

Título :

COMPARAÇÃO DE MATRIZES
DE
DEPENDÊNCIA E CODEPENDÊNCIA

Este exemplar corresponde a redação final
da tese devidamente corrigida e defendida
por Cibele Queiroz da Silva e aprovada pela
Comissão Julgadora

Campinas, 30 de julho de 1991

p/ Hermann Rohrer
Prof. Dr. Antonio José Cordeiro

Orientador

Hermann Rohrer
Prof. Dr. Hermann G. Rohrer
Co-Orientador

Dissertação apresentada ao Instituto de
Matemática, Estatística e Ciências da
Computação, UNICAMP, para obtenção do
Título de Mestre em Estatística

8c

330/BC

02.7109193

UNICAMP
BIBLIOTECA CENTRAL

Ao meu pai José Maria, homem sábio
que me ensinou a amar os livros

Agradecimentos

Aos profs. Drs. José Antonio Cordeiro e Hermann G. Rohrer
por suas críticas e sugestões.

Às professoras Dras. Regina C. P. Moran e Cicília Yuko Wada

Ao meu companheiro Eduardo e aos amigos Aloísio J. F.
Ribeiro e José de Sousa Ramos.

E a todos que direta ou indiretamente contribuíram para a realização
deste trabalho.

Índice

Capítulo 1 - Introdução 1	01
Capítulo 2 - A distância de Hellinger	20
2.1 - Introdução	20
2.2 - Divergências, Distâncias e Informação Discriminatória	22
2.3 - Desigualdades	25
2.4 - Afinidade e Distância - Definições Básicas	26
2.5 - A Distância de Hellinger como membro de uma Família de Estatísticas de Divergência em Potência	34
2.6 - Distribuições Qui-quadrado como Limite	36
Capítulo 3 - Inferência Baseada na Distância de Hellinger	39
3.1 - Teoremas Fundamentais	39
3.2 - Testes de Hipóteses e de Aderências	42
3.3 - Tamanhos de Amostras	49
3.4 - Aplicação do Teorema V - Teste de Independência	50

3.5 - Discussão	54
3.6 - Exemplo Numérico	55
3.7 - Coeficiente de Associação	
Baseado na Distância de Hellinger	58
3.8 - Teste de Independência para Tabelas de Contingência	
Multidimensionais baseado na Distância de Hellinger	67
3.9 - Dependência e Codependência	71
3.10 - Análise de Componentes Principais utilizando	
Coeficientes de Dependência e Codependência	80
 Capítulo 4 - Comparação de Matrizes de Dependência e Codependência	
4.1 - Descrição do problema e Metodologia	88
4.2 - Geração de Matrizes de Dependência e Codependência	
para o cálculo do Critério de Comparação ξ	97
4.3 - Resultados	102
Apêndice1 -	121
Apêndice2 -	130
Referências Bibliográficas	148

Capítulo 1

Introdução

1.1 TABELAS DE CONTINGÊNCIA

Dados coletados em áreas das ciências exatas representam frequentemente resultados experimentais de medidas com variáveis *quantitativas*. Em outras áreas de estudo, no entanto, os dados raramente resultam de experimentos controlados. Além disto, as observações usualmente podem ser expressas somente em termos *qualitativos* ou *categóricos*, tais como sim - não, vivo - morto, concorda - discorda, classe A - classe B - classe C, etc. Com este tipo de *resposta* obtém-se tabelas de frequências ou contagens.

Um indivíduo de dada população pode ser *classificado* por exemplo, por sexo, raça, hábito de fumar, por nível de renda, etc; para uma amostra de indivíduos provenientes desta população, dos indivíduos classificados segundo as *características* acima resulta uma *Tabela de Contingência*, com tantas dimensões quantas forem as *classificações* ou *características*.

Uma Tabela de Contingência é uma forma de apresentação de dados agrupados. Em um caso simples, um grupo de N itens pode ser classificado em dois grupos de acordo com, digamos, presença e ausência de certas características. Em uma característica fixa, os diferentes grupos de classificação são denotados por *categorias*. Por exemplo, N indivíduos podem ser classificados de acordo com a Cor dos cabelos (característica) e as categorias serão preto, castanho, louro, etc. As categorias podem ser *qualitativas*, como as últimas, ou *quantitativas* como, por exemplo, na classificação de Peso em quilogramas consistindo de quatro categorias :

40 - 50, 50 - 60, 60 - 70 e 70 - 80.

Quando existir sómente uma característica de acordo com a qual os indivíduos são classificados, obtem - se o que se denomina por *Tabela de Frequências*. Se existirem duas formas de classificação, digamos por linhas e colunas, a tabela resultante é denominada por *Tabela de duas Entradas*, ou tabela *IxJ*. Esta última notação proporciona o número de categorias em cada classificação, desta feita, uma tabela *IxJxK* terá três características de classificação, onde a primeira terá *I* categorias, a segunda terá *J* categorias e a terceira terá *K* categorias de classificação.

EXEMPLOS

As tabelas 1.1 e 2.1 são exemplos de Tabelas de uma Entrada, e podem ser encontradas em Kullback e Gokhale [35]. Na tabela 1.1 existe apenas uma característica de classificação, Área Geográfica, e quatro categorias. Tal tabela proporciona a distribuição de Estudantes por Área Geográfica :

Tabela 1.1.- Distribuição de Estudantes por Área Geográfica

LESTE	NORTE	OESTE	SUL	TOTAL
4201	4552	2840	5130	16723

Tabela 2.1 - Distribuição de vinte bolas em seis caselas

Casela	1	2	3	4	5	6	Total
Frequência	2	4	4	5	1	4	20

Pode ser observado neste ponto, que a situação da distribuição de *N* bolas em *k* caselas é adequadamente descrita pela distribuição Multinomial. É de se esperar portanto, que a distribuição Multinomial tenha um papel importante na análise de Tabelas de Contingência.

A tabela 3.1 é um exemplo de Tabela de duas Entradas. A distribuição de estudantes por Área Geográfica e Sexo resulta em uma Tabela de Contingência 2x4 :

Tabela 3.1 - Distribuição de estudantes por Área Geográfica e Sexo

SEXO	ÁREA GEOGRÁFICA				TOTAIS
	LESTE	NORTE	OESTE	SUL	
MASCULINO	2201	2350	1400	3100	9051
FEMININO	2000	2202	1440	2030	7672
TOTAIS	4201	4552	2840	5130	16723

Cada uma das combinações das categorias de Sexo e Área Geográfica, resulta em uma casela da Tabela de Contingência. A tabela 3.1 possui $2 \times 4 = 8$ caselas. As frequências de ocorrência das caselas resultam da classificação de um número fixo de indivíduos (no caso da tabela acima, $N = 16723$) de acordo com uma certa distribuição de probabilidade, em geral desconhecida. Em outras palavras, um indivíduo é classificado em uma dada casela de acordo com uma probabilidade associada a esta casela. Neste exemplo, a distribuição associada aos indivíduos amostrados pode ser considerada como uma Multinomial com oito caselas.

Neste ponto introduzimos algumas notações gerais para Tabelas de Contingência $I \times J$. Estas notações podem ser extendidas para Tabelas de Contingência de maiores dimensões.

A Forma Geral de uma Tabela de Contingência $I \times J$ é dada pela tabela 4.1, na qual uma amostra aleatória simples de N indivíduos é classificada com respeito a duas variáveis (características) qualitativas, uma tendo I categorias e a outra tendo J categorias.

A frequência observada na i -ésima categoria da Variável - Linha e j -ésima categoria da Variável - Coluna, i.é., a frequência da ij -ésima casela da Tabela de Contingência, é representada por n_{ij} .

O número Total de observações na i -ésima categoria de Variável-Linha, é denotado por $n_{i.}$, e o número Total de observações na j -ésima categoria da Variável-Coluna por $n_{.j}$. Estes são conhecidos por *Totais Marginais*, e em termos das frequências de casela, n_{ij} , são dados por : $n_{i.} = \sum_j n_{ij}$ e $n_{.j} = \sum_i n_{ij}$.

Tabela 4.1 Forma Geral de uma Tabela de Contingência $I \times J$

		COLUNAS (VARIÁVEL 2)				TOTAIS
		1	2	...	J	
LINHAS (VARIÁVEL 1)	1	n_{11}	n_{12}	...	n_{1J}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2J}	$n_{2.}$
	:					
	:					
	I	n_{i1}	n_{i2}	...	n_{iJ}	$n_{i.}$
TOTAIS		$n_{.1}$	$n_{.2}$...	$n_{.J}$	$n_{..} = N$

$n_{..}$ representa o Número Total de indivíduos da amostra e usualmente é denotado por N .

Na Tabela de Contingência $I \times J$, representaremos cada casela por (i,j) com $i=1,...,I$ e $j=1,...,J$. A probabilidade de um indivíduo ser classificado na casela (i,j) é denotada por p_{ij} ; as *probabilidades Marginais* são, conseqüentemente, dadas por :

$$p_{i.} = \sum_j p_{ij}, \quad (1.1.1)$$

$$p_{.j} = \sum_i p_{ij} \quad (1.1.2)$$

com

$$\sum_i \sum_j p_{ij} = 1 \quad (1.1.3)$$

Neste estudo, assumiremos que as probabilidades de classificação dos indivíduos nas caselas permanecem constantes de indivíduo para indivíduo e que os mesmos são amostradas independentemente um ao outro.

Na Tabela de Contingência acima, seja e_{ij} a frequência esperada na ij -ésima casela com

$$e_{ij} = Np_{ij}, \quad (1.1.4)$$

onde $e_{ij} = E(n_{ij})$ é a Esperança Matemática de n_{ij} sob a suposição de que as Frequências Observadas seguem *Distribuição Multinomial* com probabilidades p_{ij} .

A seguir consideraremos alguns *esquemas amostrais* para a geração de *Tabelas de Classificação Cruzada*. Este termo é genérico e pode ser utilizado para quaisquer das Tabelas geradas pelos esquemas amostrais abaixo .

1.2 - MODELOS AMOSTRAIS

Existem três *modelos* que são comumente utilizados para a amostragem de dados de contagem :

(1) *Poisson* : Observa-se um conjunto de processos de Poisson (totalmente independentes), um para cada casela da Tabela de Classificação Cruzada, sob um período fixo de tempo e sem nenhum conhecimento a priori do número total de observações na tabela.

Cada processo produz uma frequência observada da correspondente casela e, no caso de Tabelas de Classificação Cruzada $I \times J$, as frequências observadas n_{ij} têm distribuição de Poisson com frequências esperadas λ_{ij} . A probabilidade de uma ocupação ($n_{ij} : 1 \leq i \leq I, 1 \leq j \leq J$) para este esquema amostral é dado pelo produto das probabilidades em todas as caselas :

$$\prod_{i,j} \frac{\lambda_{ij}^{n_{ij}} \exp(-\lambda_{ij})}{n_{ij}!} \quad (1.2.1)$$

Como as caselas têm distribuições de Poisson totalmente independentes, a frequência total da tabela, $n_{..}$, tem também distribuição de Poisson com um *Número Total Esperado* $\lambda_{..} = \sum_{i,j} \lambda_{ij}$.

(2) *Multinomial* : Toma-se uma amostra de tamanho fixo N e procede-se à classificação cruzada dos indivíduos. Este modelo foi assumido em todas as tabelas acima. A probabilidade de ocupação para este esquema amostral é dado por :

$$\frac{N!}{\prod_{i,j} n_{ij}!} \prod_{i,j} p_{ij}^{n_{ij}} \quad (1.2.2)$$

onde $\sum_{i,j} p_{ij} = 1$ e $\sum_{i,j} n_{ij} = N$.

(3) *Produto de Multinomiais* : No caso de Tabelas $I \times J$, para cada categoria da variável-Linha tem-se uma amostra (Multinomial) de tamanho $n_{i.}$ independente das outras e classifica-se cada membro da amostra de acordo com sua categoria para a variável-Coluna (ou vice-versa). A probabilidade de ocupação para este esquema amostral é dado por :

$$\prod_i \left[\frac{n_{i.}!}{\prod_j n_{ij}!} \prod_j p_{ij}^{n_{ij}} \right]. \quad (1.2.3)$$

Neste esquema tem-se I amostras Multinomiais independentes de tamanhos n_i , $i=1,\dots,I$.

Estes esquemas amostrais podem ser extendidos para Tabelas de Classificação Cruzada de maiores dimensões. O termo Tabela de Classificação Cruzada é geral, enquanto o termo Tabela de Contingência é conveniente para o *esquema amostral* (2) (vide Everitt [09] cap. 1).

1.3 - INVESTIGAÇÕES COM TABELAS DE CONTINGÊNCIA $I \times J$ - TESTES

Uma análise de Tabelas de Contingência consiste, geralmente, de um estudo sobre o relacionamento entre as classificações ou variáveis. Uma questão de muito interesse é saber se as classificações ou características são *independentes* ou não.

Para responder a esta questão é necessário esclarecer o que vem a ser independência entre duas características. No caso de uma Tabela 2×2 , isto é relativamente fácil.

Para a tabela 5.1 (vide Everitt [09]) é de se esperar que a proporção de homens que morrem de Tuberculose do Sistema Respiratório seja igual à proporção de mulheres que morrem da mesma causa se a Forma de Tuberculose da qual a pessoa morre for independente de seu sexo. Se as proporções diferem, morte por Tuberculose do Sistema Respiratório tende a estar mais vinculada a um dos sexos do que com o outro. Logo, a independência em Tabelas 2×2 implica na igualdade de duas proporções.

Tabela 5.1 Mortes por Tuberculose

	HOMENS	MULHERES	TOTAIS
TUBERCULOSE DO SIST. RESPIRATÓRIO	3534	1319	4853
OUTRAS FORMAS DE TUBERCULOSE	270	252	522
TOTAIS	3804	1571	5375

Examinaremos o conceito de Independência de maneira mais formal, para Tabelas de Contingência $I \times J$.

A independência em probabilidade entre as variáveis Linha e Coluna implica em

$$p_{ij} = p_{i.} p_{.j}, \quad (1.3.1)$$

em termos das Frequências Esperadas e_{ij} , independência implica em

$$e_{ij} = N p_{i.} p_{.j}. \quad (1.3.2)$$

As probabilidades p_{ij} são geralmente desconhecidas mas é possível estimá-las de maneira simples a partir das frequências observadas n_{ij} .

Em Agresti [01] cap. 3, é mostrado que as *melhores* estimativas para p_{ij} , $p_{i.}$ e $p_{.j}$, respectivamente, f_{ij} , $f_{i.}$ e $f_{.j}$, são dadas por :

$$f_{ij} = n_{ij}/N, \quad (1.3.3)$$

$$f_{i.} = n_{i.}/N \text{ e} \quad (1.3.4)$$

$$f_{.j} = n_{.j}/N. \quad (1.3.5)$$

Esta nomenclatura é facilmente estendida para Tabelas de Contingência de maiores dimensões.

Utilizando as estimativas de $p_{i.}$ e $p_{.j}$ é possível estimar a frequência esperada e_{ij} , na ij -ésima casela, se as variáveis Linha e Coluna forem independentes. A Frequência Esperada e_{ij} é estimada por :

$$E_{ij} = N f_{i.} f_{.j} = \frac{n_{i.} \cdot n_{.j}}{N} \quad (1.3.6)$$

Quando as variáveis Linha e Coluna são independentes, as Frequências Estimadas E_{ij} e as Frequências Observadas n_{ij} devem diferir por fatores devidos somente ao acaso. Se, no entanto, as duas variáveis não são independentes, pode-se esperar diferenças sistematicamente maiores entre E_{ij} e n_{ij} . Consequentemente, é coerente basear um *Teste de Independência* entre as variáveis Linha e Coluna no tamanho das diferenças entre os dois conjuntos de frequências E_{ij} e n_{ij} . Um Teste possível é o denominado Teste Qui-quadrado discutido entre outros, na secção 2.1.

No caso de não haver independência em probabilidade entre as variáveis Linha e Coluna significa que existe alguma *forma de associação* entre estas.

Algumas *formas de associação* mais conhecidas podem ocorrer em Tabelas de Contingência quadradas, ou seja, $I \times I$. Estas tabelas podem surgir de diferentes situações (vide Everitt [09] cap. 6):

(I) Quando uma amostra de indivíduos é classificada de acordo com duas variáveis categóricas similares; por exemplo, o Grau de visão do olho direito e esquerdo.

(II) Quando amostras de pares de indivíduos tais como esposo e esposa, ou pais e filhos, são classificados de acordo com alguma variável categórica de interesse, comum a cada metade do par.

(III) Quando indivíduos que podem ser, por exemplo, objetos de uma Linha de Produção são submetidos a algum tipo de exame por dois examinadores independentemente, e as notas que cada indivíduo obtém são dadas de acordo com uma escala comum e pré-estabelecida.

Para Tabelas de Contingência resultantes dos esquemas I, II ou III, é de interesse testar a *simetria e homogeneidade das marginais*. A simetria em tabelas quadradas significa que :

$$p_{ij} = p_{ji} \quad (i \neq j) \quad (1.3.7)$$

e a homogeneidade das marginais significa que

$$p_{i.} = p_{.i} \quad (\text{para } i = 1, \dots, I). \quad (1.3.8)$$

Em tabelas 2x2 estes esquemas são equivalentes; em tabelas maiores, a simetria definida em (1.3.7) implica na homogeneidade de marginais definida em (1.3.8).

É possível a utilização de testes Qui-quadrado para *testar hipóteses* relativas a simetria e homogeneidade de marginais, bastando para tanto que as frequências esperadas de casela sejam convenientemente definidas. Para maiores detalhes vide Everitt [09] cap. 6.

1.4 - TABELAS MULTIDIMENSIONAIS

Os métodos para análise de Tabelas de Contingência $I \times J$ são geralmente bem conhecidos mesmo por pesquisadores não-estatísticos. No entanto, métodos para análise de tabelas com três ou mais variáveis envolvem conhecimentos mais técnicos. Tais tabelas podem surgir de muitas áreas de estudo, por exemplo, a tabela 6.1 encontrada em Bishop et al [03] descreve os dados relativos à sobrevivência de recém nascidos de acordo com os cuidados pré-natal (menos - mais) recebidos pela mãe. Os dados foram coletados nas Clínicas A e B.

Tabela 6.1 - Tabela *tri-dimensional* relatando a sobrevivência de recém nascidos de acordo com os cuidados recebidos pela mãe em duas clínicas - A e B.

		SOBREVIVÊNCIA DO RÉCEM-NASCIDO			
		VIVO		MORTO	
CUIDADOS PRÉ-NATAL RECEBIDOS PELA MÃE		MENOS	MAIS	MENOS	MAIS
LOCAL ONDE RECEBEU A ASSISTÊNCIA	CLINICA A	3	4	176	293
	CLINICA B	17	2	197	23

O esquema amostral utilizado na tabela acima foi o de Produto de Multinomiais, onde duas amostras foram coletadas independentemente nas Clínicas A e B. A distribuição de probabilidade de cada amostra é Multinomial com quatro caselas.

Na análise desta tabela, é intuitivo que se pense em simplificar o problema reduzindo este caso ao de tabelas $I \times J$. Isto seria feito somando-se os dados, por exemplo, das duas Clínicas. O resultado seria uma tabela combinada com quatro caselas. Uma razão pela qual este procedimento não deveria ser feito é que pode levar a conclusões equivocadas.

Segundo Everitt [09] cap. 2 e 3, a combinação de tabelas, digamos, para o exemplo acima, somente pode ser feita se as proporções obtidas nas Tabelas de Contingência das Clínicas A e B forem muito próximas. Consequentemente, se as proporções variam de tabela para tabela, este procedimento não deve ser utilizado, pois os dados combinados não refletem a informação contida nas tabelas originais.

Para ilustrar o que o problema significa, tomemos como exemplo a tabela 6.1. As estatísticas Qui-quadrado, χ^2_A e χ^2_B , calculadas sob a hipótese de independência em cada uma das clínicas, são aproximadamente zero. Se combinarmos os dados das duas tabelas com respeito às clínicas, i.é., não levando-se em conta o Local onde as mães receberam o atendimento, a estatística Qui-quadrado combinada χ^2_c iguala a 5.26, que é *significativa* com 1 grau de liberdade ao nível de 5% (vide Everitt cap. 4).

Na análise de tabelas separadas conclui-se que a sobrevivência não está associada aos cuidados pré-natais, enquanto na análise da tabela combinada conclui-se justamente o contrário. As proporções das caselas em cada clínica são :

CLINICA A	0.006	0.008	0.370	0.616
-----------	-------	-------	-------	-------

CLINICA B	0.059	0.009	0.860	0.100
-----------	-------	-------	-------	-------

Como se observa, as proporções nas clínicas diferem consideravelmente, e de acordo com os critérios para a combinação de tabelas descritos por Everitt, a tabela 6.1 é um caso onde a combinação das sub-tabelas não deveria ser feita. Este exemplo esclarece a observação de que tabelas multidimensionais não devam ser tratadas como simples combinação de tabelas bi-dimensionais.

1.5 - AJUSTES DE MODELOS

Mencionamos anteriormente algumas maneiras para analisar a associação em Tabelas de Contingência IxJ. A maioria dos estudos, no entanto, inclui muitas variáveis (digamos K) e o tratamento do problema via cálculo das $\binom{K}{2}$ tabelas é inadequado. Para descrever as associações existentes entre todas as variáveis de maneira segura, a construção de Modelos é a técnica adequada.

Esta metodologia é comum em Estatística. Um exemplo é a Análise de Regressão, onde na maioria dos modelos lineares os valores esperados das observações são dados por uma combinação linear de um certo número de *parâmetros*. Técnicas tais como Máxima Verossimilhança e Mínimos Quadrados (vide Bishop et al [03]) podem ser utilizadas para *estimar* os *parâmetros* e para identificar quais variáveis são de maior importância na determinação dos valores das observações. Os modelos para Tabelas de Classificação Cruzada são similares aos utilizados para dados quantitativos.

1.5.1 - MODELOS PARA TABELAS DE CLASSIFICAÇÃO CRUZADA

Modelos para Tabelas de Classificação Cruzada estão descritos em Bishop et al [03], Fienberg [11], Everitt [09], entre outros. As maiores vantagens obtidas com o uso destas técnicas, segundo Fienberg, são primeiro: o tratamento que é dado às tabelas é sistemático e segundo : estimativas da magnitude dos *efeitos* (ou parâmetros do modelo) de mais interesse podem ser obtidos. Consequentemente, é possível julgar a importância relativa dos diferentes efeitos para a escolha de um modelo. Por exemplo, na tabela 6.1 que é 2x2x2, quais seriam os efeitos de maior interesse ?. Para responder a esta pergunta, deve-se introduzir a notação necessária :

$$n_{1j.} = n_{1j1} + n_{1j2} \quad (1.5.1.1)$$

$$n_{1..} = n_{111} + n_{112} + n_{121} + n_{122} = n_{11.} + n_{12.}, \quad (1.5.1.2)$$

e

$$n_{...} = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^2 n_{ijk} = \sum_{i=1}^2 n_{i..} \quad (1.5.1.3)$$

Se as três variáveis, Sobrevivência, Cuidados Pré-parto e Local, que denominaremos por variáveis 1, 2 e 3, nesta ordem, fossem independentes, então valeria a relação

$$p_{ijk} = p_{i..} p_{.j.} p_{..k} \quad (1.5.1.4)$$

e analogamente a (1.3.6), sob a independência, a estimativa para a frequência esperada da casela (i,j,k) seria dada por :

$$E_{ijk} = \left(\frac{n_{i..}}{N} \right) \left(\frac{n_{.j.}}{N} \right) \left(\frac{n_{..k}}{N} \right) N \quad (1.5.1.5)$$

Tomando o logaritmo nos dois lados da igualdade, tem-se

$$\log E_{ijk} = \log n_{i..} + \log n_{.j.} + \log n_{..k} - 2\log N \quad (1.5.1.6)$$

Esta forma aditiva para os logaritmos dos valores esperados é muito utilizada na construção destes tipos de modelos. Tomando o valor esperado de $\log E_{ijk}$ em (1.5.1.6) e lembrando que $e_{ijk} = E(n_{ijk})$, temos

$$v = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^2 \log e_{ijk} = 4 \sum_{i=1}^2 e_{i..} + 4 \sum_{j=1}^2 e_{.j.} + 4 \sum_{k=1}^2 e_{..k} - 8 \times 2 \log N$$

Fazendo

$$u = v/8 = -2\log N + \frac{1}{2} \left[\sum_{i=1}^2 e_{i..} + \sum_{j=1}^2 e_{.j.} + \sum_{k=1}^2 e_{..k} \right], \quad (1.5.1.7)$$

e

$$u_{1(i)} = \frac{1}{4} \sum_{j=1}^2 \sum_{k=1}^2 \log e_{ijk} - u, \quad (1.5.1.8)$$

e de maneira análoga para $u_{2(j)}$ e $u_{3(k)}$, pode-se escrever $\log e_{ijk}$ na forma

$$\log e_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)}. \quad (1.5.1.9)$$

Este modelo é muito conhecido em Análise de Variância (ANOVA) e é conhecido como Modelo Log-linear.

Sendo

$$\sum_{i=1}^2 u_{1(i)} = \sum_{j=1}^2 u_{2(j)} = \sum_{k=1}^2 u_{3(k)} = 0, \quad (1.5.1.10)$$

os termos $u_{1(i)}$, $u_{2(j)}$ e $u_{3(k)}$ representam os desvios da Média Geral u e (1.5.1.10) representam as *restrições* do *Modelo Log-linear de Independência* entre as variáveis 1, 2 e 3.

Supondo que as variáveis 1, 2 e 3 não sejam independentes, surgem outros tipos de modelos que podem ser expressos em termos do *relacionamento estrutural* entre as três variáveis. O *Modelo Log-linear Geral* é dado por :

$$\begin{aligned} \log e_{ijk} = & u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} \\ & + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} \end{aligned} \quad (1.5.1.11),$$

e as restrições do modelo geral são :

$$\begin{aligned} \sum_{i=1}^2 u_{1(i)} &= \sum_{j=1}^2 u_{2(j)} = \sum_{k=1}^2 u_{3(k)} = 0, \\ \sum_{i=1}^2 u_{12(ij)} &= \sum_{j=1}^2 u_{12(ij)} = \sum_{i=1}^2 u_{13(ik)} = \sum_{k=1}^2 u_{13(ik)} \\ &= \sum_{j=1}^2 u_{23(jk)} = \sum_{k=1}^2 u_{23(jk)} = 0, \end{aligned}$$

e

$$\sum_{j=1}^2 u_{123(1jk)} = \sum_{j=1}^2 u_{123(1jk)} = \sum_{k=1}^2 u_{123(1jk)} = 0,$$

vide Fienberg [11] cap 3.

Os termos $u_{1(i)}$, $u_{2(j)}$ e $u_{3(k)}$ representam os *efeitos principais* das variáveis 1, 2 e 3, ao passo que $u_{12(ij)}$, $u_{13(ik)}$ e $u_{23(jk)}$ representam os *efeitos de interação de primeira ordem* e $u_{123(ijk)}$ representa o *efeito de interação de segunda ordem*.

No caso de

$$u_{12(ij)} = u_{13(ik)} = u_{23(jk)} = u_{123(ijk)} = 0,$$

o modelo acima se reduz ao modelo (1.5.1.9) de independência completa.

Modelos descritivos com tantos parâmetros independentes quantos forem as caselas de uma Tabela de Classificação Cruzada são chamados de *Saturados*. O modelo (1.5.1.11) é um exemplo de modelo saturado.

Estes modelos são úteis na *redução da complexidade* de um problema, pois proporcionam meios seguros para que se possa analisar a possibilidade da redução do número de variáveis na tabela.

Em termos de modelos, a complexidade de uma problema é refletida pelo número de parâmetros que devem estar presentes para que se tenha um ajuste satisfatório quando se observa as divergências entre os valores ajustados e observados. Quando o relacionamento estrutural é simples, o modelo possui poucos parâmetros.

A estatística Qui-quadrado, entre outras, é utilizada para testes de ajuste (vide Bishop [03]).

Quando o modelo possui menos parâmetros do que o número de caselas, diz-se que o modelo é *Insaturado*. Em alguns modelos insaturados pode-se reduzir o número de caselas na Tabela, sem que haja distorção na estrutura de associação das variáveis.

A redução da complexidade de um problema via modelos insaturados, é norteadada por algumas condições discutidas por Bishop et al [03] cap.2, e Fienberg [11] cap. 3, e é conhecida pelo termo *COLLAPSING*.

Esta técnica proporciona *ferramentas* para que se possa estudar o problema da combinação de tabelas discutido na seção 1.4.

Suponha, para o caso de tabelas $I \times J \times K$, que estejamos interessados em estudar o relacionamento entre as variáveis 1 e 2, utilizando um modelo Log-linear de frequências esperadas das caselas $\{e_{ijk}\}$, ou seja, deseja-se estudar os termos de interação $u_{12(ij)}$. Se agregarmos a terceira variável, produzimos uma tabela de frequências $\{e_{ij}\}$, queremos saber se os termos de interação $u_{12(ij)}^*$, para esta tabela marginal, são os mesmos $u_{12(ij)}$, calculados de acordo com uma tabela tri-dimensional.

Fienberg [11] cap. 3, apresenta um Teorema que indica os casos onde é razoável proceder o *collapsing* :

Teorema 1.0 - As interações entre duas variáveis em uma Tabela tri-dimensional podem ser medidas a partir de uma tabela marginal obtida agregando a terceira variável se a terceira variável for independente de no mínimo uma das duas variáveis restantes.

Isto significa para o caso da tabela 6.1, que podemos medir u_{12} a partir de uma tabela marginal bi-dimensional se e somente se $u_{13} = 0$ ou $u_{23} = 0$. Bishop [03], ajustou o modelo (1.5.1.11) para a tabela 6.0 em questão e encontrou que $u_{13} \neq 0$, $u_{23} \neq 0$ e $u_{12} = 0$. O autor conclui que u_{12} não pode ser avaliada em termos da tabela combinada, mas segundo o ajuste do modelo (1.5.1.11), as variáveis 1 e 2 não são associadas.

Nosso estudo, neste trabalho, é voltado para a redução de dimensão em Tabelas de Contingência do tipo $I \times J \times 2$. A abordagem, no entanto, é distinta da que descrevemos acima, pois não faremos ajustes de Modelos, mas sim, o que se pretende é investigar a possibilidade da construção de um teste para a hipótese

$$H_0 : D_1 = D_2$$

onde D_1 e D_2 , são Matrizes de Dependência e Codependência, definidas por Cordeiro [05].

Roy [31] propôs, em 1957, o critério das Raízes Características Máxima e Mínima, para comparar matrizes V_1 e V_2 de Variância e Covariância onde os autovalores extremos de $V_2^{-1}V_1$ são encontrados e se espera que sejam iguais quando $H_0 : V_1 = V_2$ for verdadeira. O mesmo enfoque foi utilizado por Flury [12].

Os elementos de matrizes de Dependência e Codependência D_1 e D_2 , mencionados anteriormente, são obtidos em função das divergências medidas por meio da distância de Hellinger (vide Cordeiro [05]) entre a distribuição observada na tabela e sua distribuição sob a hipótese de independência. Ter-se-ia portanto as matrizes de Dependência e Codependência D_1 e D_2 para cada nível da variável 3. O caso $D_1 = D_2$ implica que as tabelas de contingência para cada nível da variável 3 podem ser combinadas, produzindo desta forma uma terceira tabela que deverá refletir as tabelas originais.

Para testar a hipótese

$$H_0 : D_1 = D_2$$

é necessário a construção da distribuição nula de um critério que compare D_1 a D_2 . Este critério deve ser invariante, pelo menos, ao tipo de associação entre linhas e colunas.

O critério utilizado neste trabalho está descrito no capítulo 4, e foi baseado nas idéias de Roy, mencionadas acima.

De acordo com os resultados que obtivemos as distribuições simuladas indicam que os dados se ajustam a famílias de distribuições Beta. Estes resultados, no entanto, não são definitivos. Devido a limitações de espaço em disco apenas um número reduzido de simulações de Monte Carlo puderam ser empreendidas. Naturalmente ao aumentarmos consideravelmente o tamanho das amostras simuladas teremos uma idéia mais clara da distribuição do critério ξ . Além disto, um estudo da distribuição teórica do critério ξ deve ser feito para que possamos explicar questões que no momento encontram-se em aberto.

As questões acima motivaram este trabalho que está dividido em três partes :

Na primeira, apresentamos a medida de Hellinger e suas propriedades. A segunda parte é uma revisão bibliográfica de artigos em quais a medida de Hellinger é utilizada com fins inferenciais, sem a pretensão de ser exaustiva, e a última parte é destinada à descrição e resultados da comparação entre matrizes de Dependências e Codependências.

Capítulo 2

A distância de Hellinger

2.1 Introdução

Quando se pretende estudar as discrepâncias entre duas populações (traduzidas por suas distribuições de probabilidade) utiliza-se muito frequentemente a estatística Qui - quadrado de Pearson. Tal estatística é um dos exemplos mais comuns de Distância Euclideana Ponderada. Para estudarmos algumas das medidas de discrepância existentes, introduziremos algumas definições e notações.

Seja $X = (X_1, \dots, X_k)^t$ uma variável aleatória com distribuição Multinomial de parâmetros n e $p = (p_1, \dots, p_k)^t$, onde $\sum_{i=1}^k p_i = 1$ e $p_i > 0$ para $1 \leq i \leq k$. A esperança matemática do vetor X é o vetor $E(X)$ com componentes $E(X_i) = np_i$, $i = 1, \dots, k$.

O vetor $p = (p_1, \dots, p_k)^t$ de probabilidades da distribuição da variável aleatória X representa uma dada população estatística Π . É natural portanto que sejam definidas discrepâncias entre duas populações Π_1 e Π_2 em termos de vetores p e q de probabilidades das distribuições de variáveis aleatórias X e Y .

Seja $O = (x_1, \dots, x_k)^t$ uma observação de X , com $\sum_{i=1}^k x_i = n$. Os estimadores de Máxima Verossimilhança de p_i são dados por $f_i = x_i/n$ (vide Agresti pgs. 40 e 41 [01]), resultando no vetor $f = (f_1, \dots, f_k)^t$.

A estatística Qui-quadrado de Pearson para testar a existência de discrepâncias entre \mathbf{f} e \mathbf{p} é dada por :

$$Q = n(\mathbf{f} - \mathbf{p})^t \mathbf{D}_{\mathbf{p}}^{-1} (\mathbf{f} - \mathbf{p}) = n \sum_1^k (f_i - p_i)^2 / p_i, \quad (2.1.1)$$

onde $\mathbf{D}_{\mathbf{p}}$ é a matriz diagonal do vetor \mathbf{p} .

A estatística Q pode ser re-escrita na forma simplificada :

$$Q = n \left\{ \sum_{i=1}^k (f_i^2 / p_i) - 1 \right\}, \quad (2.1.2)$$

Supondo-se que \mathbf{X} provenha da população Π_1 e \mathbf{O} da população Π_2 , e que a distribuição de \mathbf{X} tenha um vetor de probabilidades dado por \mathbf{p} e a de \mathbf{O} , dado por \mathbf{f} , a quantidade

$$\sum_{i=1}^k (f_i^2 / p_i) - 1, \quad (2.1.3)$$

pode ser considerada segundo Mathai e Rhatie [16] pg 96, uma medida de discrepância entre as populações discretas Π_1 e Π_2 . Tal quantidade mede a "separação" entre as distribuições de \mathbf{f} e \mathbf{p} , mesmo que (2.1.3) não seja distância no sentido matemático por não ser métrica, pois

$$\sum_{i=1}^k (f_i^2 / p_i) - 1 \neq \sum_{i=1}^k (p_i^2 / f_i) - 1.$$

Ao aceitar Q como uma quantidade que expressa a discrepância entre duas distribuições devemos observar que as discrepâncias em eventos com pequena probabilidade são responsáveis por uma excessiva contribuição. Isto revela que em geral a estatística Q não deva ser utilizada em problemas de ajuste, sem que se atente para esses detalhes.

2.2 Divergências, Distâncias e Informação Discriminatória

I - Divergências

Vários autores consideraram definições de distância, divergência, ou Informação discriminatória, entre populações Π_1 e Π_2 .

Cressie e Read [05] em 1984, e Aickin [02] em 1983, derivaram independentemente famílias de medidas de Divergências para testes de Ajuste.

Tais famílias englobam várias estatísticas de teste como a Qui-quadrado de Pearson, Qui-quadrado de Neyman, Qui-quadrado da Razão de Verossimilhança, Freeman-Tukey, entre outras. A família derivada pelos primeiros autores mencionados será tratada na secção 2.5.

II - Medidas de distância baseadas em idéias de teoria da Informação

Sejam as populações Π_1 e Π_2 e as variáveis aleatórias contínuas X_1 e X_2 em \mathbb{R}^n tais que $F_1(x)$ e $F_2(x)$ sejam funções distribuição de X_1 e X_2 respectivamente, em Π_1 e Π_2 , enquanto $p_1(x)$ e $p_2(x)$ são suas correspondentes funções densidade.

Vários autores propuseram definições de distância, ou informação discriminatória $\Delta(\Pi_1, \Pi_2)$ entre as populações Π_1 e Π_2 :

i) Kullback e Leibler [14] em 1951 apresentaram as medidas de informação discriminatória

$$\Delta_{Ka}(\Pi_1, \Pi_2) = \int p_1(x) \log[p_1(x)/p_2(x)] dx, \quad (2.2.1)$$

$$\Delta_{Kb}(\Pi_1, \Pi_2) = \int p_2(x) \log[p_2(x)/p_1(x)] dx. \quad (2.2.2)$$

Δ_{Ka} e Δ_{Kb} são sempre não-negativas, atingindo o valor zero quando e somente quando p_1 e p_2 forem a mesma distribuição e portanto quando Π_1 e Π_2 forem a mesma população.

A distância de Kullback-Leibler entre f e p é relacionada à estatística Qui-quadrado da Razão de Verossimilhança, G^2 , como segue :

sendo

$$\Delta_{Ka}(f, p) = \sum_{i=1}^k x_i/n \log(x_i/np_i),$$

então

$$\Delta_G^2(f, p) = n\Delta_{Ka}(f, p),$$

vide Bishop et al [03] pg 346.

ii) Foi observado por Kullback e Leibler em 1951 que a medida de Jeffreys [14] é obtida como uma combinação das medidas acima :

$$\Delta_J(\Pi_1, \Pi_2) = \int \{p_2(x) - p_1(x)\} \log[p_2(x)/p_1(x)] dx, \quad (2.2.3)$$

isto é :

$$\Delta_J(\Pi_1, \Pi_2) = \Delta_{Ka}(\Pi_1, \Pi_2) + \Delta_{Kb}(\Pi_1, \Pi_2) \quad (2.2.4)$$

A medida de Jeffreys descreve simetricamente a divergência entre as populações Π_1 e Π_2 .

III - Medidas de distância relacionadas à Afinidade de Bhattacharyya

Bhattacharyya propôs em 1943 a medida ρ de afinidade como segue

$$\rho = \int [p_1(x)p_2(x)]^{1/2} dx, \quad (2.2.5)$$

e em 1946, definiu $\Delta_B(\Pi_1, \Pi_2) = \cos^{-1}\rho$.

Matusita [17], em 1955, definiu

$$\Delta_M(\Pi_1, \Pi_2) = \left\{ \int [\sqrt{p_1(x)} - \sqrt{p_2(x)}]^2 dx \right\}^{1/2} = (2 - 2\rho)^{1/2} \quad (2.2.6)$$

de onde se vê que $0 \leq \rho \leq 1$.

Mas,

$$\left\{ \int [\sqrt{p_1(x)} - \sqrt{p_2(x)}]^2 dx \right\}^{1/2} \quad (2.2.7)$$

é a distância de Hellinger, vide Cordeiro [04].

Segundo Mathai e Rhatie [16] pg. 100, da expressão

$$\Delta(\pi_1, \pi_2) = \left\{ \int | [p_1(x)]^{1/s} - [p_2(x)]^{1/s} |^r dx \right\}^{1/r} \quad (2.2.8)$$

onde $r \geq 1$ e $s \geq 1$, derivam-se

se $r=s=2$, a distância de Hellinger;

se $r=s=1$, a distância variacional V de Kolmogorov :

$$\text{onde} \quad V = \int |p_1(x) - p_2(x)| dx. \quad (2.2.9)$$

2.3 Desigualdades

Segundo Mathai e Rhatie [16] pg. 112, Toussaint em 1972 provou que

$$\Delta_j(\pi_1, \pi_2) \geq 4(1 - \rho), \quad (2.3.1)$$

e Matusita [17] em 1955 mostrou que

$$2(1 - \rho) \leq V \leq 2\{2(1 - \rho)\}^{1/2}, \quad (2.3.2)$$

o mesmo autor em 1956, provou que

$$V \leq 2(1 - \rho^2)^{1/2} \leq 2\{2(1 - \rho)\}^{1/2}. \quad (2.3.3)$$

A medida de Hellinger (2.2.7) possui propriedades ótimas que passaremos a descrever. Além disso, daremos alguns exemplos de sua aplicação.

2.4 Afinidade e distância - Definições Básicas

Estudaremos a medida de Hellinger apenas no caso discreto pois o nosso maior interesse está no estudo de dados agrupados em tabelas de contingência e na investigação de medidas de Dependência e Codependência descritas no capítulo 3. Um estudo mais geral da distância de Hellinger pode ser encontrado em Dias [06].

Para estudarmos a distância entre distribuições discretas podemos nos valer da definição matemática de distância entre elementos de um conjunto.

Definição 2.4.1 Dado um conjunto $M \neq \emptyset$, seja $\delta : M \times M \longrightarrow \mathbb{R}^+$. A função δ é uma distância sobre M se as seguintes condições se verificam para quaisquer x, y e $z \in M$:

$$(M1) \quad \delta(x, y) = 0 \iff x = y$$

$$(M2) \quad \delta(x, y) = \delta(y, x)$$

$$(M3) \quad \delta(x, y) \leq \delta(x, z) + \delta(z, y).$$

A propriedade M3 é conhecida como desigualdade triangular.

Consideremos agora uma sequência de n execuções independentes de um experimento onde em cada uma das quais somente um de k eventos mutuamente exclusivos $\mathcal{E}_1, \dots, \mathcal{E}_k$ pode ser observado. A probabilidade de ocorrência do evento \mathcal{E}_j é p_j , as probabilidades não variam de execução para execução e ainda $p_j \geq 0$ com $\sum_j^k p_j = 1$ e $k \geq 2$.

A distribuição conjunta F , das variáveis aleatórias X_1, \dots, X_k representando o número de ocorrências dos eventos $\mathcal{E}_1, \dots, \mathcal{E}_k$ nas n execuções do experimento é Multinomial de parâmetros n e $\mathbf{p} = (p_1, \dots, p_k)^t$ onde $\sum_j^k X_j = n$.

Lema 2.4.1 Seja F_1 a distribuição conjunta discreta de probabilidades $p = (p_1, \dots, p_k)^t$ do vetor aleatório $X = (X_1, \dots, X_k)^t$. Seja Ω_k o conjunto das distribuições conjuntas discretas F_1 onde k determina o número de variáveis aleatórias de X .

Uma função $H : \Omega_k \times \Omega_k \rightarrow \mathbb{R}^+$ tal que

$$H^2 = \|F_1 - F_2\|^2 = \sum_{i=1}^k \left(\sqrt{p_{1i}} - \sqrt{p_{2i}} \right)^2 = 2 \left[1 - \sum_{i=1}^k \sqrt{p_{1i} p_{2i}} \right] = 2(1 - \rho), \quad (2.4.1)$$

é uma distância sobre Ω_k . Então Ω_k é um espaço métrico com respeito à distância $\|\cdot\|$.

Observação : A distância $\|\cdot\|$ é denominada distância de Hellinger e ρ é denominado afinidade entre F_1 e F_2 .

A prova do Lema 2.4.1 é dada a seguir :

A quantidade $H(F_1, F_2) = \|F_1 - F_2\|$ é distância sobre Ω_k se as características (M1), (M2) e (M3) forem válidas; analisaremos cada uma delas :

$$(M1): H(F_1, F_2) = 0 \iff F_1 = F_2$$

$$\text{Se } H(F_1, F_2) = 0, \text{ então } \sum_{i=1}^k \left(\sqrt{p_{1i}} - \sqrt{p_{2i}} \right)^2 = 0, \text{ por ser soma de}$$

números não negativos, então para todo i ,

$$\left(\sqrt{p_{1i}} - \sqrt{p_{2i}} \right)^2 = 0 \implies \sqrt{p_{1i}} = \sqrt{p_{2i}}, \text{ portanto } F_1 = F_2.$$

Reciprocamente, se $F_1 = F_2 \implies \sqrt{p_{11}} = \sqrt{p_{21}}$ e $Z_1 = \sqrt{p_{11}} - \sqrt{p_{21}} = 0$ para todo i , logo,

$$H(F_1, F_2) = \left[\sum_{i=1}^k Z_i^2 \right]^{1/2} = 0.$$

$$(M2): H(F_1, F_2) = H(F_2, F_1) :$$

$$H(F_1, F_2) = \left\{ \sum_{i=1}^k \left(\sqrt{p_{11}} - \sqrt{p_{21}} \right)^2 \right\}^{1/2} = \left\{ \sum_{i=1}^k \left(\sqrt{p_{21}} - \sqrt{p_{11}} \right)^2 \right\}^{1/2} = H(F_2, F_1).$$

(M3): Sejam F_1, F_2 e F_3 distribuições discretas $\in \Omega_k$, com as densidades respectivas $p_{11}, p_{12}, \dots, p_{1k}$ ($i = 1, 2, 3$),

$$\begin{aligned} H^2(F_1, F_3) &= \sum_{i=1}^k \left(\sqrt{p_{11}} - \sqrt{p_{31}} \right)^2 = \sum_{i=1}^k \left(\sqrt{p_{11}} - \sqrt{p_{21}} + \sqrt{p_{21}} - \sqrt{p_{31}} \right)^2 \\ &\leq \sum_{i=1}^k \left(\sqrt{p_{11}} - \sqrt{p_{21}} \right)^2 + 2 \sum_{i=1}^k \left| \sqrt{p_{11}} - \sqrt{p_{21}} \right| \left| \sqrt{p_{21}} - \sqrt{p_{31}} \right| + \sum_{i=1}^k \left(\sqrt{p_{21}} - \sqrt{p_{31}} \right)^2, \end{aligned}$$

De acordo com a desigualdade de Cauchy-Schwarz temos

$$\sum_{i=1}^k \left| \sqrt{p_{11}} - \sqrt{p_{21}} \right| \left| \sqrt{p_{21}} - \sqrt{p_{31}} \right| \leq \left\{ \sum_{i=1}^k \left(\sqrt{p_{11}} - \sqrt{p_{21}} \right)^2 \right\}^{1/2} \left\{ \sum_{i=1}^k \left(\sqrt{p_{21}} - \sqrt{p_{31}} \right)^2 \right\}^{1/2}$$

então

$$H^2(F_1, F_3) \leq \left\{ \left[\sum_{i=1}^k \left(\sqrt{p_{11}} - \sqrt{p_{21}} \right)^2 \right]^{1/2} + \left[\sum_{i=1}^k \left(\sqrt{p_{21}} - \sqrt{p_{31}} \right)^2 \right]^{1/2} \right\}^2,$$

logo

$$H^2(F_1, F_3) \leq \{H(F_1, F_2) + H(F_2, F_3)\}^2$$

de forma que

$$H(F_1, F_3) \leq H(F_1, F_2) + H(F_2, F_3) .$$

Como as propriedades (M1) a (M3) são satisfeitas $\| \cdot \|$ é distância sobre Ω_k .

2.4.2 Algumas das propriedades da distância $\| \cdot \|$ e afinidade ρ :

$$1) 0 \leq \rho \leq 1 \quad (2.4.2.1)$$

$$2) \|F_1 - F_2\|^2 = 2(1 - \rho(F_1, F_2)) \leq 2 \quad (2.4.2.2)$$

$$3) \|F_1 - F_2\|^2 \leq \sum_{i=1}^k |p_{1i} - p_{2i}| \leq 2\|F_1 - F_2\| \quad (2.4.2.3)$$

4) Se a sequência de distribuições $\{F_n\}$ em Ω_2 converge para a distribuição $F \in \Omega_2$ em termos da distância $\| \cdot \|$, então a sequência $\{F_n\}$ também converge para F uniformemente.

Provas : 1) $0 \leq \rho \leq 1$. Como

$$\|F_1 - F_2\|^2 = \sum_{i=1}^k (\sqrt{p_{1i}} - \sqrt{p_{2i}})^2 = \sum_{i=1}^k p_{1i} + \sum_{i=1}^k p_{2i} - 2 \sum_{i=1}^k \sqrt{p_{1i} p_{2i}} \geq 0$$

$$\|F_1 - F_2\|^2 = 2(1 - \rho(F_1, F_2)) \geq 0, \text{ tem-se } \rho \leq 1.$$

A afinidade é o produto interno entre as raízes quadradas das densidades de F_1 e F_2 isto é, $\rho = \langle \sqrt{p_1}, \sqrt{p_2} \rangle$ é não negativa pois $p_{1i} \geq 0$ e $p_{2i} \geq 0$, $i = 1, 2, \dots, k$, e portanto $0 \leq \rho \leq 1$.

Tem-se que $\cos \theta = \langle \sqrt{p_1}, \sqrt{p_2} \rangle$, pois $\|\sqrt{p_1}\| = \|\sqrt{p_2}\| = 1$, portanto, ρ iguala o cosseno do ângulo θ entre os vetores $\sqrt{p_1} = (\sqrt{p_{11}}, \dots, \sqrt{p_{1k}})^t$ e $\sqrt{p_2} = (\sqrt{p_{21}}, \dots, \sqrt{p_{2k}})^t$.

A interpretação da afinidade é que à medida que esta se aproxima da unidade $\sqrt{p_1}$ e $\sqrt{p_2}$ ficam tanto mais próximas. Um valor de $\rho=1$ ocorre quando $\sqrt{p_1}$ e $\sqrt{p_2}$ forem coincidentes. $\rho = 0$ ocorre quando F_1 e F_2 têm suportes disjuntos.

$$2) \left\| \sqrt{F_1} - \sqrt{F_2} \right\|^2 = 2(1 - \rho(F_1, F_2)) \leq 2, \text{ é trivial decorrência de 1)}$$

$$3) \left\| \sqrt{F_1} - \sqrt{F_2} \right\|^2 \leq \sum_{i=1}^k |p_{1i} - p_{2i}| \leq 2 \left\| \sqrt{F_1} - \sqrt{F_2} \right\|$$

Para provar que $\left\| \sqrt{F_1} - \sqrt{F_2} \right\|^2 \leq \sum_{i=1}^k |p_{1i} - p_{2i}|$, verifica-se que

$$|\sqrt{p_{11}} - \sqrt{p_{21}}| \leq |\sqrt{p_{11}} + \sqrt{p_{21}}|,$$

dai

$$\left(\sqrt{p_{11}} - \sqrt{p_{21}} \right)^2 \leq |\sqrt{p_{11}} - \sqrt{p_{21}}| |\sqrt{p_{11}} + \sqrt{p_{21}}| = |p_{11} - p_{21}|,$$

portanto

$$\left\| \sqrt{F_1} - \sqrt{F_2} \right\|^2 = \sum_{i=1}^k \left(\sqrt{p_{1i}} - \sqrt{p_{2i}} \right)^2 \leq \sum_{i=1}^k |p_{1i} - p_{2i}|.$$

Para provar a segunda desigualdade de 4) temos que

$$\sum_{i=1}^k |p_{1i} - p_{2i}| = \sum_{i=1}^k |\sqrt{p_{1i}} - \sqrt{p_{2i}}| |\sqrt{p_{1i}} + \sqrt{p_{2i}}|,$$

e por desigualdade de Cauchy-Schwarz temos

$$\sum_{i=1}^k |\sqrt{p_{1i}} - \sqrt{p_{2i}}| |\sqrt{p_{1i}} + \sqrt{p_{2i}}| \leq \left[\sum_{i=1}^k (\sqrt{p_{1i}} - \sqrt{p_{2i}})^2 \right]^{1/2} \left[\sum_{i=1}^k (\sqrt{p_{1i}} + \sqrt{p_{2i}})^2 \right]^{1/2}$$

por definição de $\| \cdot \|$ e (1) fica

$$\begin{aligned} \left[\sum_{i=1}^k (\sqrt{p_{1i}} - \sqrt{p_{2i}})^2 \right]^{1/2} \left[\sum_{i=1}^k (\sqrt{p_{1i}} + \sqrt{p_{2i}})^2 \right]^{1/2} &= \|F_1 - F_2\| \left[2 + 2\rho(F_1, F_2) \right]^{1/2} \\ &\leq \|F_1 - F_2\| \left[2 + 2 \right]^{1/2} \end{aligned}$$

portanto

$$\sum_{i=1}^k |p_{1i} - p_{2i}| \leq 2 \|F_1 - F_2\|.$$

- 4) Sejam X_n e X variáveis aleatórias com distribuição Binomial de parâmetros n e p_n e n e p respectivamente, ou seja as distribuições F_n e F de X_n e X respectivamente pertencem a Ω_2 . Sejam q_{n1}, \dots, q_{nk} e q_1, \dots, q_k as probabilidades de X_n e X assumirem os valores $x_1 < x_2 < \dots < x_k$ então

$$F_n(x) = \sum_{x_1 \leq x} q_{ni} \quad \text{e} \quad F(x) = \sum_{x_1 \leq x} q_i$$

Sendo

$$|F_n(x) - F(x)| = \left| \sum_{x_i \leq x} q_{ni} - \sum_{x_i \leq x} q_i \right| \leq \sum_{x_i \leq x} |q_{ni} - q_i| \leq 2 \|F_n - F\|,$$

de acordo com a propriedade 3), e notando que $F_n - F$ independe de x , então $F_n - F \rightarrow 0$ para $n \rightarrow \infty$ implica em $|F_n(x) - F(x)| \rightarrow 0$ quando $n \rightarrow \infty$ uniformemente em x .

2.4.1 Relacionamento entre a distância de Hellinger e a distância de Fréchet

A distância $\Delta_F(F_1, F_2)$ de Fréchet entre as funções distribuição F_1 e F_2 em Ω_2 é dada por

$$\Delta_F(F_1, F_2) = \sup_x |F_1(x) - F_2(x)| \text{ onde } -\infty < x < \infty.$$

Observemos que

$$|F_1(E) - F_2(E)| \leq \sup_{x \in E} |F_1(x) - F_2(x)| ; E \subseteq \mathbb{R}$$

o lado direito desta expressão sendo a menor das cotas superiores de $|F_1(E) - F_2(E)|$ é majorado por $2\|F_1 - F_2\|$ (vide propriedade 4 secção 2.4) isto é

$$\sup_{x \in E} |F_1(x) - F_2(x)| \leq 2\|F_1 - F_2\| \quad (2.4.1.1)$$

Kolmogorov em 1933, foi o primeiro a propor a distância de Fréchet como estatística de teste para Bom Ajuste de S_n a F (vide Kendall e Stuart cap. 30) usando

$$D_n = \sup_x |F_n(x) - F(x)|, \quad -\infty < x < \infty,$$

isto posto que

$$\sup_x |F_n(x) - F(x)| \xrightarrow{P} 0 ,$$

ou seja, para grandes amostras F_n é muito próxima de F . Nesta ocasião Kolmogorov obteve a distribuição assintótica de D_n pressupondo F contínua.

Quanto ao cálculo do poder de D_n , são vários os trabalhos existentes no caso de F contínua. Um *survey* a respeito é encontrado em Kendall e Stuart cap. 30. Segundo Gleser (1985) [13] o caso do poder de D_n para F discreto foi tratado por poucos, entre estes o próprio Gleser e Conover em 1972. Tanto Gleser quanto Conover avaliaram o poder exato de D_n para F discreto apenas para distribuições univariadas.

É de interesse investigar o poder exato de D_n e compará-lo ao de alguma estatística baseada em $\| \cdot \|$, como por exemplo a estatística M proposta por Khan e Ali [21] (vide secção 3.7) para o caso de F ter distribuição Multinomial e n finito.

As limitações existentes para tal estudo são em primeiro lugar devidas ao fato do poder de D_n ter sido avaliado somente para distribuições univariadas. Isto nos restringiria a estudar o caso de variáveis aleatórias com distribuição Binomial. Este caso, embora tenha sido o que expressamos em (2.4.1.1), talvez não seja muito informativo quanto à questão colocada.

Em segundo lugar, testes como o de Khan e Ali e assemelhados (vide secção 3.7) foram desenvolvidos apenas para o caso assintótico, mas para n finito, situação em que temos real interesse, desconhecemos a existência de alguma estatística envolvendo $\| \cdot \|$ em que possamos obter o poder exato do teste.

Todas estas questões são complexas e de real interesse, exigindo um estudo aprofundado que não levaremos adiante para não nos desviarmos excessivamente do escopo de nosso trabalho.

2.5 A distância de Hellinger como membro de uma família de estatísticas de Divergência em Potência

A distância de Hellinger pode ser enquadrada em uma família de estatísticas de divergências em potências apresentada por Cressie e Read [05], para testes de Bom Ajuste, e está diretamente relacionada à estatística de Freeman - Tukey, que é membro desta família.

De acordo com propriedades que esta família possui, pode ser mostrado que a distância de Hellinger entre dada distribuição F em Ω_k (vide Lema 2.4.1) e a distribuição empírica de uma variável aleatória com distribuição Multinomial com base em uma amostra de tamanho n e distribuição F , possui distribuição assintótica χ^2_{k-1} , como veremos adiante.

Definição 2.5.1 Sejam $X = (X_1, \dots, X_k)^t$ uma variável aleatória com densidade Multinomial de parâmetros n e $p = (p_1, \dots, p_k)^t$ com $\sum_{i=1}^k p_i = 1$ e o vetor de frequências esperadas de X , $E(X) = (e_1, \dots, e_k)^t$, com $e_i = np_i$. O conjunto $\{I^\lambda; \lambda \in \mathbb{R}\}$ é a família de estatísticas de divergência em potência para testes de Bondade do Ajuste e $2nI^\lambda$ é expresso como

$$2nI^\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^k X_i \left\{ \left(X_i/e_i \right)^\lambda - 1 \right\}; \quad \lambda \in \mathbb{R}. \quad (2.5.1).$$

Para $\lambda = 0, -1$, $2nI^\lambda$ é definida por continuidade. No caso de $\lambda=0$, tomando $a_1 = X_1/e_1$ e aplicando -se a regra de L'Hospital, temos que

$$2nI^0 = \lim_{\lambda \rightarrow 0} \frac{2 \sum_{i=1}^k X_i \ln a_i \exp(\lambda \ln a_i)}{2\lambda + 1} = 2 \sum_{i=1}^k X_i \ln a_i,$$

portanto

$$2nI^0 = 2 \sum_{i=1}^k X_i \ln(X_i/np_i),$$

e no caso de $\lambda = -1$,

$$2nI^{-1} = \lim_{\lambda \rightarrow -1} \frac{2 \sum_{i=1}^k X_i \ln a_i \exp(\lambda \ln a_i)}{2\lambda + 1} = -2 \sum_{i=1}^k X_i/a_i \ln a_i,$$

logo,

$$2nI^{-1} = 2 \sum_{i=1}^k np_i \ln(X_i/np_i),$$

isto é, para $\lambda = 0$, tem-se a estatística Logarítmo da Razão de Verossimilhança, e para $\lambda = -1$, tem-se a estatística Logarítmo da Razão de Verossimilhança Modificado. Em especial, quando $\lambda = 1$, a estatística χ^2 é obtida.

Seja $\mathbf{p} = (p_1, p_2, \dots, p_k)^t$ o vetor de probabilidades de F em Ω_k , e S_n a distribuição empírica de F com base em uma amostra de tamanho n. A estatística de Freeman - Tukey (que corresponde a $\lambda = -1/2$), para testar o ajuste entre F e S_n é a que segue :

$$T^2 = 4 \sum_{i=1}^k \{ \sqrt{X_i} - \sqrt{np_i} \}^2 \quad (2.5.2)$$

Em termos da distância de Hellinger, H , a distância entre F e S_n é dada por :

$$\|F - S_n\| = \left[\sum_{i=1}^k \{ \sqrt{X_i/n} - \sqrt{p_i} \}^2 \right]^{1/2} = H.$$

Verifica - se que

$$H^2 = T^2/4n \quad (2.5.3)$$

2.6 Distribuição Qui - Quadrado como limite

Nesta secção mostraremos que as estatísticas membros de $\{I^\lambda, \lambda \in \mathbb{R}\}$, possuem distribuição assintótica Qui-quadrado com $k - 1$ graus de liberdade, quando S_n vem de F .

Sejam F em Ω_k , k o número de eventos possíveis em F , e a hipótese,

$$H_0 : p = p_0,$$

onde as probabilidades $\{p_{0i}\}$ são completamente especificadas.

Para $\lambda \neq 0, -1$, seja

$$2nI^\lambda(X/n; p_0) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^k p_{0i} \left\{ \left(1 + \frac{X_i - np_{0i}}{np_{0i}} \right)^{\lambda+1} - 1 \right\} \quad (2.6.1)$$

Seja

$$V_i = (X_i - np_{0i})/np_{0i}, \quad (2.6.2)$$

substituindo - se V_i em (2.6.1), temos para $\lambda \neq 0, -1$,

$$2nI^{\lambda}(X/n;p_0) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^k p_{0i} \left\{ (1 + V_i)^{\lambda+1} - 1 \right\} \quad (2.6.3);$$

expandindo $f(V_i) = (1 + V_i)^{\lambda+1}$ em série de McLaurin, temos :

$$f(V_i) = 1 + (\lambda + 1)V_i + \frac{\lambda(\lambda + 1)V_i^2}{2!} + R(V_i) \quad (2.6.4),$$

onde

$$R(V_i) = \frac{f^{(3)}(c)V_i^3}{3!}, \quad \text{com } 0 \leq c \leq |V_i| \quad (2.6.5),$$

Tome $|V_i| \leq \epsilon$, com $\epsilon > 0$, e tenha

$$|R(V_i)/V_i^3| \leq |(\lambda - 1)\lambda(\lambda + 1)(1 + \epsilon)^{\lambda-2}/3| = c_1.$$

Como V_i é variável aleatória, temos que

$$R(V_i) = O_p(V_i^3),$$

(vide Leite e Singer [15], pg 39), pois sendo c_1 uma variável aleatória degenerada,

$$\Pr\left\{|c_1| \geq k\right\} = \begin{cases} 0, & k > |c_1| \\ 1, & k \leq |c_1|, \end{cases}$$

ou seja, se $k > |c_1|$, $R(V_i) = o_p(V_i^3)$ e se $k \leq |c_1|$, $R(V_i) = O_p(V_i^3)$,

então, $R(V_i) = O_p(V_i^3)$.

Obtemos portanto :

$$2nI^\lambda(X/n;p_0) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{\infty} p_{0i} \left\{ (\lambda+1)V_i + \frac{\lambda(\lambda+1)V_i^2}{2!} + O_p(V_i^3) \right\},$$

Sob H_0 , X_i tem distribuição Binomial de parâmetros n e p_{0i} , e de acordo com a desigualdade de Chebyshev,

$$\begin{aligned} \Pr\left\{|R(V_i)| \geq k\right\} &= \Pr\left\{c_2|(X_i - np_{0i})^3/(np_{0i})^3| \geq k\right\} \\ &= \Pr\left\{|(X_i - np_{0i})| \geq np_{0i}k_1\right\} \leq \frac{(1 - p_{0i})}{n p_{0i} k_1^2}, \end{aligned}$$

com $k_1^3 = k/c_1$. Logo

$$\Pr\left\{|R(V_i)| \geq k\right\},$$

converge para zero quando n cresce e portanto, $R(V_i) = o_p(1)$,

donde se obtém, sob o modelo H_0 ,

$$2nI^\lambda(X/n;p_0) = 2nI^1(X/n;p_0) + o_p(1); \lambda \in \mathbb{R} \quad (2.6.6).$$

Portanto, toda estatística de divergência em potência para teste de Bom Ajuste, possui assintoticamente a mesma distribuição da estatística Qui - Quadrado de Pearson sob a hipótese

$$H_0 : p = p_0,$$

ou seja, distribuição χ_{k-1}^2 (vide Rao [20] pg 325).

Inferência baseada na distância de Hellinger

3.1 Teoremas fundamentais

Utilizando a desigualdade de Markov, Matusita [17] descreveu em 1955, regiões de confiança baseadas na distância de Hellinger para distribuições em Ω_k .

Descreveremos alguns dos teoremas de autoria de Matusita e, posteriormente, estudos desenvolvidos por outros autores baseados nas idéias de Matusita.

Seja F uma distribuição em Ω_k com vetor de probabilidades $(p_1, p_2, \dots, p_k)^t$ e S_n a distribuição empírica de uma amostra aleatória X_1, X_2, \dots, X_n de X , onde X possui distribuição F . Seja n_1 o número de vezes em quais \mathcal{E}_1 foi observado, onde \mathcal{E}_1 é um evento com $p_1 = p(\mathcal{E}_1)$. A distribuição S_n tem vetor de probabilidades estimadas $(n_1/n, n_2/n, \dots, n_k/n)^t$.

Teorema I

Quando dada variável aleatória possui distribuição F , tem-se,

$$\Pr\left\{\|F - S_n\|^2 \leq \frac{k-1}{n} t\right\} \geq 1 - 1/t, \quad (3.1.1)$$

para todo número $t \geq 1$.

Prova - Tomemos $p_1, p_2, \dots, p_k > 0$ então

$$H^2(F, S_n) = \|F - S_n\|^2 = \sum_{i=1}^k \left(\sqrt{n_i/n} - \sqrt{p_i} \right)^2;$$

sendo

$$\left(n_i/n - p_i \right) = \left(\sqrt{n_i/n} - \sqrt{p_i} \right) \left(\sqrt{n_i/n} + \sqrt{p_i} \right)$$

temos

$$\sum_{i=1}^k \left(\sqrt{n_i/n} - \sqrt{p_i} \right)^2 = \sum_{i=1}^k \left\{ n_i/n - p_i \right\}^2 / \left\{ \sqrt{n_i/n} + \sqrt{p_i} \right\}^2$$

como

$$\left\{ \sqrt{n_i/n} + \sqrt{p_i} \right\}^2 \geq p_i,$$

temos

$$\sum_{i=1}^k \left(\sqrt{n_i/n} - \sqrt{p_i} \right)^2 \leq \sum_{i=1}^k \left\{ n_i/n - p_i \right\}^2 / p_i$$

portanto

$$E [H^2] \leq \sum_{i=1}^k \frac{1}{p_i} E \left\{ n_i/n - p_i \right\}^2 = \sum_{i=1}^k \frac{1}{p_i} \text{var}(n_i/n) = \sum_{i=1}^k \frac{(1 - p_i)}{n} = \frac{k - 1}{n},$$

logo,

$$E [H^2] \leq \frac{k - 1}{n}.$$

Por desigualdade de Markov temos

$$\Pr \left\{ \|F - S_n\|^2 \leq \frac{k - 1}{n} t \right\} \geq 1 - \frac{n}{t(k-1)} E \left\{ \|F - S_n\|^2 \right\} \geq 1 - 1/t$$

A expressão (3.1.1) permite construir uma região de confiança para F com probabilidade $1 - \alpha$ de F estar dentro da bola $B(S_n, \frac{k-1}{n} t)$.

Teorema II

Quando dada variável aleatória possuir distribuição F em Ω_k com $p_i > 0$ para todo $i=1, \dots, k$, e n for grande, tem-se a igualdade assintótica :

$$\Pr\{\|F - S_n\|^2 \leq \delta^2\} \approx \Pr\{\chi_{(k-1)}^2 \leq 4n\delta^2\} \quad (3.1.2)$$

Prova :

Sendo $\|F - S_n\|^2 = T^2/4n$ (vide 1.3.3), de acordo com Cressie e Read [05],

$$T^2 \longrightarrow \chi_{(k-1)}^2 \quad \text{quando } n \longrightarrow \infty,$$

portanto, para n suficientemente grande,

$$\Pr\{\|F - S_n\|^2 \leq \delta^2\} = \Pr\{T^2 \leq 4n\delta^2\} \approx \Pr\{\chi_{k-1}^2 \leq 4n\delta^2\}.$$

Definição 3.1.1 Denote-se por $C_{n,k-1}$ a classe de distribuições $F = \{(p_1, p_2, \dots, p_k)\}$ em Ω_k para a qual existe um número positivo p_0 , tal que $p_i > p_0$ para todo i , e a estatística Q (Qui-quadrado de Pearson) calculada com base em n observações de uma variável aleatória, tenha distribuição assintótica χ_{k-1}^2 , para toda distribuição em $C_{n,k-1}$.

Segundo Bishop, Fienberg e Holland [03] pgs. 31 e 177, a incidência de celas vazias em Tabelas de Contingência pode ser devida a um resultado amostral, denominado "Zero Amostral", ou ainda, devida ao fato de o evento que descreve a cela seja impossível, determinando o denominado "Zero Estrutural".

O "Zero Amostral" ocorre quando a probabilidade associada a determinada casela for muito pequena, no entanto, tende a desaparecer com o crescimento do tamanho da amostra. O "Zero Estrutural", por sua vez, independe do tamanho da amostra, e é conhecido à priori.

O teorema II é válido para o caso de distribuições pertencentes à classe $C_{n,k-1}$. Mais precisamente, o teorema II se aplica a distribuições onde o "Zero Estrutural" não ocorre. O teorema I é mais geral, e pode ser utilizado em qualquer caso. No teorema II, em casos onde as pressuposições do mesmo são atendidas, é possível que se obtenha maior precisão na estimação da probabilidade de $\|F - S_n\| \leq \delta$, em comparação ao teorema I.

3.2 Testes de hipóteses e de Aderências

Utilizando os teoremas I e II, Matusita desenvolveu regras de decisão de acordo com as quais para dada distribuição discreta F_0 , e para todo número finito $\delta_0 > 0$, seja possível decidir se uma variável aleatória em estudo possui distribuição F_0 ou distribuição F , fora da vizinhança de F_0 , isto é, $\|F - F_0\| \geq \delta_0$. Este problema foi abordado por Matusita [17] em 1955.

Teorema III

Para todo número positivo $t \geq 1$, se X possuir distribuição F_0 , tem-se :

$$\Pr\left\{\|F_0 - S_n\|^2 \leq \frac{k-1}{n} t\right\} \geq 1 - 1/t \quad (3.2.1)$$

e se X possuir distribuição F com $\|F - F_0\| \geq \delta_0$, e se $n \geq 4(k-1)t/\delta_0^2$, então

$$\Pr\left\{\|F_0 - S_n\|^2 \geq \frac{k-1}{n} t\right\} \geq 1 - 1/t \quad (3.2.2)$$

Prova :

A primeira parte do Teorema já foi provada no Teorema I. Esta primeira desigualdade independe de n .

Quando X possuir distribuição F com $\|F - F_0\| \geq \delta_0$, temos, pela desigualdade triangular :

$$\|F_0 - S_n\| \geq \|F_0 - F\| - \|F - S_n\| \geq \delta_0 - \|F - S_n\|.$$

Quando $n \geq 4(k-1)t/\delta_0^2$,

$$\Pr\left\{\|F_0 - S_n\| \geq \delta_0 - \sqrt{\frac{k-1}{n} t}\right\} \geq \Pr\left\{\|F - S_n\| \leq \sqrt{\frac{k-1}{n} t}\right\} \geq 1 - 1/t,$$

de acordo com o Teorema I e a substituição de n em $\delta_0 - \sqrt{\frac{k-1}{n} t}$, na desigualdade acima, resulta em :

$$\delta_0 - \sqrt{\frac{k-1}{n} t} \geq \delta_0/2 \quad \text{ou} \quad \delta_0/2 \geq \sqrt{\frac{k-1}{n} t}$$

e, portanto,

$$\begin{aligned} \Pr\left\{\left\|F_0 - S_n\right\|^2 \geq \frac{k-1}{n} t\right\} &\geq \Pr\left\{\left\|F_0 - S_n\right\|^2 \geq \delta_0/2\right\} \\ &\geq \Pr\left\{\left\|F_0 - S_n\right\|^2 \geq \delta_0 - \sqrt{\frac{k-1}{n} t}\right\}. \end{aligned}$$

Quando se tem conhecimento de que F_0 , e toda distribuição alternativa F em Ω_k com densidade (p_1, p_2, \dots, p_k) estão contidas na classe $C_{n,k-1}$ (vide definição 3.1.1), aplica-se o Teorema II na demonstração do Teorema III como segue :

Teorema IV

Sejam F_0 e F distribuições discretas em Ω_k e contidas em $C_{n,k-1}$ (vide definição 3.1.1), com $\|F - F_0\| \geq \delta_0$. Seja η qualquer número positivo menor do que δ_0 .

Quando X possuir distribuição F_0 ,

$$\Pr\left\{\left\|F_0 - S_n\right\|^2 < \eta^2\right\} \approx \Pr\left\{\chi_{(k-1)}^2 < 4n\eta^2\right\} \quad (3.2.3),$$

e quando X possuir distribuição F ,

$$\Pr\left\{\left\|F_0 - S_n\right\|^2 \geq \eta^2\right\} \geq \Pr\left\{\chi_{(k-1)}^2 < 4n(\delta_0 - \eta)^2\right\} \quad (3.2.4).$$

Prova : Em (3.2.3) a prova é dada pelo Teorema II. Quando X possuir distribuição F , sendo $\|F_0 - S_n\| \geq \delta_0 - \|F - S_n\|$,

$$\Pr\left\{\left\|F_0 - S_n\right\|^2 \geq \eta^2\right\} \geq \Pr\left\{\left\|F - S_n\right\|^2 < (\delta_0 - \eta)^2\right\},$$

como

$$\|F - S_n\|^2 = T^2/4n \quad (\text{vide 2.3.3}), \text{ então}$$

$$\Pr\{\|F_0 - S_n\|^2 \geq \eta^2\} \geq \Pr\{T^2 \geq 4n(\delta_0 - \eta)^2\},$$

para n suficientemente grande,

$$\Pr\{\|F_0 - S_n\|^2 \geq \eta^2\} \approx \Pr\{\chi_{(k-1)}^2 \geq 4n(\delta_0 - \eta)^2\}.$$

Seja P uma propriedade referente aos elementos de Ω_k . Por exemplo, para $k=4$, seja F a distribuição conjunta com vetor de probabilidades $p = (p_{11}, p_{12}, p_{21}, p_{22})$ do vetor aleatório $X = (X_{11}, X_{12}, X_{21}, X_{22})$. No caso de $p_{ij} = p_{i.} p_{.j}$ para todo i e j , a propriedade P é a de F ter vetor de probabilidades $p = (p_{1.} p_{.1}, p_{1.} p_{.2}, p_{2.} p_{.1}, p_{2.} p_{.2})$. Denotemos por W , o conjunto de distribuições pertencentes a Ω_k que possuem a propriedade P , isto é,

$$W = \{F \in \Omega_k : p_{ij} = p_{i.} p_{.j}\}.$$

Teorema V

Seja F_0 uma distribuição em Ω_k e S_n a distribuição empírica de uma amostra aleatória X_1, X_2, \dots, X_n de X , onde X possui distribuição F_0 . Para todo conjunto de distribuições W e para $\eta > 0$, tem-se :

$$\Pr\left\{\inf_{F \in W} \|F - S_n\| \leq \eta\right\} \geq 1 - \frac{k-1}{n\eta^2}, \text{ quando } F_0 \in W \quad (3.2.5)$$

e

$$\Pr\left\{\inf_{F \in W} \|F - S_n\| > \eta\right\} \geq 1 - \frac{k-1}{n(\epsilon-\eta)^2}, \text{ quando } \inf_{F \in W} \|F_0 - F\| \geq \epsilon > \eta \quad (3.2.6)$$

Prova : Quando $F_0 \in W$, $\inf_{F \in W} \|F - S_n\| \leq \|F_0 - S_n\|$, então

$$\Pr\left\{ \inf_{F \in W} \|F - S_n\| < \eta \right\} \geq \Pr\left\{ \|F_0 - S_n\| < \eta \right\},$$

e pelo teorema III,

$$\Pr\left\{ \|F_0 - S_n\| < \eta \right\} \geq 1 - \frac{k-1}{n\eta^2},$$

tendo-se,

$$\Pr\left\{ \inf_{F \in W} \|F - S_n\| < \eta \right\} \geq 1 - \frac{k-1}{n\eta^2}.$$

Para demonstrar a segunda parte do teorema usaremos a desigualdade

$$\inf_{F \in W} \|F - F_0\| \leq \inf_{F \in W} \|F - S_n\| + \|F_0 - S_n\|,$$

quando $\inf_{F \in W} \|F - F_0\| \geq \varepsilon > \eta$, tem-se também $\inf_{F \in W} \|F - S_n\| \geq \varepsilon - \|F_0 - S_n\|$,

dai vem que

$$\Pr\left\{ \inf_{F \in W} \|F - S_n\| > \eta \right\} \geq \Pr\left\{ \|F_0 - S_n\| < \varepsilon - \eta \right\} \geq 1 - \frac{k-1}{n(\varepsilon-\eta)^2}.$$

O Teorema V enunciado por Matusita e Akaike [18], é uma generalização do Teorema I. Este último teorema possibilita testarmos a um nível de confiança de $1 - 1/t$, se a distribuição empírica S_n da variável aleatória $X = (X_1, \dots, X_k)^t$, onde X possui distribuição Multinomial em geral desconhecida, está a determinada distância de certa distribuição F , ou ainda, se X possui distribuição F ; ao passo

que o teorema V possibilita testarmos a um nível de confiança de $1 - \frac{k-1}{n\eta^2}$, se S_n está a determinada distância de uma classe W de distribuições, em outras palavras, se a distribuição F_0 de X pertence ou não à classe W de distribuições em questão.

3.3 Regras de Decisão baseadas no Teorema V

Para testarmos as hipóteses

$$\begin{aligned} H_0 &: F_0 \in W \\ H_a &: \inf_{F \in W} \|F - F_0\| \geq \varepsilon > \eta, \end{aligned} \quad (3.2.7)$$

utilizando as desigualdades do teorema V, o risco envolvido em (3.2.5) é :

$$\Pr\{\text{rejeitar } H_0 \text{ quando } H_0 \text{ é verdadeira}\} = \alpha \leq \frac{k-1}{n\eta^2},$$

e o risco envolvido em (3.2.6) é :

$$\Pr\{\text{não rejeitar } H_0 \text{ quando } H_a \text{ é verdadeira}\} = \beta \leq \frac{k-1}{n(\varepsilon-\eta)^2},$$

deste modo, o poder do teste é maior ou igual a $1 - \beta$. Observa-se que o poder aumenta quando n cresce.

Para que determinemos uma região crítica C , para o teste de hipóteses (3.2.7) fixemos n , k e α e de acordo com (3.2.6) obteremos

$$\eta = \sqrt{\frac{k-1}{n\alpha}} \quad (3.2.8)$$

A região crítica C de nível α do teste será :

$$C : \inf_{F \in W} \|F - S_n\| > \eta,$$

isto é, rejeitaremos H_0 com nível de confiança α se $\inf_{F \in W} \|F - S_n\| > \eta$.

Pode-se ainda determinar o nível crítico do teste para um dado valor amostral d :

$$\Pr\left\{ \inf_{F \in W} \|F - S_n\| \leq d \right\} \geq 1 - \frac{k-1}{nd^2},$$

a hipótese H_0 não deve ser rejeitada para grandes valores de $\frac{k-1}{nd^2}$.

No caso de H_0 ser rejeitada, decide-se que $\inf_{F \in W} \|F - S_n\| \geq \epsilon > d$ com um poder maior ou igual a $1 - \frac{k-1}{n(\epsilon-d)^2}$.

Resumidamente, o teorema V permite a construção da seguinte regra de decisão :

$$\text{Quando } \inf_{F \in W} \|F - S_n\| \leq \eta, \text{ decide-se que } F_0 \in W; \quad (3.2.9)$$

$$\text{Quando } \inf_{F \in W} \|F - S_n\| > \eta, \text{ decide-se que } \inf_{F \in W} \|F - F_0\| \geq \epsilon > \eta. \quad (3.2.10)$$

Sendo W um conjunto de distribuições com dada característica, se (3.2.9) se verificar, decide-se que F_0 é membro de W ao nível de $1 - \frac{k-1}{n\eta^2}$ de confiança. Em caso contrário, decide-se que F_0 não é membro de W com um poder de $1 - \frac{k-1}{n(\epsilon-d)^2}$. No caso do conjunto W

consistir de uma única distribuição, a regra de decisão é aplicável, e o problema se reduz ao teste de Bom Ajuste, que foi abordado no início desta secção.

A regra de decisão acima, reescrita em termos da afinidade, é a que segue :

Quando

$$\sup_{F \in W} \rho(F, S_n) \geq 1 - \eta^2/2, \quad \text{decide-se que } F_0 \in W ;$$

Quando

$$\sup_{F \in W} \rho(F, S_n) < 1 - \eta^2/2 \quad \text{decide-se que } \inf_{F \in W} \|F - F_0\| \geq \varepsilon > \eta.$$

3.3 Tamanhos de amostras

Os limites inferiores das probabilidades em (3.2.5) e (3.2.6), permitem a determinação do tamanho da amostra necessária para a obtenção de um teste de tamanho α . De acordo com (3.2.5), seja

$$1 - \frac{k-1}{n\eta^2} = 1 - \alpha \geq 0, \text{ com } \alpha \text{ fixo, então}$$

$$n > \frac{k-1}{\alpha\eta^2}, \quad (3.3.1)$$

desta maneira, para que se tenha $100(1 - \alpha)\%$ de certeza de que $\inf_{F \in W} \|F - S_n\|$ não exceda η , o tamanho da amostra requerida é dado por (3.3.1).

$$\text{Como } \|\cdot\|^2 \leq 2 \text{ então } \inf_{F \in W} \|F - F_0\| \leq \eta \leq \sqrt{2}, \text{ logo } n > \frac{k-1}{2\alpha},$$

então

$$n > \frac{k-1}{\alpha \eta^2} \geq \frac{k-1}{2\alpha \eta}$$

tais relações implicam em grandes valores de n e à medida que η diminui, n eleva-se dramaticamente (vide tabela I - secção 3.5).

3.4 Aplicação do Teorema V - Teste de Independência

Seja $W = \{F \in \Omega_k : p_{ij} = p_{i.} p_{.j}\}$, onde $W \subseteq [0,1]$. Temos que $W \subset \Omega_k$ é compacto :

representemos F por p ; sejam $p^{(1)}, p^{(2)}, \dots, p^{(n)} \in W$. $\lim_n p^{(n)} = p$ se e só se para todo $\epsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que para $n \geq n_0$ implica

$$H^2(p^{(n)}, p) < \epsilon^2 \text{ se e só se } \sum_{i,j} \left(\sqrt{p_{i.}^{(n)} p_{.j}^{(n)}} - \sqrt{p_{ij}} \right)^2 < \epsilon^2 \text{ implicando em}$$

$$\lim_n p_{i.}^{(n)} p_{.j}^{(n)} = p_{ij} = \lim_n p_{i.}^{(n)} \lim_n p_{.j}^{(n)} = p_{i.}^* p_{.j}^* \in W.$$

Como todo subconjunto compacto W do espaço métrico Ω_k é fechado, temos que o ínfimo e o supremo de W pertencem a W .

Temos ainda que para todo $\epsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que para $n \geq n_0$ implica $H^2(p^{(n)}, p) < \epsilon^2$ se e só se $\rho(F_j, S_n) > 1 - 2\epsilon^2$, então $\lim_n \rho(F, S_n) = 1$ onde $\rho(F_j, S_n) \in [0,1] \subset \mathbb{R}$ para todo n e todo $F_j \in W$.

Da mesma forma, $H^2(p^{(n)}, p) > \epsilon^2$ se e só se $\rho(F_j, S_n) < 1 - 2\epsilon^2$ implica em $\lim_n \rho(F, S_n) = 0$.

Como o limite de toda sequência $Y_n = \{y_1, y_2, \dots\}$ onde $y_j = \rho(F_j, S_n)$ com n fixo e $F_j \in W$ está contido no conjunto compacto $N = [0,1]$ então

$$\max_{F \in W} \rho(F, S_n) = \sup_{F \in W} \rho(F, S_n) \text{ e } \min_{F \in W} \rho(F, S_n) = \inf_{F \in W} \rho(F, S_n).$$

Sejam X e Y duas variáveis aleatórias assumindo valores respectivos E_1, E_2, \dots, E_k e $E'_1, E'_2, \dots, E'_{k'}$. Aqui o espaço Ω das distribuições F é de dimensão kk' . Denotemos por F_0 a distribuição conjunta de (X, Y) e por W ao conjunto de todas as distribuições F em $\Omega_{kk'}$ sobre os eventos conjuntos (E_i, E'_j) tal que as probabilidades de (E_i, E'_j) possam ser sempre escritos na forma $p_i q_j$, com

$$\sum_{i=1}^k p_i = 1 \text{ e } \sum_{j=1}^{k'} q_j = 1 \text{ com } p_i \geq 0 \text{ e } q_j \geq 0,$$

O problema em questão, é decidir se $F_0 \in W$ ou $\inf_{F \in W} \|F - F_0\| > \epsilon$, onde ϵ é um número positivo e fixado à priori.

Para que a regra de decisão possa ser aplicada (vide secção 3.2), deve-se mostrar que o $\sup_{F \in W} \rho(F, S_n)$ pode ser calculado.

Sejam

$$F = \{p_i q_j\} \text{ e } S_n = \{p_{ij}\},$$

e façamos

$$x_i = \sqrt{p_i}, \quad y_j = \sqrt{q_j} \quad \text{e} \quad a_{ij} = \sqrt{p_{ij}},$$

A afinidade entre F e S_n é dada por :

$$\rho(F, S_n) = \sum_{i=1}^k \sum_{j=1}^{k'} \sqrt{p_i} \sqrt{q_j} \sqrt{p_{ij}} = \sum_{i,j} x_i y_j a_{ij} \quad (3.4.1),$$

que em termos matriciais, fazendo

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1k'} \\ \vdots & \ddots & \vdots \\ a_{k1} & \dots & a_{kk'} \end{bmatrix}$$

fica :

$$\rho(F, S_n) = \langle \mathbf{A}^t \mathbf{x}, \mathbf{y} \rangle \text{ ou } \rho(F, S_n) = \langle \mathbf{x}, \mathbf{A} \mathbf{y} \rangle,$$

sendo

$$\|\mathbf{x}\| = \left(\sum_{i=1}^k x_i^2 \right)^{1/2} = 1, \quad \|\mathbf{y}\| = \left(\sum_{j=1}^{k'} y_j^2 \right)^{1/2} = 1,$$

então

$$\rho(F, S_n) = \cos \varphi \|\mathbf{A}^t \mathbf{x}\| \|\mathbf{y}\| = \cos \theta \|\mathbf{x}\| \|\mathbf{A} \mathbf{y}\|,$$

$$= \begin{cases} \cos \varphi \|\mathbf{A}^t \mathbf{x}\| \leq \|\mathbf{A}^t \mathbf{x}\| & \text{ou} \\ \cos \theta \|\mathbf{A} \mathbf{y}\| \leq \|\mathbf{A} \mathbf{y}\|, \end{cases}$$

onde θ é o ângulo entre \mathbf{x} e $\mathbf{A} \mathbf{y}$ e φ é o ângulo entre $\mathbf{A}^t \mathbf{x}$ e \mathbf{y} , temos então

$$\rho(F, S_n) = \begin{cases} \|\mathbf{A}^t \mathbf{x}\| & \text{quando } \mathbf{A}^t \mathbf{x} = c_1 \mathbf{y}, \text{ ou} \\ \|\mathbf{A} \mathbf{y}\| & \text{quando } \mathbf{A} \mathbf{y} = c_2 \mathbf{x}, \text{ com } c_1 \text{ e } c_2 \in \mathbb{R}. \end{cases}$$

De acordo com a expressão acima, a maximização de $\rho(F, S_n)$ para $F \in W$ corresponde a encontrarmos uma direção (\mathbf{x}) que maximize $\mathbf{A}^t \mathbf{x}$, isto é :

$$\max_{\substack{F \in W \\ \mathbf{x} \geq 0 \\ \|\mathbf{x}\|=1}} \rho(F, S_n) = \max_{\substack{\mathbf{x} \geq 0 \\ \|\mathbf{x}\|=1}} \|\mathbf{A}^t \mathbf{x}\| = \max_{\substack{\mathbf{y} \geq 0 \\ \|\mathbf{x}\|=1}} \|\mathbf{A} \mathbf{y}\|,$$

então

$$\max_{\substack{F \in W \\ \|x\|=1}} \rho(F, S_n) = \max_{\substack{x \geq 0 \\ \|x\|=1}} x^t A A^t x = \max_{\substack{y \geq 0 \\ \|y\|=1}} y^t A^t A y, \quad (3.4.2),$$

A maximização de (3.4.2) corresponde à maximização da expressão:

$$H(x) = x^t A A^t x - \lambda (x^t x - 1) \quad (3.4.3),$$

onde λ é multiplicador de Lagrange. A maximização de (3.4.3), corresponde a derivar $H(x)$ em função de x e igualar esta derivada a zero. Este problema se reduz a resolver a seguinte equação :

$$A A^t x = \lambda x \text{ com } x^t x = 1 ; \quad (3.4.4)$$

onde $A A^t$ pode ser escrita como

$$A A^t = \Gamma^t \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k) \Gamma \text{ com } \Gamma^t \Gamma = I,$$

onde $\lambda_1 \geq \dots \geq \lambda_k$ e $\gamma_1, \dots, \gamma_k$ são os correspondentes auto-valores e auto-vetores de $A A^t$.

Temos

$$A A^t \gamma_i = \lambda_i \gamma_i, \quad i=1, 2, \dots, k, \quad (3.4.5)$$

então

$$\|A^t \gamma_i\|^2 = \gamma_i^t A A^t \gamma_i = \lambda_i \quad (3.4.6)$$

logo :

$$\max \|A^t x\|^2 = \lambda_1.$$

A matriz AA^t é semi-definida positiva segundo (3.4.6), portanto todos os seus auto-valores são não negativos. A soma dos auto-valores de AA^t é 1, pois

$$\sum_{i=1}^I \lambda_i = \text{traço}(AA^t) = \sum_{i=1}^k \sum_{j=1}^{k'} a_{ij} a_{ij} = \sum_{i=1}^k \sum_{j=1}^{k'} p_{ij} = 1.$$

Como foi possível encontrarmos $\sup_{F \in W} \rho(F, S_n)$ as regras de decisão de Matusita podem ser utilizadas para o problema da independência entre duas variáveis aleatórias discretas.

Note que se $\max_i \lambda_i = 1$ então $\sup_{F \in W} \rho(F, S_n) = 1$, portanto $S_n \in W$ e de acordo com (3.4.2) $\max_{F \in W} \rho(F, S_n) = \sqrt{\lambda_{\max}} = (\text{raiz quadrada do máximo autovalor de } AA^t \text{ ou } A^t A)$. Como consequência temos

$$\inf_{F \in W} \|F - S_n\|^2 = \min_{F \in W} \|F - S_n\|^2 = 2(1 - \sqrt{\lambda_{\max}}).$$

3.5 Discussão

Para alguns dos Teoremas propostos por Matusita neste capítulo, avaliaremos os tamanhos amostrais calculados segundo valores fixos de precisão, η , e níveis de confiança, $1 - 1/t$. Fixamos em $k = 5$, o número de classes em $F \in \Omega_5$.

Nos teoremas I, III, e V sejam

- $1 - 1/t = 1 - \alpha_0$: Nível de Confiança
- k : Número de classes em $F \in \Omega_k$
- η : precisão
- n : Tamanho da amostra.

De acordo com os Teoremas acima, podemos compor a seguinte tabela:

Tabela I - Tamanhos amostrais para valores fixos de precisão, nível de confiança e número de classes igual a 5, para as desigualdades dos Teoremas I, III, V.

α_0	η	teoremas I e V $n = (k-1)t/\eta^2$	teorema III $n = 4(k-1)t/\eta^2$
0.01	0.5	1.600	6.400
	0.1	40.000	160.000
	0.01	4.000.000	16.000.000
0.05	0.5	320	1.280
	0.1	8.000	32.000
	0.01	800.000	3.200.000

Como se pode notar, são necessários tamanhos amostrais muito grandes para que se tenha alta precisão e riscos pequenos. Este é um dos problemas destas técnicas.

3.6 Exemplo numérico

Seja a distribuição conjunta de (X,Y) representada pela tabela abaixo. As variáveis X e Y são independentes ?

Tabela II - Distribuição amostral conjunta de (X,Y)

$y \backslash x$	1	2	3	
1	154	12	15	181
2	25	6	4	35
3	18	7	12	37
	197	25	31	253

A distribuição empírica de (X,Y) é proporcional a S_n que é dada por :

$$S_n = \begin{bmatrix} 154 & 12 & 15 \\ 25 & 6 & 4 \\ 18 & 7 & 12 \end{bmatrix}$$

De acordo com a notação utilizada na secção 3.4, seja $\sqrt{n}A = (\sqrt{S_n}_{ij})$, $i,j = 1,2,3$. O máximo auto-valor de $\sqrt{n}AA^t$, é $\lambda_{\max} = 0.9756$, e o correspondente ao auto-vetor é :

$$y_{S_n} = \begin{bmatrix} 0.854 \\ 0.370 \\ 0.366 \end{bmatrix}.$$

Sendo $d^2 = \min_{F \in W} \|F - S_n\|^2 = 0.0426$, e pelo teorema V

$$\Pr\left\{\inf_{F \in W} \|F - S_n\| < \sqrt{0.0426}\right\} \geq 1 - \frac{3 \times 3 - 1}{253 \times 0.0426} = 0.258, \text{ se } F_0 \in W,$$

a conclusão é que H_0 não deva ser rejeitada.

Pode-se ainda fixar α , digamos em $\alpha_0 = 0.08$, e calcularmos η . Se d não exceder η , decide-se que X e Y são independentes e pode-se estabelecer que com um risco menor do α_0 a distribuição de (X,Y) pertence a W . onde W é o conjunto das distribuições conjuntas de (X,Y), tais que X,Y sejam independentes.

$$\text{No exemplo acima, } d^2 = 2(1 - \lambda_{\max}) = 0.0488 \text{ e } \eta^2 = \frac{(3 \times 3 - 1)}{253 \times 0.08} = 0.395,$$

portanto não temos razão suficiente para rejeitar H_0 . Quando se aplica o teste Qui-Quadrado de Pearson obtém-se um valor calculado

$$\chi^2 = 26.955 \text{ com 4 graus de liberdade,}$$

e a hipótese de que X e Y sejam independentes é rejeitada ao nível de 0.1 % .

A conclusão nos parece estranha pois os testes pertencem à mesma classe sob a hipótese nula, ou seja, a classe dos testes com distribuição assintótica χ^2_{k-1} . Na tentativa de explicar estes resultados, calculamos as matrizes C e H a seguir, que descrevem as discrepâncias em termos da estatística Qui-quadrado e da distância de Hellinger respectivamente entre as distribuições $F = \{p_{1.}, p_{.j}\}$ e $S_n = \{p_{ij}\}$.

$$C = \begin{bmatrix} 1.210 & 1.920 & 2.320 \\ 0.185 & 1.864 & 0.019 \\ 4.050 & 3.050 & 12.320 \end{bmatrix} \quad H = \begin{bmatrix} 0.289 & 0.584 & 0.699 \\ 0.048 & 0.347 & 5.07 \times 10^{-3} \\ 1.263 & 0.534 & 1.784 \end{bmatrix}$$

onde

$$C = \{(p_{ij} - p_{1.}p_{.j})^2 / p_{1.}p_{.j}\} \text{ e } H = \{(\sqrt{p_{ij}} - \sqrt{p_{1.}p_{.j}})^2\}.$$

Obtivemos $Q = 26.955$ e $\|F - S_n\|^2 = 5.274$ isto é, $\|F - S_n\|^2 = 0.021$. O fato de $\|F - S_n\|^2$ ter sido tão inferior em relação a Q pode ser devido à transformação raiz quadrada que se efetua no cálculo de $\|\cdot\|$.

Esta transformação, assim como a transformação logarítmica, entre outras, tem a propriedade de aproximar valores atípicos como os assinalados em C, dos demais. Note que os valores em H são bem mais homogêneos comparados a C. Além disto, o teste Qui-quadrado é muito sensível para grandes valores amostrais, de forma que pequenas diferenças entre os valores esperado e observado podem levar à rejeição de H_0 , enquanto o teste proposto por Matusita leva à não rejeição de H_0 se S_n estiver "próxima" de W . Além disto, os poderes dos testes devem ser diferentes e isto é algo que precisa ser avaliado, no entanto, este problema está fora do escopo de nosso trabalho.

3.7 Coeficiente de Associação Baseado na distância de Hellinger

Com base nos trabalhos de Matusita [17] e [18], Khan e Ali [13] desenvolveram um coeficiente de associação e um teste de independência para tabelas de contingência $I \times J$.

O coeficiente de associação desenvolvido, é baseado na distância de Hellinger e possui muitas propriedades desejáveis, segundo as condições de Rényi [] tais como :

- 1) simetria nas variáveis aleatórias envolvidas;
- 2) assume valor zero quando a distribuição conjunta das variáveis aleatórias envolvidas puder ser escrita como o produto das marginais;
- 3) é invariante sob transformações um-a-um, das variáveis;
- 4) e se as variáveis aleatórias envolvidas possuírem distribuição Normal bivariada com coeficiente de correlação r , o coeficiente de Khan e Ali aumenta quando $|r|$ aumenta.

Os autores calcularam ainda a distribuição assintótica do coeficiente proposto, para tabelas $I \times J$.

Sejam X e Y variáveis aleatórias discretas cuja distribuição conjunta pode ser disposta em uma tabela de contingência $I \times J$, com

$$p_{ij} = P(X = x_i, Y = y_j) \quad i=1,2,\dots,I \quad \text{e} \quad j=1,2,\dots,J.$$

O coeficiente de associação proposto é o que segue :

$$M = \left\{ \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J \left(\sqrt{p_{ij}} - \sqrt{p_{i.} p_{.j}} \right)^2 \right\}^{1/2}, \quad (3.7.1.1)$$

portanto,

$$M = \left\{ 1 - \sum_{i=1}^I \sum_{j=1}^J \left(\sqrt{p_{ij} p_{i.} p_{.j}} \right)^2 \right\}^{1/2}, \quad (3.7.1.2)$$

onde

$$\sum_{i=1}^I p_{ij} = p_{.j}, \quad \sum_{j=1}^J p_{ij} = p_{i.},$$

e a Afinidade entre p_{ij} e $p_{i.} p_{.j}$ é dada por

$$\rho = \sum_{i=1}^I \sum_{j=1}^J \sqrt{p_{ij} p_{i.} p_{.j}} \quad (1.7.1.3)$$

portanto,

$$M^2(X,Y) = 1/2 H^2(X,Y), \quad (3.7.1.4)$$

vide definição 2.2.1.

3.7.1 Propriedades de $M(X,Y)$

Como já mencionamos, trataremos apenas de casos onde as variáveis aleatórias envolvidas forem discretas, no entanto, as proposições apresentadas a seguir podem ser facilmente extendidas para o caso contínuo.

Proposição I - Se $U = h(X)$ e $V = k(Y)$ denotam transformações um-a-um de X e Y em U e V , respectivamente, então $M(X,Y)=M(U,V)$.

Prova : Sejam (X, \mathcal{A}) e (Y, \mathcal{B}) dois espaços mensuráveis e seja $(X \times Y, \mathcal{A} \times \mathcal{B})$ respectivo espaço produto. Considere

$$U = \{u : u = h(x), x \in X\} \text{ e } V = \{v : v = k(y), y \in Y\},$$

$$Pr\{U = u, V = v\} = Pr\{h(x) = u, k(y) = v\},$$

desde que h e k são funções um a um,

$$Pr\{U = u, V = v\} = Pr\{X = h^{-1}(u), Y = k^{-1}(v)\}, \text{ ainda}$$

$$Pr\{U = u\} = Pr\{h(Y) = u\} = Pr\{X = h^{-1}(u)\} \text{ e}$$

$$Pr\{V = v\} = Pr\{k(X) = v\} = Pr\{Y = k^{-1}(v)\},$$

de acordo com a afinidade,

$$\sum_{\substack{u \in U \\ v \in V}} \{Pr(U=u, V=v)Pr(U=u)Pr(V=v)\}^{1/2} = \sum_{\substack{x \in X \\ y \in Y}} \left[Pr\{X = h^{-1}(u), Y = k^{-1}(v)\} \right]^{1/2}$$

$$\cdot \left[Pr\{X=h^{-1}(u)\}Pr\{Y=k^{-1}(v)\} \right]^{1/2},$$

subtraindo ambos os lados da igualdade de uma unidade e tomando-se a raiz quadrada, obtém-se o resultado da proposição V.

A afinidade entre distribuições contínuas Z_0 e Z_1 é dada por :

$$\rho = \int_{\mathbb{R}} \sqrt{p_0} \sqrt{p_1} dm,$$

onde m é uma medida que domina p_0 e p_1 . Matusita [18], em 1956, calculou a afinidade entre variáveis aleatórias com distribuição Normal Bivariada para vários valores do coeficiente de correlação r .

O mesmo autor em 1966, demonstrou que a afinidade entre as distribuições $N(\alpha, A^{-1})$ e $N(\alpha, b^{-1})$ é dada por :

$$\rho = \frac{\det(AB)^{1/4}}{\det(A + B/2)^{1/2}}.$$

No caso especial

$$A^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ e } B^{-1} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix},$$

sendo r o coeficiente de correlação, resulta

$$\rho = \frac{(1 - r^2)^{1/4}}{(1 - r^2/4)^{1/2}}, \text{ então } \rho^4 = \frac{16(1 - r^2)}{(4 - r^2)^2},$$

vide Khan e Ali [13].

Para este caso especial, Matusita [18], construiu uma tabela relacionando r a ρ e Khan e Ali [13], relacionaram r a $M(X,Y)$ com os mesmos valores de r utilizados por Matusita [17]. A tabela abaixo descreve o relacionamento entre r , ρ e $M(X,Y)$:

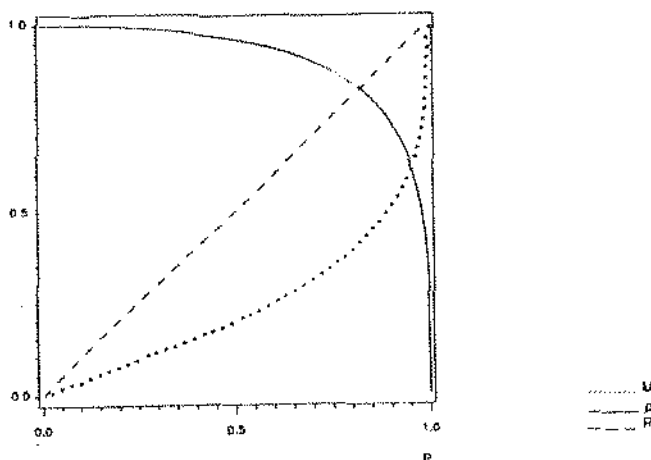
$ r $	ρ	$M(X,Y)$
0.0	1.0000	0.0000
0.1	0.9987	0.0361
0.2	0.9948	0.0721
0.3	0.9879	0.1100
0.4	0.9771	0.1513
0.5	0.9611	0.1972
0.6	0.9376	0.2498
0.7	0.9021	0.3129
0.8	0.8452	0.3935
0.9	0.7393	0.5106
1.0	0.0000	1.0000

Como se observa a propriedade 4 (vide secção 3.7) é satisfeita.

Observa-se na tabela acima que M é insensível como medida de associação pois quando é $|r| = 0.9$ por exemplo, $M = 0.5106$ e só há uma concordância maior entre as medidas para os valores mais extremos de $|r|$.

A seguinte figura ilustra as curvas descritas para M e ρ em função de r .

gráfico 3.1 - Relação entre o valor absoluto do Coef. de Correlação, Afinidade e o Coef. de Khon e Ali



Proposição II : Se (X,Y) segue distribuição Normal bivariada com coeficiente de correlação r , então $m(X,Y)$ é função crescente de $|r|$.

Verifica-se que ρ^4 é uma função monótona decrescente de r^2 ($|r| \leq 1$), então ρ é uma função monótona decrescente de r^2 . Consequentemente, $M(X,Y)$ é uma função crescente de $|r|$.

3.7.2 Distribuição Amostral de $M(X,Y)$ para Tabelas de Contingência

Consideremos uma tabela de contingência $I \times J$, com n_{ij} representando a realização de $X = x_i$ e $Y = y_j$ de um total de n realizações.

Sejam, $m(X,Y)$, o coeficiente de associação amostral para $M(X,Y)$ e as probabilidades estimadas $f_{ij} = n_{ij}/n$, $f_{i.} = n_{i.}/n$ e $f_{.j} = n_{.j}/n$,

onde,

$$m^2(X,Y) = 1/2 \sum_{i=1}^I \sum_{j=1}^J \left(\sqrt{f_{ij}} - \sqrt{f_{i.} f_{.j}} \right)^2 = (1 - \rho) \quad (3.7.2.1)$$

A Afinidade estimada (vide 3.7.1.2) pode ser expressa por :

$$\sum_{i,j} \sqrt{f_{ij} f_{i.} f_{.j}} = \sum_{i,j} f_{i.} f_{.j} \left[1 + \frac{f_{ij} - f_{i.} f_{.j}}{f_{i.} f_{.j}} \right]^{1/2} \quad (3.7.2.2)$$

Assumindo que

$$-1 < \frac{f_{ij} - f_{i.} f_{.j}}{f_{i.} f_{.j}} < 1, \quad (3.7.2.3)$$

pode-se expandir (3.7.2.2) em termos de série parcial de MacLaurin,

$$\sum_{i,j} \sqrt{f_{ij} f_{i.} f_{.j}} = \sum_{i,j} f_{i.} f_{.j} \left[1 + \frac{1}{2} \frac{f_{ij} - f_{i.} f_{.j}}{f_{i.} f_{.j}} - \frac{1}{8} \left(\frac{f_{ij} - f_{i.} f_{.j}}{f_{i.} f_{.j}} \right)^2 + \epsilon_1 \right]$$

onde $|\epsilon_1|$ é $o_p(n^{-3/2})$.

Portanto,

$$\sum_{i,j} \sqrt{f_{ij} f_{i.} f_{.j}} = \sum_{i,j} f_{i.} f_{.j} \left[1 + \frac{1}{2} \frac{f_{ij} - f_{i.} f_{.j}}{f_{i.} f_{.j}} - \frac{1}{8} \left(\frac{f_{ij} - f_{i.} f_{.j}}{f_{i.} f_{.j}} \right)^2 + o_p(n^{-3/2}) \right]$$

Sendo

$$\sum_{i,j} f_{ij} = 1 \quad \text{e} \quad \sum_{i,j} f_{i.} f_{.j} = 1,$$

então

$$\sum_{i,j} \sqrt{f_{ij} f_{i.} f_{.j}} = 1 - \frac{1}{8} \sum_{i,j} \frac{(f_{ij} - f_{i.} f_{.j})^2}{f_{i.} f_{.j}} + o_p(n^{-1/2}) \quad (3.7.2.4),$$

logo

$$\sum_{i,j} \sqrt{f_{ij} f_{i.} f_{.j}} = 1 - \frac{1}{8n} Q + o_p(n^{-1/2}) \quad (3.7.2.5),$$

onde

$$Q = \sum_{i,j} \frac{(n_{ij} - n_{i.} n_{.j}/n)^2}{n_{i.} n_{.j}/n}$$

então segue que

$$m^2(X,Y) \cong Q/8n, \quad (3.7.2.6)$$

onde Q é a estatística Qui-quadrado de Pearson sob a hipótese de independência.

A distribuição assintótica de Q , sob independência na tabela, é $\chi^2_{(I-1)(J-1)}$ (vide Rao pg 338 [20]).

Seja,

$$A = \chi^2_{(I-1)(J-1), \alpha} / 8n \quad (3.7.2.7)$$

onde $\chi^2_{(I-1)(J-1), \alpha}$ denota quantil de ordem $1 - \alpha$ da distribuição Qui-quadrado com $(I-1)(J-1)$ graus de liberdade. Define-se, então a seguinte regra de decisão:

Se $m^2 \leq A$, conclui-se que a associação entre X e Y não é significativa. Em caso contrário, admitimos que há uma associação entre X e Y. Este teste possui o mesmo comportamento que o teste Qui-quadrado : Pode-se encontrar sempre um n_0 tal que $n \geq n_0$ qualquer hipótese nula seja rejeitada.

Sendo

$$m^2(X,Y) = 1 - \sum_{i,j} \sqrt{f_{ij} f_{i.} f_{.j}} \quad (3.7.2.8)$$

a estatística m^2 , envolve um menor esforço computacional comparada à estatística Q, Qui-quadrado de Pearson.

3.7.3 Considerações com respeito à desigualdade (3.7.2.3)

No desenvolvimento exposto na secção 3.7.2 os autores assumiram (3.7.2.3) como verdadeiro. Observa-se que o lado esquerdo de (3.7.2.3) leva simplesmente a $f_{ij} > 0$ ao passo que o lado direito da expressão conduz a $f_{ij} < 2f_{i.}f_{.j}$. A existência da desigualdade

$$\frac{f_{ij} - f_{i.}f_{.j}}{f_{i.}f_{.j}} \geq 1$$

indica uma configuração patológica de associação entre linhas e colunas de uma tabela de contingência. Sob condições normais

$$\Pr \left[\frac{f_{ij} - f_{i.}f_{.j}}{f_{i.}f_{.j}} \geq 1 \right]$$

tende rapidamente para zero

Tais afirmações são baseadas em resultados de simulação que empreendemos e que descreveremos a seguir.

Seja a matriz 2x2 de frequências $M = \begin{pmatrix} x & a \\ c & b \end{pmatrix}$. Se para qualquer uma das caselas de M ocorrer $f_{ij} \geq 2f_{1.}f_{.j}$ a desigualdade (3.7.2.3) não é válida. Digamos que para $i=j=1$ $f_{11} \geq 2f_{1.}f_{.1}$, isto implica em $b \geq x + a + c + 2\frac{ac}{x}$.

Matrizes de maiores ordens podem ser reduzidas a matrizes 2x2 ao somarmos elementos de mesma linha ou coluna. Por exemplo, a matriz M 3x3 de elementos m_{ij} , $1 \leq i, j \leq 3$ pode ser reduzida a matriz M 2x2 se fizermos $x=m_{11}$, $a = m_{12} + m_{13}$, $c = m_{21} + m_{31}$ e b sendo a soma dos demais elementos de $M_{3 \times 3}$.

O procedimento que empreendemos foi o de simular matrizes 3x3 sob o modelo de independência :

$$I = \begin{bmatrix} p^2 & \alpha p^2 & \alpha p^2 \\ \alpha p^2 & \alpha^2 p^2 & \alpha^2 p^2 \\ \alpha p^2 & \alpha^2 p^2 & \alpha^2 p^2 \end{bmatrix} \begin{matrix} p \\ \alpha p \\ \alpha p \end{matrix}$$

nestas condições $p = 1/(2\alpha + 1)$.

Para α fixo, cada matriz I 3x3 foi simulada 1000 vezes e reduzida a uma matriz 2x2 (de acordo com o descrito acima). Registrou-se o número de ocorrências de $f_{11} \geq 2f_{1.}f_{.1}$ que observamos crescer para valores de α maiores ou iguais a 4.5. A medida que α cresce os elementos p^2 e αp^2 de I tendem a zero, o que indica que o teste de Khan e Ali não deva ser utilizado para matrizes com configurações deste tipo.

3.8.1 Teste de Independência para Tabelas de Contingência Multidimensionais baseado na distância de Hellinger

Para o caso de testar hipóteses de independência em tabelas de contingência multidimensionais, Goldstein, Wolf e Dillon [10] propuseram um teste que é baseado na distância de Hellinger. Estes autores investigaram o caso de tabelas do tipo $2 \times 2 \times 2$ e $2 \times 2 \times 3$.

Supondo-se X , Y e Z variáveis aleatórias cuja distribuição conjunta é dada por

$$P\{X = x_i, Y = y_j, Z = z_k\} = p_{ijk}, \text{ com } \begin{array}{l} i = 1, \dots, r_1; \\ j = 1, \dots, r_2; \\ k = 1, \dots, r_3, \end{array}$$

denotemos por

$$p_{i..} = \sum_j \sum_k p_{ijk},$$

$$p_{.j.} = \sum_i \sum_k p_{ijk} \quad \text{e}$$

$$p_{..k} = \sum_i \sum_j p_{ijk},$$

A afinidade ρ , entre $\{p_{ijk}\}$ e $\{p_{i..} p_{.j.} p_{..k}\}$ (vide Matusita [17]) é dada por :

$$\rho = \sum_i \sum_j \sum_k \sqrt{p_{ijk}} \sqrt{p_{i..} p_{.j.} p_{..k}}. \quad (3.8.1.1)$$

Define-se a distância $\| \cdot \|$, entre $\{p_{ijk}\}$ e $\{p_{i..} p_{.j.} p_{..k}\}$ por :

$$\left\| \sqrt{p_{ijk}} - \sqrt{p_{i..} p_{.j.} p_{..k}} \right\|^2 = 2(1 - \rho) \quad (3.8.1.2)$$

Se X, Y e Z são totalmente independentes então $p_{ijk} = p_{i..} p_{.j.} p_{..k}$ para todo i, j e k; portanto ρ será igual à unidade. Obtendo-se uma estimativa para ρ , digamos $\hat{\rho}$, pode-se concluir para valores próximos da unidade que a hipótese de independência entre X,Y e Z deve ser consistente com a informação amostral.

3.8.2 Distribuição Assintótica de $\hat{\rho}$:

Para encontrar a distribuição assintótica de $\hat{\rho}$, Goldstein et al [10], utilizaram as expansões em série MacLaurin, desenvolvidas por Khan e Ali [13] que foram apresentadas na secção 3.7.2. Consideremos uma tabela de contingência $R_1 \times R_2 \times R_3$ e frequência da casela (ijk), denotado por n_{ijk} .

Sejam os totais marginais $n_{i..}$, $n_{.j.}$ e $n_{..k}$ definidos por :

$$\begin{aligned} n_{i..} &= \sum_j \sum_k n_{ijk}, \\ n_{.j.} &= \sum_i \sum_k n_{ijk} \quad \text{e} \\ n_{..k} &= \sum_i \sum_j n_{ijk}, \end{aligned}$$

As estimativas de Máxima Verossimilhança de p_{ijk} , $p_{i..}$, $p_{.j.}$ e $p_{..k}$, são respectivamente :

$$\begin{aligned} f_{ijk} &= \frac{n_{ijk}}{n}, \\ f_{i..} &= \frac{n_{i..}}{n}, \end{aligned}$$

$$f_{.j.} = \frac{n_{.j.}}{n} \quad e$$

$$f_{..k} = \frac{n_{..k}}{n} ,$$

$$\text{onde } n = \sum_i \sum_j \sum_k n_{ijk} .$$

Para estimar ρ , substitui-se p_{ijk} , $p_{i..}$, $p_{.j.}$ e $p_{..k}$ por seus respectivos estimadores de Máxima Verossimilhança e obtém-se :

$$\hat{\rho} = \sum_i^{r_1} \sum_j^{r_2} \sum_k^{r_3} \sqrt{f_{ijk} f_{i..} f_{.j.} f_{..k}} , \quad (3.8.2.1)$$

De acordo com os mesmos argumentos utilizados em 3.7.2 sob a hipótese de Independência na tabela,

$$\hat{\rho} = 1 - \frac{1}{8n} Q_1$$

onde,

$$Q_1 = \sum_{i,j,k} \frac{(n_{ijk} - n_{i..} n_{.j.} n_{..k} / n^2)^2}{n_{i..} n_{.j.} n_{..k} / n^2}$$

A distribuição assintótica de V_1 é $\chi^2_{(r_1 r_2 r_3 - r_1 - r_2 - r_3 + 2)}$, então assintoticamente,

$$\hat{\rho} \cong 1 - \frac{1}{8n} \chi^2_{(r_1 r_2 r_3 - r_1 - r_2 - r_3 + 2)} . \quad (3.8.2.2)$$

O lado esquerdo de (3.8.2.2) pode ser expresso por :

$$\hat{\rho} = \frac{1}{n} \sum_i^{r_1} \sum_j^{r_2} \sum_k^{r_3} \sqrt{O_{ijk} E_{ijk}} \quad (3.8.2.3)$$

onde O_{ijk} é a frequência observada na casela (ijk) , enquanto E_{ijk} é a frequência esperada, sob a hipótese de independência, isto é,

$$E_{ijk} = n_{i..} n_{.j.} n_{...k} / n^2.$$

Pode-se estabelecer a seguinte

Regra de Decisão : Rejeita-se a hipótese de independência ao nível de significância α se :

$$8n \left\{ 1 - \frac{1}{n} \sum_i \sum_j \sum_k \sqrt{O_{ijk} E_{ijk}} \right\} > \chi^2_{(r_1 r_2 r_3 - r_1 - r_2 - r_3 + 2), \alpha} \quad (3.8.2.4)$$

A estatística m (vide 3.7.2.6), para tabelas $I \times J$, aqui tem sua extensão imediata para tabelas de maior ordem.

Suponha que se queira analisar a hipótese de independência em uma tabela K -dimensional, digamos $R_1 \times R_2 \times \dots \times R_k$, então a região crítica do teste, é dada por :

$$8n \left\{ 1 - \frac{1}{n} \sum_{i_1} \sum_{i_2} \dots \sum_{i_k} \sqrt{O_{i_1 i_2 \dots i_k} E_{i_1 i_2 \dots i_k}} \right\} > \chi^2_{\left(\prod_{i=1}^k r_i - \sum_{i=1}^k r_i + k - 1 \right), \alpha}$$

(3.8.2.5)

O teste é também aplicável a diferentes tipos de independência, por exemplo, a hipótese

$$H_0 : p_{ijk} = p_{ij.} p_{...k},$$

indica independência entre (X, Y) e Z . A região crítica é dada por (3.8.2.4), onde $E_{ijk} = (n_{ij.} n_{...k}) / n^2$ e os graus de liberdade são $(r_1 r_2 - 1)(r_3 - 1)$.

Goldstein et al [10] compararam através de simulações (para tamanhos amostrais 100, 150 e 200) o poder da estatística m de Khan e Ali [13] com o poder das estatísticas Q e G^2 , respectivamente, Qui-quadrado de Pearson e Razão de verossimilhança. Os autores observaram que o poder destas estatísticas é praticamente o mesmo, de forma que o esforço computacional requerido por estas pode ser um fator de importância na escolha de alguma.

3.9 Dependência e Codependência

Baseado nos trabalhos de Matusita [17] - [18] e Khan e Ali [13], Cordeiro [04], desenvolveu algumas expressões para medir a dissimilaridade entre distribuições Multinomiais. Estas expressões foram obtidas em função das divergências medidas por meio da distância de Hellinger entre uma distribuição multinomial qualquer e dada distribuição sob a condição de independência. A utilização destas expressões permite a realização de análises do tipo Componentes Principais e testes de Independência.

A vantagem que se tem com a utilização de tais medidas de dissimilaridade, está no ganho de informação com respeito aos dados em estudo. Esta última afirmação ficará mais clara na seção 3.10 a seguir.

Sejam A e B duas características de classificação de indivíduos em uma população (suposta infinita), com A_1, A_2, \dots, A_I e B_1, B_2, \dots, B_J sendo os níveis de A e B respectivamente.

Seja

$$P_{I \times J} = (p_{ij}) \quad \text{matriz das probabilidades,} \quad (3.9.1)$$

onde p_{ij} é a probabilidade de um indivíduo escolhido ao acaso em uma população (suposta infinita) possuir os níveis i e j com $1 \leq i \leq I$ e $1 \leq j \leq J$ das características A e B respectivamente. Utilizaremos as seguintes notações

$$P_{.i} = \sum_{j=1}^J p_{ij} \quad (3.9.2)$$

$$P_{.j} = \sum_{i=1}^I p_{ij} \quad (3.9.3)$$

$$r_{.i}^t = (p_{i1}, p_{i2}, \dots, p_{iJ}) / P_{.i} \quad - \text{ i-ésimo perfil linha} \quad (3.9.4)$$

$$c_j^t = (p_{1j}, p_{2j}, \dots, p_{Ij}) / P_{.j} \quad - \text{ j-ésimo perfil coluna} \quad (3.9.5)$$

$$r_{.+}^t = (P_{.1}, P_{.2}, \dots, P_{.J}) \quad - \text{ perfil linha marginal} \quad (3.9.6)$$

$$c_{.+}^t = (P_{.1}, P_{.2}, \dots, P_{.J}) \quad - \text{ perfil coluna marginal} \quad (3.9.7)$$

onde $p_{ij} > 0$, e $1 \leq i \leq I$, $1 \leq j \leq J$.

As expressões (3.9.4) a (3.9.7) são distribuições de probabilidade.

Sejam $1 \leq j, j^* \leq J$ e $1 \leq i, i^* \leq I$. Definem-se por :

Codependência entre as colunas j e j^* :

$$d_{jj^*} = (1/2) \sum_{i=1}^I (\sqrt{p_{i,j}} - \sqrt{p_{i.} p_{.j}})(\sqrt{p_{i,j^*}} - \sqrt{p_{i.} p_{.j^*}}) \quad (3.9.8)$$

Dependência da coluna j :

$$\begin{aligned}
 {}_c\partial_{jj}^2 &= {}_c d_{jj} = (1/2) \sum_{i=1}^I (\sqrt{p_{ij}} - \sqrt{p_{i.} p_{.j}})^2 & (3.9.9) \\
 &= (1/2) P_{.j} (\sqrt{{}_c P_j} - \sqrt{{}_c P_+})^t (\sqrt{{}_c P_j} - \sqrt{{}_c P_+}) \\
 &= (1/2) P_{.j} \left\| \sqrt{{}_c P_j} - \sqrt{{}_c P_+} \right\|^2
 \end{aligned}$$

aqui $\left\| \cdot \right\|$ é a distância Euclideana.

Dependência Padrão da coluna j :

$${}_c\partial_j = \sqrt{{}_c\partial_{jj}^2} \quad (3.9.10)$$

Coefficiente de Codependência entre as colunas j e j*

(Definido se ${}_c\partial_j > 0$ e ${}_c\partial_{j^*} > 0$) :

$${}_c\delta_{jj^*} = \frac{{}_c d_{jj^*}}{{}_c\partial_j {}_c\partial_{j^*}} ; \quad (3.9.11)$$

Matriz de Dependência e Codependência para as colunas :

$$D = \Delta^t \Delta, \text{ onde } \Delta_{ixj} = \{ (\sqrt{p_{ij}} - \sqrt{p_{i.} p_{.j}}) / \sqrt{2} \} \quad (3.9.12)$$

vide 3.9.1, 3.9.6 e 3.9.7 .

Equivalentemente, para as linhas, definem-se :

Codependência entre as linhas i e i* :

$$d_{ii^*} = (1/2) \sum_{j=1}^J (\sqrt{p_{1j}} - \sqrt{p_{1, p_{\cdot j}}})(\sqrt{p_{1^*j}} - \sqrt{p_{1^*, p_{\cdot j}}}) \quad (3.9.13)$$

Dependência da linha i :

$$\begin{aligned} \delta_i^2 = d_{ii} &= (1/2) \sum_{j=1}^J (\sqrt{p_{1j}} - \sqrt{p_{1, p_{\cdot j}}})^2 \quad (3.9.14) \\ &= (1/2) P_{1,} (\sqrt{P_{1,}} - \sqrt{P_{+,}})^t (\sqrt{P_{1,}} - \sqrt{P_{+,}}) \\ &= (1/2) P_{1,} \left\| \sqrt{P_{1,}} - \sqrt{P_{+,}} \right\|^2 \end{aligned}$$

aqui $\left\| \cdot \right\|$ é a distância Euclídeana.

Dependência Padrão da linha i :

$$\delta_i = \sqrt{\delta_i^2} \quad (3.9.15)$$

Coefficiente de Codependência entre as linhas i e i*

(definido de $\delta_i > 0$ e $\delta_{i^*} > 0$)

$$\delta_{ii^*} = \frac{d_{ii^*}}{\delta_i \delta_{i^*}} ; \quad (3.9.16)$$

Matriz de Dependência e Codependência para as linhas :

$$D = \Delta \Delta^t, \text{ onde } \Delta_{ixj} = \{(\sqrt{p_{1j}} - \sqrt{p_{1, p_{\cdot j}}})/\sqrt{2}\} \quad (3.9.17)$$

vide 3.9.1, 3.9.6 e 3.9.7 .

3.9.1 Interpretação das definições 3.9.8 a 3.9.16

As definições 3.9.8 a 3.9.11, e similarmente 3.9.13 a 3.9.17, têm interpretações semelhantes à covariância, variância, desvio-padrão e coeficiente de correlação nesta ordem, para variáveis aleatórias.

A codependência entre as colunas j e j^* (definição 3.9.8), pode ser re-escrita como :

$$d_{jj^*} = (1/2)\sqrt{p_{.j}p_{.j^*}}(\sqrt{p_{.j}} - \sqrt{p_{.+}})(\sqrt{p_{.j^*}} - \sqrt{p_{.+}}), \quad (3.9.11)$$

Portanto, a codependência (3.9.8) é proporcional ao produto interno entre os desvios das raízes quadradas dos perfis probabilísticos $P_{.j}$ e $P_{.j^*}$ em relação à raiz quadrada do perfil probabilístico marginal $P_{.+}$.

Para analisar-se a semelhança entre a codependência e a covariância, seja a definição a seguir :

Definição 3.9.1.1 A coluna j , ou nível B_j , $1 \leq j \leq J$, da característica B , é independente das linhas, ou seja, da característica A , se :

$$P_{ij} = P_{i.}P_{.j} \quad \text{para } 1 \leq i \leq I.$$

Tal definição traz o conceito de independência local em tabelas de contingência. Se todas as categorias de uma característica, digamos B , forem independentes das categorias da característica A , então $d_{jj^*} = 0$ para todo j , ou seja, tem-se independência total na tabela.

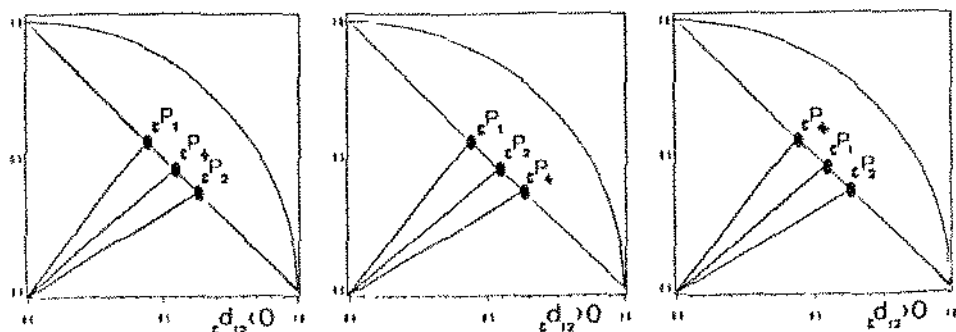
Se a coluna j for independente das linhas então $\sigma_j^2 = 0$ e também $d_{jj^*} = 0$ para $j \neq j^*$. Isto quer dizer que o perfil p_j não se diferencia do perfil marginal p_+ (vide 3.9.8), isto é, o nível B_j da característica B não é diferente, segundo a característica A , da coluna marginal. Quando isto acontece, o nível B_j pode ser ignorado no estudo de associações entre características A e B . Como se observa, esta é uma técnica que permite a redução dos níveis de uma tabela de contingência.

Há situações em que $d_{jj^*} = 0$ e nem a coluna j ou a coluna j^* é independente das linhas, da mesma forma que uma covariância nula não implica, em geral, em independência estocástica, vide Cordeiro [04] pg 3.

Se a codependência d_{jj^*} não for nula, o quanto é positiva ou negativa, vai depender também dos pesos marginais das colunas j e j^* (vide expressão 3.9.1.1). Da mesma forma, se a coluna j for dependente das linhas, sua dependência σ_j^2 será proporcional ao seu peso marginal. Desta maneira, colunas com pesos marginais maiores, tendem a ser mais dependentes.

Valores positivos de d_{jj^*} indicam que os perfis coluna j e j^* em \mathbb{R}^I tendem a estar do mesmo lado quando comparados ao perfil coluna marginal. Um valor negativo de d_{jj^*} indica tendência contrária, isto é, os perfis coluna j e j^* em \mathbb{R}^I tendem a estar em posições opostas em relação ao perfil coluna marginal. Vide gráfico 3.2 a seguir.

Gráfico 3.2 Posição relativa dos Perfis-coluna 1 e 2 em \mathbb{R}^2
com respeito ao Perfil Marginal



3.9.2 O Coeficiente de Codependência

O Coeficiente de Codependência (vide 3.9.11 e 3.9.16) elimina como medida de associação, a influência dos pesos marginais, da mesma forma que o coeficiente de correlação elimina a influência das dispersões para variáveis aleatórias.

Proposição 3.9.2.1 : O valor absoluto de δ_{cJJ^*} é no máximo igual a 1.

Prova - Tem-se que $|d_{cJJ^*}| \leq \delta_{cJ} \delta_{cJ^*}$, por desigualdade de Cauchy - Schwarz, logo $|\delta_{cJJ^*}| \leq 1$.

Proposição 3.9.2.2 : Se $\delta_{cJ} > 0$ e $\delta_{cJ^*} > 0$, então $\delta_{cJJ^*} = 1$ se e só se

$$P_{cJ} = P_{cJ^*}.$$

Prova : Note que ${}_c\delta_{jj^*}$ é exatamente o cosseno do ângulo entre

$$(\sqrt{{}_cP_j} - \sqrt{{}_cP_+}) \text{ e } (\sqrt{{}_cP_{j^*}} - \sqrt{{}_cP_+}),$$

portanto, $\cos\alpha = {}_c\delta_{jj^*} = 1$, implica $\alpha = 0$, isto é, os vetores $\sqrt{{}_cP_j}$, $\sqrt{{}_cP_+}$ e $\sqrt{{}_cP_{j^*}}$ pertencentes ao R^I , são colineares.

Como ${}_c\delta_j > 0$ e ${}_c\delta_{j^*} > 0$, isto é, ${}_cP_j \neq {}_cP_+$ e ${}_cP_{j^*} \neq {}_cP_+$, a colinearidade só é possível se ${}_cP_j = {}_cP_{j^*}$.

Reciprocamente, se ${}_cP_j = {}_cP_{j^*}$, por (3.9.1.1), ${}_c\delta_{jj^*} = 1$.

O Coeficiente de Codependência pode ser utilizado na redução de dimensão em tabelas de contingência, ou seja, se o coeficiente de codependência entre duas colunas, digamos j e j^* for próximo de 1, isto indica que é razoável a aglutinação das mesmas em uma única.

Para tanto, deve-se agregar as frequências respectivas das colunas j e j^* para cada nível das linhas da tabela. O Coeficiente de Codependência entre duas colunas será nulo, se o ângulo que seus perfis formarem com o perfil marginal for de 90° .

Cordeiro [04], apresentou uma série de exemplos utilizando cruzamentos fictícios em tabelas quadradas $I \times I$, com associação completa e balanceamento marginal, e foram derivadas as seguintes relações :

$${}_c d_{jj^*} = \frac{1}{2I} [1 - 2\sqrt{I}] ; \quad j \neq j^* \quad (3.9.2.1)$$

$${}_c \delta_j^2 = \frac{1}{I} [1 - 1/\sqrt{I}] ; \quad 1 \leq j \leq J \quad (3.9.2.2)$$

$${}_c\delta_{jj^*} = \frac{1}{2} \frac{[\sqrt{I} - 2]}{[\sqrt{I} - 1]} ; \quad j \neq j^* . \quad (3.9.2.3)$$

Donde conclui-se que para aqueles casos :

$$(1) \quad {}_c\delta_{jj^*} = 0 \text{ se e só se } I = 4,$$

(2) Para $I > 4$, a Codependência é positiva.

(3) A cota superior para o Coeficiente de Codependência

$$\text{é o } \lim_{I \rightarrow \infty} {}_c\delta_{jj^*} = 1/2$$

Para $I = 4$, o fato de ${}_c\delta_{jj^*} = 0$, e no entanto, existir associação completa entre linhas e colunas, não invalida o uso do Coeficiente de Codependência como medida de associação pois situações semelhantes ocorrem com a covariância, e esta é amplamente utilizada.

Pondo

$$M = 1/2 \left\{ \sum_{i=1}^I \sum_{j=1}^J \left(\sqrt{p_{ij}} - \sqrt{p_{i.} p_{.j}} \right)^2 \right\}^{1/2} ,$$

tem-se que

$$M^2 = \sum_{j=1}^J {}_c\delta_j^2 = \sum_{i=1}^I {}_r\delta_i^2 \quad (3.9.2.4)$$

Portanto, o Coeficiente de Associação de Khan e Ali [13], pode ser decomposto na soma das Dependências de linhas ou colunas.

Definição 3.9.2.1 Define-se $d^2 = M^2$, a Dependência Total de uma tabela de contingência.

3.9.3 Dependência e Codependência Amostral

Seja

$$T = (n_{ij}; 1 \leq i \leq I, 1 \leq j \leq J),$$

uma tabela de frequências absolutas de classificação de $n = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$ indivíduos amostrados de uma população estratificada de acordo com as características A e B, onde $A = (A_1, A_2, \dots, A_I)$ e $B = (B_1, B_2, \dots, B_J)$, sendo n_{ij} o número de indivíduos que foram amostrados nos níveis A_i de A e B_j de B.

Fazendo $f_{ij} = n_{ij}/n$, os estimadores de Máxima Verossimilhança dos parâmetros em (3.9.1) a (3.9.8), são obtidos substituindo-se p_{ij} por f_{ij} , $p_{i.}$ por $f_{i.}$ e $p_{.j}$ por $f_{.j}$.

A Dependência Total amostral d^2 satisfaz :

$$Q = 8nd^2, \quad (3.9.3.1)$$

onde n é o tamanho da amostra em uma tabela de contingência e Q é a estatística Qui-quadrado de Pearson.

3.10 Análise de Componentes Principais utilizando coeficientes de Dependência e Codependência

Antes de nos dirigirmos diretamente ao tema, traçaremos um breve histórico sobre o surgimento de Análise de Componentes Principais, situando-a em um conjunto de técnicas de análise de dados denominada Escalonamento Dual.

De acordo com Nishisato [27], uma vez que se tenham dados dispostos em forma de tabela (por exemplo, de contingência), a análise a ser empregada envolve quantificação das linhas e colunas da tabela, de modo a otimizar certo critério. Por exemplo, em uma tabela de incidência de três níveis de ansiedade e cinco categorias de aproveitamento escolar, em um grupo de estudantes, podem-se derivar duas variáveis reais, em função das categóricas, de modo que a correlação entre as variáveis derivadas seja máxima. Esta abordagem não é nova, já possui mais de 57 anos. Durante este tempo, esta técnica tem sido descoberta e redescoberta, independentemente, por vários investigadores e tem sido advogada sob uma variedade de nomes, tais como Método das Médias Recíprocas, *Additive Scoring*, *Appropriate Scoring*, *Canonical Scoring*, Análise de Componentes Principais, Análise de Correspondência, entre outros.

Alguns destes nomes foram derivados de um algoritmo utilizado em particular por seus proponentes e outros são específicos demais para caracterizar uma técnica. Para evitar confusão entre usuários destas técnicas, em áreas de estudo distintas, Nishisato [27] propôs o nome Escalonamento Dual.

Fénelon e Lebart [10] cap. IV, apresentam uma Análise Geral, que consta de resultados básicos para a construção de análises como a de Componentes Principais e Análise de Correspondência, entre outras. Segundo os mesmos autores, pg. 206, a natureza dos dados a serem analisados, impõem transformações preliminares nas variáveis envolvidas, após as quais, a Análise Geral é empregada.

3.10.1 Descrição da Análise de Componentes Principais

Segundo Lebart e Fénelon [10], a Análise de Componentes Principais (A.C.P.) pode ser descrita resumidamente como segue :

Dispõe-se de dados estatísticos que representam um espaço de dimensão elevada, por exemplo, 60 variáveis observadas em 1000 indivíduos, ou 8 características econômicas medidas em 21 regiões, e se deseja representar, na medida do possível, os dados estatísticos em um espaço de menor dimensão com perda mínima de informação. Por exemplo, desejando resumir as 60 variáveis medidas nos 1000 indivíduos, a 5 variáveis, determina-se desta forma, um subespaço do espaço das 60 variáveis.

Na A.C.P., o subespaço em questão é formado por auto-vetores correspondentes aos maiores auto-valores da matriz V de Variância e Covariância das variáveis. Estes autovetores são denominados de Eixos Principais.

Em termos gerais, sendo $X_{n \times p}$ uma matriz de dados onde X_{ij} é a i -ésima observação da j -ésima variável e \bar{X}_j a média das n observações da j -ésima variável em \mathbb{R}^p , efetuando a transformação (dissimétrica entre os índices i e j)

$$R_{n \times p} = (r_{ij}) = (X_{ij} - \bar{X}_j), \quad (3.10.1.1)$$

e aplicando a Análise Geral mencionada anteriormente, obtemos uma forma quadrática em termos da matriz V de variância e covariância das p variáveis, que deve ser maximizada sujeito à restrição de que o vetor u da forma quadrática seja unitário.

Sendo n o número de observações em X , verificamos que a matriz de Variância e Covariância V é dada por :

$$V = \frac{1}{n} R^t R. \quad (3.10.1.2)$$

Os vetores unitários que maximizam a forma quadrática de V , os denominados eixos principais, são os auto-vetores de V dados por :

$$\text{no } \mathbb{R}^p : w_q = \frac{1}{\sqrt{n\lambda_q}} R^t v_q \quad (3.10.1.3)$$

$$\text{no } \mathbb{R}^n : v_q = \frac{1}{\sqrt{n\lambda_q}} R w_q \quad (3.10.1.4)$$

com w_q e v_q unitários, e respectivamente auto-vetores de $R^t R$ e RR^t . λ_q é auto-valor de w_q e v_q .

Ao projetarmos R sobre w_q , obtendo o produto Rw_q , temos as coordenadas do q -ésimo Fator ou Componente Principal do espaço \mathbb{R}^n . E de maneira análoga, $R^t v_q$ são as coordenadas do q -ésimo Fator ou Componente Principal do espaço \mathbb{R}^p . Estes Fatores explicam como se dão as associações entre as covariáveis em estudo.

A variância do q -ésimo fator do espaço linha - \mathbb{R}^n , é segundo Fénelon e Lebart [10] pg. 215 dada por :

$$\text{var}(Rw_q) = \frac{1}{n} w_q^t R^t R w_q = \lambda_q. \quad (3.10.1.5)$$

e de maneira análoga, a variância do q -ésimo fator do espaço coluna - \mathbb{R}^p , dado por :

$$\text{var}(R^t v_q) = \frac{1}{n} v_q^t R R^t v_q = \lambda_q, \quad (3.10.1.6)$$

estas variâncias informam as dispersões dos pontos - matriz R ou R^t - projetados sobre os eixos principais.

A matriz V em (3.10.1.2) é sempre semi definida positiva. Este fato significa que o número de fatores é menor do que o posto da matriz de dados, e em consequência a matriz V contém menos informação

que a matriz original X. Esta situação esta descrita em Lebart e Fénelon [10] pg. 235.

Cordeiro [05], desenvolveu uma técnica de Análise de Componentes Principais onde os fatores de ponderação para linhas e colunas da matriz de dados (raiz quadrada dos perfis-linha ou coluna) são encontrados em função dos auto-vetores da matriz de Dependência e Codependência D (para linhas ou colunas).

Baseado na Análise Geral de Fénelon e Lébart [10], Cordeiro verificou que a raiz quadrada de todo perfil-linha (ou coluna), sob o modelo de independência, ao ser projetada sobre um vetor unitário \underline{u} , resulta na projeção do perfil-marginal (linha ou coluna) como segue para os perfis linha :

sob o modelo geral,

$$\underline{u}^t \sqrt{r p_i} = \sum_j u_j \sqrt{p_{ij} / p_i} ,$$

e sob o modelo de independência,

$$\underline{u}^t \sqrt{r p_i} = \sum_j u_j \sqrt{p_{.j}} = \underline{u}^t \sqrt{r p_{+}} ,$$

para $i=1, \dots, I$.

Para tabelas de contingência $I \times J$, ao definir vetores dados por

$$z_k = \left\{ \sqrt{\frac{p_{k1}}{p_{k.}}} - \sqrt{\frac{p_{k.} p_{.1}}{p_{k.}}}, \dots, \sqrt{\frac{p_{kJ}}{p_{k.}}} - \sqrt{\frac{p_{k.} p_{.J}}{p_{k.}}} \right\} ,$$

para $k = 1, \dots, I$, obteve que S^2 , soma dos quadrados das projeções dos

pontos \underline{z}_k ponderados por $p_k/2$ sobre um vetor unitário \underline{u} para $k=1,\dots,I$, resulta em uma forma quadrática em termos da matriz de Dependências e Codependências para colunas como segue :

$$S^2 = \frac{1}{2} \underline{u}^t \left\{ \sum_{i=1}^I p_{i.} \underline{z}_i \underline{z}_i^t \right\} \underline{u} = \underline{u}^t \Delta^t \Delta \underline{u}, \quad (3.10.1.7)$$

onde Δ foi definido na equação (3.9.12).

S^2 foi denominado por Cordeiro [05] por Dependência de um vetor unitário \underline{u} , d_u , como segue pela definição abaixo.

Definição Seja T uma tabela de frequências absolutas de classificação, como descrito na secção 3.9.5. Se $\underline{u} \in \mathbb{R}^J$ é um vetor unitário, então a dependência do vetor \underline{u} , d_u é a que segue :

$$d_u = 1/2 \sum_{i=1}^I p_{i.} \left[\sum_{j=1}^J u_j (\sqrt{p_{ij}/p_{i.}} - \sqrt{p_{.j}}) \right]^2. \quad (3.10.1.8)$$

Se houver independência entre as categorias da tabela, então $d_u = 0$ para todo \underline{u} , isto quer dizer que $u^t(\sqrt{P_{r1}} - \sqrt{P_{r+}}) = 0$, para $i = 1, \dots, I$, isto é, $P_{r1} = P_{r+}$ para todo i .

$d_u > 0$, implica em dependência na tabela, neste caso, é intuitivo procurar os vetores $\underline{u} \in \mathbb{R}^J$, unitários, com maior dependência pois estes indicarão direções para se olharem os dados (perfis-linha) no espaço coluna de forma mais favorável para se perceberem as associações.

O problema que se coloca é, portanto, o de maximizar d_u sujeito à restrição $u^t u = 1$.

Segundo Cordeiro [05], obtem-se a matriz Γ_k dos k primeiros "Fatores de Dependência" no espaço coluna,

$$\Gamma_k = (\gamma_1 | \gamma_2 | \dots | \gamma_k) : j \times k$$

onde γ_1 auto-vetores de $\Delta' \Delta$, e correspondentemente,

$$\psi_k = (\psi_1 | \psi_2 | \dots | \psi_k)$$

os k primeiros "Fatores de Dependência" no espaço linha.

Os eixos principais (Componentes de Dependência) são relacionados pelas expressões :

$$\text{no } R^J : \gamma_q = \frac{1}{\sqrt{\lambda_q}} \Delta^t \psi_q \quad (3.10.1.8)$$

$$\text{no } R^I : \psi_q = \frac{1}{\sqrt{\lambda_q}} \Delta \gamma_q \quad (3.10.1.9)$$

e a Dependência de γ_q , d_{γ_q} , é dada por :

$$d_{\gamma_q} = \gamma_q^t \Delta^t \Delta \gamma_q = \lambda_q \quad (3.10.1.10)$$

Há um auto-vetor (direção) que aponta ao setor da esfera onde estão as raízes quadradas dos perfis-linha e tem todas as suas j coordenadas com o mesmo sinal.

Sua Dependência mede o "quanto" as raízes quadradas dos perfis-linha estão espalhados sobre a esfera, ao passo que as outras componentes de dependência dizem "como" esses pontos estão distribuídos. Este auto-vetor é único pois $\Delta^t \Delta$ sendo simétrica implica em todos os seus auto-vetores serem ortogonais.

Em outras técnicas, como na Análise de Componentes Principais e de Correspondências, todas as componentes informam "como" se dão as associações entre as covariáveis em estudo através dos "contrastes" descritos pelos fatores (vide Cordeiro pgs 20 e 21 [05]).

A transformação a seguir, denominada em Cordeiro [05], de transformação ANADEP, permite a visualização das posições relativas dos perfis-linha em relação à linha marginal :

transformação ANADEP de ordem k do i-ésimo perfil-linha :

$${}_r y_i = \Gamma_k^t ({}_r p_i - {}_r p_+) , \quad (3.10.1.11)$$

transformação ANADEP de ordem k do j-ésimo perfil-coluna :

$${}_c x_j = \psi_k^t ({}_c p_j - {}_c p_+) ; \quad (3.10.1.12)$$

A leitura das posições relativas dos perfis-linha em relação à linha marginal produz interpretações sobre as associações na tabela cruzada.

Cordeiro [05] apresenta uma série de exemplos demonstrativos da utilização da técnica descrita.

Comparação de Matrizes de Dependência e Codependência

4.1 Descrição do problema e Metodologia

Neste capítulo nosso objetivo principal é o de investigar a viabilidade da construção de uma tabela de pontos percentuais que possibilite testar a igualdade entre duas matrizes de Dependência e Codependência.

Mais precisamente, no tocante à tabelas de classificação cruzada do tipo $(I \times J) \times K$ onde $2 \leq J \leq I \leq 3$ e $K = 2$, se o último critério de classificação não modificar a maneira na qual linhas e colunas (referentes aos critérios I e J) se associam, a cada nível do último critério (K), é razoável que este seja eliminado da análise, restando portanto duas tabelas de contingência, T_1 e T_2 , que podem ser agregadas.

Sabendo que tabelas de classificação cruzada próximas deverão ter matrizes de Dependência e Codependência próximas, para estudarmos se o critério K é influente ou não, é razoável a hipótese nula

$$H_0 : D_1 = D_2, \quad (4.1.1)$$

onde D_1 e D_2 são as matrizes de Dependência e Codependência para as tabelas $T_1 = (I \times J)|_{k=1}$ e $T_2 = (I \times J)|_{k=2}$, respectivamente.

Seja ξ um critério de comparação entre D_1 e D_2 . Suponhamos que sob a hipótese H_0 , $\xi \equiv 1$. Um questionamento pertinente à questão é se o tipo de associação entre linhas e colunas em T_1 e T_2 ou as dimensões das tabelas T_1 e T_2 causam ou não diferenças na distribuição de ξ .

Caso a distribuição de ξ mantenha-se a mesma, digamos para todas as dimensões, é possível tabelar os pontos percentuais da distribuição de ξ no intuito de testar H_0 .

Roy em 1957 [31] estabeleceu regiões de confiança (segundo o Princípio da União e Intersecção) para hipóteses compostas.

No caso de vetores de variáveis aleatórias com as distribuições

$$\underline{X}_{1(m \times p)} \sim N(\underline{\mu}_1, \underline{\Sigma}_1) \text{ e}$$

$$\underline{X}_{2(m \times p)} \sim N(\underline{\mu}_2, \underline{\Sigma}_2),$$

consideremos as hipóteses,

$$H_0 : \underline{\Sigma}_1 = \underline{\Sigma}_2$$

$$H_a : \underline{\Sigma}_1 \neq \underline{\Sigma}_2$$

Segundo Roy,

$$\bigcap_{\underline{a} \neq \underline{0}} H(\underline{a}^t \underline{\Sigma}_1 \underline{a} = \underline{a}^t \underline{\Sigma}_2 \underline{a}) \text{ equivale a } H(\underline{\Sigma}_1 = \underline{\Sigma}_2).$$

Definimos uma função $F_{\underline{a}}$ como

$$F_{\underline{a}} = \frac{\underline{a}^t \underline{S}_1 \underline{a}}{\underline{a}^t \underline{S}_2 \underline{a}}, \quad (4.1.2)$$

onde \underline{S}_1 e \underline{S}_2 , são as matrizes de Variância e Covariância amostrais de \underline{X}_1 e \underline{X}_2 respectivamente, com \underline{S}_1 pelo menos Semi-Definida Positiva (s.d.p.) e \underline{S}_2 Positiva Definida (p.d.).

Ao dizermos que Y é uma matriz definida positiva estamos dizendo que todos os seus auto-valores são positivos, enquanto Y semi-definida positiva possui pelo menos um auto-valor nulo.

O Teste baseado no Princípio da União e Intersecção de Roy utiliza o fato de que sob H_0

$$\min \underline{a}^t \Sigma_1 \underline{a} / \underline{a}^t \Sigma_2 \underline{a} = \max \underline{a}^t \Sigma_1 \underline{a} / \underline{a}^t \Sigma_2 \underline{a} = 1,$$

para todo $\underline{a} \in \mathbb{R}^p$ (vide Flury [12]).

Então o teste de Roy [31] se baseia na maximização e minimização de $F_{\underline{a}}$ que conduz ao problema de se encontrar as raízes características extremas de $S_2^{-1} S_1$.

Denominando por l_1 e l_p o máximo e mínimo auto - valores, respectivamente, de $S_2^{-1} S_1$, a estatística de Roy é o par (l_1, l_p) , denotado por "Critério das Raízes Máxima e Mínima".

A hipótese

$$H_0 : \Sigma_1 = \Sigma_2,$$

é rejeitada se l_1 for muito grande com relação a l_2 .

Seguindo a mesma linha de raciocínio de Roy, inicialmente o critério de comparação adotado foi o de proceder à maximização e minimização de

$$F_1(\underline{a}) = \underline{a}^t D_1 \underline{a} / \underline{a}^t D_2 \underline{a}, \quad (4.1.3)$$

para todo \underline{a} não nulo $\in \mathbb{R}^J$, pois D_1 e D_2 são simétricas e pelo menos semi-definidas positivas :

sendo $D_{(JxJ)} = \Delta^t \Delta$, com

$$\Delta_{(IxJ)} = \left[\frac{1}{\sqrt{2}} \left\{ \sqrt{p_{ij}} - \sqrt{p_{i.} p_{.j}} \right\} \right], \quad (4.1.4)$$

e $p_{i.}$, $p_{.j}$ e p_{ij} definidos na secção 3.10, temos as seguintes propriedades :

1) $D = \Delta^t \Delta$ é simétrica :

$$D^t = (\Delta^t \Delta)^t = \Delta^t \Delta = D ;$$

2) Para todo vetor $\underline{x} \in \mathbb{R}^J$,

$$\underline{x}^t D \underline{x} = (\Delta \underline{x})^t (\Delta \underline{x}) = \underline{y}^t \underline{y} = \|\underline{y}\|_2^2 \geq 0, \text{ então}$$

D é Semi Definida Positiva (pelo menos) .

Ao calcularmos D_1 e D_2 para T_1 e T_2 podemos ter os seguintes casos :

(i) D_1 e D_2 semi-definidas positivas ;

(ii) D_1 definida positiva e D_2 semi-definida positiva ou
 D_2 definida positiva e D_1 semi-definida positiva e

(iii) D_1 e D_2 definidas positivas.

Sejam D_1 e D_2 matrizes $J \times J$ de dependência e codependência (que são reais, simétricas e ao menos semi-definidas positivas) com D_2 definida positiva. Nestas condições, de acordo com Rao [28] pg.41, existe $R_{J \times J}$ tal que

$$D_1 = (R^{-1})^t \Lambda R^{-1} \quad \text{e} \quad D_2 = (R^{-1})^t R^{-1} \quad (4.1.5)$$

isto é,

$$R^t D_1 R = \Lambda \quad \text{e} \quad R^t D_2 R = I, \quad (4.1.6)$$

onde Λ é a matriz diagonal dos auto-valores de $D_2^{-1} D_1$ e R_1 de R satisfaz a equação $D_1 R_1 = \lambda_1 D_2 R_1$. De acordo com estas igualdades temos o seguinte lema.

Lema 4.1 - Sejam D_1 e D_2 matrizes $J \times J$ de dependência e codependência com D_2 definida positiva. A matriz $D_2^{-1} D_1$ será semi-definida positiva ou definida positiva dependendo de assim o ser D_1 .

Prova : Seja dado o sistema de equações

$$D_2^{-1} D_1 X = \Lambda X \quad \text{então} \quad X^t D_2^{-1} D_1 X = X^t \Lambda X = X^t R^t D_1 R X \quad \text{por (4.1.6),}$$

fazendo $Y = RX$, então

$$X^t D_2^{-1} D_1 X = Y^t D_1 Y = Y^t \Lambda_1^t \Lambda_1 Y = Z^t Z = \|Z\|^2 \geq 0, \quad \text{com } Z = \Lambda_1 Y.$$

O lema 4.1 nos garante que os auto-valores de $D_2^{-1} D_1$ serão todos não negativos.

Dos argumentos acima surge a seguinte definição:

Definição 4.1: Com a maximização e minimização de $F(\underline{a})$ (4.1.3), definimos heurísticamente como critério de comparação entre D_1 e D_2

$$\xi = \frac{\lambda_{\max}}{\lambda_{\min}}, \quad \text{com } \lambda_{\min} > 0 \quad (4.1.7)$$

onde λ_{\max} e λ_{\min} são os autovalores máximo e mínimo de $D_2^{-1}D_1$ ou de $D_1^{-1}D_2$, dependendo para qual produto ξ for menor. Sob a hipótese nula $H_0 : D_1 = D_2$, $\xi \equiv 1$.

Para que as condições da definição 4.1 sejam garantidas devemos ter D_1 ou D_2 definidas positivas.

Em (4.1.7) $\xi \geq 1$, a igualdade só é atingida no caso de $D_1 = D_2$.

Um argumento que motivou a consideração de um critério de comparação entre grandezas de mesma escala baseado na razão destas (como o da definição 4.1) e não por exemplo, na diferença, pode ser ilustrado como segue :

sejam

$$\begin{array}{ll} x_1 = 0.0001 & x_3 = 100.000.000 \\ x_2 = 0.0002 & x_4 = 200.000.000 , \end{array}$$

então

$$\frac{x_2}{x_1} = \frac{x_4}{x_3} = 2, \text{ no entanto,}$$

$$x_2 - x_1 = 0.0001 \quad \text{e} \quad x_4 - x_3 = 100.000.000 ,$$

Observa-se que o critério da razão é invariante com respeito a ordens de grandeza.

Lema 4.2 - Seja $D = \Delta^t \Delta$ uma matriz de dependência e codependência de colunas. Se D provém de uma tabela de contingência em que pelo menos uma coluna é independente das linhas então o mínimo autovalor de D é nulo.

O lema é naturalmente válido ao substituirmos "colunas" por "linhas".

Prova : Suponha que $T_{I \times J}$ seja uma tabela de contingência cuja k -ésima coluna é independente das linhas. De acordo com (4.1.4) temos

$$\Delta = (\Delta_1 | \Delta_2 | \dots | \Delta_k | \dots | \Delta_J) \text{ com } \Delta_k = \underline{0}_{I \times 1},$$

então

$$\Delta^t \Delta = \begin{bmatrix} \Delta_1^t \Delta_1 & \Delta_1^t \Delta_2 \dots \Delta_1^t \Delta_{k-1} & 0 & \Delta_1^t \Delta_{k+1} \dots \Delta_1^t \Delta_J \\ \Delta_2^t \Delta_1 & \Delta_2^t \Delta_2 \dots \Delta_2^t \Delta_{k-1} & 0 & \Delta_2^t \Delta_{k+1} \dots \Delta_2^t \Delta_J \\ \vdots & \vdots & \vdots & \vdots \\ \Delta_{k-1}^t \Delta_1 & \Delta_{k-1}^t \Delta_2 \dots \Delta_{k-1}^t \Delta_{k-1} & 0 & \Delta_{k-1}^t \Delta_{k+1} \dots \Delta_{k-1}^t \Delta_J \\ 0 & 0 & \dots & 0 \\ \Delta_{k+1}^t \Delta_1 & \Delta_{k+1}^t \Delta_2 \dots \Delta_{k+1}^t \Delta_{k-1} & 0 & \Delta_{k+1}^t \Delta_{k+1} \dots \Delta_{k+1}^t \Delta_J \\ \vdots & \vdots & \vdots & \vdots \\ \Delta_J^t \Delta_1 & \Delta_J^t \Delta_2 \dots \Delta_J^t \Delta_{k-1} & 0 & \Delta_J^t \Delta_{k+1} \dots \Delta_J^t \Delta_J \end{bmatrix}$$

ou seja, quando $\Delta_k = \underline{0}$ então a k -ésima linha e k -ésima coluna de $\Delta^t \Delta$ serão nulas implicando em $\det(\Delta^t \Delta) = 0$. Então

$$0 = \det(\Delta^t \Delta) = \det(\Gamma^t \text{diag}(\lambda_1, \dots, \lambda_J) \Gamma) = \prod_{i=1}^J \lambda_i,$$

sendo $\Gamma^t \Gamma = I_J$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_J$, e $\lambda_i \geq 0$ para todo $i \leq J$, pois $\Delta^t \Delta$ é simétrica. Então existe pelo menos um auto-valor nulo de $D = \Delta^t \Delta$ implicando em $\lambda_J = 0$.

Observações:

1) - Note que $\lambda_i > 0$ para todo $i=1,\dots,J$ implica em $\Delta_i \neq 0$ para todo i e $\Delta_i \neq \Delta_j$ para todo $i \neq j$, ou seja, $D = \Delta^t \Delta$ é definida positiva. Nestas condições, T é uma tabela de contingência irredutível em que todas as linhas são dependentes de todas as colunas.

2) - De acordo com 1) a recíproca do lema não é verdadeira pois $\det(\Delta^t \Delta) = 0$ pode implicar na existência de pelo menos duas colunas (ou linhas) idênticas e não nulas em T , que podem ser aglutinadas. O determinante da matriz de dependência e codependência resultante da aglutinação pode ser não nulo implicando em dependência entre todas as linhas e todas as colunas da tabela de contingência reduzida.

3) - Como consequência do lema 4.1, se todas as linhas forem independentes de todas as colunas então $\Delta = (Q_1 | \dots | Q_k | \dots | Q)$ de modo que todos os auto-valores de $\Delta^t \Delta$ são nulos.

4) - Quando houver "colinearidade geodésica" entre os perfis-coluna (ou perfis-linha) de T então $\det(\Delta^t \Delta) = 0$.

Do lema 4.2 e suas observações, concluímos que o critério de comparação ξ para matrizes de dependência e codependência D_1 e D_2 (vide definição 4.1) deve ser utilizado somente para o caso em que exista associação entre todas as linhas e todas as colunas de T_1 e T_2 , ou no caso de existirem linhas ou colunas em T_1 e T_2 que sejam independentes, pudermos reduzir T_1 e T_2 a tabelas T'_1 e T'_2 de mesma ordem. Em outros casos podemos ter indeterminações em ξ .

Na tentativa de encontrar um melhor critério de comparação entre D_1 e D_2 , menos restritivo e que não tivesse o problema da indeterminação de ξ propomos o seguinte critério :

$$\xi = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max}} = 1 - \frac{\lambda_{\min}}{\lambda_{\max}} \quad (4.1.8)$$

Como já mencionamos, ao calcularmos D_1 e D_2 para T_1 e T_2 podemos ter os seguintes casos :

- (i) D_1 e D_2 semi-definidas positivas ;
- (ii) D_1 definida positiva e D_2 semi-definida positiva ou
 D_2 definida positiva e D_1 semi-definida positiva e
- (iii) D_1 e D_2 definidas positivas.

Para este novo critério, só o caso da independência total entre linhas e colunas nas tabelas de contingência T_1 ou T_2 trará o problema da indeterminação de ξ . No entanto, nestas circunstâncias, definiremos ξ igualando a unidade.

No caso da independência local de linhas ou colunas de T_i , $i=1,2$, podemos avaliar ξ quando D_1 ou D_2 for definida positiva, isto é, no caso (ii).

Se D_1 for definida positiva podemos encontrar os auto-valores máximo e mínimo de $D_1^{-1}D_2$ e de maneira análoga, no caso de D_2 ser definida positiva, os auto-valores máximo e mínimo de $D_2^{-1}D_1$. Sabemos de antemão (vide lema 4.1) que $\lambda_{\min} = 0$ e desta forma $\xi = 1$.

O caso (i) não pode ser prontamente avaliado pois D_1 e D_2 são não inversíveis.

Por (4.1.8)

$$0 \leq \xi \leq 1,$$

$\xi = 0$, isto é, $\lambda_{\max} = \lambda_{\min}$, pode ocorrer sómente no caso (iii) indicando antes de mais nada que as linhas e colunas das tabelas T_1 e T_2 são totalmente associadas. Naturalmente quando $\lambda_{\max} = \lambda_{\min}$ teremos $D_1 = D_2$.

É possível que $\xi = 1$ ocorra no caso (iii), além de, obviamente, nos casos (i) e (ii). No caso (iii) $\xi = 1$ poderia indicar que o auto-valor máximo (λ_{\max}) de $D_2^{-1}D_1$ (ou de $D_1^{-1}D_2$) fosse arbitrariamente grande em relação a λ_{\min} .

Resultados assim ocorreriam se D_1 (ou D_2) tivesse determinante muito próximo de zero, isto é, quando as associações entre as linhas e colunas de T_1 (e de T_2) fossem muito fracas.

Como nosso objetivo é estudar a distribuição nula de ξ e o caso (ii) implica em modos de associação distintos em T_1 e T_2 , não o avaliaremos. Quando nos referirmos à distribuição nula de ξ nas próximas secções, estaremos tratando do caso (iii).

Com resultados de simulação pretendemos observar sob a hipótese nula de que D_1 seja igual a D_2 a forma da distribuição de ξ sob vários tipos de associação entre linhas e colunas de T_1 e T_2 .

4.2 Geração de Matrizes de Dependência e Codependência para o cálculo do critério de comparação ξ

Como foi mencionado na secção 4.1, o objetivo principal deste trabalho é o de investigar a viabilidade da construção de uma tabela para a distribuição de ξ que possibilite testar a proximidade entre duas matrizes de Dependência e Codependência. Geramos, então, duas tabelas de contingência com a mesma estrutura de associação entre linhas e colunas e tamanho amostral 500.

Para cada par de tabelas de contingência geradas calculamos as respectivas matrizes de Dependência e Codependência e posteriormente o critério de comparação (4.1.8) de ξ . O procedimento foi repetido 500 vezes e construímos distribuições de frequência para ξ .

Na geração de números pseudo-aleatórios para a construção das tabelas de contingência, utilizamos o gerador G05CAF da Biblioteca de Algoritmos NAG. Segundo Ripley (1989) [30], entre os geradores de números aleatórios congruenciais, tais como o RANDU, o gerador do produto GLIM, ou do IMSL, o gerador G05CAF da NAG é o mais aceitável, tendo em vista seu período da ordem de 2^{59} .

De acordo com o mesmo autor, um gerador é recomendável quando seu período for superior a 2^{32} . Tendo em vista o que diz Ripley a respeito do gerador G05CAF da NAG, pode-se assegurar a validade dos resultados a serem apresentados. Contudo, verificamos a equi-distribuição dos números gerados utilizando um teste proposto em 1989 por Kalos e Whitlock [18]. O resultado nos pareceu satisfatório (vide apêndice).

4.2.1 Apresentação do Método utilizado para a geração de Tabelas de Contingência

Já que o tamanho da amostra é dado, estamos no caso em que \underline{X} é uma variável aleatória com densidade Multinomial de parâmetros n e $\underline{P} = \{p_{11}, \dots, p_{IJ}\}$, isto é, $\underline{X} \approx M(n, \underline{P})$, onde $\sum_{i=1}^I \sum_{j=1}^J p_{ij} = 1$ (vide capítulo 1).

O método convencional para gerar tabelas de contingência $I \times J$, segundo a densidade Multinomial $M(n, \underline{P})$, consiste dos seguintes procedimentos :

a) Construir a distribuição acumulada de $M(n, \underline{p})$, ou seja,

$$F = (\phi_{11}, \dots, \phi_{IJ}) \text{ com } \phi_{kj} = \sum_{r=1}^k \sum_{s=1}^J p_{rs}, \quad 1 \leq k \leq I \text{ e } 1 \leq j \leq J.$$

b) Para cada número gerado G_s , $s = 1, \dots, n$, com distribuição uniforme $(0,1)$, é verificado qual das seguintes desigualdades é verdadeira :

$$\begin{aligned} 0 &< G_s \leq \phi_{11}, \text{ ou} \\ \phi_{11} &< G_s \leq \phi_{12}, \text{ ou} \\ &\vdots \\ \phi_{I-1, J-1} &< G_s \leq \phi_{IJ}; \end{aligned}$$

Ao intervalo ao qual G_s pertence acumula-se uma unidade.

Este método, segundo Kemp e Kemp [19] em 1987, é muito demorado, razão pela qual utilizaremos o Método de Geração Rápida de tabelas de frequência, sugerido por estes mesmos autores.

Resumidamente, este método consiste em ordenar as probabilidades de uma distribuição multinomial e com distribuição definida em ordem decrescente de magnitude e gerar uma amostra pseudo-aleatória através distribuições Binomiais condicionadas. A razão pela qual é feita a ordenação é de que se pode minimizar o número de chamadas ao gerador.

A afirmação que fizemos de que uma amostra com densidade Multinomial pode ser gerada através de densidades Binomiais, pode ser verificada de acordo com o teorema seguinte.

Teorema 4.2.1 Seja $\underline{X} = \{X_1, X_2, \dots, X_k\}$ uma variável aleatória com densidade Multinomial de parâmetros n e $\underline{P} = \{p_1, \dots, p_k\}$ isto é, $\underline{X} \approx M(n, \underline{P})$. A densidade de X_1 é Binomial com parâmetros n e p_1 , isto é, $X_1 \approx B(n, p_1)$.

Seja $X_1 = x_1$,

então a densidade de X_2 condicionada a $X_1 = x_1$, $X_2|X_1$, é

$$B(n - x_1, p_2 / (1 - p_1)).$$

De maneira análoga, a densidade de $X_3|(X_2 = x_2, X_1 = x_1)$, é

$$B(n - x_2 - x_1, p_3 / (1 - p_2 - p_1)),$$

portanto, a densidade de

$$X_k|(X_{k-1} = x_{k-1}, \dots, X_1 = x_1) \text{ é } B\left(n - \sum_{i=1}^{k-1} x_i, p_k / \left\{1 - \sum_{i=1}^{k-1} p_i\right\}\right).$$

Verifica-se que a densidade conjunta de

$$\{X_1 = x_1, X_2|X_1 = x_1, \dots, X_k|X_{k-1} = x_{k-1}, \dots, X_1 = x_1\},$$

é uma densidade Multinomial $M(n, p_1, \dots, p_k)$.

A prova do teorema acima é encontrada no Apêndice.

Para gerarmos (de acordo com o teorema acima) o resultado de um experimento pseudo-aleatório Multinomial, isto é, gerarmos os valores de $X_1=x_1$, onde $X = (X_1, \dots, X_k)$ é $M(n, p_1, \dots, p_k)$, os seguintes passos devem ser seguidos :

1 - Faça $p_1 \geq p_2 \geq \dots \geq p_k$.

2 - Gerar n números pseudo-aleatórios com distribuição Uniforme $(0,1)$ e contar quantos dos n números gerados é menor ou igual a p_1 .

3 - Digamos que x_1 dos n números gerados foram menores ou iguais a p_1 , isto é, obtivemos $X_1 = x_1$. Devemos gerar uma nova amostra de tamanho $n - x_1$ de números pseudo-aleatórios com distribuição uniforme $(0,1)$ e contar quantos destes são menores ou iguais a $p_2/(1 - p_1)$. Digamos que obtivemos $(X_2 | X_1=x_1) = x_2$ valores.

4 - Devemos gerar outra amostra de tamanho $n - x_1 - x_2$ e contar quantos destes são menores ou iguais a $p_3/(1 - p_1 - p_2)$. Digamos que obtivemos $(X_3 | X_1=x_1, X_2=x_2) = x_3$ valores. Repetiremos indutivamente estes procedimentos até obtermos $(X_{k-1} | (X_1=x_1, \dots, X_{k-2}=x_{k-2})) = x_{k-1}$.

5 - Finalmente, o valor de $X_k | (X_1=x_1, \dots, X_{k-1}=x_{k-1}) = n - \sum_{i=1}^{k-1} x_i$.

A seguir, descreveremos os resultados obtidos nas simulações.

4.3 RESULTADOS

Como foi mencionado na secção 4.2, calculamos o critério de comparação ξ , vide (4.1.8), a partir da geração de tabelas de classificação cruzada sob a hipótese nula em (4.1.1) e algumas especificações : dimensão das tabelas T_1 e T_2 (vide secção 4.1) que denominaremos por DIM e tipo de associação entre linhas e colunas de T_1 e T_2 que denominaremos por ASSOCIAÇÕES.

Para as dimensões 2x2 e 3x3 trabalhamos com 12 tipos de associações entre as linhas e colunas de T_1 e T_2 e com 8 tipos para a dimensão 3x2.

Ao longo do trabalho utilizamos uma única semente : 23, escolhida ao acaso.

Para cada par (DIM,ASSOCIAÇÃO) fixamos em 500 o tamanho amostral para a geração das tabelas T_1 e T_2 e fizemos 500 simulações de Monte Carlo. Com os 500 valores gerados de ξ construimos sua distribuição de frequências relativas.

Para cada dimensão apresentaremos os Diagramas de Linha de Frequências Relativas de ξ mais representativos. Estes diagramas são oriundos de associações (entre as linhas e colunas de T_1 e T_2) que variam de muito fortes a muito fracas. Para os diagramas escolhidos, isto é, para as associações escolhidas, utilizamos 500 valores gerados de ξ para ajustar distribuições Beta.

A princípio escolhemos estas distribuições ao observarmos as diferentes formas que os Diagramas de Linha de Frequências Relativas de ξ apresentaram. Os diagramas são todos unimodais e além disto ξ pertence ao intervalo fechado $[0,1]$; isto resulta em uma boa indicação de que a distribuição de ξ poderia ajustar-se a uma distribuição Beta.

Os ajustes às distribuições Beta foram efetuados com o auxílio do procedimento **CAPABILITY** do produto **SAS**.

Os resultados das simulações, diagramas de linha, testes de ajustes e gráficos superpostos das distribuições empírica e ajustada de ξ podem ser consultados de acordo com a tabela a seguir.

Tabela 4.0 - Organização dos resultados

DIM	ASSOCIAÇ.	ESTAT.BÁSICAS	DIAGRAMAS	AJUSTES	GRÁFICOS
2x2	tabela 4.1 pg. 108	tabela 4.4 pg. 111	pg. 114	tab. 4.7 pg. 117	pg. 118
3x2	tabela 4.2 pg. 109	tabela 4.5 pg. 112	pg. 115	tab. 4.8 pg. 117	pg. 119
3x3	tabela 4.3 pg. 110	tabela 4.6 pg. 113	pg. 116	tab. 4.9 pg. 117	pg. 120

Ao falarmos nas associações envolvendo as tabelas de contingência T_1 e T_2 , temos implicitamente dois *efeitos* a considerar : o efeito *entre* as tabelas e o efeito *dentro* das tabelas. O efeito *entre* foi anulado ao simularmos T_1 e T_2 sob a hipótese H_0 em (4.1.1). O efeito *dentro* é caracterizado pelas ASSOCIAÇÕES entre as linhas e colunas de T_1 e T_2 que mencionamos no início desta secção; este efeito pode variar de fraco a forte, dependendo das relações entre as linhas e colunas de T_1 e T_2 assim o serem. Deste modo, quando nos referirmos a "efeito dentro" estaremos falando das ASSOCIAÇÕES entre as linhas e colunas de T_1 e T_2 .

Para cada dimensão podemos separar quanto à *forma* as distribuições de frequências geradas de ξ em três grupos: distribuições assimétricas à direita, distribuições simétricas e distribuições assimétricas à esquerda. Na tabela 4.0.1 relacionamos as Associações que trabalhamos (vide tabelas 4.1 a 4.3) com a forma das distribuições e as dimensões.

Tabela 4.0.1 Forma da distribuição de ξ em função da dimensão e tipo de associação entre linhas e colunas de T_1 e T_2

Distribuições		Assim. direita	Simétricas	Assim. esquerda
Dimensão	2x2	1,2,7,11,12	3,5,6,8,10	4,9
	3x2	1	2,3,5,6,7	4,8
	3x3	1,2,5	3,6,7,8,9,10,12	4,11

* vide tabelas 4.1 a 4.3.

Para cada uma das dimensões, os três grupos que descrevemos na tabela acima se separam de acordo com a "distância" das proporções nas diversas associações com respeito às proporções que teriam sob o modelo de independência. As associações fortes por exemplo, discordam radicalmente de seu modelo independência.

No caso da dimensão 3x3 por exemplo, as distribuições simétricas na tabela 4.0.1 englobam as associações moderadas do tipo homogeneidade de marginais e simetria. As distribuições de ξ assimétricas à direita relacionam associações fortes enquanto as distribuições assimétricas à esquerda, associações fracas. Este comportamento é comum a todas as dimensões em estudo, de forma que a dimensão não parece influenciar a forma da distribuição de ξ .

Podemos concluir disto que a força da associação entre as linhas e colunas de T_1 e T_2 e não a forma (homogeneidade de marginais, simetria, etc) da associação é o fator essencial para a determinação dos grupos na tabela 4.0.1. Desta maneira observamos que o critério de comparação ξ é afetado pela força da associação, isto é, pela intensidade do efeito *dentro* (vide Diagramas de Linha pgs. 114 a 117).

Nos casos de associações fracas, as distribuições de ξ tiveram valor modal um, enquanto nos casos de associações fortes, o valor modal das distribuições de ξ foi zero. Estes resultados contrariam nossas expectativas quanto a estatística em questão pois os valores esperados de ξ em cada um dos casos é $\xi = 0$, levando em conta que as distribuições foram geradas sob a hipótese $H_0 : D_1 = D_2$.

A forma geral da densidade da distribuição Beta, ou ainda, a distribuição Beta do primeiro tipo, segundo Johnson e Kotz [17] é dada por :

$$f(x) = c(x+A_1)^{(A_1+a)/b2(A_1+A_2)}(A_2-x)^{(A_2-a)/b2(A_1+A_2)}$$

onde c é escolhido de modo que $\int f(x)dx = 1$ e $-A_1 < x < A_2$.

De acordo com nossos resultados, temos que $0 \leq x \leq 1$, de modo que podemos fixar $A_1 = 0$ e $A_2 = 1$ e utilizar a forma padronizada da distribuição Beta para ajustar a distribuição de ξ , isto é :

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

com $\alpha, \beta > 0$, $0 < x < 1$ e $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$.

Para as três dimensões : 2x2, 3x2 e 3x3, para que fosse possível procedermos a ajustes utilizando o procedimento **Capability** do produto **SAS** para ajustar distribuições Beta padronizadas, substituímos os valores 1 de ξ (mais frequentes em associações fracas) por 0.999 e os valores 0 de ξ , que foram mais frequentes em associações fortes, por 0.001. Desta maneira contornamos o problema de necessitarmos conhecer previamente os parâmetros da distribuição a ser ajustada.

A seguir temos os valores ajustados para os parâmetros α e β da distribuição Beta padronizada que obtivemos para a distribuição de ξ segundo as associações 1.1 a 4.1, 1.2 a 4.2 e 1.3 a 4.3 (vide tabelas 4.1 a 4.3, 4.7 a 4.8 e os gráficos nas três últimas páginas deste capítulo) :

Tabela 4.0.1 - Valores ajustados para os parâmetros α e β da distribuição Beta padronizada

Associação	α	β
1.1	1.60	51.60
2.1	2.39	8.46
3.1	2.66	2.33
4.1	2.97	1.30
1.2	2.59	37.30
2.2	12.90	86.00
3.2	2.98	3.49
4.2	2.55	0.52
1.3	3.04	23.40
2.3	2.11	10.80
3.3	7.53	6.12
4.3	5.22	0.70

Para distribuições assimétricas à direita (associações fortes) a distribuição Beta ajustada aos valores gerados de ξ , a primeira vista, parecem não resultar em bons ajustes (vide tabelas 4.7 a 4.9). Devido ao reduzido número de simulações de Monte Carlo que empreendemos, apenas 500, no momento não podemos afirmar que a distribuição Beta não seja realmente adequada para ajustar as distribuições de ξ neste caso. Para termos uma idéia mais clara do problema deveremos aumentar

sensivelmente o número de simulações de Monte Carlo. Além disto, é necessário que determinemos a distribuição teórica de ξ , que no momento, estamos inclinados a pensar que se trate uma distribuição Beta. Esta tarefa exige pesquisas que não pudemos empreender por motivos diversos, mas que certamente será mais um passo para a elucidação de nossa pesquisa.

Considerando as questões expostas, concluímos que o critério ξ para a comparação das matrizes D_1 e D_2 de Dependência e Codependência com o intuito descrito no início deste capítulo não é afetado pelas dimensões das tabelas de contingência T_1 e T_2 , no entanto, ainda que sob o mesmo tipo de associação entre as linhas e colunas destas duas sub-tabelas, a distribuição de ξ é descrita diferentemente nos casos de associações forte, moderada ou fraca entre linhas e colunas de T_1 e T_2 , isto é, ξ depende do tipo de associação entre as linhas e colunas de T_1 e T_2 . Até o presente momento não temos idéia de como "retirar" este "efeito" de associação de modo a tornar o uso desta estatística viável do ponto de vista prático.

Tabela 4.1 de Associações

1.1 : Assoc. Total
não Balanceada

0.80	0.00	0.80
0.00	0.20	0.20
0.80	0.20	1.00

2.1 : Simetria -
Assoc. Forte

0.45	0.03	0.48
0.03	0.49	0.52
0.48	0.52	1.00

3.1 : Assoc.
Moderada

0.18	0.34	0.52
0.30	0.18	0.48
0.48	0.52	1.00

4.1 : Simetria -
Assoc. Fraca

0.19	0.29	0.48
0.29	0.23	0.52
0.48	0.52	1.00

5.1 : Assoc.
Moderada

0.16	0.36	0.52
0.32	0.36	0.48
0.48	0.52	1.00

6.1 : Assoc.
Moderada

0.35	0.17	0.52
0.13	0.35	0.48
0.48	0.52	1.00

7.1 : Assoc.
Moderada

0.45	0.07	0.52
0.03	0.45	0.48
0.48	0.52	1.00

8.1 : Simetria -
Assoc. Moderada

0.10	0.38	0.48
0.38	0.14	0.52
0.48	0.52	1.00

9.1 : Simetria -
Assoc. Fraca

0.28	0.20	0.48
0.20	0.32	0.52
0.48	0.52	1.00

10.1 : Simetria -
Assoc.

0.17	0.31	0.48
0.31	0.21	0.52
0.48	0.52	1.00

11.1 : Assoc. Total
não Balanceada

0.20	0.00	0.20
0.00	0.80	0.80
0.20	0.80	1.00

12.1 : Assoc. Total
Balanceada

0.50	0.00	0.50
0.00	0.50	0.50
0.50	0.50	1.00

Tabela 4.2 de Associações

1.2 : Assoc. Forte			2.2 : Assoc. Moderada			3.2 : Assoc. Moderada		
0.00	0.40	0.40	0.40	0.00	0.40	0.08	0.32	0.40
0.30	0.00	0.30	0.00	0.30	0.30	0.24	0.06	0.30
0.00	0.30	0.30	0.30	0.00	0.30	0.28	0.02	0.30
0.30	0.70	1.00	0.70	0.30	1.00	0.60	0.40	1.00
4.2 : Assoc. Fraca			5.2 : Assoc. Moderada			6.2 : Assoc. Moderada		
0.21	0.19	0.40	0.40	0.00	0.40	0.30	0.10	0.40
0.20	0.10	0.30	0.10	0.20	0.30	0.02	0.28	0.30
0.19	0.11	0.30	0.10	0.20	0.30	0.28	0.02	0.30
0.60	0.40	1.00	0.60	0.40	1.00	0.60	0.40	1.00
7.2 : Assoc. Moderada			8.2 : Assoc. Fraca					
0.10	0.30	0.40	0.26	0.14	0.40			
0.28	0.02	0.30	0.20	0.10	0.30			
0.22	0.08	0.30	0.14	0.16	0.30			
0.60	0.40	1.00	0.60	0.40	1.00			

Tabela 4.3 de Associações

1.3 : Assoc. Total Balanceada				2.3 : Assoc. Total não Balanceada				3.3 : Simetria			
0.33	0.00	0.00	0.33	0.10	0.00	0.00	0.10	0.03	0.20	0.07	0.30
0.00	0.33	0.00	0.33	0.00	0.20	0.00	0.20	0.20	0.05	0.08	0.33
0.00	0.00	0.34	0.34	0.00	0.00	0.70	0.70	0.07	0.08	0.22	0.37
0.33	0.33	0.34	1.00	0.10	0.20	0.70	1.00	0.30	0.33	0.37	1.00
4.3 : Assoc. próxima da independência				5.3 : Assoc. Total não Balanceada				6.3 : Simetria			
0.10	0.10	0.15	0.35	0.70	0.00	0.00	0.70	0.25	0.03	0.02	0.30
0.05	0.13	0.12	0.30	0.00	0.20	0.00	0.20	0.03	0.18	0.12	0.33
0.15	0.10	0.10	0.35	0.00	0.00	0.10	0.10	0.02	0.12	0.23	0.37
0.30	0.33	0.37	1.00	0.70	0.20	0.10	1.00	0.30	0.33	0.37	1.00
7.3 : Homogeneidade de Marginais				8.3 : Simetria				9.3 : Assoc. Moderada			
0.05	0.05	0.20	0.30	0.14	0.13	0.03	0.30	0.13	0.15	0.07	0.35
0.05	0.18	0.10	0.33	0.13	0.05	0.15	0.33	0.02	0.12	0.16	0.30
0.20	0.10	0.07	0.37	0.03	0.15	0.19	0.37	0.15	0.06	0.14	0.35
0.30	0.33	0.37	1.00	0.30	0.33	0.37	1.00	0.30	0.33	0.37	1.00
10.3 : Assoc. Moderada				11.3 : Assoc. Moderada				12.3 : Simetria com Homog. Marginais			
0.06	0.20	0.09	0.35	0.03	0.17	0.15	0.35	0.15	0.15	0.05	0.35
0.17	0.05	0.08	0.30	0.18	0.05	0.07	0.30	0.15	0.00	0.15	0.30
0.07	0.08	0.20	0.35	0.09	0.11	0.15	0.35	0.05	0.15	0.15	0.35
0.30	0.33	0.37	1.00	0.30	0.33	0.37	1.00	0.35	0.30	0.35	1.00

Tabela 4.4

Estatísticas Básicas para $\xi = 1 - \lambda_{\min}/\lambda_{\max}$

Semente da geracao : 23

Dimensão da tabela : 2 x 2

T. assoc.	minimo	maximo	media	variância
1.1	0.0000	0.1138	0.0319	0.0005
2.1	0.0190	0.5524	0.2207	0.0130
3.1	0.0549	0.9386	0.5304	0.0365
4.1	0.1111	0.9947	0.7126	0.0402
5.1	0.0593	0.8984	0.4582	0.0329
6.1	0.0294	0.8445	0.4251	0.0328
7.1	0.0000	0.6165	0.2389	0.0166
8.1	0.0093	0.7355	0.3363	0.0251
9.1	0.0635	0.9821	0.6424	0.0431
10.1	0.0598	0.9549	0.5676	0.0441
11.1	0.0000	0.1174	0.0286	0.0005
12.1	0.0000	0.1336	0.0292	0.0005

Tabela 4.5

Estatísticas Básicas para $\xi = 1 - \lambda_{\min}/\lambda_{\max}$

Semente da geracao : 23

Dimensão da tabela : 3 x 2

T. Assoc.	minimo	maximo	media	variância
1.2	0.0018	0.2368	0.0707	0.0024
2.2	0.2545	0.4156	0.3166	0.0006
3.2	0.0180	0.9196	0.4666	0.0378
4.2	0.2067	1.0000	0.8438	0.0287
5.2	0.0178	0.8161	0.3219	0.0249
6.2	0.0198	0.9556	0.5053	0.0359
7.2	0.0029	0.7882	0.3401	0.0246
8.2	0.1962	1.0000	0.9594	0.0131

Tabela 4.6

Estatísticas Básicas para $\xi = 1 - \lambda_{\min}/\lambda_{\max}$

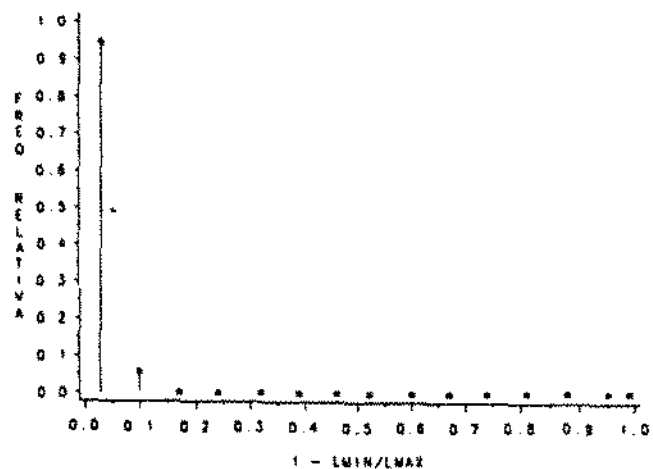
Semente da geracao : 23

Dimensão da tabela : 3 x 3

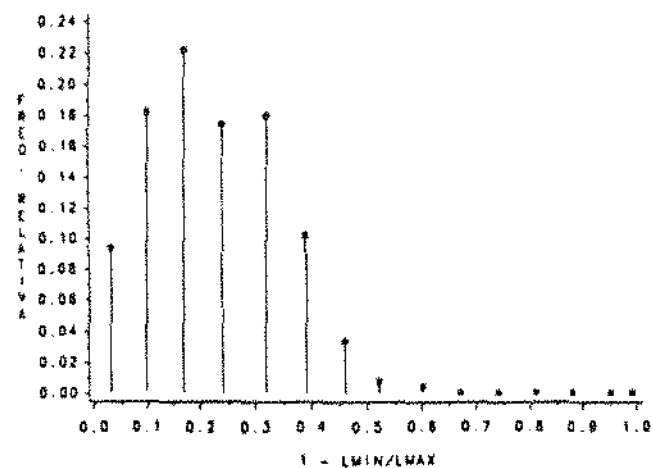
T. Assoc.	minimo	maximo	media	variância
1	0.0070	0.3178	0.1177	0.0037
2	0.0107	0.4535	0.1567	0.0083
3	0.1456	0.8624	0.5381	0.0187
4	0.3682	1.0000	0.8731	0.0147
5	0.0082	0.5219	0.1601	0.0099
6	0.0625	0.9262	0.5959	0.0238
7	0.0962	0.9118	0.5699	0.0173
8	0.1800	0.9166	0.5948	0.0183
9	0.2450	0.9624	0.6579	0.0187
10	0.1771	0.9325	0.6058	0.0213
11	0.3612	1.0000	0.8852	0.0187
12	0.1280	0.8835	0.5057	0.0208

Tam. Amostra = 500

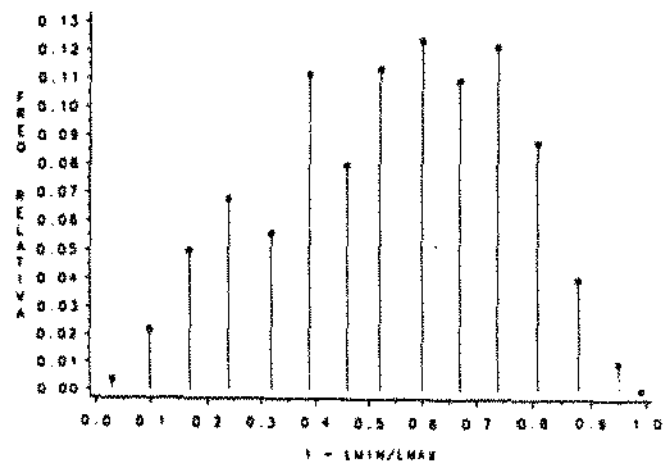
ASSOCIACAO 1.1



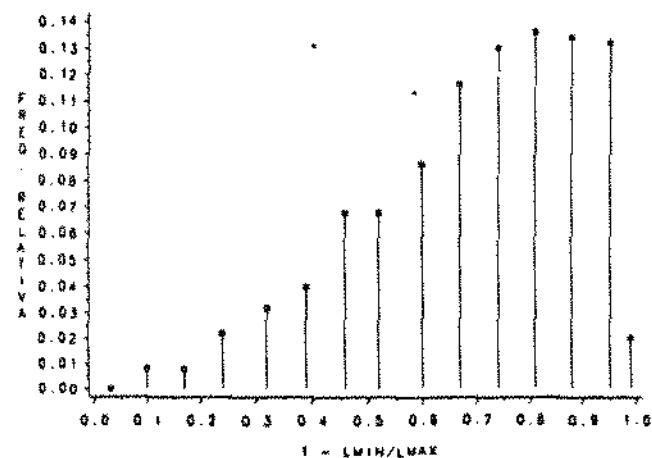
ASSOCIACAO 2.1



ASSOCIACAO 3.1

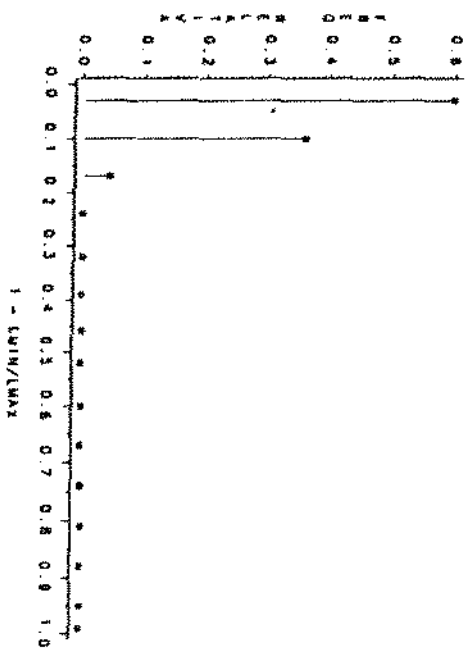


ASSOCIACAO 4.1

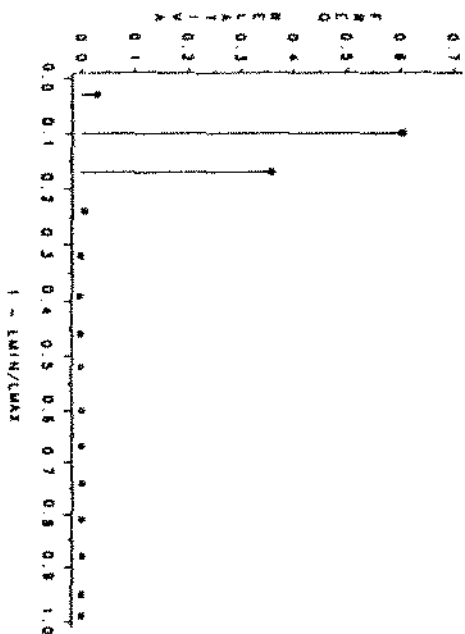


Tam. Amostra = 500

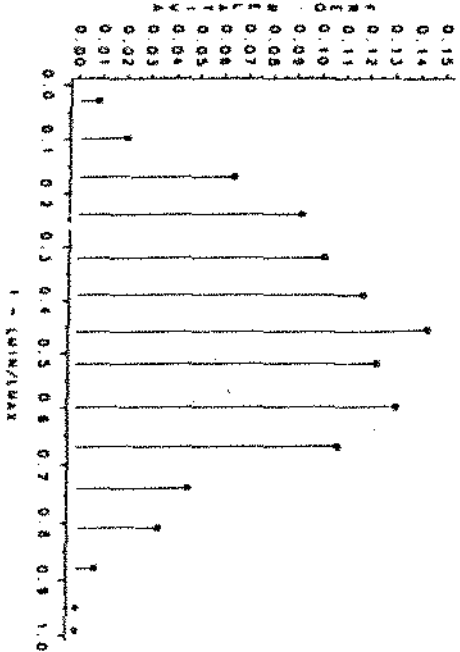
ASSOCIACAO 1.2



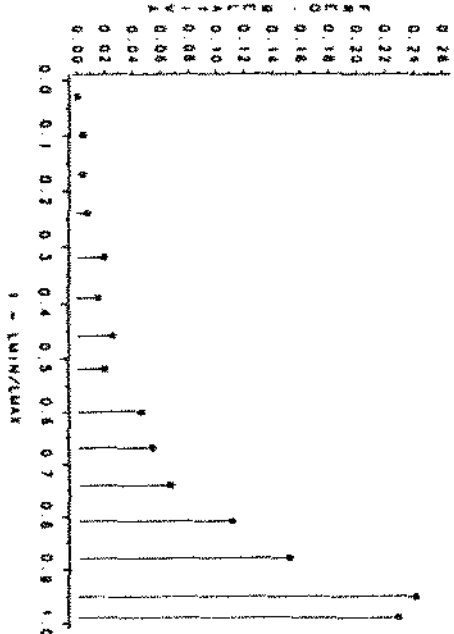
ASSOCIACAO 2.2



ASSOCIACAO 3.2

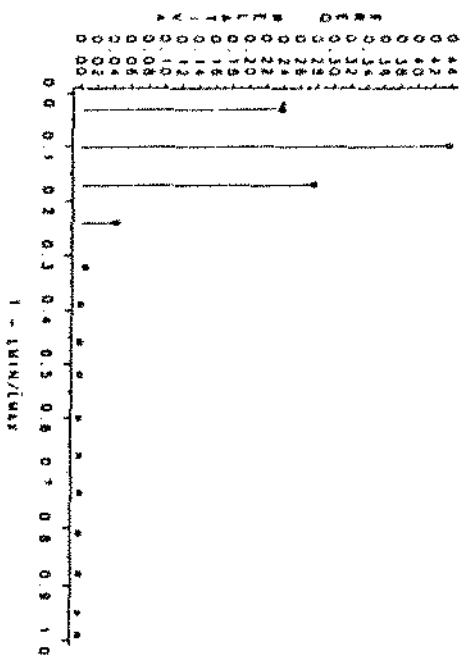


ASSOCIACAO 4.2

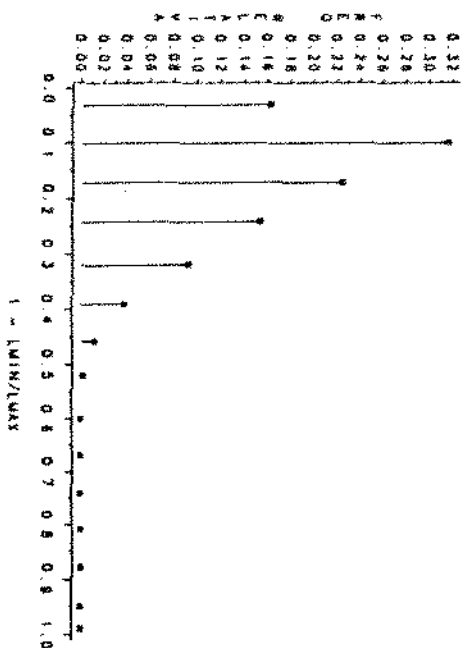


Tam. Amostra = 500

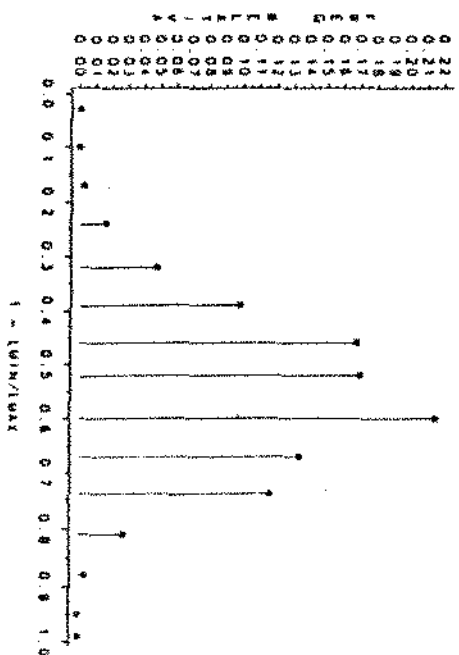
ASSOCIACAO 1.3



ASSOCIACAO 2.3



ASSOCIACAO 3.3



ASSOCIACAO 4.3

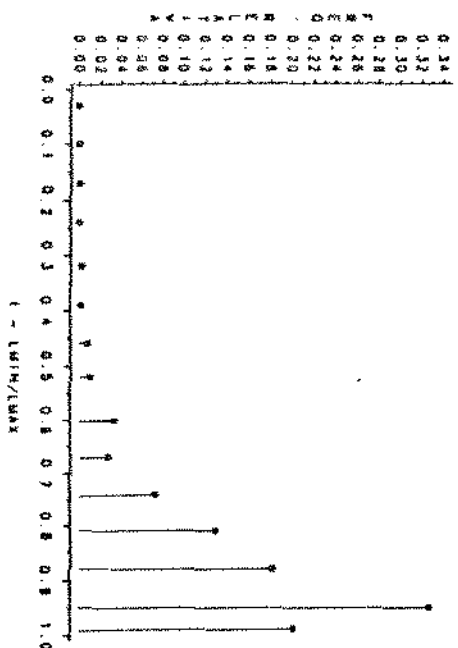


Tabela 4.7 Ajustes de curvas Beta padronizada : dimensão 2x2

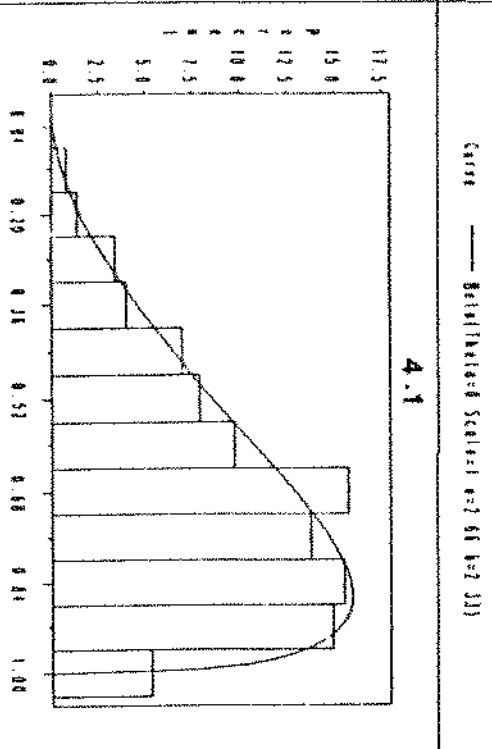
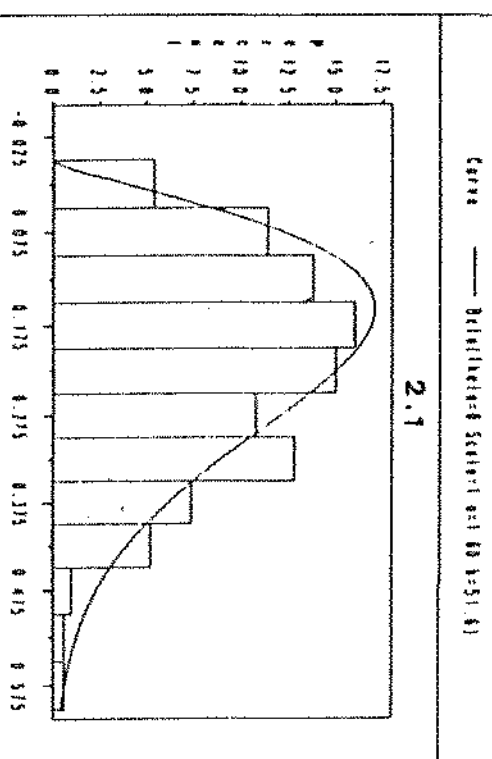
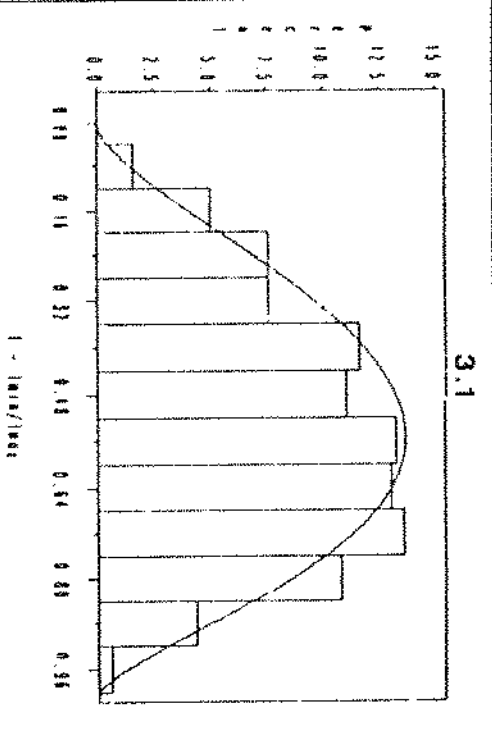
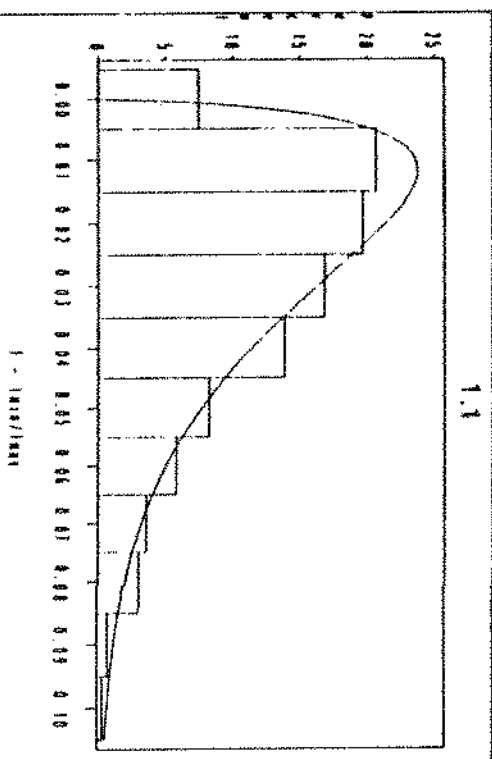
Associação	Chi-Quadrado	G.L.	Pr. > Chi-Quadrado
1.1	9.4224724	7	0.2237298
2.1	22.4204073	8	0.0041938
3.1	13.5400001	8	0.0945714
4.1	5.5980952	8	0.6921493

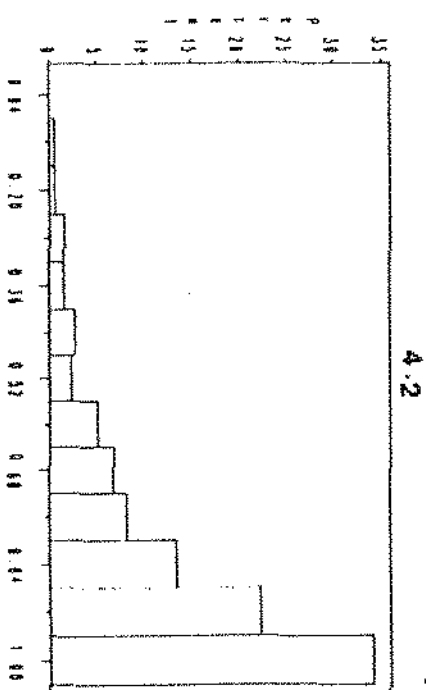
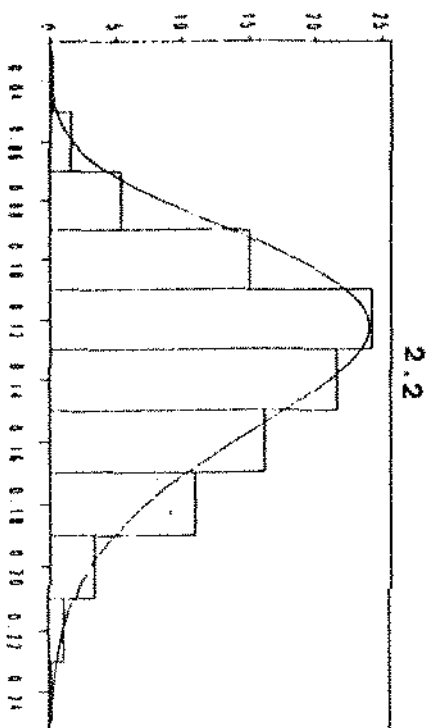
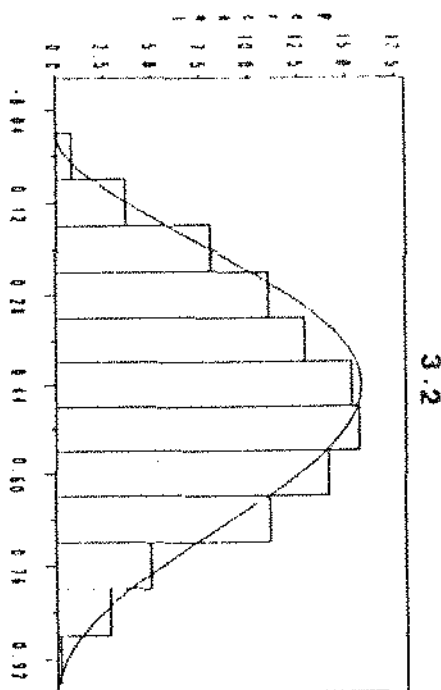
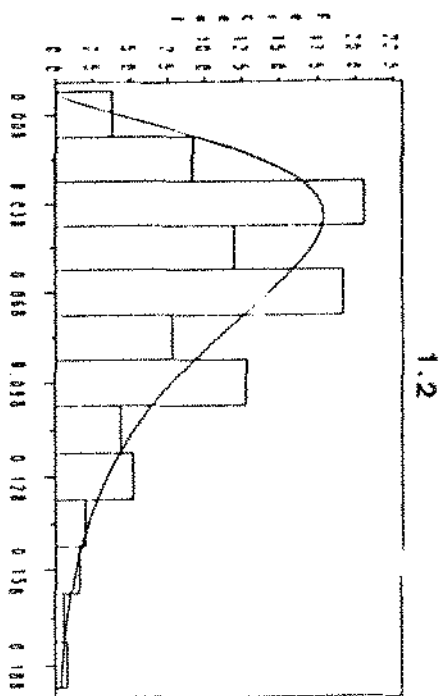
Tabela 4.7 Ajustes de curvas Beta padronizada : dimensão 3x2

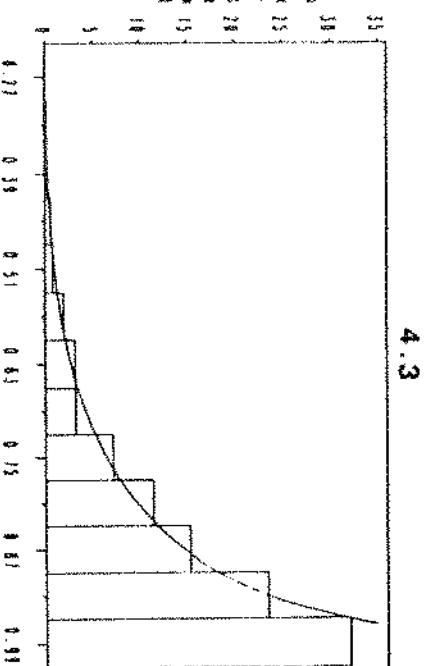
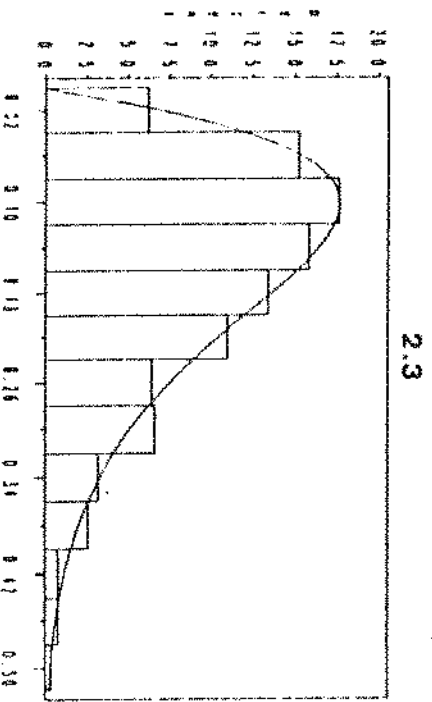
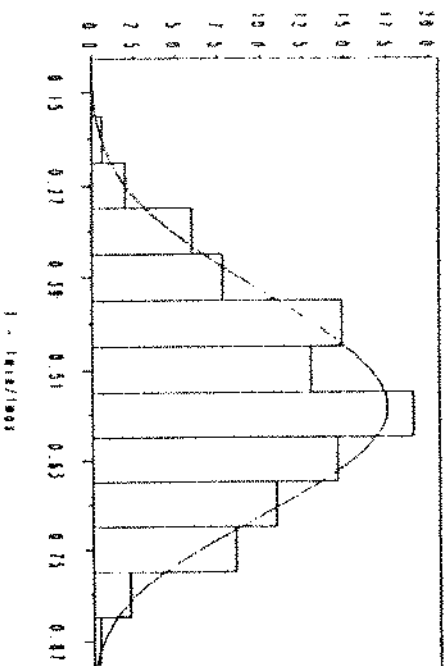
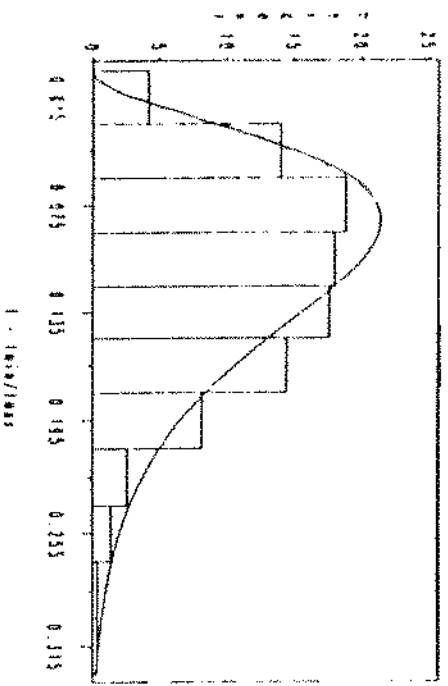
Associação	Chi-Quadrado	G.L.	Pr. > Chi-Quadrado
1.2	52.2032298	9	4.1374E-8
2.2	20.5383422	7	0.00451703
3.2	6.3818396	8	0.60454168
4.2	6.7028858	8	0.56900176

Tabela 4.7 Ajustes de curvas Beta padronizada : dimensão 3x3

Associação	Chi-Quadrado	G.L.	Pr. > Chi-Quadrado
1.3	16.9976355	7	0.01741144
2.3	4.8916121	9	0.84365245
3.3	11.8856765	8	0.15637809
4.3	4.6928763	8	0.78984067







Apêndice I

De acordo com Rao [28] pg. 50, a maximização ou minimização da forma quadrática $\underline{x}^t A \underline{x}$, onde $A_{(m \times m)}$ é simétrica, e \underline{x} é um vetor unitário, é dada por :

$$\sup_{\underline{x}} \frac{\underline{x}^t A \underline{x}}{\underline{x}^t \underline{x}} = \lambda_1 \text{ e } \inf_{\underline{x}} \frac{\underline{x}^t A \underline{x}}{\underline{x}^t \underline{x}} = \lambda_m$$

Para resolver

$$\sup_{\underline{x}} \frac{\underline{x}^t A \underline{x}}{\underline{x}^t B \underline{x}} \text{ e } \inf_{\underline{x}} \frac{\underline{x}^t A \underline{x}}{\underline{x}^t B \underline{x}}$$

onde $A_{(m \times m)}$ é simétrica e $B_{(m \times m)}$ é simétrica e positiva definida, deve-se reduzir este caso ao anterior. Sendo B é simétrica e definida positiva, a decomposição de Choleski de B é $B = L^t L$, onde L é uma matriz triangular superior.

Seja a transformação,

$$\underline{y} = L \underline{x} \implies \underline{x} = L^{-1} \underline{y},$$

portanto,

$$\frac{\underline{x}^t A \underline{x}}{\underline{x}^t B \underline{x}} = \frac{\underline{x}^t A \underline{x}}{\underline{x}^t L^t L \underline{x}} = \frac{\underline{y}^t (L^t)^{-1} A^{-1} L^{-1} \underline{y}}{\underline{y}^t \underline{y}} = \frac{\underline{y}^t C \underline{y}}{\underline{y}^t \underline{y}}$$

onde $C = (L^t)^{-1} A^{-1} L^{-1}$.

Pelos resultados anteriores,

$$\sup_{\underline{x}} \frac{\underline{x}^t A \underline{x}}{\underline{x}^t B \underline{x}} = \sup_{\underline{y}} \frac{\underline{y}^t C \underline{y}}{\underline{y}^t \underline{y}} = \lambda_1 \text{ e } \inf_{\underline{x}} \frac{\underline{x}^t A \underline{x}}{\underline{x}^t B \underline{x}} = \inf_{\underline{x}} \frac{\underline{y}^t C \underline{y}}{\underline{y}^t \underline{y}} = \lambda_m$$

onde λ_1 e λ_m são autovalores máximo e mínimo de C .

A solução do sistema,

$$C \underline{y} = \lambda \underline{y}$$

implica na solução de

$$B^{-1}A\underline{x} = \lambda \underline{x} :$$

$$C \underline{y} = \lambda \underline{y} \implies (L^t)^{-1}A L^{-1} \underline{y} = \lambda \underline{y} ,$$

como $\underline{y} = L\underline{x}$, tem-se

$$(L^{-1})^t A L^{-1} \underline{x} = \lambda \underline{x} ,$$

então

$$L^{-1}(L^{-1})^t A \underline{x} = \lambda L^{-1} \underline{x} ,$$

ou seja,

$$B^{-1}A \underline{x} = \lambda \underline{x} ,$$

no caso de $B = A$, todos os autovalores de $B^{-1}A$ são iguais a 1.

O problema do mal condicionamento em matrizes de Dependência e Codependência

Na secção 4.1 comentamos ser razoável a utilização do critério de comparação ξ para matrizes de Dependência e Codependência, nos casos onde a matriz D_2 de (3.1.3) for positiva definida. A seguir traçaremos resumidamente, os problemas que podem ocorrer quando D_2 for quase singular.

O mal condicionamento de matrizes $A_{(n \times n)}$, pode ser traduzido por determinantes muito próximos de zero, onde pequenas perturbações em A podem ocasionar grandes imprecisões na matriz inversa. Para ilustrar a situação, descreveremos um exemplo de Shayle Searle [32] pg. 146 :

$$\begin{bmatrix} 2.042.49 \\ 2.493.04 \end{bmatrix}^{-1} = \begin{bmatrix} 2026 & \frac{2}{3} & -1660 \\ -1660 & & 1360 \end{bmatrix}$$

enquanto,

$$\begin{bmatrix} 2.0 & 2.5 \\ 2.5 & 3.0 \end{bmatrix}^{-1} = \begin{bmatrix} -12 & 10 \\ 10 & -8 \end{bmatrix}$$

As perturbações às quais nos referimos, podem ser devidas à limitações da representação de números pelo computador. De acordo com Wilkinson [34] pg. 344, a solução do sistema $A\mathbf{x} = \lambda B\mathbf{x}$ que se reduz ao cálculo dos autovalores de $B^{-1}A$, é insatisfatória se B for mal condicionado com respeito à inversão. Neste caso, pode-se encontrar autovalores finitos e infinitos.

Exemplo :

$$\text{Para } A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

$\det(A - \lambda B) = 3 - 2\lambda$, os autovalores encontrados são $1 \frac{1}{2}$ e ∞ .

Pequenas perturbações em A e B fazem com que ocorram pequenas alterações nos autovalores finitos (este são insensíveis às perturbações) enquanto autovalores infinitos podem tornar-se finitos, porém muito grandes.

Os elementos de matrizes de Dependência e Codependência, são em geral pequenos, para que estes números pudessem ser expressos com melhor aproveitamento de seus dígitos significativos pelo computador, as matrizes D_1 e D_2 para o cálculo de ξ , foram multiplicadas cada uma por um fator constante, de acordo com o respectivo inverso da menor ordem dos elementos de cada matriz.

Esta operação não altera o resultado final de ξ , desde que se tomarmos c_1 e c_2 como sendo os fatores constantes mencionados acima para D_1 e D_2 respectivamente, teremos a seguinte situação :

para o sistema de equações

$$D_1 \underline{x} = \lambda D_2 \underline{x} \text{ com } \underline{x} \in \mathbb{R}^J \text{ e } \underline{x}^t D_2 \underline{x} = 1,$$

sejam c_1 e $c_2 \in \mathbb{R}$ e façamos

$$A = c_1 D_1 \text{ e } B = c_2 D_2,$$

definamos o sistema de equações

$$A \underline{y} = \gamma B \underline{y} \text{ com } \underline{y} \in \mathbb{R}^J \text{ e } \underline{y}^t B \underline{y} = 1,$$

este sistema é equivalente a,

$$B^{-1} A \underline{y} = \gamma \underline{y},$$

que por sua vez implica em,

$$D_2^{-1} D_1 \underline{y} = \frac{c_2}{c_1} \gamma \underline{y},$$

portanto $\lambda = c_2/c_1 \gamma$ (como os autovalores de uma matriz são únicos), isto mostra nossa afirmação.

Teorema 4.2.1

Seja $\underline{X} = \{X_1, X_2, \dots, X_k\}$ uma variável aleatória com densidade Multinomial de parâmetros n e $\underline{P} = \{p_1, p_2, \dots, p_k\}$, isto é, $\underline{X} \approx M(n, \underline{P})$. A densidade de X_1 é Binomial com parâmetros n e p_1 , isto é, $X_1 \approx B(n, p_1)$.

Seja $X_1 = x_1$, então a densidade de X_2 condicionada a $X_1 = x_1$,

$$X_2 | X_1 = x_1, \text{ é } B(n - x_1, p_2 / (1 - p_1)).$$

De maneira análoga, a densidade de

$$X_3 | X_2 = x_2, X_1 = x_1, \text{ é } B(n - x_2 - x_1, p_3 / (1 - p_2 - p_1)),$$

e portanto, a densidade de

$$X_k | X_{k-1} = x_{k-1}, \dots, X_1 = x_1, \text{ é } B\left(n - \sum_{i=1}^{k-1} x_i, p_k / \left\{1 - \sum_{i=1}^{k-1} p_i\right\}\right).$$

Verifica-se que a densidade conjunta de

$$\{X_1, X_2 | X_1 = x_1, \dots, X_k | X_{k-1} = x_{k-1} \dots X_1 = x_1\},$$

é uma densidade Multinomial $M(n, p_1, \dots, p_k)$.

Prova :

Provaremos (sem perda de generalidade) o caso de $k = 3$.

Para verificar que X_1 tem distribuição Binomial com parâmetros n e p_1 , seja,

$$\begin{aligned}
 p_{X_1}(x_1) &= \sum_{x_2} \sum_{x_3}^{n-x_1-x_2} p_{X_1, X_2, X_3}(x_1, x_2, x_3) = \sum_{x_2=0}^{n-x_1} p_{X_1, X_2, X_3}(x_1, x_2, n-x_1-x_2) \\
 &= \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1! x_2! (n-x_1-x_2)!} p_1^{x_1} p_2^{x_2} (1-p_1-p_2)^{n-x_1-x_2} \\
 &= \frac{n!}{x_1! (n-x_1)!} p_1^{x_1} \left[\sum_{x_2=0}^{n-x_1} \binom{n-x_1}{x_2} p_2^{x_2} (1-p_1-p_2)^{(n-x_1)-x_2} \right] \\
 &= \frac{n!}{x_1! (n-x_1)!} p_1^{x_1} [(p_2 + 1 - p_1 - p_2)^{n-x_1}]
 \end{aligned}$$

então,

$$p_{X_1}(x_1) = \frac{n!}{x_1! (n-x_1)!} p_1^{x_1} (1-p_1)^{n-x_1}, \text{ logo } X_1 \approx B(n, p_1) \text{ .(A.2.1)}$$

Para provar a segunda parte do teorema, seja

$$\begin{aligned}
 p_{X_2, X_3 | X_1=x_1}(x_2, x_3 | x_1) &= \frac{p_{X_1, X_2, X_3}(x_1, x_2, x_3)}{p_{X_1}(x_1)} \\
 &= \frac{(n-x_1)!}{x_2! x_3!} [p_2/(p_2+p_3)]^{x_2} [p_3/(p_2+p_3)]^{x_3},
 \end{aligned}$$

com $x_2 + x_3 = n - x_1$,

então,

$p_{x_2, x_3 | x_1 = x_1}^{(x_2, x_3)}$, possui distribuição multinomial com parâmetros

$$n - x_1 \text{ e } \theta = [p_2 / (p_2 + p_3), p_3 / (p_2 + p_3)] ; \quad (\text{A.2.2})$$

A densidade condicional

$$p_{x_2 | x_1}^{(x_2)} = \frac{p_{x_2, x_1}^{(x_1, x_2)}}{p_{x_1}^{(x_1)}} = \sum_{x_3} \frac{p_{x_1, x_2, x_3}^{(x_1, x_2, x_3)}}{p_{x_1}^{(x_1)}} ,$$

mas,

$$\begin{aligned} p_{x_1, x_2, x_3}^{(x_1, x_2, x_3)} &= p_{x_2, x_3 | x_1 = x_1}^{(x_2, x_3 | x_1)} p_{x_1}^{(x_1)}, \text{ com } x_2 + x_3 = n - x_1 \\ &= p_{x_2, x_3 | x_1 = x_1}^{(x_2, n - x_1 - x_2 | x_1)} , \\ &= \frac{(n - x_1)!}{x_2! (n - x_1 - x_2)!} [p_2 / (p_2 + p_3)]^{x_2} [1 - p_2 / (p_2 + p_3)]^{n - x_1 - x_2} \end{aligned}$$

por (A.2.2),

$$X_2 | X_1 = x_1 \approx B[(n - x_1), p_2 / (1 - p_1)] , \text{ pois } p_2 / (p_2 + p_3) = p_2 / (1 - p_1);$$

finalmente,

a densidade de

$$X_3 | X_1 = x_1, X_2 = x_2 = \begin{cases} 1, & x_3 = n - x_1 - x_2 \\ 0, & \text{caso contrário} \end{cases} ,$$

portanto,

$$\begin{aligned}
 p_{x_1, x_2, x_3}(x_1, x_2, x_3) &= p_{x_3|x_1, x_2}(x_3|x_1, x_2) p_{x_1, x_2}(x_1, x_2) \\
 &= p_{x_3|x_1, x_2}(x_3|x_1, x_2) p_{x_2|x_1}(x_2|x_1) p_{x_1}(x_1) \\
 &= \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}.
 \end{aligned}$$

Teste de equidistribuição para o gerador de números aleatórios G05CAF da NAG

Nas simulações empreendidas utilizamos o gerador de números aleatórios G05CAF da Biblioteca de Algoritmos Numéricos NAG (FORTRAN). Para testar a existência de equidistribuição dos números gerados pelo gerador G05CAF, utilizamos o procedimento descrito por Kalos e Whitlock pgs. 174 a 175, (1986) [18]. O procedimento é o que segue :

1) Gerar 10.000 números com distribuição uniforme (0,1) e distribuí-los em 100 intervalos equiespaçados em (0,1). O número esperado em cada intervalo é 100. Calcula-se o valor da estatística χ^2 , isto é,

$$\chi^2 = \frac{k}{N} \sum_{j=1}^K \left(N_j - \frac{N}{k} \right)^2, \text{ onde}$$

k = número de intervalos

N = Tamanho da amostra (no caso, 10.000).

2) O procedimento 1) é repetido 1.000 vezes para que sejam obtidos 1.000 valores χ^2 . Se a sequência de números gerados for uniforme, então $\chi_1^2, \chi_2^2, \dots, \chi_{1000}^2$ se comportará como uma amostra aleatória proveniente de uma distribuição $\chi^2_{(k-1)}$.

3) Os 1.000 valores χ^2 são dispostos em 100 intervalos igualmente prováveis :

$(0, \chi_{(99,0.01)}^2), (\chi_{(99,0.01)}^2, \chi_{(99,0.02)}^2), \dots, (\chi_{(99,0.99)}^2, \chi_{(99,1.0)}^2)$ e o

número B_j , esperado em cada intervalo é 10.

4) Calcula-se a estatística χ^2 total,

$$\chi_T^2 = \sum_{j=1}^{100} (B_j - 10)^2 / 10 ,$$

De acordo com Kalos e Whitlock [18] pg 175, o valor de χ_T^2 nos melhores geradores está em torno de 80 a 82.

Ao empreendermos os passos descritos acima, o valor de χ_T^2 obtido foi de 103,54. No entanto, um valor desta ordem parece ser razoável, desde que $\chi_{(99,0.01)}^2 = 134,16$.

Observação : O valores de $\chi_{(99,\alpha)}^2$, com $\alpha = 0.01, \dots, 0.99$, foram obtidos utilizando-se a função INVCHI do produto SAS .

Apêndice2 : Programa

- c Programa para gerar a distribuição de frequências do critério de comparação $\xi = 1 - \lambda_{\min}/\lambda_{\max}$ de acordo com distribuições Multinomiais especificadas

```

real*8      a(5,5),aux(2,3,3),zero,cmx,quo(2),te(20)
real*8      aux1(2,3,3),aux1t(2,3,3),b(5,5),cmn,rinc
real*8      d(2,3,3),acl(2,10),tel(20),sum(3),  rmat(50,10)
real*8      pc(2,5),pl(2,5),t(2,3,3),tp(2,3,3),hf(5)
real*8      rr(5),ri(5),wkspce(12),c(2200),aux2(2,5,5)
integer     n(2,12),nc(2,3,3),nf(100),ncl(3)
integer     i,ia,iunit,ipt(20),ifail,ins,idim1,idim2,num,idig
integer     nv,ita,isl,ig,idi,id2,ident,kd,iflagl(2),ndis

```

- c O primeiro DATA a seguir especifica o numero de classes para as distribuições de frequência, o intervalo de classes, o número de distribuições e a semente a ser utilizada.

- c O segundo DATA especifica o numero de simulações de MONTE CARLO, as dimensões das tabelas de contingência a serem geradas segundo as associacoes em dados.arq e o tamanho da amostra.

```

data      ncl,hf,ndis,isl,/15,0.07,12,23/
data      ins,idim1,idim2,ita/500,2,2,500/

```

O arquivo dados.arq deve conter as distribuições Multinomiais de interesse para a geração de tabelas de contingência. A leitura destes dados é feita pelo READ a seguir. Cada distribuição Multinomial deverá ser registrada em uma linha do arquivo externo.

No arquivo graf.dat serão gravados os valores gerados de ξ .

No arquivo resumo.dat serão gravados os pontos médios e frequência da distribuição de ξ para cada distribuição Multinomial em estudo.

No arquivo stat.dat serão gravados estatísticas básicas tais como os valores minimo, maximo, média e variância da distribuição de ξ .

```

open(unit=33,file='dados.arq',status='old')
open(unit=11,file='graf.dat',status='new')
open(unit=88,file='resumo.dat',status='new')
open(unit=99,file='stat.dat',status='new')

```



```

zero = dble(0.0)
write(90,*)'a semente desta geracao e : '
write(99,*)'a semente desta geracao e : '
write(90,37)isl
write(99,37)isl
37   format(/,5x,i3,/)
write(99,*)'Valores mini,max,media e var da Distribuicao : '
read(33,34,end=35)((rmat(i,j),j=1,4),i=1,ndis)

c     A subrotina g05cbf envia a semente escolhida para o gerador
de números pseudo-aleatórios g05caf da NAG a ser utilizado na
subrotina gera.

35   call g05cbf(isl)
ig = idim1*idim2
idig = 2*ig
do i=1,5
    hf(i) = zero
end do

c
write(90,34)((rmat(i,j),j=1,4),i=1,ndis)
c     write(11,34)((rmat(i,j),j=1,4),i=1,ndis)
34   format(4(f6.4,1x))

do kd = 1,ndis

c     Rotinas de inicializacao

    call point(idim1,idim2,idig,ipt)
    call start(kd,ins,ig,rmat,t,tp,aux2,te,tel,
*       ia,iunit,ifail,ident,idim1,idim2,c)
    call margl(idim1,idim2,pl,tp)
    call margc(idim1,idim2,pc,tp)
    call sums(ita,kd,idim1,idim2,aux2,pl,pc)
    call codepe(idim1,idim2,tp,d)

c     Rotinas para inicializacao do gerador Multinomial

    call decres(ig,tel,idig,ipt)
    call acc(ig,tel,aci)
    call mxmn(ita,cmx,cmn,nf,nv)

    do i = 1,ins

c     Simulacao - Contagem

        call gera(idim1,idim2,ita,tel,aci,idig,ipt,n)
        call freq(i,ita,idim1,idim2,n,nc)

```

c Calculo das matrizes de Dependencia
 c e Codependencia D1 e D2

```

call    prop(idim1,idim2,ita,tp,nc)
call    margl(idim1,idim2,pl,tp)
call    margc(idim1,idim2,pc,tp)
call    provet(pl,pc,idim1,idim2,aux)
call    criaux(tp,aux,idim1,idim2,aux1)
call    transp(aux1,idim1,idim2,aux1t)
call    produt(aux1,aux1t,idim2,idim1,idim2,d)

```

c Calculo dos autovalores maximo(lmax) e mínimo(lmin) de
 inv(D2)*D1

```

call    inicia(idim2,d,a,b)
call    zera(rr,ri,wkspce)
k = 1
call    nag(a,b,idim2,rr,ri)
call    calc(quo,k,zero,rr,ri,idim2,iflag1)
call    zera(rr,ri,wkspce)
k = 2
call    nag(b,a,idim2,rr,ri)
call    calc(quo,k,zero,rr,ri,idim2,iflag1)
if(iflag1(1).eq.0.and.iflag1(2).eq.0) then
    if(quo(1).lt.quo(2)) go to 111
    c(i) = quo(2)
    go to 222
111    c(i) = quo(1)
    go to 222
    else
    endif
    if(iflag1(1).eq.0)then
    c(i) = quo(1)
    go to 222
    else
    endif
    if(iflag1(2).eq.0)then
    c(i) = quo(2)
    go to 222
    else
    go to 333
    endif
c      write(90,*)'C(i) = '
c      write(90,*)c(i)
222    if(c(i).ge.zero)go to 444
    write(90,*)rr(1),rr(2)
    stop
444    if(c(i).gt.cmx)cmx=c(i)
    if(c(i).lt.cmn)cmn=c(i)
    nv=nv+1
333  end do

```

c

c Distribuição de 1 - lmin/lmax

```
do j=1,ins
write(11,61)c(j)
61     format(5x,f10.5)
end do
```

c

```
call     saidas(num,rinc,ident,nv,c,cmn,cmx,ins)
end do
end
```

c

c Subrotinas

c

```
subroutine     start(kd,ins,ig,rmat,t,tp,aux2,te,tel,
*     ia,iunit,ifail,ident,idim1,idim2,c)
real*8     tp(2,3,3),t(2,3,3),aux2(2,5,5),tel(20)
real*8     c(1100),zero,te(20),rmat(50,10)
integer     kd,ins,ig,ia,iunit,ifail,ident,idim1,idim2
zero = dble(0.0)
ident=kd
ia =5
iunit = 5
ifail=1
do i =1,20
te(i) =zero
end do
do i=1,ins
c(i) = zero
end do
```

c

```
do kf = 1,ig
te(kf) = rmat(kd,kf)
mx = kf + ig
te(mx) = rmat(kd,kf)
end do
```

c

```
write(90,40)(te(i),i=1,8)
40     format(4(2x,f4.2))
```

c

```
ll = 2*ig
do i = 1,ll
tel(i) = te(i)
end do
```

c

```
l=1
do m = 1,2
do j = 1,idim1
do k =1,idim2
tp(m,j,k) = zero
tp(m,j,k) = te(l)
t(m,j,k) = te(l)
```

```

        l = l + 1
    end do
end do
end do
return
end

```

c

```

subroutine    point(idim1,idim2,idig,ipt)
integer idim1,idim2,idig,ipt(20)
ig = idim1*idim2
do i=1,ig
    ipt(i) = i
    mx = i + ig
    ipt(mx) = i
end do
return
end

```

c

```

subroutine    sums(ita,kd,idim1,idim2,aux2,pl,pc)
real*8      zero,aux2(2,5,5),pl(2,5),pc(2,5),sum(3)
integer      kd,idim1,idim2,id1,id2,ita
zero = dble(0.0)
do m=1,2
    sum(m)=zero
    do j=1,idim1
        sum(m) = sum(m) + pl(m,j)
    end do
end do

id1 = idim1 + 1
id2 = idim2 + 1

```

c

```

do m=1,2
    do j=1,id1
        do k=1,id2
            aux2(m,id1,k) = zero
            aux2(m,j,id2) = zero
            pl(m,id1) = zero
            pc(m,id2) = zero
        end do
    end do
end do

```

c

```

do m=1,2
    do i=1,id1
        aux2(m,i,id2) = pl(m,i)
    end do
end do

```

c

```

do m=1,2
    do j=1,id2

```

```

        end do
    end do
c
    do m=1,2

        aux2(m,id1,id2) = sum(m)
    end do
c
    write(90,2)
2    format(/,15x,'Tabela de Associac. Linha/Coluna')
    write(90,3)((aux2(1,i,j),j=1,id2),i = 1,id1)
3    format(/,3(5x,f6.4),/)
c    write(88,33)((aux2(1,i,j),j=1,id2),i = 1,id1)
c33    format(/,3(5x,f6.4),/)
        return
    end
c
    subroutine acc(ig,tel,acl)
    real*8 tel(20),acl(2,10),zero
    integer ig
    zero = dble(0.0)
    l = ig + 1
    j = 0
    do m = 1,2
        acl(m,1) = zero
        do i=2,l
            acl(m,i) = zero
            acl(m,i) = acl(m,i-1) + tel(i-1+j)
        end do
        j = ig
    end do
    return
    end
c
    subroutine mxmn(ita,cmx,cmn,nf,nv)
    real*8 cmx,cmn,zero
    integer nv,nf(100)
    zero = dble(0.0)

    cmx = zero
    cmn = dble(70000.00)
    nv = 0
    do j = 1,42
        nf(j) = 0
    end do
    write(90,6) ita
c    write(99,6) ita
6    format(/,32hTamanho da amostra p/simulacao :lh , 5x,i4)
    return
    end
c
    subroutine gera(idim1,idim2,ita,tel,acl,idig,ipt,n)
c    As dimensoes de IA e XA devem estar de acordo
c    com o tamanho da amostra a ser gerada.
    integer n(2,12),ipt(idig),ia(2000),nt(4),ita,ig,idig,ifail

```

```

ig = idim1*idim2
l = ig - 1
j2 = 0
ifail=0

do j = 1,2
  nt(j) = 0
  isz = ita
  ix = 0
  do j1 = 1,1
    ii = j1 + j2
    j3 = ipt(ii)
    n(j,j3) = 0
    if(isz.eq.0) go to 100
    p = tel(j1+j2)/(1.0 - ac1(j,j1))
    kk = 0
    kk = isz-1
    do k = 1,isz
      kk = kk - 1
      aux = g05caf(r)
      if(kk.ne.ita) go to 22
      if(j.ne.1) go to 22
      aux=g05caf(r)
22      if(j1.ne.1) go to 21
      if(j.ne.2) go to 21
      if(k.eq.(isz-1)) call      g05cff(ia,ita,xa,ita,ifail)
21      if(aux.gt.p) go to 20
      n(j,j3) = n(j,j3) + 1
20    end do
      ix = ix + n(j,j3)
      isz = ita - ix
    end do
100    nt(j) = ix
      j2 = ig
  end do
  l1 = 2*ig
  n(1,ipt(ig)) = ita - nt(1)
  n(2,ipt(l1)) = ita - nt(2)
c      write(90,90)((n(i,j),j=1,ig),i=1,2)

c90      format(/,15x,2(i3,5x))
c      call g05cgf(ia,ita,xa,ita,ifail)
return
end

c

subroutine      freq(i,ita,idim1,idim2,n,nc)
integer n(2,10),nc(2,3,3)
zero = dble(0.0)
do j =1,2
  kk = 0
  do l =1,idim1
    do k =1,idim2
      nc(j,l,k) = 0
      kk = kk + 1
      nc(j,l,k) = n(j,kk)

```

```

        end do
    end do
c      write(90,*)'Tabela das Frequencias'
c      write(90,8)((nc(j,k,l),l =1,idim1),k =1,idim2),j=1,2)

c8      format(/,2(5x,i3),/)
    return
end

c
c      Calculo das Proporcoes Amostrais
c
    subroutine prop(idim1,idim2,ita,tp,nc)
    integer nc(2,3,3)
    real*8 tp(2,3,3),zero
    zero = dble(0.0)
    do j = 1,2
        do k = 1,idim1
            do l =1,idim2
                tp(j,k,l) = zero
                tp(j,k,l) = dfloat(nc(j,k,l))/dfloat(ita)
            end do
        end do
    end do
c      write(90,*)'Tabela das Proporcoes Amostrais'
c      write(90,13)((tp(j,k,l),l = 1,idim2),k =1,idim1),j =1,2)
c13     format(/,3(5x,f7.5))
    return
end

c
c      Perfis Marginais de Linha
c
    subroutine margl(idim1,idim2,pl,tp)
    real*8 pl(2,5),tp(2,3,3),zero
    zero = dble(0.0)
    do j = 1,2
        do k = 1,idim1
            pl(j,k) = zero
            do l =1,idim2
                pl(j,k) = pl(j,k) + tp(j,k,l)
            end do
        end do
    end do
c      write(90,*)'Perfis Marg. de Linhas'
c      write(90,208)((pl(j,k),k = 1,idim1),j =1,2)
208     format(/,3(5x,f7.5),/)
    return
end

c
c      Perfis Marginais de Coluna
c
    subroutine margc(idim1,idim2,pc,tp)
    real*8 pc(2,5),tp(2,3,3),zero
    zero = dble(0.0)
    do j = 1,2
        do k = 1,idim2

```

```

        do l = 1,idim1
          pc(j,k) = pc(j,k) + tp(j,l,k)
        end do
      end do

c      write(90,*)'Perfis Marg. de Colunas'
c      write(90,209)((pc(j,k),k = 1,idim2),j = 1,2)
c209    format(/,3(5x,f7.5),/)
      return
      end

c
      subroutine provet(e,e1,k1,k2,res1)
      real*8 e(2,3) ,e1(2,3) ,res1(2,3,3)
        do jl = 1,2
          do j = 1,k1
            do l = 1,k2
              res1(jl,j,l) = dble(0.0)
              res1(jl,j,l) = e(jl,j)*e1(jl,l)
            end do
          end do
        end do
c      write(90,*)'Produto dos Perfis Marg.'
c      write(90,105)((res1(jl,j,l),l = 1,k2),j = 1,k1),jl = 1,2)
c105    format(/,3(5x,f7.5),/)
      return
      end

c
      subroutine criaux(tp,aux,idim1,idim2,aux1)
      real*8 sq,tp(2,3,3),aux(2,3,3),aux1(2,3,3)
c      Criando aux1 = sqrt(2)*(tp#.5 - (pl*pc)#.5)
c
        sq = dble(0.5)
        do j = 1,2
          do k = 1,idim1
            do l = 1,idim2
              aux1(j,k,l) = dble(0.0)
              aux1(j,k,l) = dsqrt(tp(j,k,l)) - dsqrt(aux(j,k,l))
              aux1(j,k,l) = dsqrt(sq) * aux1(j,k,l)
            end do
          end do
        end do
c      write(90,*)'Matriz Delta'
c      write(90,212)((aux1(j,k,l),l = 1,idim2),k = 1,idim1),j= 1,2)
c212    format(/,2(5x,f7.5),/)
      return
      end

c
      subroutine transp(x,i1,i2,res)
      real*8 x(2,3,3) , res(2,3,3)
        do j = 1,2
          do k = 1,i1
            do l = 1,i2
              res(j,k,l) = dble(0.0)
              res(j,k,l) = x(j,l,k)

```



```

c 218          format(//,3(3x,i3),3x,f7.5)
              end do
            end do
          end do
        return

      end

c
      subroutine  produt(x1,x2,k1,k2,k3,d)
      real*8      x1(2,3,3),x2(2,3,3),d(2,3,3)
      do jj = 1,2
        do j = 1,k1
          do k = 1,k3
            d(jj,j,k) = dble(0.0)
            do l = 1,k2
              d(jj,j,k) = d(jj,j,k) + x2(jj,j,l)*x1(jj,l,k)
            end do
          end do
        end do
      end do

c      write(90,*)'Tabelas de Dep.e Codep.'
c      write(90,*)(((d(jj,j,k),k =1,k3),j =1,k1),jj =1,2)
c      write(90,209)(((d(jj,j,k),k =1,k3),j =1,k1),jj =1,2)
c209      format(/,2(5x,f7.5),/)
      return
    end

c
    subroutine  inicia(idim2,d,a,b)
    real*8      d(2,3,3),a(5,5),b(5,5),vmax
    real*8      zero,vmin,const,const1,dez
    integer     idim2,ita
    zero = dble(0.0)
    dez = dble(10.0)
    do k =1,idim2
      do l = 1,idim2
        a(k,l) = zero
        b(k,l) = zero
        a(k,l) = d(1,k,l)
        b(k,l) = d(2,k,l)
      end do
    end do

    call  fatcom(a,idim2,vmax)
    const = vmax
    call  fatcom(b,idim2,vmax)
    const1 = vmax

c      write(90,*)'fatores comuns'
c      write(90,*)const,const1
    do j=1,idim2
      do k =1,idim2
        a(j,k) = dez**const*(a(j,k))
        b(j,k) = dez**const1*(b(j,k))
      end do
    end do

c      write(90,3)((a(j,k),k=1,idim2),j=1,idim2)

```

```

c          write(90,4)((b(j,k),k=1,idim2),j=1,idim2)
c4         format(2hD2/(1h,2(5x,f10.5)))
          return
          end

c

subroutine   nag(a1,b1,idim2,rr,ri)
real*8      a1(5,5),b1(5,5),ans(5,5),ans1(5,5)
real*8      rr(5),ri(5),rinvb1(5,5),zero
integer     ia,idim2,ifail,intger(5)
zero = dble(0.0)
          ia=5
          do j = 1,5
            intger(j) = 0
            do jj = 1,5
              rinvb1(j,jj) = zero
              ans(j,jj) = zero
              ans1(j,jj) = zero
            end do
          end do
          call inversa(b1,idim2,rinvb1)
          call pro(a1,rinvb1,idim2,ans,ans1)
c          write(90,*)'PRODUTO'
c          write(90,92)((ans1(k,j),j=1,idim2),k=1,idim2)
c92         format(1h,2(5x,f12.5))
          ifail = 1
          call f02aff(ans,ia,idim2,rr,ri,intger,ifail)
          if(ifail.eq.0) goto 20
          write(90,*)'erro na subrot. f02aff'
          stop
c20         write(90,93)(rr(j),ri(j),j=1,idim2)
c93         format(11hautovalores/(2h,(2(5x,f10.5,1h,f10.5)))
20          return
          end

c

subroutine   calc(quo,k,zero,rr,ri,idim2,iflag1)
real*8      vmin,vmax,rr(5),ri(5),zero,quo(2)
integer     k,idim2,iflag,iflag1(2)
          iflag = 0
          do j =1,idim2
            if(ri(j).ne.zero) iflag = 1
          end do
          quo(k) = zero
          iflag1(k)=iflag
          if(iflag.eq.0) go to 10
          write(90,*)'Autovalor com componente complexo'
          goto 88
10          call minmax(rr,idim2,vmin,vmax)
          quo(k) = dble(1.0) - vmin/vmax
c          write(90,*)quo(k)
c88         return
          end

c          Autovalores Maximo e Minimo
c
subroutine   minmax(rr,idim2,vmin,vmax)

```

```

vmin = rr(1)
vmax = rr(1)
j = 1
34  j = j + 1

if (j.le.idim2) then
  if (rr(j).le.vmin) then
    vmin = rr(j)
    go to 34
  else
    endif
    if(rr(j).ge.vmax) then
      vmax = rr(j)
      go to 34
    else
      go to 34
    endif
  endif
endif
c    write(90,35)
c35  format(/,15x,'Autovalores Minimo e Maximo')
c    write(90,36) vmin,vmax
c36  format(/,2(3x,f9.5))
return
end

c
subroutine codepe(idim1,idim2,tp,d)
real*8    d(2,3,3),tp(2,3,3),pl(2,3),pc(2,3)
real*8    aux(2,3,3),aux1(2,3,3),aux1t(2,3,3)
call margl(idim1,idim2,pl,tp)
call margc(idim1,idim2,pc,tp)
call provet(pl,pc,idim1,idim2,aux)
call criaux(tp,aux,idim1,idim2,aux1)
call transp(aux1,idim1,idim2,aux1t)
call produt(aux1,aux1t,idim2,idim1,idim2,d)
write(90,*)'Tabelas de Dep. e Codepend. p/assoc. linha - coluna'
write(90,10)((d(1,j,k),k=1,idim2),j = 1,idim2)
10  format(/,2(5x,f7.5))
return
end

c
subroutine decres(ig,tel,idig,ipt)
real*8 tel(20),b
integer  ig,bl,idig,ipt(idig)
jl = 0
l = 1
do i = 1,2
  il = ig + jl
  nl0 = il - 1
  do j = 1,nl0
    do k = j,il
      if(tel(j) - tel(k))30,40,40
30    b = tel(j)
      tel(j) = tel(k)
      tel(k) = b
      bl = ipt(j)
      ipt(j) = ipt(k)

```

```

40         end do
           end do
           j1 = ig

           l = ig + 1
         end do
c         write(90,11)(tel(i),ipt(i),i=1,11)
c11        format(/,5x,f7.5,5x,i3)
           return
           end

c
subroutine somat(alfa,idim1,idim2,ig,t,tel)
real*8    t(2,3,3),tel(20),alfa
integer   idim1,idim2,ig
write(90,*)'subrotina somat'
l = 1
  do j = 1,idim1
    do k = 1,idim2
      tel(l) = alfa*t(1,j,k) + (dble(1.0) - alfa)*t(2,j,k)
      ll = ig + 1
      tel(ll) = tel(l)
      write(90,1)    alfa,tel(l),tel(ll)
1      format(/,3(5x,f7.5))
      l = l + 1
    end do
  end do
  return
  end

c
subroutine verify(idim1,idim2,nc,t,ita)
real*8    t(2,3,3),q(5),v1
integer   nc(2,3,3),idim1,idim2
l = 0
q(1) = dble(0.0)
q(2) = dble(0.0)
  do l=1,2
    do j=1,idim1
      do k=1,idim2
        v1 = dfloat(ita)*t(l,j,k)
        q(l) = q(l)+((dfloat(nc(l,j,k))- v1)**2)
      end do
    end do
  end do
  write(90,4)(q(j),j=1,2)
4  format(7hQuiQuad/(1h,2(5x,f10.4)))
  return
  end

c
subroutine pro(help1,unit,idim2,ans,ans1)
real*8    help1(5,5),unit(5,5),ans(5,5),ans1(5,5)
integer   idim2
  do j=1,idim2
    do k=1,idim2
      ans(j,k) = dble(0.0)
      ans1(j,k) = dble(0.0)

```

```

        ans(j,k) = ans(j,k) + unit(j,l)*help1(l,k)
        ans1(j,k) = ans(j,k)
    end do
end do
end do
return
end

```

c

```

subroutine fatcom(aux3,idim2,vmax)
real*8 aux3(5,5),v,dez,fat(12),x,vmin,vmax
real*8 zero
integer k,idim2,l
dez = dble(10.0)
zero = dble(0.0)
l = 0
do i=1,idim2
    do j = 1,idim2
        l = l + 1
        k = 0
        v = aux3(i,j)
        if(v.lt.zero) v = dabs(v)
        x = v
1      x = x*dez
        k = k + 1
        if(x.gt.dez) goto 2
        go to 1
2      k = k - 1
        fat(l) = dfloat(k)
    end do
end do
call minmax(fat,l,vmin,vmax)
return
end

```

c

```

subroutine inversa(bl,idim2,rinvbl)
real*8 bl(5,5),aux1(5,5),aux2(5,5),p(5),same(5,5)
real*8 prod(5,5),ans(5,5),ans1(5,5),rinvbl(5,5),zero

integer id,idim2

```

c ROTINA DE FADDEZEW

```

zero = dble(0.0)
id = idim2 - 1
do j=1,idim2
    p(j) = zero
    do jj = 1,idim2
        aux1(j,jj) = zero
        aux2(j,jj) = zero
        aux2(j,j) = dble(1.0)
        prod(j,jj) = zero
        same(j,jj) = zero
        same(j,j) = dble(1.0)
        aux1(j,jj) = bl(j,jj)
        rinvbl(j,jj) = zero
    end do
end do

```

```

end do

do j = 1,id
  call pro(aux2,b1,idim2,ans,ans1)
  do i=1,idim2
    do ii=1,idim2
      aux1(i,ii) = ans(i,ii)
    end do
  end do
  do i = 1,idim2
    p(j) = p(j) + aux1(i,i)
  end do
  p(j) = p(j)/dfloat(j)
  do i=1,idim2
    prod(i,i) = p(j)*same(i,i)
  end do
  do i=1,idim2
    do ii=1,idim2
      aux2(i,ii) = aux1(i,ii) - prod(i,ii)
    end do
  end do
end do
call pro(aux2,b1,idim2,ans,ans1)
do i=1,idim2
  p(idim2) = p(idim2) + ans(i,i)
end do
p(idim2) = (dble(1.0)/dfloat(idim2))*p(idim2)
do i=1,idim2
  do ii=1,idim2
    rinvbl(i,ii) = (dble(1.0)/p(idim2))*aux2(i,ii)
  end do
end do
c write(90,*)'INVERSA'
c write(90,1)((rinvbl(i,j),j=1,idim2),i=1,idim2)
cl format(/,2(5x,f16.10))
return
end

c

subroutine saidas(num,rinc,ident,nv,c,cmn,cmx,ins)
real*8 rli(100),rls(100),xmed(100),zero
real*8 cmn,cmx,sq,sc,h,media,var,c(1100),rinc
integer nv,ins,ident,num,nf(100)
zero = dble(0.0)
um = dble(1.0)
sq =zero
sc=zero
h = zero
h = rinc
do i=1,30
  xmed(i) = zero
  rli(i) =zero
  rls(i) = zero
  nf(i) = 0
end do
rli(1) = 0

```

```

      xmed(1) = (rli(1) + rls(1))/2

      rli(num) = 14.0*h
      rls(num) = dble(1.00)
      xmed(num) = (rli(num) + rls(num))/2
      rls(num) = rls(num) + 0.01
      ncont = num - 1
      do j=2,ncont
        ii = j - 1
        rli(j) = rls(ii)
        rls(j) = rli(j) + h
        xmed(j) = (rli(j) + rls(j))/2
      end do
      write(90,*)'limites inf. e sup. da dist. de C'
      i=0
10     i=i+1
      j = 0
      if(i.gt.nv) goto 20
      sc = sc + c(i)
      sq = sq + c(i)**2
30     j= j + 1
      if(j.gt.num) go to 10
      if(c(i).ge.rli(j).and.c(i).lt.rls(j)) then
        nf(j) = nf(j) + 1
        go to 10
      else
        go to 30
      end if
c
20     i = 0
      rls(num) = dble(1.0)
      do j = 1,num
        write(90,3) j,rli(j),rls(j),xmed(j),nf(j)
3       format(i3,3x,f12.4,2x,'[----',2x,f12.4,3x,f12.4,3x,i4)
        write(88,7)j,ident,xmed(j),nf(j)
7       format(i3,1x,i2,1x,f12.2,1x,i3)
      end do
c
      media = sc/dfloat(nv)
      var = (sq - (sc)**2.0/dfloat(nv))/(dfloat(nv) - 1)
      write(90,*)'Valores mini,max,media e var da Distribuicao : '
      write(90,5)cmn,cmx,media,var
5       format(/,2(5x,f12.4),2(5x,f12.4))
      write(99,9)cmn,cmx,media,var
9       format(2(3x,f12.4),2(2x,f12.4))
      write(90,*)'numero de simulacoes validas : '
      write(90,*)nv
      return
      end
c
      subroutine zera(rr,ri,wkspce)
      real*8 rr(5),ri(5),wkspce(12),zero
      zero = dble(0.0)
      do j=1,12
        wkspce(j) = zero
      end do

```

```
        rr(j) = zero
        ri(j) = zero
    end do
return
end
```


- 19 - Kemp, A. W. and Kemp, C. D. - (1987)
Rapid Generation of Frequency Tables - Appl. Stat. 36 277
- 20 - Kendall and Stuart - (1970)
Capítulo 30
- 21 - Khan, A. H. and Ali, S. M. - (1973)
A New Coefficient of Association - Ann. Inst. Statist. Math.,
vol 25, 41 - 50
- 22 - Kullback S., Leibler A. - (1951)
On Information and Sufficiency - Ann. Math. Statist., vol. 22
pp. 79-86
- 23 - Leite J.G., Singer J.M. - (1990)
Métodos Assintóticos em Estatística - Fundamentos e Aplicações,
9.^o Simpósio Brasileiro de Probabilidade e Estatística / IME - USP
- 24 - Mathai A. M. , Rathie P.N. - (1975)
Basic Concepts in Information Theory - New Delhi, Wiley Eastern
Limited
- 25 - Matusita, Kameo - (1955)
Decision Rules based on the Distance for problems of Fit, Two
Samples, and Estimation - Ann. Inst. Statist. Math. vol 26, 631 - 640
- 26 - Matusita, Kameo - (1956)
Decision Rules based on the Distance for problems of
Independence, Invariance and Two Samples - Ann. Inst. Statist. Math.
vol 7, 66 - 74
- 27 - Nishisato, Shizuiko - (1980)
Analysis of Categorical data : dual scaling and its applications
Toronto - University of Toronto press

- 28 - Rao, C. R. - (1968)
Linear Statistical Inference and its applications - New York - J. Wiley
- 29 - Rényi A. - (1987)
On Diary on Information Theory - Budapest, J. Wiley & Sons
- 30 - Ripley, Brian D. - (1987)
Stochastic Simulation - New York - J. Wiley
- 31 - Roy, S. N. - (1957)
Some Aspects of Multivariate Analysis - New York - J. Wiley
- 32 - Searle, Shayle R. - (1982)
Matrix Algebra Useful for Statistics - New York - J. Wiley
- 33 - Wigner, E. P. - (1958)
On the Distribution of the roots of certain symmetric matrices -
Annals of Mathematics - vol. 67 No.2, pp 325 - 327, march, 1958
- 34 - Wilkinson, J. H. - (1966)
The Algebraic Eigenvalue Problem - London - Oxford University Press