(22,9); writeln('DA INFORMACAO MINIMA DE DISCRIMINACAO');
  gotoxy(40,13); writeln('Lusbi A.Herrera');
  gotoxy(40,14); writeln('Gabriel A.Sarries (CIAGRI/USP)');
  gotoxy(40,15); writeln('Marcelo Z.Silva (CIAGRI/USP)');
  gotoxy(40,20); writeln('Piracicaba, Maio de 1987');
  gotoxy(1,22);
  TeclEnter;
  clrscr;
end;

Procedure RISCO (TIPO:Char;ALTURA,COMEÇO,FIM:Byte;SIMBOLO:integer);
Var
  I:Byte;
Begin
  If (TIPO='H') Or (TIPO='h') Then
    For I:=COMEÇO To FIM Do
      Begin
        GoToXY(I,ALTURA);Write(chr(SIMBOLO));
      End;
  If (TIPO='V') Or (TIPO='v') Then
    For I:=COMEÇO To FIM Do
      Begin

```

```

    GoToXY(ALTURA,I);Write(chr(SIMBOLO));
End:
writeln;
End:

Procedure QUADRO(X0,Y0,X1,Y1:Byte);
Begin
    GoToXY(X0,Y0);Write(#201);RISCO('H',Y0,X0+1,X1-1,205);
    GoToXY(X1,Y0);Write(#187);RISCO('V',X1,Y0+1,Y1-1,186);
    GoToXY(X1,Y1);Write(#188);RISCO('H',Y1,X0+1,X1-1,205);
    GoToXY(X0,Y1);Write(#200);RISCO('V',X0,Y0+1,Y1-1,186);
End:

```

```

function space(n:integer) : str80;
var t : integer;
    s : string[80];
begin
    s := '';
    for t:=1 to N do
        s := concat(s, ' ');
    space := s;
end;

```

```

function AbraArquivo(NomeArq : str80) : boolean;
var Ok : boolean;
begin
    assign(Arquivo,NomeDoArquivo);
    {$I-} reset(Arquivo) {$I+};
    Ok := (IOresult = 0);
    if not Ok then
        begin
            gotoxy(20,14);
            write('Esse arquivo nao existe');
            delay(1500);
            delline;
        end;
    AbraArquivo := Ok;
end;{function}

```

```

procedure TesteParar;
begin
    cont := cont + 1;
    if cont >= 14 then
        begin
            TeclEnter;
            cont := 0;
        end;
end;

```

```

procedure EntraDadosArquivo;
begin
    repeat
        gotoxy(20,12);
        write('Nome do arquivo:');
        gotoxy(37,12);
        readln(NomeDoArquivo);
    until AbraArquivo(NomeDoArquivo);
    for I := 1 to R do
        for J := 1 to S do
            for K := 1 to T do

```

```

begin
  readln(Arquivo,A[I,J,K]);
end;
clrscr;
close(Arquivo);
end;{procedure}

procedure EntraDadosTeclado;
var Confirma : char;
begin
  repeat
    clrscr;
    for I := 1 to R do
      for J := 1 to S do
        for K := 1 to T do
          begin
            write('X[I,',I,',',J,',',K,'] = ');
            readln(A[I,J,K]);
          end;
          writeln;
          write('Confirma os dados? (s/n)');
          readln(Confirma);
          Confirma := upcase(Confirma);
        until Confirma = 'S';
      clrscr;
    end;

procedure EntradaDeDados;
var ConfirmaTudo, Dispo : char;
begin
  clrscr;
  gotoxy(32,2); writeln('ENTRADA DE DADOS');
  quadro(30,1,49,3);
  gotoxy(20,5); write('Numero de niveis da variavel "R" = '); readln(R);
  gotoxy(20,7); write('Numero de niveis da variavel "S" = '); readln(S);
  gotoxy(20,9); write('Numero de niveis da variavel "T" = '); readln(T);
  writeln;
  repeat
    gotoxy(20,12);
    write('Entrada de dados arquivo ou teclado? (A/T) ');
    readln(Dispo);
    Dispo := upcase(Dispo);
  until Dispo in ['A','T'];
  if Dispo = 'A' then
    EntraDadosArquivo
  else
    EntraDadosTeclado;
  end;

procedure ImprimeC1C2C3 (aux : char);
begin
  for I := 1 to R do
    for J := 1 to S do
      begin
        write('X [',I,J,','] = ');
        if aux='A' then
          writeln(C1[I,J]:9:4)
        else
          writeln(C1a[I,J]:9:4);
      end;
    end;
  end;
end;

```

```

    TesteParar;
end;
for I := 1 to R do
  for J := 1 to T do
    begin
      write('X [' ,I, ',' ,J, ' ] = ');
      if aux='A' then
        writeln(C2[I,J]:9:4)
      else
        writeln(C2a[I,J]:9:4);
      TesteParar;
    end;
  for I := 1 to S do
    for J := 1 to T do
      begin
        write('X [' ,I, ',' ,J, ' ] = ');
        if aux='A' then
          writeln(C3[I,J]:9:4)
        else
          writeln(C3a[I,J]:9:4);
        TesteParar;
      end;
    writeln;writeln;
    writeln('X [...] = ',SomaTotal:9:4);
    writeln;
    TeclEnter;
    clrscr;
    cont := @;
  end;{procedure}

procedure CalcularSomatoriasMarginaisUnidimensionais(X: MatA);
var st,valor : str80;
    MaiorNivel : integer;
begin
  SomaTotal := @;
  for L := 1 to 3 do
    begin
      case L of
        1 : begin VI := R; VJ := S; VK := T;end;
        2 : begin VI := S; VJ := R; VK := T;end;
        3 : begin VI := T; VJ := S; VK := R;end;
      end;{case}
      for I := 1 to VI do
        begin
          Soma := @;
          for J := 1 to VJ do
            begin
              for K :=1 to VK do
                begin
                  case L of
                    {R,S,T}
                    1 : Soma := Soma + X[I,J,K];      { Xi..}
                    2 : Soma := Soma + X[J,I,K];      { X.j.}
                    3 : Soma := Soma + X[K,J,I];      { X..k}
                  end;{case}
                  if L = 1 then SomaTotal := SomaTotal + X[I,J,K];
                end;{K}
              end;{J}
            case L of
              1 : B1[I] := Soma;
              2 : B2[I] := Soma;
            end;
          end;
        end;
      end;
    end;
  end;
end;

```

```

        3 : B3[I] := Soma;
      end;{case}
    end;{I}
  end;{L}
  if (R > S) and (R > T) then
    MaiorNivel := R
  else
    if S > T then
      MaiorNivel := S
    else
      MaiorNivel := T;
    for I := 1 to MaiorNivel do LinhaImp[I] := space(80);
    for I := 1 to R do
      begin
        str(B1[I]:9:4,valor);
        str(I,st);
        LinhaImp[I] := concat('X [',st,'..] = ',valor,' ');
      end;
    for I := 1 to S do
      begin
        str(B2[I]:9:4,valor);
        str(I,st);
        if LinhaImp[I][17] = ' ' then
          LinhaImp[I] := space(22);
        LinhaImp[I] := concat(LinhaImp[I],'X [',st,'.] = ',valor,' ');
      end;
    for I := 1 to T do
      begin
        str(B3[I]:9:4,valor);
        str(I,st);
        if (LinhaImp[I][38] = ' ') then
          if (LinhaImp[I][17] = ' ') then
            LinhaImp[I] := concat(space(44),'X [..',st,'] = ',valor)
          else
            LinhaImp[I] := concat(LinhaImp[I],space(22),'X [..',st,'] = ',valor)
          else
            LinhaImp[I] := concat(LinhaImp[I],'X [..',st,'] = ',valor);
      end;
    for I := 1 to MaiorNivel do
      writeln(LinhaImp[I]);
    TeclaEnter;
    clrscr;
  end;

procedure CalcularSomatoriasMarginaisBidimensionais(var X:MatA;aux:char);
begin {gerando C1,C2,C3}
  Soma := 0;
  for L := 1 to 3 do
    begin
      case L of
        1 : begin VI := R; VJ := S; VK := T; end;
        2 : begin VI := R; VJ := T; VK := S; end;
        3 : begin VI := S; VJ := T; VK := R; end;
      end;{case}
      for I := 1 to VI do
        begin
          for J := 1 to VJ do
            begin
              for K := 1 to VK do
                begin

```

```

      case L of
        {R,S,T}
        1 : Soma := Soma + X[I,J,K ];           {Xi.i.}
        2 : Soma := Soma + X[I,K,J];           {Xi.k}
        3 : Soma := Soma + X[K,I,J];           {X.jk}
      end;{case}
    end;{K}
  if aux='A' then
    case L of
      1 : C1[I,J] := Soma;           {Xi.j.}
      2 : C2[I,J] := Soma;           {Xi.k}
      3 : C3[I,J] := Soma;           {X.jk}
    end {case}
  else
    case L of
      1 : C1a[I,J] := Soma;
      2 : C2a[I,J] := Soma;
      3 : C3a[I,J] := Soma;
    end;{case}
    Soma := 0;
  end;{J}
end;{I}
end;{L}
if aux='A' then
  ImprimeC1C2C3('A')
else
  ImprimeC1C2C3('B');
end;{procedure}

```

procedure EstimativasDasFrequenciasEsperadasSobModeloDeIndependenciaMutuaCompleta;

```

type
  MatA1 = array[1..10,1..10,1..10] of real;
  MatB1 = array[1..10] of real;

var
  B10,B20,B30           :MatB1;
  I,J,K,L,Aux          :integer;
  Soma1, Soma2, Soma3  :real;

begin
  for L := 1 to 3 do
    begin
      case L of
        1 : begin VI := R; VJ := S; VK := T; end;
        2 : begin VI := S; VJ := R; VK := T; end;
        3 : begin VI := T; VJ := S; VK := R; end;
      end;{case}
      if L=1 then
        begin
          for I:=1 to VI do
            for J:=1 to VJ do
              for K:=1 to VK do
                A1[I,J,K]:=SomaTotal/(VI*VJ*VK);           {Xijk(o)}
              end;{se L =1}
            end;
          for I := 1 to VI do
            begin
              Soma1 := 0;
              Soma2 := 0;
              Soma3 := 0;
              for J := 1 to VJ do

```

```

begin
  for K := 1 to VK do
    begin
      case L of
        {R,S,T}
        1 : Soma1 := Soma1 + A1[I,J,K]; {Xi..(o)}
        2 : Soma2 := Soma2 + A1[J,I,K]; {X..j.(1)}
        3 : Soma3 := Soma3 + A1[K,J,I]; {X..k.(2)}
      end;{case}
    end;{K}
  end;{J}
  case L of
    1 : B1@[I] := Soma1; {Xi..(o)}
    2 : B2@[I] := Soma2; {X..j.(1)}
    3 : B3@[I] := Soma3; {X..k.(2)}
  end;{case}
end;{I}
for I:=1 to VI do {Obtencao de Xijk(*)}
  for J:=1 to VJ do
    for K:=1 to VK do
      begin
        case L of
          {R,S,T}
          1 : A1[I,J,K]:=(B1[I] / B1@[I]) * A1[I,J,K]; {Xijk(1)}
          2 : A1[J,I,K]:=(B2[I] / B2@[I]) * A1[J,I,K]; {Xijk(2)}
          3 : A1[K,J,I]:=(B3[I] / B3@[I]) * A1[K,J,I]; {Xijk(3)}
        end;{case}
      end;{K}
    end;{L}
  for I:=1 to R do
    for J:=1 to S do
      for K:=1 to T do
        begin
          writeln('X(1) [' ,I ,',',J ,',',K ,'] = ',A1[I,J,K]:9:4);
          TesteParar;
        end;
      TeclaEnter;
      clrscr;
      cont := 0;
    end;{procedure}

procedure CalculoDaEIMDisobIndependenciaMutuaCompleta(Z,Z1 : MatA);
var H,G,F : integer;
begin
  {Obtencao de ZI(X:X(1))}
  EIMD1 := 0;
  for F :=1 to R do
    for G :=1 to S do
      for H := 1 to T do
        if (Z[F,G,H] > 0) then
          EIMD1 := EIMD1 + 2*(Z[F,G,H] * ln(Z[F,G,H] / Z1[F,G,H]));
      GL1 := R * S * T - R - S - T +2;
    end;

procedure TesteDoModeloDeIndependenciaMutuaCompleta;
begin
  clrscr;
  quadro(14,1,64,3);
  gotoxy(16,2);
  writeln('TESTE DO MODELO DE INDEPENDENCIA MUTUA COMPLETA');
  gotoxy(1,5);
  write('Entre com qui-quadrado com ',GL1:3:0,' Graus de Liberdade = ');
  readln(quiquadrado);

```

```

gotoxy(44,7); writeln('EIMD(1) = ',EIMD1:7:4); gotoxy(10,13);
if (EIMD1 >= quiquadrado) then
  writeln('Rejeita-se o Modelo de Independencia Mutua Completa')
else
  writeln('Nao rejeita-se o Modelo de Independencia Mutua Completa');
writeln;writeln;
endi;

procedure EstimativasDasFrequenciasEsperadasSobModeloDeIndependenciaParcial;

var H,G,F : integer;
    Soma : real;

begin
  menuopt := ' ';
  TeclEnter;
  clrscr;
  quadro(22,2,57,4);
  gotoxy(24,3); writeln('MODELOS DE INDEPENDENCIA PARCIAL');
  gotoxy(25,7); writeln('2. R independente de S e T');
  gotoxy(25,9); writeln('3. S independente de R e T');
  gotoxy(25,11);writeln('4. T independente de R e S');
  repeat
    gotoxy(25,15);write('Pressione n' da opcao escolhida : ');
    readln(menuopt);
  until menuopt in ['2'..'4'];
  clrscr;
  quadro(8,2,74,6);
  gotoxy(10,3);
  writeln('FREQUENCIAS ESTIMADAS E FREQUENCIAS MARGINAIS UNIDIMENSIONAIS E');
  gotoxy(15,5);
  writeln('BIDIMENSIONAIS SOB MODELO DE INDEPENDENCIA PARCIAL');
  writeln;writeln;
  for I:=1 to R do
    for J:=1 to S do
      for K:=1 to T do
        Aa[I,J,K]:=SomaTotal/(R*S*T);          {Xijk(o)}
  case menuopt of
    '2' : begin VI := R; VJ := S; VK := T; end;
    '3' : begin VI := S; VJ := R; VK := T; end;
    '4' : begin VI := T; VJ := S; VK := R; end;
  end;
  for I := 1 to VI do
  begin
    Soma := 0;
    for J := 1 to VJ do
    begin
      for K :=1 to VK do
        case menuopt of          {R,S,T}
          '2' : Soma := Soma + Aa[I,J,K];
          '3' : Soma := Soma + Aa[J,I,K];
          '4' : Soma := Soma + Aa[K,J,I];
        end;
      end;
    end;
    case menuopt of
      '2' : B1a@[I] := Soma;   { Xi.. (o)}
      '3' : B2a@[I] := Soma;   { X.j (o)}
      '4' : B3a@[I] := Soma;   { X..k (o)}
    end;
  end;
end;

```



```

                                { ***** obtencao de Xijk (1) ***** }
for I:= 1 to VI do
  for J := 1 to VJ do
    for K := 1 to VK do
      begin
        case menuopt of
          {R,S,T}
          '2' : Aa[I,J,K] := (B1[I] / B1a0[I]) * Aa[I,J,K];
          '3' : Aa[J,I,K] := (B2[I] / B2a0[I]) * Aa[J,I,K];
          '4' : Aa[K,J,I] := (B3[I] / B3a0[I]) * Aa[K,J,I];
        end;
      end; {K}
    case menuopt of
      '2' : begin VI := S; VJ := T; VK := R; end;
      '3' : begin VI := R; VJ := T; VK := S; end;
      '4' : begin VI := R; VJ := S; VK := T; end;
    end;{case}
  for I := 1 to VI do
    begin
      for J := 1 to VJ do
        begin
          Soma := 0;
          for K := 1 to VK do
            begin
              case menuopt of
                {R,S,T}
                '2' : Soma := Soma + Aa[K,I,J];
                '3' : Soma := Soma + Aa[I,K,J];
                '4' : Soma := Soma + Aa[I,J,K];
              end;{case}
            end;{K}
          case menuopt of
            '2' : C3a1[I,J] := Soma;
            '3' : C2a1[I,J] := Soma;
            '4' : C1a1[I,J] := Soma;
          end;{case}
        end;{J}
      end;{I}

for I := 1 to VI do
  for J := 1 to VJ do
    for K := 1 to VK do
      begin
        case menuopt of
          '2' : Aa[K,I,J] := (C3[I,J] / C3a1[I,J]) * Aa[K,I,J];
          '3' : Aa[I,K,J] := (C2[I,J] / C2a1[I,J]) * Aa[I,K,J];
          '4' : Aa[I,J,K] := (C1[I,J] / C1a1[I,J]) * Aa[I,J,K];
        end;{case}
      end;{K}
    end;{procedure}

procedure ImprimeXa;
begin
  for I:=1 to R do
    for J:=1 to S do
      for K:=1 to T do
        begin
          writeln('X(a) [' ,I,' ,',J,' ,',K,' ] =', Aa[I,J ,K:9:4]);
          TesteParar;
        end;
      TeclaEnter;
    cont := 0;

```

```

  clrscr;
end;

procedure CalculoDaEIMDaSobIndependenciaParcial(Z,Z1 : MatA);
var H,G,F : integer;
begin
  {Obtencao de 2I(X:X(a))}
  EIMDa := 0;
  for F :=1 to R do
    for G :=1 to S do
      for H := 1 to T do
        if (Z[F,G,H] > 0) then
          EIMDa := EIMDa + 2*(Z[F,G,H] * ln(Z[F,G,H] / Z1[F,G,H]));
        case menuopt of
          '2' : GLa := (R - 1)*(S * T - 1);
          '3' : GLa := (S - 1)*(R * T - 1);
          '4' : GLa := (T - 1)*(R * S - 1);
        end;
      end;
    end;
  end;

procedure TesteDoModeloDeIndependenciaParcial;
begin
  clrscr;
  quadro(13,1,56,3);
  gotoxy(15,2);
  writeln('TESTE DO MODELO DE INDEPENDENCIA PARCIAL');
  gotoxy(1,5);
  write('Entre com qui-quadrado com ',GLa:3:0,' Graus de Liberdade = ');
  readln(quiquadrado);
  gotoxy(43,7); writeln('EIMD (a) = ',EIMDa:7:4); gotoxy(16,13);
  if (EIMDa >= quiquadrado) then
    writeln('Rejeita-se o Modelo de Independencia Parcial')
  else
    writeln('Nao rejeita-se o Modelo de Independencia Parcial');
  writeln;writeln;
  TeclEnter;
end;

procedure EstimativasDeFrequenciasEsperadassobModelosDeIndependenciaCondicional;
var
  mens : string[35];
  C1b0,C2b0,C3b1,C2b1 : MatC;
  F,G,H : integer;

begin
  Submenuopt := ' ';
  mens := 'Pressione n' da opcao escolhida : ';
  clrscr;
  quadro(20,1,59,3);
  gotoxy(22,2); writeln('MODELOS DE INDEPENDENCIA CONDICIONAL');
  writeln;writeln;
  case menuopt of
    '2' : begin
      gotoxy(15,5);
      writeln('5. R e T sao independentes em cada nivel de S');
      gotoxy(15,7);
      writeln('6. R e S sao independentes em cada nivel de T');
      repeat
        gotoxy(15,10);
        write(mens);
        readln(Submenuopt);
      until Submenuopt = '2';
    end;
  end;

```

```

        until Submenuopt in ['5','6']:
            end;
'3' : begin
    gotoxy(15,5);
    writeln('7. S e T sao independentes em cada nivel de R');
    gotoxy(15,7);
    writeln('6. R e S sao independentes em cada nivel de T');
    repeat
        gotoxy(15,10);
        write(mens);
        readln(Submenuopt);
    until Submenuopt in ['6','7'];
    end;
'4' : begin
    gotoxy(15,5);
    writeln('7. S e T sao independentes em cada nivel de R');
    gotoxy(15,7);
    writeln('5. R e T sao independentes em cada nivel de S');
    repeat
        gotoxy(15,10);
        write(mens);
        readln(Submenuopt);
    until SubMenuopt in ['7','5'];
    end;
end; {case}
clrscr;
quadro(8,2,74,6);
gotoxy(10,3);
writeln('FREQUENCIAS ESTIMADAS E FREQUENCIAS MARGINAIS UNIDIMENSIONAIS E');
gotoxy(13,5);
writeln('BIDIMENSIONAIS SOB MODELO DE INDEPENDENCIA CONDICIONAL');
writeln;writeln;
for I:=1 to R do
    for J:=1 to S do
        for K:=1 to T do
            Ab[I,J,K]:=SomaTotal/(R*S*T);          {*** Xijk(o) ***}
        case Submenuopt of
            '5','7' : begin VI := R; VJ := S; VK := T; end;
            '6' : begin VI := R; VJ := T; VK := S; end;
        end;{case}
    for I := 1 to VI do
        for J := 1 to VJ do
            begin
                Soma := 0;
                for K := 1 to VK do
                    begin
                        case Submenuopt of
                            '5','7' : Soma := Soma + Ab[I,J,K];
                            '6' : Soma := Soma + Ab[I,K,J];
                        end;{case}
                    end;
                case Submenuopt of
                    '5','7' : C1b@[I,J]:= Soma;          { Xij. (o) }
                    '6' : C2b@[I,J]:= Soma;          { Xi.k (o) }
                end; {case}
            end; {J}

                { ***** Obtencao de Xijk (i) ***** }

for I := 1 to VI do

```

```

for J := 1 to VJ do
  for K := 1 to VK do
    begin
      case Submenuopt of
        '5','7' : Ab[I,J,K] := (C1[I,J] / C1b0[I,J]) * Ab[I,J,K];
        '6' : Ab[I,K,J] := (C2[I,J] / C2b0[I,J]) * Ab[I,K,J];
      end;{case}
    end;{K}

    { ***** obtencao de Xi.k (1) e X.jk (1) ***** }
  case Submenuopt of
    '5','6' : begin VI := S; VJ := T; VK := R; end;
    '7' : begin VI := R; VJ := T; VK := S; end;
  end;{case}
  for I := 1 to VI do
    for J := 1 to VJ do
      begin
        Soma := 0;
        for K := 1 to VK do
          begin
            case Submenuopt of
              '5','6' : Soma := Soma + Ab[K,I,J];
              '7' : Soma := Soma + Ab[I,K,J];
            end;{case}
          end; {K}
          case Submenuopt of
            '5','6' : C3b1[I,J] := Soma; { X.jk (1) }
            '7' : C2b1[I,J] := Soma; { Xi.k (1) }
          end; {case}
        end; {J}

        { ***** Obtencao de Xijk (2) ***** }
      for I := 1 to VI do
        for J := 1 to VJ do
          for K := 1 to VK do
            begin
              case Submenuopt of
                '5','6' : Ab[K,I,J] := (C3[I,J] / C3b1[I,J]) * Ab[K,I,J];
                '7' : Ab[I,K,J] := (C2[I,J] / C2b1[I,J]) * Ab[I,K,J];
              end;{case}
            end;{K}
          end;{procedure}

        procedure ImprimeXb;
        begin
          for I := 1 to R do
            for J := 1 to S do
              for K := 1 to T do
                begin
                  writeln('X(b) [',I,',',J,',',K,'] = ',Ab[I,J,K]:9:4);
                  TesteParar;
                end;
              TecleEnter;
            clrscr;
            cont := @;
          end;
        end;

        procedure CalculoDaEIMDbSobIndependenciaCondicional(Z,Z1 : MatA);
        var H,G,F : integer;
        begin
          {Obtencao de 2I(X:X(b))}
          EIMDb := 0;

```

```

for F :=1 to R do
  for G :=1 to S do
    for H := 1 to T do
      if (ZCF,G,H) <> 0) then
        EIMDb := EIMDb + 2*(ZCF,G,H) * ln(ZCF,G,H) / Z1CF,G,H));
case Submenuopt of
  '5' : GLb := S * ( R - 1) * ( T - 1);
  '6' : GLb := T * ( R - 1) * ( S - 1);
  '7' : GLb := R * ( S - 1) * ( T - 1);
end;{case}
end;{procedure}

procedure TesteDoModeloDeIndependenciaCondicional;
begin
  quadro(11,1,50,3);
  gotoxy(13,2);
  writeln('TESTE DO MODELO DE INDEPENDENCIA CONDICIONAL');
  gotoxy(1,5);
  write('Entre com qui-quadrado com ',GLb:3:0,' Graus de Liberdade = ');
  readln(qui-quadrado);
  gotoxy(43,7); writeln('EIMDb (b) = ',EIMDb:7:4); gotoxy(16,13);
  if (EIMDb >= qui-quadrado) then
    writeln('Rejeita-se o Modelo de Independencia Condicional')
  else
    writeln('Nao rejeita-se o Modelo de Independencia Condicional');
  writeln;writeln;
  TeclaEnter;
  clrscr;
end;{procedure}

procedure EstimativasDeFrequenciasEsperadasSobModeloDeNaoInteracaoDeSegundaOrdem;

Type
  MatA2 = array[1..10,1..10,1..10] of real;
  MatC1 = array[1..10,1..10] of real;

Var
  C120,C220,C320,C12,C22,C32 : MatC1;
  I,J,K,L,M,contador,Aux : Integer;
  Soma : Real;
  AuxTesteFin,TesteFin : Boolean;

begin
  for I := 1 to R do
    for J := 1 to S do
      for K := 1 to T do
        A2[I,J,K] := SomaTotal/(R*S*T) ;
      contador := 0;
    repeat
      TesteFin:= True;
      for L := 1 to 3 do
        begin
          case L of
            1 : begin VI :=R; VJ := S; VK := T; end;
            2 : begin VI :=R; VJ := T; VK := S; end;
            3 : begin VI :=S; VJ := T; VK := R; end;
          end;{case}
          for I := 1 to VI do
            begin
              for J := 1 to VJ do

```

```

begin
  Soma := 0;
  for K := 1 to VK do
    begin
      case L of
        1 : Soma := soma + A2[I,J,K];      {Xij.(o)}
        2 : Soma := soma + A2[I,K,J];      {Xi.k(1)}
        3 : soma := soma + A2[K,I,J];      {X.jk(2)}
      endi{case}
    endi{K}
    case L of
      1 : C120[I,J] := soma;                {Xij.(o)}
      2 : C220[I,J] := soma;                {Xi.k(1)}
      3 : C320[I,J] := soma;                {X.jk(2)}
    endi{case}
  endi{J}
endi{I}
for I := 1 to VI do
  for J := 1 to VJ do
    for K := 1 to VK do
      begin
        case L of
          1 : A2[I,J,K]:= (C1[I,J]/C120[I,J])*A2[I,J,K];  {Xijk(1)}
          2 : A2[I,K,J]:= (C2[I,J]/C220[I,J])*A2[I,K,J];  {Xijk(2)}
          3 : A2[K,I,J]:= (C3[I,J]/C320[I,J])*A2[K,I,J];  {Xijk(3)}
        endi{case}
      endi{K}
    endi{L}
  }
  {Apenas para conferir passo a passo :
  for I := 1 to R do
    for J := 1 to S do
      for K := 1 to T do
        begin
          writeln('X(2) [' ,I ,',' ,J ,',' ,K ,']: ' ,A2[I,J,K]:9:3);
          TesteParar;
        end;
      cont := 0;
    TeclerEnter;
    CalcularSomatoriasMarginaisBidimensionais(A2,'B'); }
  for M := 1 to 3 do
    begin
      case M of
        1 : begin VI := R; VJ := S; VK := T; end;
        2 : begin VI := R; VJ := T; VK := S; end;
        3 : begin VI := S; VJ := T; VK := R; end;
      endi{case}
      for I := 1 to VI do
        begin
          for J := 1 to VJ do
            begin
              soma := 0;
              for K := 1 to VK do
                begin
                  case M of
                    1 : soma := soma + A2[I,J,K];
                    2 : soma := soma + A2[I,K,J];
                    3 : soma := soma + A2[K,I,J];
                  endi{case}
                endi{J}
              case M of

```

```

1 : begin
  C12[I,J] := soma;
  if (abs(C1[I,J]-C12[I,J]) <=0.001)
    then AuxTesteFin:= true
    else AuxTesteFin:= false;
  if (TesteFin=true) and (AuxTesteFin=true)
    then TesteFin:=true
    else TesteFin:=false;
  end;
2 : begin
  C22[I,J] := soma;
  if (abs(C2[I,J]-C22[I,J]) <=0.001)
    then AuxTesteFin:= true
    else AuxTesteFin:= false;
  if (TesteFin=true) and (AuxTesteFin=true)
    then TesteFin:=true
    else TesteFin:=false;
  end;
3 : begin
  C32[I,J] := soma;
  if (abs(C3[I,J]-C32[I,J]) <=0.001)
    then AuxTesteFin:= true
    else AuxTesteFin:= false;
  if (TesteFin=true) and (AuxTesteFin=true)
    then TesteFin:=true
    else TesteFin:=false;
  end;
end;{case}
end;{J}
end;{I}
end;{M}
contador := contador + 1;
until (TesteFin=true);
for I := 1 to R do
  for J := 1 to S do
    for K := 1 to T do
      begin
        writeln('X(2) [' ,I,',',J,',',K,'] = ',A2[I,J,K]:9:4);
        TesteParar;
      end;
    cont := 0;
    TeclEnter;
    gotoxy(5,30);writeln('Numero de ciclos = ',contador:3);
    TeclEnter;
  end;{procedure}

procedure CalculoDaEIMD2sobNaoInteracaoDeSegundaOrdem(Z,Z1 : MatA);
var H,G,F : integer;
begin
  {Obtencao de Z1(X:Z(2))}
  EIMD2 := 0;
  for F :=1 to R do
    for G :=1 to S do
      for H := 1 to T do
        if (Z[F,G,H] <> 0) then
          EIMD2 := EIMD2 + 2*(Z[F,G,H] * ln(Z[F,G,H] / Z1[F,G,H]));
        GL2 := (R - 1)* (S - 1) * (T - 1);
      end;
    end;

procedure TesteDoModeloDeNaoInteracaoDeSegundaOrdem;
begin

```

```

clrscr;
quadro(14,1,66,3);
gotoxy(16,2);
writeln('TESTE DO MODELO DE NAO INTERACAO DE SEGUNDA ORDEM');
gotoxy(1,5);
write('Entre com qui-quadrado com ',GL2:3:0,' Graus de Liberdade = ');
readln(quiquadrado);
gotoxy(44,7); writeln('EIMD(2) = ',EIMD2:7:4); gotoxy(10,13);
if (EIMD2 >= quiquadrado) then
  writeln('Rejeita-se o Modelo de Nao Interacao de Segunda Ordem')
else
  writeln('Nao rejeita-se o Modelo de Nao Interacao de Segunda Ordem');
writeln;writeln;
TeclEnter;
clrscr;
end;{procedure}

```

```

procedure EscreverQuadroDeAnaliseDeInformacao;
begin
  clrscr;
  quadro(27,1,51,3);
  gotoxy(29,2);writeln('ANALISE DE INFORMACAO');
  gotoxy(11,4);writeln('COMPONENTES');
  gotoxy(38,4);writeln('INFORMACAO');
  gotoxy(68,4);writeln('G.L. ');
  risco('h',5,1,71,196);
  gotoxy(14,6);writeln('R,S,T');
  gotoxy(37,6);writeln('2I[X:X(1)] = ',EIMD1:7:4);
  gotoxy(67,6);writeln(GL1:3:0);
  case menuopt of
    '2' : begin gotoxy(12,9);writeln('Efeito de ST'); end;
    '3' : begin gotoxy(12,9);writeln('Efeito de RT'); end;
    '4' : begin gotoxy(12,9);writeln('Efeito de RS'); end;
  end;
  gotoxy(34,9);writeln('2I[X(a):X(1)] = ',(EIMD1 - EIMDa):7:4);
  gotoxy(67,9);writeln((GL1 - GLa):3:0);
  case menuopt of
    '2' : begin gotoxy(7,10);writeln('R independente de ST');end;
    '3' : begin gotoxy(7,10);writeln('S independente de RT');end;
    '4' : begin gotoxy(7,10);writeln('T independente de RS');end;
  end;
  gotoxy(37,10);writeln('2I[X:X(a)] = ',EIMDa:7:4);
  gotoxy(67,10);writeln(GLa:3:0);
  case menuopt of
    '2' : begin
      if submenuopt = '5' then
        begin
          gotoxy(11,13);writeln('Efeito de RS/ST')
        end
      else
        begin
          gotoxy(11,13);writeln('Efeito de RT/ST')
        end;
      end;
    '3' : begin
      if submenuopt = '7' then
        begin
          gotoxy(11,13);writeln('Efeito de RS/RT')
        end
      end;
    end;
  end;
end;

```



```

        else
            begin
                gotoxy(11,13);writeln('Efeito de ST/RT')
            end;
        end;
    '4' : begin
        if submenuopt = '7' then
            begin
                gotoxy(11,13);writeln('Efeito de RT/RS')
            end
        else
            begin
                gotoxy(11,13);writeln('Efeito de ST/RS')
            end;
        end;
    end;
end;
gotoxy(34,13);writeln('2I[X(b):X(a)] = ',(EIMDa - EIMDb):7:4);
gotoxy(67,13);writeln((GLa -GLb):3:0);
gotoxy(3,14);writeln('Independencia Condicional de');
case submenuopt of
    '5' : begin gotoxy(11,15);writeln('(R x T)/S');end;
    '6' : begin gotoxy(11,15);writeln('(R x S)/T');end;
    '7' : begin gotoxy(11,15);writeln('(S x T)/R');end;
end;
gotoxy(37,15);writeln('2I[X:X(b)] = ',EIMDb:7:4);
gotoxy(67,15);writeln((GLb):3:0);
case submenuopt of
    '5' : begin gotoxy(10,18);writeln('Efeito de RT/ST,RS');end;
    '6' : begin gotoxy(10,18);writeln('Efeito de RS/ST,RT');end;
    '7' : begin gotoxy(10,18);writeln('Efeito de ST/RT,RS');end;
end;
gotoxy(34,18);writeln('2I[X(2):X(b)] = ',(EIMDb -EIMD2):7:4);
gotoxy(67,18);writeln((GLb - GL2):3:0);
gotoxy(5,19);writeln('Interacao de segunda ordem');
gotoxy(37,19);writeln('2I[X:X(2)] = ',EIMD2:7:4);
gotoxy(67,19);writeln((GL2):3:0);
risco('h',20,1,71,196);
end;

{Programa Principal}
var ret : char;
begin
    ret := ' ';
    TelaInicial;
    EntradaDeDados;
    quadro(22,2,57,4);
    gotoxy(24,3);
    writeln('FREQUENCIAS MARGINAIS OBSERVADAS');
    writeln;
    CalcularSomatoriasMarginaisUnidimensionais(A);
    CalcularSomatoriasMarginaisBidimensionais(A, 'A');
    quadro(8,2,72,6);
    gotoxy(10,3);
    writeln('FREQUENCIAS ESTIMADAS E FREQUENCIAS MARGINAIS UNIDIMENSIONAIS');
    gotoxy(10,5);
    writeln('SOB O MODELO DE INDEPENDENCIA MUTUA COMPLETA');
    writeln;writeln;
    EstimativasDasFrequenciasEsperadasSobModeloDeIndependenciaMutuaCompleta;
    CalcularSomatoriasMarginaisUnidimensionais(A1);
    CalculoDaEIMD1SobIndependenciaMutuaCompleta(A,A1);

```

```

TesteDoModeloDeIndependenciaMutuaCompleta:
repeat
  EstimativasDasFrequenciasEsperadasSobModeloDeIndependenciaParcial;
  ImprimeXa;
  CalcularSomatoriasMarginaisUnidimensionais(Aa);
  CalcularSomatoriasMarginaisBidimensionais(Aa, 'B');
  CalculoDaEIMDaSobIndependenciaParcial(A,Aa);
  TesteDoModeloDeIndependenciaParcial;
  EstimativasDeFrequenciasEsperadasSobModelosDeIndependenciaCondicional;
  ImprimeXb;
  CalcularSomatoriasMarginaisUnidimensionais(Ab);
  CalcularSomatoriasMarginaisBidimensionais(Ab, 'B');
  CalculoDaEIMDbSobIndependenciaCondicional(A,Ab);
  TesteDoModeloDeIndependenciaCondicional;
  quadro(16,2,58,6);
  gotoxy(18,3);
  writeln('FREQUENCIAS ESTIMADAS SOB MODELO DE NAO');
  gotoxy(25,5);
  writeln('INTERACAO DE SEGUNDA ORDEM');
  writeln;writeln;
  EstimativasDeFrequenciasEsperadasSobModeloDeNaoInteracaoDeSegundaOrdem;
  CalculoDaEIMD2SobNaoInteracaoDeSegundaOrdem(A,A2);
  TesteDoModeloDeNaoInteracaoDeSegundaOrdem;
  EscreverQuadroDeAnaliseDeInformacao;
  writeln;
  write('Quer testar outra sequencia de modelos hierarquicos ? (s/n) ');
  readln(ret); ret := upcase(ret);
until(ret = 'N')
end.{programa}

```