UM METODO DE SELEÇÃO
PARA VARIAVEIS POLITOMICAS

MAURO SERGIO DE FREITAS MARQUES

Orientador:
J. NORBERTO W. DACHS

Dissertação apresentada no Instituto de Matemática, Estatística e Ciência da Computação da Universidade Estadual de Campinas, como requisito parcial para obtenção do título de Mestre em Estatística.

Campinas
[illegible]

S U M A R I O

Este trabalho tem como finalidade:

1) Revisão da teoria de algumas medidas de associação para classificações cruzadas.

2) Estudo da medida de associação $\lambda(L/C)$, baseada na redução da probabilidade do erro de previsão.

3) Proposição de um método para selecionar de um conjunto de politomias $\{X^{(1)}, X^{(2)}, \ldots, X^{(p)}\}$ - que em princípio estão associadas com uma politomia Y - um subconjunto menor, sem a pressuposição de uma relação funcional a politomia Y e as politomias X'S.

S U M M A R Y

The scope of this work is:

1) To review the theory of some measures of association for cross classification.

2) To study the measure of association $\lambda(L/C)$ based on the reduction of the error probability of prediction.

3) To propose a method of selecting from a set of politomic variables $\{X^{(1)}, X^{(2)}, \ldots, X^{(p)}\}$ - which at the beginning are associated with a politomic variable Y - a subset, without the assumption of a functional relationship between the politomic variable Y and the politomic variables X'S.

AGRADECIMENTOS:

Adalberto O. Tardelli

J. Norberto W. Dachs

I N D I C E

INTRODUÇÃO

O problema de construir um modelo de previsão que relacione p variáveis independentes com uma variável dependente, e a utilidade de selecionar as variáveis que devem fazer parte do modelo tem sido vastamente discutido na literatura. Para o caso em que a variável dependente é quantitativa o problema está bastante estudado e há uma grande variedade de técnicas para selecionar as variáveis a serem incluídas no ajuste. Para discussões sobre o problema de seleção, neste caso, pode-se citar o livro de Draper e Smith (1966), onde é apresentado o método por passos (stepwise), e um elevado número de artigos com propostas de outras técnicas: Anscombe (1967), Tukey (1967), Allen (1971), Mallows (1973), Furnival e Wilson (1974). Um problema que aparece frequentemente é aquele em que a variável dependente é politômica. Neste caso o problema está bem menos estudado e as técnicas pressupõem o conhecimento de um modelo à priori. Veja-se por exemplo Goodman (1971) e (1973).

O objetivo desse trabalho é apresentar - para o caso em que temos uma politomia dependente Y e um conjunto de politomias independentes $X^{(1)}, X^{(2)}, \ldots, X^{(p)}$ - um método de seleção das politomias independentes, sem pressupor um modelo de predição a priori. Ao invés do modelo, trabalharesmos com uma estrutura onde as politomias são classificadas de maneira cruzada. Este método pode ser usado, por exemplo, quando desejamos saber quais as politomias independentes que melhor podem ser usadas para prever a dependente e não de que forma esta previsão pode ou deve ser feita. Esta seleção será feita apenas através de uma medida que nos indique, no senso desejado - o de previsão - o grau de associação entre a politomia dependente e as independentes. Também não será imposta nenhuma condição adicional sobre estas, que poderão ser dos mais variados tipos.

Na escolha da medida usada no processo de seleção, levamos em conta: a ausência de restrições mais fortes sobre as politomias, a indicação de associação no senso de previsão e facilidade de interpretação. Para uma melhor compreensão da escolha a ser usada apresentaremos, no capítulo 1, um estudo de algumas medidas de associação propostas para classificações cruzadas. Um melhor detalhamento, em termos de demonstração, será feito apenas para a medida - denotada por $\lambda(L/C)$ - usada no método de seleção.

Visando o método de seleção reservamos o capítulo 2 para o estudo do estimador da medida e suas propriedades assintóticas. A teoria assintótica foi usada dada a complexidade, teórica e prática, da teoria exata. Neste capítulo nos referiremos, várias vezes, os teoremas de convergência enunciados, sem demonstração mas com referência, no apêndice 1.

O método proposto será apresentado no capítulo 3 e um exemplo de aplicação será discutido no capítulo 4.

No apêndice 2 estão fluxogramas e listagens dos programas com as instruções de uso dos mesmos.

CAPITULO 1

MEDIDAS DE ASSOCIAÇÃO PARA CLASSIFICAÇÕES CRUZADAS

1.1 - Preliminares:

O termo classificação cruzada é usado quando trabalhamos com populações, onde os indivíduos são classificados através de duas ou mais politomias. O caso mais comum é aquele em que temos uma dupla politomia: L e C. A politomia L, que particiona a população em r classes (L_1, L_2, ... L_r) e a politomia C, que particiona a população em s classes (C_1, C_2, ..., C_s). Assim, qualquer indivíduo da população pertencerá, a uma e somente uma, classificação L e a uma e somente uma, classificação C, isto é, a um e somente um cruzamento (L_i, C_j). Um cruzamento (L_i, C_j) será chamado cela.

Uma dupla politomia pode ser representada por uma tabela, chamada tabela de classificação cruzada, da seguinte forma

L/C	C_1	C_2	...	C_j	...	C_s	Total
L_1	ρ_{11}	ρ_{12}	...	ρ_{1j}	...	ρ_{1s}	$\rho_{i.}$
L_2	ρ_{21}	ρ_{22}	...	ρ_{2j}	...	ρ_{2s}	$\rho_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
L_i	ρ_{i1}	ρ_{i2}	...	ρ_{ij}	...	ρ_{is}	$\rho_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
L_r	ρ_{r1}	ρ_{r2}	...	ρ_{rj}	...	ρ_{rs}	$\rho_{r.}$
Total	$\rho_{.1}$	$\rho_{.2}$	...	$\rho_{.j}$	...	$\rho_{.s}$	1

onde:

i) ρ_{ij} é a probabilidade de um indivíduo da população pertencer a cela (L_i, C_j).

ii) $\rho_{i.} = \sum_{j=1}^{s} \rho_{ij}$, é a probabilidade de um indivíduo pertencer a classe L_i.

iii) $\rho_{.j} = \sum_{i=1}^{r} \rho_{ij}$, é a probabilidade de um indivíduo pertencer a classe C_j.

A questão que aparece frequentemente é sobre o grau de associação existente entre as politomias. Para que este grau de associação seja medido é necessário que definamos uma regra, que associe cada tabela de classificação cruzada a um número real; assim sendo, formalmente, uma medida de associação será uma função $f(\underset{\sim}{\rho})$ do espaço de todos os vetores de probabilidade, $\underset{\sim}{\rho} = (\rho_{ij})$, nos reais. Devemos observar que uma medida de associação é um parâmetro populacional.

Várias medidas de associação para classificações cruzadas são propostas na literatura; as que aqui serão apresentadas, no entanto, podem ser divididas em cinco classes:

i) Medidas baseadas na estatística do teste de χ^2 para independência.

ii) Medidas baseadas em produtos cruzados; usadas para o caso de duas dicotomias.

iii) Medidas baseadas na proporção de redução do erro de previsão.

iv) Medidas baseadas na ordem de ocorrência das politomias.

v) Medidas baseadas na proporção de variação explicada.

Quando trabalhamos com medidas de associação devemos, a priori, fazer algumas distinções sobre as politomias. Nós podemos ou não pensar que nossa politomia surge de uma variação contínua; por exemplo, idade pode por conveniência ser dividida em 10 classes, mas é claro, que ela varia continuamente; o mesmo não - acontece se quisermos classificar um indivíduo pela cor do cabelo. Uma politomia pode ser ordenada ou não; algumas vezes a ordem pode ser importante, mas não seu sentido. Pode existir ou não uma relação causal, em uma certa direção,entre as duas politomias; - quando existir uma relação causal em uma certa direção, estaremos no caso assimétrico e caso contrário no caso simétrico. Estas distinções se fazem necessárias, pois estaremos interessados em trabalhar com a medida mais apropriada para uma situação específica.

Outro aspecto para o qual devemos estar atentos, é o da definição das classes de uma politomia, pois esta definição pode afetar nossa medida de associação. No entanto, é desejável que uma medida de associação reflita a maneira como as classes foram definidas. Nós não deveríamos falar, por exemplo, sobre a associação entre nível de renda e nível de educação sem especificar uma particular definição de classes. Vários exemplos deste "efeito de definição de classes" são mostrados em Yule (1912).

É convencional e frequentemente conveniente, que a medida de associação:

i) Tome valores entre -1 e 1 inclusive, com -1 ou +1 no caso de completa associação e 0 no caso de independência, ou

ii) Tome valores entre 0 e 1 inclusive, com 1 no caso de completa associação e 0 no caso de independência.

A convenção (i) é desejável quando a associação tem um sinal (sentido), caso contrário, a convenção (ii) é satisfatória.

Outro cuidado, a ser tomado, é sobre o uso do termo associação completa ; enquanto que independência tem um significado formal - $\rho_{ij} = \rho_{i.} \, \rho_{.j}$ para todo par (i, j) - a associação completa é algo ambíguo e deve ser interpretado de acordo com a medida de associação usada.

1.2 - Medidas de Associação:

1.2.1 - Medidas Tradicionais:

A maior parte das medidas tradicionais são baseadas na estatística do teste χ^2 para independência e na razão dos produtos cruzados no caso de tabelas 2 x 2.

Para uma população finita com ν indivíduos classificados de acordo com duas politomias L e C, definimos $\nu_{ij} = \nu\rho_{.j}$, $\nu_{i.} = \nu\rho_{i.}$, $\nu_{.j} = \nu\rho_{.j}$, etc..

O χ^2 é então dado por

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\nu_{ij} - \nu_{i.} \ \nu_{.j} \ / \ \nu)^2}{\nu_{i.} \ \nu_{.j} \ / \ \nu} =$$

$$= \nu \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\rho_{ij} - \rho_{i.} \ \ \rho_{.j})^2}{\rho_{i.} \ \rho_{.j}} =$$

$$= \nu \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\rho_{ij}^{\ 2}}{\rho_{i.} \ \rho_{.j}} - \nu \quad .$$

Uma grande atenção tem sido dada para o caso $r=s=2$, onde diversas medidas de associação foram propostas baseadas na razão dos produtos cruzados, $\alpha =(\rho_{11} \ \ \rho_{22}) / (\rho_{11} \ \ \rho_{21})$.

Yule, baseado em α, define o seguinte coeficiente de associação:

$$Q = \frac{\nu_{11} \ \nu_{12} - \nu_{12} \ \nu_{21}}{\nu_{11} \ \nu_{22} + \nu_{12} \ \nu_{21}} = \frac{\alpha - 1}{\alpha + 1}$$

cujo numerador ao quadrado é o mesmo do χ^2 para o caso $r=s=2$.

Outro coeficiente sugerido por Yule, também baseado em α, é :

$$Y = \frac{\nu_{11} \ \nu_{22} - \ \ \nu_{12} \ \nu_{21}}{\nu_{11} \ \nu_{22} + \ \ \nu_{12} \ \nu_{21}} = \frac{\alpha - 1}{\alpha + 1}$$

Um coeficiente frequentemente usado para o caso r x s, algumas vezes chamado de contingência quadrática média, é:

$$\phi^2 = \frac{\chi^2}{\nu} .$$

Uma variação sugerida por Karl Pearson (1904) é:

$$C = \sqrt{\frac{\chi^2 / \nu}{1 + \chi^2 / \nu}}$$

que tem sido chamado de coeficiente de contingência ou coeficiente da contingência quadrática média. Outra variação proposta por Tschuprow é

$$T = \sqrt{\frac{\chi^2 / \nu}{(r-1)\ (s-1)}} .$$

Estas duas últimas sugestões, segundo Kendall (1948), foram feitas para normalizar o χ^2, ou seja, tomar valores entre 0 e 1. A primeira porém, não assume o valor 1 (completa associação); por exemplo, em uma tabela r x r com $\rho_{ij} = 0$ para $i \neq j$, nós temos - $\chi^2 = (r-1)$ e $C = (r-1) / r$. Portanto sob completa associação, o valor máximo de C depende do número de linhas e de colunas. A segunda, assume o valor 1 para o caso de completa associação em tabelas r x r, mas não no caso geral r x s. Quando a associação é completa - no sentido de probabilidades positivas apenas em uma das diagonais de uma tabela r x s, - o valor máximo do χ^2 é o min $[(r-1), (s-1)]$, e como $[(r-1)(s-1)]^{1/2} \geq \min [(r-1),(s-1)]$, o máximo de T é menor do que 1. Baseado nestes fatos Cramer (1946) sugere a seguinte variação:

$$V = (\chi^2 / \nu) / \min (r-1, s-1),$$

que toma valores entre 0 e 1 inclusive, e assume os valores 0 e 1 apropriadamente .

O fato que um excelente teste de independência possa ser feito através do χ^2, ou de uma função deste, não significa que o mesmo é uma medida apropriada do grau de associação. Goodman e Kruskal (1954) discutiram a inexistência de defesas convincentes do χ^2 como medida de associação. Outro problema com o uso de medidas tradicionais, é a sua difícil interpretabilidade operacional.

Fisher (1948) discutiu este aspecto no caso amostral, onde um mesmo grau de associação pode ser - em termos do teste de independência - significante para grandes amostras e não significante para pequenas amostras. Se é não significante, nós não temos razão, baseados nos dados, para suspeitar de qualquer grau de associação e é inútil tentar medi-lo. Se por outro lado, o teste é significante, o valor do χ^2 indica o fato, mas não mede o grau de associação. Não temos, portanto, razão para assumir que diferentes valores significantes do χ^2 indiquem diferentes graus de associação, não importando se esta é significante a 1% ou 10%. Para medirmos o grau de associação, é necessário que tenhamos uma hipótese, que indique a natureza do desvio da independência e isto não acontece no teste de χ^2.

1.2.2 - Medidas baseadas na proporção de redução do erro de previsão.

a) Caso Assimétrico:

Consideremos a seguinte situação:

i) Duas politomias L e C.

ii) Nenhuma condição sobre continuidade.

iii) Nenhuma condição de ordem.

iv) Existência de assimetria. A classificação C precede a classificação L, causalmente ou não.

Seja então o seguinte modelo probabilístico: um indivíduo é selecionado aleatoriamente da população; e nós queremos prever sua classe L quando,

1. nenhuma informação sobre sua classe C é dada, ou

2. dada sua classe C.

Sejam $\rho_{(m).} = \max_i \rho_{i.}$ - a maior probabilidade marginal entre as linhas - e $\rho_{(m)j} = \max_i \rho_{ij}$ - a maior probabilidade da coluna j e suponhamos a princípio que $\rho_{(m).}$ e os $\rho_{(m)j}$'s sejam únicos. No caso 1, nossa melhor previsão será para aquela classe L_i, tal que $\rho_{i.} = \rho_{(m)}$, ou seja, a classe com maior probabilidade marginal. No caso 2, nossa melhoı previsão será, dado que ocorreu a classe C_j, prever i para a classe L, tal que $\rho_{ij} = \rho_{(m)j}$; ou seja,a classe L que tem maior probabilidade dentro de coluna j. No caso 1, nossa probabilidade de erro é $1 - \rho_{(m).}$. No caso 2, nossa probabilidade de erro dado C_j é $1 - \rho_{(m)j}$ e a probabilidade total de er ro $1 - \sum_j \rho_{(m)j}$. Nossa suposição de unicidade dos máximos pode ser retirada, pois no caso da existência de outros máximos, qualquer que seja nossa escolha entre eles, os erros não serão afetados.

Baseado neste modelo Louis Guttman (1941) sugeriu a seguinte medida de associação

$$\lambda(L/C) = \frac{(\text{Prob. de erro no caso 1}) - (\text{Prob. de erro no caso 2})}{(\text{Prob. de erro no caso 1})}$$

$$= \frac{\sum_{j=1}^{r} \rho_{(m)j} - \rho_{(m).}}{1 - \rho_{(m).}}$$

que é exatamente a redução relativa na probabilidade do erro entre prever L, não conhecendo C e conhecendo C ou a proporção do erro que pode ser eliminada conhecendo-se C.

Esta medida de associação tem as seguintes propriedades:

i) $\lambda(L/C)$ é indeterminado, se e somente se, a população está concentrada (com probabilidade 1) em uma única linha.

<u>Demonstração:</u>

$\lambda(L/C)$ é indeterminada $\Leftrightarrow 1 - \rho_{(m).} = o$

$\Leftrightarrow \rho_{(m).} = 1$, max $\rho_{(m).} = \rho_{i.}$ para algum i

logo $\rho_{k.} = 0$ para $k \neq i$ pois $\sum_{i=1}^{j} \rho_{i.} = 1$.

ii) Se $\lambda(L/C)$ não é indeterminado, então

$$0 \leq \ (L/C) \leq 1 \ .$$

<u>Demonstração:</u>

$(1-\rho_{(m).}) - (1-\sum_j \rho_{(m)j}) \leq (1-\rho_{(m).})$ pois

$(1 - \sum_j \rho_{(m)j}) \geq 0$

portanto

$$\lambda(L/C) = \frac{(1-\rho_{(m).}) - (1-\sum_j \rho_{(m)j})}{1 - \rho_{(m).}} \leq 1 \ .$$

Agora $0 \leq \rho_{ij} \leq \rho_{(m)j}$ qualquer que seja (i, j)

$\Rightarrow \sum_j \rho_{(m)j} \geq \sum_j \rho_{ij} = \rho_{i.} \quad \forall\ i$

$\Rightarrow \sum_j \rho_{(m)j} \geq \max_i \rho_{i.} = \rho_{(m).}$

$\Rightarrow \sum_j \rho_{(m)j} - \rho_{(m).} \geq 0$

como $\lambda(L/C)$ é determinado, $1 - \rho_{(m).} > 0$, segue-se dos resultados acima que $0 \leq \lambda(L/C) \leq 1$.

iii) $\lambda(L/C) = 0$ se e somente se o conhecimento de C não ajuda na predição de L, isto é, a probabilidade de erro no caso 2 é igual a probabilidade de erro no caso 1. Isto ocorre se todos os máximos de coluna estiverem numa mesma linha.

Este resultado pode ser formalizado da seguinte maneira:

$$\lambda(L/C) = 0 \iff \exists\, k;\ \rho_{kj} = \rho_{(m)j} \quad \forall\, j\ .$$

Demonstração:

(=>) suponhamos que, $\forall\, k, \exists j;\ \rho_{kj} < \rho_{(m)j}$.

$$\Rightarrow \sum_j \rho_{kj} < \sum_j \rho_{(m)j} \quad \forall\, k, \text{ pois } \rho_{ij} \leq \rho_{(m)j}\ \forall\, j$$

$$\Rightarrow \rho_{k.} < \sum_j \rho_{(m)j} \quad \forall\, k$$

$$\Rightarrow \max \rho_{k.} < \sum_j \rho_{(m)j}$$

$$\Rightarrow \rho_{(m).} < \sum_j \rho_{(m)j}$$

$$\Rightarrow 0 < \sum_j \rho_{(m)j} - \rho_{(m).}$$

$\Rightarrow 0 < \lambda(L/C)$ o que é absurdo

logo $k;\ \rho_{kj} = \rho_{(m)j} \quad \forall\, j$.

(<=) $\exists\, k;\ \rho_{kj} = \rho_{(m)j} \quad \forall\, j$

$$\Rightarrow \sum_j \rho_{kj} = \sum_j \rho_{(m)j}$$

$$\Rightarrow \rho_{k.} = \sum_j \rho_{(m)j}$$

$$\Rightarrow \sum_j \rho_{(m)j} \leq \rho_{(m).} = \max_k \rho_{k.}$$

$$\Rightarrow \sum_j \rho_{(m)j} - \rho_{(m).} \leq 0$$

mas $\lambda(L/C) \geq 0$.

Portanto $\sum_j \rho_{(m)j} - \rho_{(m).} = 0$,

ou seja, $\lambda(L/C) = 0$.

iv) $\lambda(L/C) = 1$, se e somente se, o conhecimento da classe C especifica a classe L, isto é, o erro no caso 2 é zero. Isto ocorre se cada coluna tem no máximo um ρ_{ij} positivo. Assim $\lambda(L/C) = 1 \Leftrightarrow$ para cada j, $\exists |\ k = k(j);\ \rho_{ij} = 0 \quad k \neq i$.

Demonstração:

$$(\Rightarrow)\ \lambda(L/C) = \frac{\sum_j \rho_{(m)j} - \rho_{(m).}}{1 - \rho_{(m).}} = 1 \Leftrightarrow$$

$$\Leftrightarrow \sum_j \rho_{(m)j} - \rho_{(m).} = 1 - \rho_{(m).} \Leftrightarrow$$

$\Leftrightarrow \sum_j \rho_{(m)j} = 1$, ou seja a probabilidade de erro no caso 2 é zero.

Como $\sum_i \sum_j \rho_{ij} = 1$ e $\rho_{(m)j} = \rho_{kj}$,

para algum k (que depende j);

$\rho_{ij} = 0$ para $\rho_{ij} \neq \rho_{(m)j}$

$(\Leftarrow)$ Se $\rho_{ij} = 0$ para $\rho_{ij} \neq \rho_{(m)j}$ então

$\sum_j \rho_{(m)j} = 1 \Rightarrow \lambda(L/C) = 1$

v) Em caso de independência estatística, $\lambda(L/C)$, quando determinado, é zero. A recíproca porém não é verdadeira conforme (iii).

Demonstração:

$$\text{Se } \rho_{ij} = \rho_{i.}\,\rho_{.j} \Rightarrow \rho_{(m)j} = \max_i \rho_{ij} =$$

$$= \max_i \rho_{i.}\,\rho_{.j} = \rho_{.j} \max_i \rho_{i.} = \rho_{.j}\,\rho_{(m).}$$

$$\Rightarrow \sum_j \rho_{(m)j} = \sum_j \rho_{.j}\,\rho_{(m).} = \rho_{(m).} \sum_j \rho_{.j} =$$

$$= \rho_{(m).} \Rightarrow \sum_j \rho_{(m)j} - \rho_{(m).} = 0 \Rightarrow$$

$$\Rightarrow \lambda(L/C) = 0\ .$$

vi) $\lambda(L/C)$ é invariante sob permutações de linhas ou de colunas.

Demonstração:

Dado que as expressões envolvidas em $\lambda(L/C)$ - máximos e somas - são invariantes sob a permutação de seus argumentos e - como a permutação destes argumentos é obtida por permutações de linhas e colunas, segue a invariância de $\lambda(L/C)$.

O fato de $\lambda(L/C)$ poder assumir o valor zero sem independência estatística não deve ser considerado como uma desvantagem, pois nós consideramos $\lambda(L/C)$ como uma medida de associação -

em termos de predição (*). Se não existe associação neste sentido, mesmo que exista em outro, nós queremos que $\lambda(L/C)$ seja zero. Além disto, esta propriedade é observada também na maior parte das medidas de associação.

Devemos observar a vantagem desta medida em relação à sua interpretabilidade e as poucas restrições necessárias para sua aplicação. Estes aspectos foram responsáveis pela escolha desta medida para a parte principal deste trabalho que será exposta no capítulo 3.

De maneira análoga podemos definir

$$\lambda(C/L) = \frac{\sum_i \rho_{i(m)} - \rho_{.(m)}}{1 - \rho_{.(m)}}$$

onde

$$\rho_{i(m)} = \max_j \rho_{.j}$$

$$\rho_{.(m)} = \max_j \rho_{.j} \quad ,$$

que tem interpretação e propriedades similares a $\lambda(L/C)$.

b) <u>Caso Simétrico</u>:

Baseado nas medidas (L/C) e (C/L) Goodman e Kruskal (1954) propuseram uma medida de associação para a situação em que as politomias são simétricas.

As restrições sobre ordem e continuidade permanecem não sendo necessárias. O modelo no entanto, passa a ser o seguin

(*) Redução de probabilidade de erro de predição.

te: um indivíduo é selecionado aleatoriamente de uma população; e nós queremos prever ora sua classe A, ora sua classe B (aleatoriamente) nos seguintes casos:

1. Nenhuma informação é dada, ou
2. É dada informação sobre a classe que não está sendo prevista, isto é, se queremos prever a classe L a classe C é dada e vice-versa.

No caso 1, nossa probabilidade de erro de previsão será $1 - 1/2\ (\rho_{(m).} + \rho_{.(m)})$. No caso 2 a probabilidade de erro será $1 - 1/2\ (\sum_j \rho_{(m)j} + \sum_i \rho_{i(m)})$. Considerando a redução obtida nas probabilidades de erro, define-se a seguinte medida

$$\lambda = \frac{1/2\ (\sum_i \rho_{i(m)} + \sum_j \rho_{(m)j} - \rho_{.(m)} - \rho_{(m).})}{1 - 1/2\ (\rho_{.(m)} - \rho_{(m).})} .$$

Esta medida tem as seguintes propriedades:

i) λ é determinado exceto quando toda a população está concentrada em uma única cela.

ii) Se λ é determinado, então $0 \leq \lambda \leq 1$.

iii) $\lambda = 1$, se e somente se, a população está concentrado no máximo em uma unica cela por linha e por coluna.

iv) $\lambda = 0$ em caso de independência estatística, mas a recíproca não é verdadeira.

v) λ é invariante sob permutações de linhas e de colunas.

vi) min $(\lambda(L/C), \lambda(C/L)) \leq \lambda \leq$ max $(\lambda(L/C), \lambda(L/L))$.

As demonstrações dessas propriedades são feitas de maneira análoga às do caso assimétrico.

c) Ponderação das probabilidades:

Em algumas situações, particularmente em comparações entre populações, as medidas $\lambda(L/C)$, $\lambda(C/L)$ e não são aplicáveis por sua dependência das probabilidades marginais. Para melhor ilustrar esta situação, consideramos o exemplo do cruzamento entre duas dicotomias idênticas para duas populações distintas

	C_1	C_2	
L_1	.84	.03	.87
L_2	.04	.09	.13
	.88	.12	1

	C_1	C_2	
L_1	.42	.14	.56
L_2	.02	.42	.44
	.44	.56	1

Apesar das probabilidades condicionais, de L dado C, em ambas as populações serem as mesmas, os valores de $\lambda(L/C)$ diferem.

$$\lambda(L/C) \text{ para população 1} = \frac{.93 - .87}{.13} = .462$$

$$\lambda(L/C) \text{ para população 2} = \frac{.84 - .56}{4} = .636$$

Isso é explicado pela diferença nas probabilidades marginais de coluna entre as duas populações.

Goodman e Kruskal (1954) propuseram para estas situações uma medida similar a $\lambda(L/C)$, onde o modelo assume que todas as classes da politomia C são equiprováveis, e que as probabilidades condicionais da politomia são iguais às probabilidades da população original, ρ_{ij}. Torna-se portanto razoável, repassar as probabilidades ρ_{ij} por

$$\rho_{ij}^* = \frac{1}{s} \frac{\rho_{ij}}{\rho_{.j}} \quad ,$$

e agora com estas probabilidades, ρ_{ij}^*, calcular (L/C). Temos então uma nova medida dada por:

$$\lambda^*(L/C) = \frac{(1/s) \sum_j (\rho_{(m)j} / \rho_{.j}) - (1/s) \max_i \sum_j (\rho_{ij} / \rho_{.j})}{1 - (1/s) \max_i \sum_j (\rho_{ij} / \rho_{.j})}$$

$$= \frac{\sum_j \rho^*_{(m)j} - \rho^*_{(m).}}{1 - \rho^*_{(m).}} \quad .$$

Procedendo desta maneira, obteremos para as duas populações do exemplo dado anteriormente a mesma tabela de classificação cruzada e $\lambda^*(L/C) = .628$.

De modo análogo podemos definir $\lambda^*(C/L)$ e λ^*. A mesma idéia pode ser usada para qualquer outra distribuição marginal - (artificial) apropriada. Yule (1912) diz ser desejável que uma medida de associação seja invariante sob transformações do tipo

$$\rho_{ij} \longrightarrow a_i\, b_j\, \rho_{ij};\ a_i > 0,\ b_j > 0;\ i = 1, \ldots, r;\ j = 1, \ldots, s.$$

Essa transformação pode ser usada para, quando possível, fazer todas as marginais iguais a 1/s (ou 1/r).

d) Associação Parcial:

Quando existem mais que duas politomias em estudo, é natural pensarmos na associação parcial entre duas delas com o efeito das restantes corrigidos ou ajustados de alguma forma.

Por conveniência de notação consideremos o caso de três politomias A, B e C. A probabilidade de um indivíduo da população ser classificado segundo a cela (A_i, B_j, C_k) será denotada por ρ_{ijk} e as probabilidades marginais e máximas de maneira análoga ao caso de duas politomias.

Considerando o modelo de predição ótima assimétrico, duas medidas de associação parcial entre A e B, ajustados para C, são propostas.

A primeira delas considera para cada classe C uma dupla classificação A x B com probabilidades $\rho_{ij.k} = \rho_{ijk} / \rho_{..k}$ ou seja, a probabilidade da cela $(A_i B_j)$ dada a classe C_k. Para cada uma destas classificações podemos calcular $\lambda(B/A)$ (ou $\lambda(A/B)$) e denotá-la por $\lambda_k(B/A)$ (ou $\lambda_k(A/B)$) para mostrarmos sua dependência da classe C_k. Podemos então usar como uma medida de associação parcial entre A e B ajustado para C, a média dos $\lambda_k(B/C)$ ponderados por $\rho_{..k}$, isto é:

$$\lambda_C(B/A) = \sum_{k}^{t} \rho_{..k} \, \lambda_k(B/A) \; .$$

A segunda medida, é baseada diretamente nas probabilidades de erros de previsão. Consideremos o caso onde dada uma classe C, queremos prever B não sendo dada nenhuma informação sobre a classe A. Seguindo a mesma estratégia de previsão baseada em probabilidades máximas, nossa probabilidade de erro será $1 - (\max_j \rho_{.jk}) / \rho_{..k}$. Se agora nos é dada a classe A, nossa probabilidade de erro será $1 - (\sum_i \max_j \rho_{ijk}) / \rho_{..k}$. Tomando a soma para toda classe C e considerando a redução relativa na probabilidade de erro de previsão, teremos uma medida de associação parcial entre A e B ajustados para C dada por:

$$\lambda'_C\,(B/A) = \frac{\sum_k \sum_i \max_j \rho_{ijk} - \sum_k \max_j \rho_{.jk}}{1 - \sum_k \max_j \rho_{.jk}} .$$

Estas duas medidas de associação parcial podem ser estendidas, sem dificuldades, para um número qualquer de politomias.

c) <u>Associação Múltipla</u>:

Dado um conjunto de $p + 1$ politomias Y, $X^{(1)}$, $X^{(2)}, \ldots , X^{(p)}$, pode ser de interesse pensarmos na associação entre uma delas e um subconjunto qualquer das restantes; por exemplo, a associação múltipla entre Y e $(X^{(i_1)}, X^{(i_2)}, \ldots, X^{(i_k)})$, onde $\{i_1, i_2, \ldots , i_k\}$ é um subconjunto de $\{1, 2, \ldots , p\}$. Para medirmos essa associação podemos construir a classificação cruzada de $Y \times (X^{(i_1)}, X^{(i_2)}, \ldots , X^{(i_k)})$, onde o vetor

$(X^{(i_1)}, X^{(i_2)}, \dots, X^{(i_k)})$ pode ser considerado - através de uma renumeração - como uma nova politomia e o problema se reduz à associação entre duas politomias. Por exemplo, consideremos $k = 2$ e as politomias Y, $X^{(i_1)}$ e $X^{(i_2)}$ com r, s e t classes, respectivamente. Para calcularmos a associação múltipla entre Y e $(X^{(i_1)}, X^{(i_2)})$ podemos construir a tabela

	$X^{(i_1)}_1\ X^{(i_2)}_1$	$X^{(i_1)}_1\ X^{(i_2)}_2$	...	$(X^{(i_1)}_1\ X^{(i_2)}_t)$	$(X^{(i_1)}_2\ X^{(i_2)}_1$	...	$X^{(i_1)}_2\ X^{(i_2)}_t$	...	$X^{(i_1)}_s\ X^{(i_2)}_t$
Y_1	ρ_{111}	ρ_{112}	...	ρ_{11t}	ρ_{121}	...	ρ_{12t}	...	ρ_{1st}
Y_2	ρ_{211}	ρ_{212}	...	ρ_{21t}	ρ_{221}	...	ρ_{22t}	...	ρ_{2st}
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
Y_r	ρ_{r11}	ρ_{r12}	...	ρ_{r1t}	ρ_{r21}	...	ρ_{r2t}	...	ρ_{rst}

e tomar os pares $(X^{(i_1)}_i, X^{(i_2)}_j)$ como sendo as st classes de uma politomia $X^{(i_1, i_2)}$. A associação mútlipla entre Y e $(X^{(i_1)}, X^{(i_2)})$ será então, a associação simples entre Y e $X^{(i_1, i_2)}$.

Devemos observar, no entanto, que mesmo que exista uma ordenação entre as politomias originais, esta ordenação é perdida quando da agregação. Isto nos induz a usar a medida $\lambda(L/C)$ - para medirmos a associação múltipla, pois como vimos anteriormente esta medida não pressupõe ordenação. Neste caso a notação a ser usada será $\lambda(Y / X^{(i_1)}, X^{(i_2)}, \dots, X^{(i_k)})$.

1.2.3 - <u>Medida baseada na ordem de ocorrência das politomias.</u>

Até aqui foram consideradas medidas invariantes sob permutações de linhas e colunas nas tabelas de classificação cruzada, ou seja, invariantes sob a ordem em que as politomias eram classificadas. Goodman e Kruskal (1954) também propuseram uma medida de associação para o caso onde a ordenação é de interesse.

Suponhamos a seguinte situação:

i) Duas politomias L e C.

ii) Nenhuma condição sobre a continuidade.

iii) A direção de ordenação é de interesse.

iv) As duas politomias aparecem de maneira simétrica.

Por (iii) desejamos dizer que deve ser feita a distinção entre as tabelas

$$\begin{array}{c|c|c} \rho_{11} & 0 & 0 \\ \hline 0 & \rho_{12} & 0 \\ \hline 0 & 0 & \rho_{13} \end{array} \qquad \text{e} \qquad \begin{array}{c|c|c} 0 & 0 & \rho_{13} \\ \hline 0 & \rho_{22} & 0 \\ \hline \rho_{31} & 0 & 0 \end{array}$$

em termos de associação. Na primeira diremos que existe uma associação completa e na segunda uma contra associação completa. Assim, é desejável que a medida proposta tome os valores +1 e -1, respectivamente.

Para a definição de uma tal medida consideremos o seguinte modelo: dois indivíduos são selecionados aleatoriamente, com reposição, de uma população; sejam (L_{I1}, C_{J1}) e (L_{I2}, C_{J2}) - suas respectivas celas. Devemos observar que I1, J1, I2, J2 são variáveis aleatórias.

Se existir independência entre as politomias L e C isto se refletirá na existência de uma conexão entre a ordem das variáveis aleatórias I's e a ordem das J's. Se existir associação, a ordem dos I's deverá ser a mesma das J's. Se existir contra associação o inverso ocorrerá.

Consideremos as seguintes probabilidades:

$$\Pi_m = P\left[(L_{I1} < L_{I2} \text{ e } C_{J1} < C_{J2}) \text{ ou } (L_{I1} > L_{I2} \text{ e } C_{J1} > C_{J2})\right]$$

$$\Pi_d = P\left[(L_{I1} < L_{I2} \text{ e } C_{J1} > C_{J2} \text{ ou } (L_{I1} > L_{I2} \text{ e } C_{J1} < C_{J2})\right]$$

$$\Pi_i = P\left[L_{I1} = L_{I2} \text{ ou } C_{J1} = C_{J2}\right] .$$

Para evitar ambiguidades tomemos estas probabilidades de "ordem" e "não ordem" condicionadas à ausência de empates, ou seja, $\Pi_m / 1 - \Pi_i$ e $\Pi_d / 1 - \Pi_i$.

A associação neste sentido, pode então ser medida considerando a diferença entre as probabilidades condicionais da mesma ordem e ordem contrária; ou em outras palavras, o quanto é mais provável se obter mesma ordem do que ordens contrárias em duas classificações quando dois indivíduos são escolhidos aleatoriamente, com reposição, de uma população. Esta medida de associação nos é dada então por,

$$\gamma = \frac{\Pi_m - \Pi_d}{1 - \Pi_i} .$$

Dado que $\Pi_m + \Pi_d + \Pi_i = 1$, podemos escrever

$$\gamma = \frac{2\,\Pi_m - 1 + \Pi_i}{1 - \Pi_i}$$

e usar a relação

$$\Pi_m = 2 \sum_i \sum_j \rho_{ij} \left[\sum_{i'>i} \sum_{j'>j} \rho_{i'j'}\right]$$

$$\Pi_i = \sum_i \rho_{i.}^{\,2} + \sum_j \rho_{.j} - \sum_i \sum_j \rho_{ij}^{\,2} \quad .$$

Algumas propriedades importantes de γ são:

i) γ é indeterminado se a população está concentrada em uma mesma linha ou coluna da tabela de classificação cruzada.

ii) γ é 1 se a população está concentrada na diagonal principal da tabela e -1 na diagonal secundária.

iii) γ é 0 em caso de independência, mas a recíproca não é verdadeira, exceto no caso 2 x 2.

No caso 2 x 2, γ coincide com o coeficiente de associação de Yule.

1.2.4 - Medida Baseada na Proporção de Variação Explicada.

Gini (1912) definiu a variação total entre r possíveis categorias (classes) de ν respostas categóricas como

$$\frac{\nu}{2} - \frac{1}{2\nu} \sum_{i=1}^{r} \nu_i$$

onde ν_i é o número de respostas na i-ésima classe, i = 1,2,...,r. Light e Margolin (1971) mostraram que esta medida de variação de Gini tem duas propriedades esperadas de uma medida de variação. Primeira, ela é zero se e somente se, toda a população está concentrada em uma única classe. Segunda, ela é máxima quando as respostas estão igualmente distribuídas entre as classes. Uma terceira propriedade é que: a agregação de duas categorias em uma não aumenta a variação.

Usando a definição de variação de Gini e a terminologia de análise de variância, temos que a soma de quadrados total em uma classificação cruzada r x s será :

$$TSS = \frac{\nu}{2} - \frac{1}{2\nu} \sum_{i=1}^{r} \nu_{i.}^{2} \quad .$$

Esta variação é então particionada em uma soma de quadrados dentro de grupos e uma soma de quadrados entre grupos

$$WSS = \frac{\nu}{2} - \frac{1}{2} \sum_{i=1}^{r} \sum_{j}^{s} \left(\frac{\nu_{ij}^{2}}{\nu_{.j}}\right)$$

$$BSS = TSS - WSS =$$

$$= \frac{1}{2} \sum_{i} \sum_{j} \left(\frac{\nu_{ij}^{2}}{\nu_{.j}}\right) - \frac{1}{2\nu} \sum_{i} \nu_{i.}^{2} =$$

$$= \frac{1}{2} \sum_{i} \sum_{j} \left[\frac{1}{\nu_{.j}} \left(\nu_{ij} - \frac{\nu_{i.}\ \nu_{.j}}{\nu}\right)^{2}\right]$$

Fazendo uma analogia qualitativa para o coeficiente - de determinação para dados contínuos, Light e Margolin propuseram uma medida da proporção de variação nas linhas devido as colunas, como sendo a razão BSS/TSS; que após algumas simplificações pode ser escrito em termos de probabilidade como

$$\tau(L/C) = \frac{\sum_i \sum_j (\rho_{ij}^2 / \rho_{.j}) - \sum_i \rho_{i.}^2}{1 - \sum_i \rho_{i.}^2}$$

$$= \frac{\sum_i \sum_j \left[(\rho_{ij} - \rho_{i.} \ \rho_{.i})^2 / \ \rho_{.j}\right]}{1 - \sum_i \rho_{i.}^2} \quad .$$

É interessante observar a semelhança com o χ^2. Quando $\rho_{i.} = 1/r$ para $i = 1, \ldots, r$, $\tau(L/C) = \chi^2 / (r-1)$. Quando $r = 2$ $\tau(L/C) = \chi^2$, para todo valor de $\rho_{1.}$ e $\rho_{2.}$.

Esta medida tem as seguintes propriedades:

i) $\tau(L/C)$ é indeterminado se existe $\rho_{i.} = 1$.

ii) Se $\tau(L/C)$ é determinado, então $0 \leq \tau(L/C) \leq 1$.

iii) $\tau(L/C)$ é zero em caso de independência.

iv) $\tau(L/C)$ é um se existe j tal que $\rho_{ij} = \rho_{.j}$.

v) $\tau(L/C)$ é invariante sob permutações de linhas e colunas.

De maneira similar podemos definir a medida $\tau(C/L)$.

CAPÍTULO 2

TEORIA AMOSTRAL E ASSINTÓTICA

2.1. - Preliminares:

No capítulo anterior, tratamos de medidas de associação para classificações cruzadas em termos populacionais; nesse capítulo trataremos do aspecto amostral desta teoria.

Quando trabalhamos com amostras, tabelas de classificação cruzada também recebem o nome de tabelas de contingência; as interpretações porém são análogas.

Admitiremos nesse capítulo, que o tamanho da amostra - denotado por n - é conhecido; e que, exceto quando explicitado, estaremos trabalhando com duas politomias, L e C, com r e s classes, respectivamente.

2.2. - Métodos Amostrais:

Existem vários métodos de se amostrar as rs celas de uma população classificada de acordo com duas politomias. Podemos por exemplo, retirar uma amostra com reposição (*) ou sem reposição. No primeiro caso, nossa distribuição amostral será uma multinomial e no segundo uma multi-hipergeométrica.

Outra variação possível é obtida, se ao invés de amostrarmos conjuntamente as rs celas, retirarmos s amostras indepen

(*) (ou de uma população infinita).

dentes para cada classe C; nossa distribuição amostral será então, o produto de s distribuições multinomiais ou multi-hipergeométrica, respectivamente para amostragens com e sem reposição.

Nesse trabalho, assumiremos populações infinitas, ou seja, tecnicamente amostragem com reposição.

2.3. - Estimadores:

Como vimos no capítulo anterior, uma medida de associação é uma função $f(\underset{\sim}{\rho})$, que leva todos os vetores de probabilidades, $\underset{\sim}{\rho} = (\rho_{ij})$, nos reais. A questão que aparece é de como estimar $f(\underset{\sim}{\rho})$. Como estamos trabalhando com vetores de rs parâmetros - na verdade rs-1 parâmetros independentes - adotaremos a seguinte convenção para estimar $f(\underset{\sim}{\rho})$: se $\underset{\sim}{\hat{\rho}}$ é o estimador de máximo verossimilhança de $\underset{\sim}{\rho}$, então $f(\underset{\sim}{\hat{\rho}})$ é o "estimador de máximo verossimilhança" de $f(\underset{\sim}{\rho})$. Essa convenção é justificada pela invariância dos estimadores deste tipo sob reparametrizações.

É resultado bastante conhecido, que a frequência relativa é o estimador de máximo verossimilhança dos parâmetros de uma distribuição multinomial.

2.4. - Notação:

A colocação de um tópico reservado à notação nesse capítulo, se justifica no sentido de um melhor esclarecimento sobre os tipos de variação, que podem ocorrer quando trabalhamos com quantidades amostrais.

No geral, letras latinas maiúsculas serão usadas para denotar variáveis aleatórias, exceto o estimador de $\lambda(L/C)$

que será denotado por $\hat{\lambda}(L/C)$; as minúsculas para quantidades fixas e as letras gregas, coerentemente com o capítulo anterior, para parâmetros populacionais.

Assim, para uma amostra de tamanho n, denotaremos o número de indivíduos na cela (L_i, C_j) por N_{ij}, e portanto

$$\sum_i \sum_j N_{ij} = n .$$

Para o caso de uma amostragem multinomial nas rs celas, as quantidades marginais serão denotadas por

$$N_{i.} = \sum_j N_{ij} \quad \text{e} \quad N_{.j} = \sum_i N_{ij} .$$

O mesmo não acontece se forem feitas s amostras independentes para cada classe C, pois neste caso, as marginais desta politomia - não serão aleatórias e devem ser denotadas por

$$n_{.j} = \sum_i N_{ij} .$$

A proporção de indivíduos na cela (L_i, C_j), N_{ij} / n, será denotada por P_{ij} e as marginais por $P_{i.}$ e $P_{.j}$ ou $p_{i.}$ e $p_{.j}$, dependendo da aleatoriedade ou não.

Esta notação básica é estendida então, para os análogos amostrais de $\rho_{(m)j}$, $\rho_{(m).}$, etc. , ou seja, $P_{(m)j} = \max_i P_{ij}$, $P_{(m).} = \max_i P_{i.}$ etc..

A rigor, principalmente quando tratarmos de convergência, nossa notação deveria ser indexada pelo tamanho da amostra, mas por simplicidade este índice será omitido.

2.5. - Teoria Assintótica

Como foi dito na introdução, estudaremos a distribuição do estimador $\lambda(L/C)$ em termos assintóticos. Trabalhos de Greenwood (1950) e Kozelkc (1952 e 1956), sobre distribuição exata de frequência máxima observada em uma amostra de uma distribuição multinomial, mostram a complexidade da teoria exata e justificam plenamente o uso de métodos assintóticos. Este problema existe também quando trabalhamos com as medidas tradicionais e outros tipos de medidas de associação estudadas no capítulo anterior.Por outro lado, a teoria assintótica é comumente usada em problemas relacionados com classificação cruzada e na maioria das vezes plenamente satisfatória.

Considerando as rs variáveis aleatórias N_{ij} sendo regidas por uma distribuição multinomial com parâmetros n e $\underset{\sim}{\rho}$ - amostragem conjunta em todas as celas - temos que, para qualquer tamanho de amostra n:

$$E\left[\sqrt{n}\,(P_{ij}-\rho_{ij})\right]=0\ ,$$

$$Var\left[\sqrt{n}\,(P_{ij}-\rho_{ij})\right]=\rho_{ij}\,(1-\rho_{ij})$$

$$Cov\left[\sqrt{n}\,(P_{ij}-\rho_{ij})\ ,\ \sqrt{n}\,(P_{k\ell}-\rho_{k\ell})\right]=$$

$$=-\rho_{ij}\,\rho_{k\ell}\quad\text{para } (i,j)\neq(k,\ell)\ .$$

Usando estes resultados da distribuição multinomial e os teoremas (*) A1 e A6, segue que:

(*) *Os teoremas referidos neste capítulo se encontram enunciados com referência no Apêndice 1.*

i) P_{ij} converge em probabilidade para ρ_{ij} qualquer que seja (i, j), e

ii) O vetor $U_n = (\sqrt{n}(P_{ij} - \rho_{ij}))$, $i = 1,2,\ldots,r$; $j = 1,2,\ldots,s$

converge em distribuição para uma normal multivariada singular com vetor média zero e matriz de covariância $\Sigma = \Lambda_\rho - \rho\rho'$, onde Λ_ρ é uma rs x rs matriz diagonal baseada em $\underset{\sim}{\rho}$, isto é, os elementos da diagonal são os elementos do vetor ρ; e a singularidade de Σ se deve a restrição que $\sum_i \sum_j \rho_{ij} = 1$.

Para estudarmos as propriedades assintóticas do estimador de $\lambda(L/C)$, três suposições devem ser feitas sobre a população:

S1 - Para cada j, seja único o índice i tal que

$\rho_{(m)j} = \rho_{ij}$,

S2 - Seja único o índice i tal que $\rho_{(m).} = \rho_{i.}$

S3 - $\rho_{(m).}$ seja diferente de 1.

Os particulares índices de S1 e S2 serão denotados por i_j (para mostrar a dependência de j) e por i* respectivamente. A suposição S3 nos garante que $\lambda(L/C)$ não é indeterminado.

Sob estas suposições, pela convergência em probabilidade dos P_{ij} e usando os teoremas A2 e A3 temos que:

$(P_{ij} - P_{i_j j})$ converge em probabilidade para $(\rho_{ij} - \rho_{i_j j})$. Assim

$$\max_i (P_{ij} - P_{i_j j}) = \max_i (P_{ij}) - P_{i_j j} = P_{(m)j} - P_{i_j j}$$

converge em probabilidade para

$$\max_i (\rho_{ij} - \rho_{i_j j}) = \max_i (\rho_{ij}) - \rho_{i_j j} =$$

$$= \rho_{(m)j} - \rho_{i_j j} = \rho_{i_j j} - \rho_{i_j j} = 0 \; .$$

Segue do teorema A4 que $(P_{(m)j} - P_{i_j j})$ tem distribuição limite degenerada com massa no zero, ou em outras palavras, $P(P_{(m)j} = P_{i_j j})$ converge a um.

De maneira análoga podemos mostrar que $P(P_{(m).} = P_{i*.})$ converge a um.

2.6. - Distribuição Assintótica do Estimador de λ(L/C)

Como vimos no capítulo 1, a medida λ(L/C) é dada por

$$\frac{\sum_j \rho_{(m)j} - \rho_{(m).}}{1 - \rho_{(m).}} \; .$$

Pela convenção dada em 2.3, o seu estimador será:

$$\hat{\lambda}(L/C) = \frac{\sum_j P_{(m)j} - P_{(m).}}{1 - P_{(m).}} \; .$$

Esse estimador é definido, exceto quando $P_{(m).} = 1$. Pela suposição S3, $\rho_{(m).} \neq 1$, a probabilidade que $P_{(m).} = 1$ convergirá a zero e o teorema A5 permite que este fato, $P_{(m).} = 1$, seja desprezado em termos assintóticos.

Para analisarmos o comportamento assintótico de $\hat{\lambda}(L/C)$, consideremos as seguintes igualdades

$$\sqrt{n}\,[\hat{\lambda}(L/C) - (L/C)] = \sqrt{n}\left(\frac{\sum_j P_{(m)j} - P_{(m).}}{1 - P_{(m).}} - \frac{\sum_j \rho_{(m)j} - \rho_{(m).}}{1 - \rho_{(m).}}\right)$$

$$= \frac{\sqrt{n}\left[\sum_j (P_{(m)j} - \rho_{(m)j})(1-\rho_{(m).}) - (P_{(m).} - \rho_{(m).})(1 - \sum_j \rho_{(m)j})\right]}{(1-\rho_{(m).})^2}\;\frac{1-\rho_{(m).}}{1-P_{(m).}}$$

Pela convergência em probabilidade dos $P_{i.}$ e pela suposição S3, o teorema A2 nos garante que $(1 - \rho_{(m).})/(1 - P_{(m).})$ converge em probabilidade para 1. Segue pelo teorema A6, que assintoticamente $\sqrt{n}(\hat{\lambda}(L/C) - \lambda(L/C))$ e

$$\frac{\sqrt{n}\left[\sum_j (P_{(m)j} - \rho_{(m)j})\;(1-\rho_{(m).}) - (P_{(m).} - \rho_{(m).})\;(1 - \sum_j \rho_{(m)j})\right]}{(1-\rho_{(m).})^2}$$

tem a mesma distribuição.

Na expressão acima os termos $(1 - \rho_{(m).})$ e $(1 - \sum_j \rho_{(m)j})$ são constantes, assim por simplicidade de notação passaremos a escrever.

$$\sqrt{n}\,\left[a \sum_j (P_{(m)j} - P_{(m)j}) - b\,(P_{(m).} - \rho_{(m).})\right]$$

onde

$$a = \frac{(1 - \rho_{(m).})}{(1 - \rho_{(m).})^2} = \frac{1}{1 - \rho_{(m).}}$$

$$b = \frac{1 - \sum_j \rho_{(m)j}}{(1 - \rho_{(m).})^2} \quad .$$

Consideremos para x, um número real, i_j e i^* com os significados anteriormente definidos, os seguintes eventos

$$V' = \{ \sqrt{n}\left[a \sum_j (P_{(m)j} - \rho_{(m)j}) - b(P_{(m).} - \rho_{(m).})\right] \leq x\}$$

$$W_j = \{P_{(m)j} = P_{i_j j}\}$$

$$W^* = \{P_{(m).} = P_{i^*.}\}$$

$$V = \{ \sqrt{n}\left[a \sum_j (P_{i_j j} - \rho_{i_j j}) - b(P_{i^*.} - \rho_{i^*.})\right] \leq x\} \quad .$$

É fácil verificar que

$$V' = V \cap (\bigcap_j W_j) \cap W^* \quad .$$

Como vimos anterioemente as probabilidades dos eventos W_j e W^* convergem a 1, segue que a probabilidade do evento $(\bigcap_j W_j) \cap W^*$ também converge a 1. Com estes resultados o teorema A6 nos diz que a probabilidade do evento V' converge para o mesmo limite da probabilidade de V. Em outras palavras, o conjunto de resultados até agora vistos, nos permite afirmar que a distribuição assintótica de $\sqrt{n}(\hat{\lambda}(L/C) - \lambda(L/C))$ é igual a distribuição assintótica de

$$\sqrt{n}\ \left[a \sum_{j} (P_{i_j j} - \rho_{i_j j}) - b\ (P_{i*.} - \rho_{i*.})\right] .$$

Esta última expressão pode ser escrita como

$$\sqrt{n}\ \left[a \sum_{j} (P_{i_j j} - \rho_{i_j j}) - b \sum_{j} (P_{i*j} - \rho_{i*j})\right]$$

ou em notação matricial por

$$\Pi'\ U_n$$

onde Un é o vetor de dimensão rs.

$$(\ \sqrt{n}\ (P_{ij} - \rho_{ij})\ :\ i = 1,2,\ldots,r\ ;\ j = 1,2,\ldots,s)$$

e Π é o vetor de dimensão rs em que a k-ésima coordenada pode ser escrita, usando o delta de Kronecker, por

$$\Pi_k = a\ \delta_{k\ell} - b\ \delta_{k\ell*}$$

sendo que:

$$\ell = i_1,\ r + i_2,\ 2r + i_3, \ldots,\ (s - 1)r + i_s$$

e

$$\ell* = i*,\ r + i*,\ 2r + i*, \ldots,\ (s - 1)r\ + i*$$

Como vimos anteriormente, U_n tem uma distribuição assintótica normal multivariada singular com média zero e matriz de covariância Σ. Sabemos então que $\Pi'U_n$ tem assintoticamente uma - distribuição normal com média zero e variância dada por $\Pi'\Sigma\Pi$.

Para encontrarmos uma forma explícita para a variância - ao invés de trabalharmos diretamente sobre $\Pi'\Sigma\Pi$ - consideremos, para um valor fixo de n, a expressão

$\sqrt{n}\ [a \sum_j (P_{i_j j} - \rho_{i_j j}) - b \sum (P_{i^*j} - \rho_{i^*j})]$ temos que:

$$Var\ \{\sqrt{n}\ [a \sum_j (P_{i_j j} - \rho_{i_j j}) - b \sum_j (P_{i^*j} - \rho_{i^*j})]\} =$$

$$= n\ [a^2\ Var\ (\sum_j P_{i_j j}) + b^2\ Var\ (\sum_j P_{i^*j}) - 2\ Cov\ (\sum_j P_{i_j j}\ \sum P_{i^*j})]$$

Podemos então considerar $\sum_j P_{i_j j}$ como sendo a propor ção de sucessos em n ensaios de Bernouilli com probabilidade de sucesso igual a $\sum_j \rho_{i_j j}$; analogamente $P_{i^*.} = \sum_j P_{i^*j}$ como a propor ção de sucessos em ensaios com probabilidade $\rho_{i^*.} = \sum_j \rho_{i^*j}$. Assim sendo:

$$Var\ (\sum_j P_{i_j j}) = n^{-1} \sum_j \rho_{i_j j}\ (1 - \sum_j \rho_{i_j j})$$

e

$$Var\ (P_{i^*.}) = Var\ (\sum_j P_{i^*j}) = n^{-1} \sum_j \rho_{i^*j}\ (1-\sum_j \rho_{i^*j}).$$

Para calcularmos a covariância entre $\sum_j P_{i_j j}$ e P_{i^*}, devemos tomar cuidado com o fato que i_j pode ser igual a i^* para alguns valores de j. Tomemos então R_1 como sendo a soma dos $P_{i_j j}$ para todos os valores de j onde $i_j = i^*$ (esta soma será denotada por $\sum^* P_{i_j j}$), R_2 como a soma para todo j dos $P_{i_j j}$ menos R_1 e R_3 como $P_{i^*.}$ menos R_1. Com este artifício temos que R_1, R_2, R_3 são proporções multinomiais com probabilidades, respectivamente:

$$\sum_{\{j:i_j=i^*\}} \rho_{i_j j} = \sum^* \rho_{i_j j}\ ,\quad \sum_j \rho_{i_j j} - \sum^* \rho_{i_j j} \qquad e$$

$\sum_j P_{i*j} - \Sigma^* \rho_{i_j j}$.

Além disso $\sum_j P_{i_j j} = R_1 + R_2$ e $\sum_j P_{i*j} = R_1 + R_3$. Usando então os resultados da distribuição multinomial temos:

$$\mathrm{Cov}\ (\sum_j P_{ijj} , \sum_j P_{i*j}) =$$

$$= \mathrm{Cov}\ (R_1 + R_2 , R_1 + R_3) =$$

$$= \mathrm{Var}\ (R_1) + \mathrm{cov}\ (R_1, R_2) + \mathrm{cov}\ (R_1, R_3) + \mathrm{cov}\ (R_2, R_3)$$

$$= n^{-1} [\Sigma^* \rho_{i_j j} (1 - \Sigma^* \rho_{i_j j}) - \Sigma^* \rho_{i_j j} (\sum_J \rho_{i_j j} - \Sigma^* \rho_{i_j j}) +$$

$$- \Sigma^* \rho_{i_j j} (\sum_j \rho_{i_j j} - \Sigma^* \rho_{i_j j}) - (\sum_j \rho_{i_j j} - \Sigma^* \rho_{i_j j}) (\sum_j \rho_{i*j} - \Sigma^* \rho_{i_j j})]$$

$$= n^{-1} [\Sigma^* \rho_{i_j j} - \sum_j \rho_{i_j j} \sum_j \rho_{i*j}] .$$

Substituindo todos estes resultados, após algumas simplificações, encontramos:

$$\mathrm{Var}\ \{ \sqrt{n} [a \sum_j (P_{i_j j} - \rho_{i_j j}) - b \sum_j (P_{i*j} - \rho_{i*j})]\} =$$

$$= (1 - \sum_j \rho_{i_j j}) (\sum_j \rho_{i_j j} + \sum_j \rho_{i*j} - 2\Sigma^* \rho_{i_j j}) / (1 - \sum_j \rho_{i*j})^3 .$$

Esta variância é zero se e somente se

$\Sigma^* \rho_{i_j j} = \sum_j \rho_{i_j j} = \sum_j \rho_{i^* j}$ ou $\sum_j \rho_{i_j j} = 1$. A primeira situação ocorre se $i_j = i^*$ para todo j, ou seja, se na população todos os máximos de coluna estão em uma mesma linha ($\lambda(L/C) = 0$); e a segunda se cada coluna tem no máximo um ρ_{ij} diferente de zero ($\lambda(L/C) = 1$). Por outro lado, sob a suposição S3, esta variância nunca será indeterminada. Por último devemos observar que esta variância, calculada para um valor fixo de n, não depende do tamanho da amostra sendo portanto igual a variância assintótica.

Resumindo nossos resultados até agora, podemos afirmar que: sob as suposições S1, S2 e S3, se $\lambda(L/C)$ é diferente de zero ou um, $\sqrt{n}\ (\hat{\lambda}(L/C) - \lambda(L/C))$ tem distribuição assintótica normal com média zero e variância, usando agora nossa notação original, dada por:

$$(1 - \sum_j \rho_{(m)j})\ (\Sigma\ \rho_{(m)j} + \rho_{(m).} - 2\Sigma^*\ \rho_{(m).})/(1-\rho_{(m).})^3$$

De acordo com 2.3, podemos tomar para $0 < \lambda(L/C) < 1$,

$$(1 - \sum_j P_{(m)j})\ (\sum_j P_{(m)j} + P_{(m).} - 2\ \Sigma^*\ P_{(m)j})\ /\ (1 - P_{(m).})^3$$

como o estimador da variância assintótica de $\sqrt{n}\ (\hat{\lambda}(L/C) - \lambda(L/C))$. Sob esta condição, pelo teorema A2 , este estimador é consistente, ou seja, converge em probabilidade para a variância assintótica. Aplicando-se o teorema A2 segue que

$$\sqrt{n}(\hat{\lambda}(L/C)-\lambda(L/C))\ x\quad (1-P_{(m).})^3/(1-\sum_j P_{(m)j})(\sum_j P_{(m)_j}+P_{(m).}-2\Sigma P_{(m)j})$$

é assintoticamente normal com média zero e variância um.

Estes resultados nos permitem trabalhar (em termos assintóticos) com intervalos de confiança e testes de hipótese sobre $\lambda(L/C)$ quando este é diferente de zero ou um. Se $\lambda(L/C) = 0$, a probabilidade de $\hat{\lambda}(L/C) = 0$ converge a 1. Se $\lambda(L/C) = 1$ $\hat{\lambda}(L/C) = 1$ sem qualquer aproximação assintótica. Hipóteses do tipo $\lambda(L/C) = 0$ ou 1, com qualquer nível de significância, serão rejeitadas sempre que $\hat{\lambda}(L/C) \neq 0$ ou 1, respectivamente; intervalos de confiança devem excluir os pontos 0 e 1, a menos que $\hat{\lambda}(L/C) = 0$ ou 1 quando os intervalos consistirão somente dos pontos 0 e 1, respectivamente.

2.7 - Adequacidade da Aproximação

Goodman e Kruskal (1963) mostraram a adequacidade da aproximação de $\hat{\lambda}(L/C)$ para as seguintes situações:

Tipo de Classificação	$\lambda(L/C)$	Tamanho da Amostra	Número de Amostras	Figura
3 x 4	1/3	200	50	A
3 x 4	1/3	100	50	B
3 x 4	1/3	100	100	C
4 x 3	$\simeq$ 4/10	200	50	D
4 x 3	$\simeq$ 4/10	100	50	E
4 x 3	$\simeq$ 4/10	100	100	F
2 x 3	1/10	200	50	G
3 x 2	4/10	200	50	H

Os desvios da reta podem ser atribuídos a duas fontes:

inadequacidade da aproximação e flutuações amostrais. Podemos observar nos gráficos apresentados que a aproximação não é satisfatória para as probabilidades de cauda principalmente para λ(L/C) distante de 0.5. Este fato era esperado tendo em vista uma possível assimetria na distribuição de $\hat{\lambda}$(L/C) para o caso finito. No geral, uma comparação com os gráficos mostrados por Daniel e Wood (1971), para pequenas amostras de uma distribuição normal, nos indica que nossa aproximação é satisfatória.

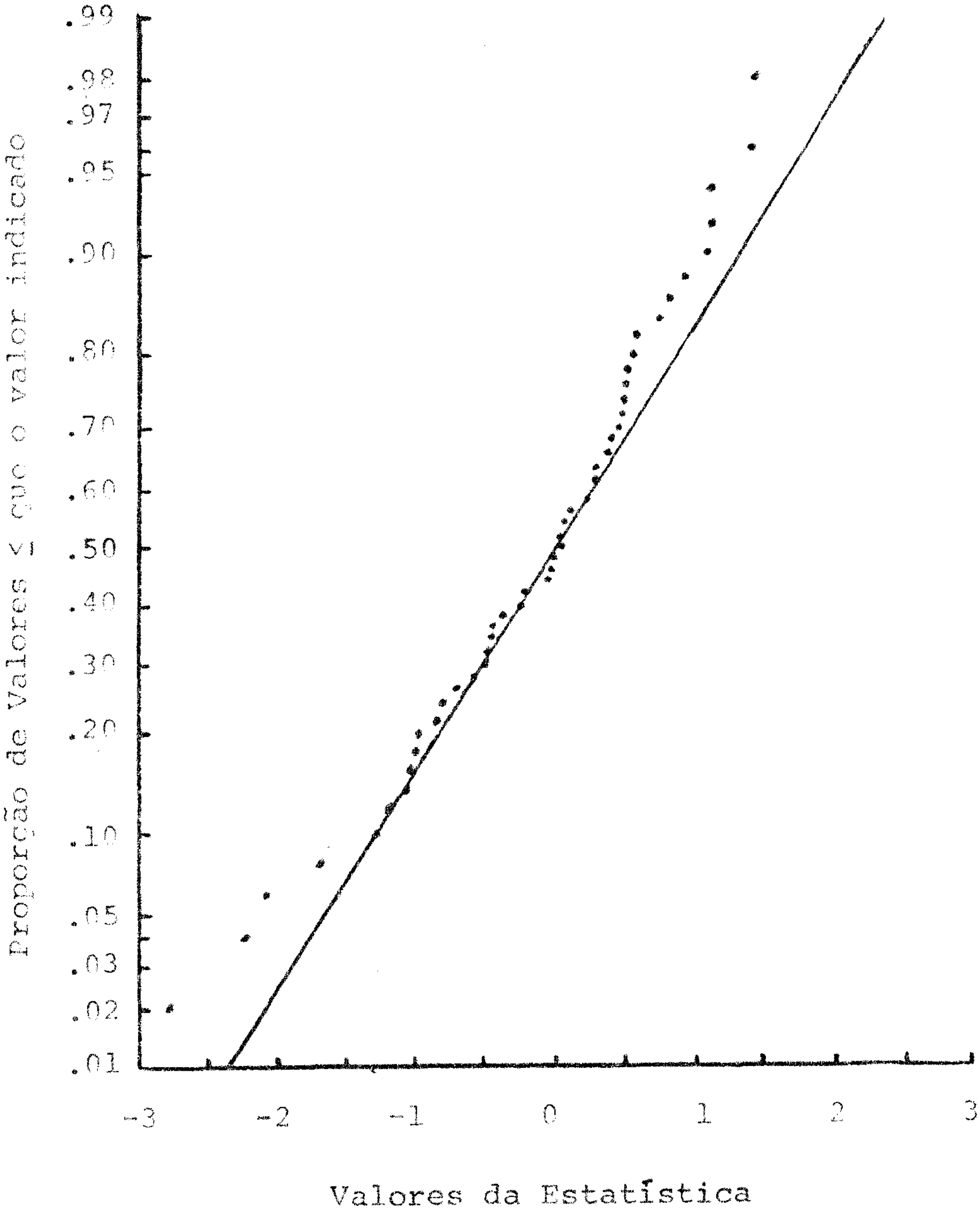

Proporção de Valores ≤ que o valor indicado
.99
.98
.97
.95
.90
.80
.70
.60
.50
.40
.30
.20
.10
.05
.03
.02
.01
-3
-2
-1
0
1
2
3
Valores da Estatística

FIGURA A

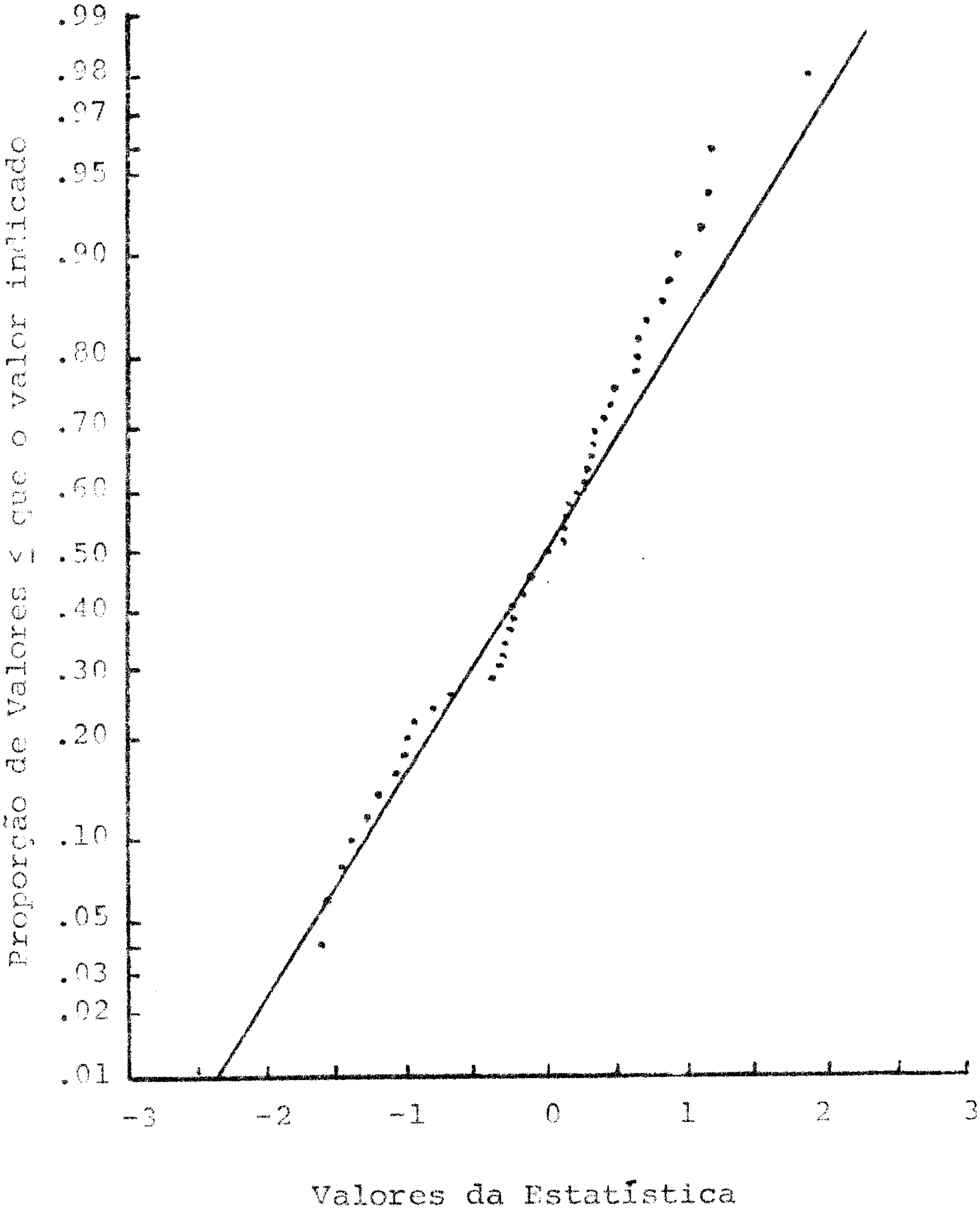
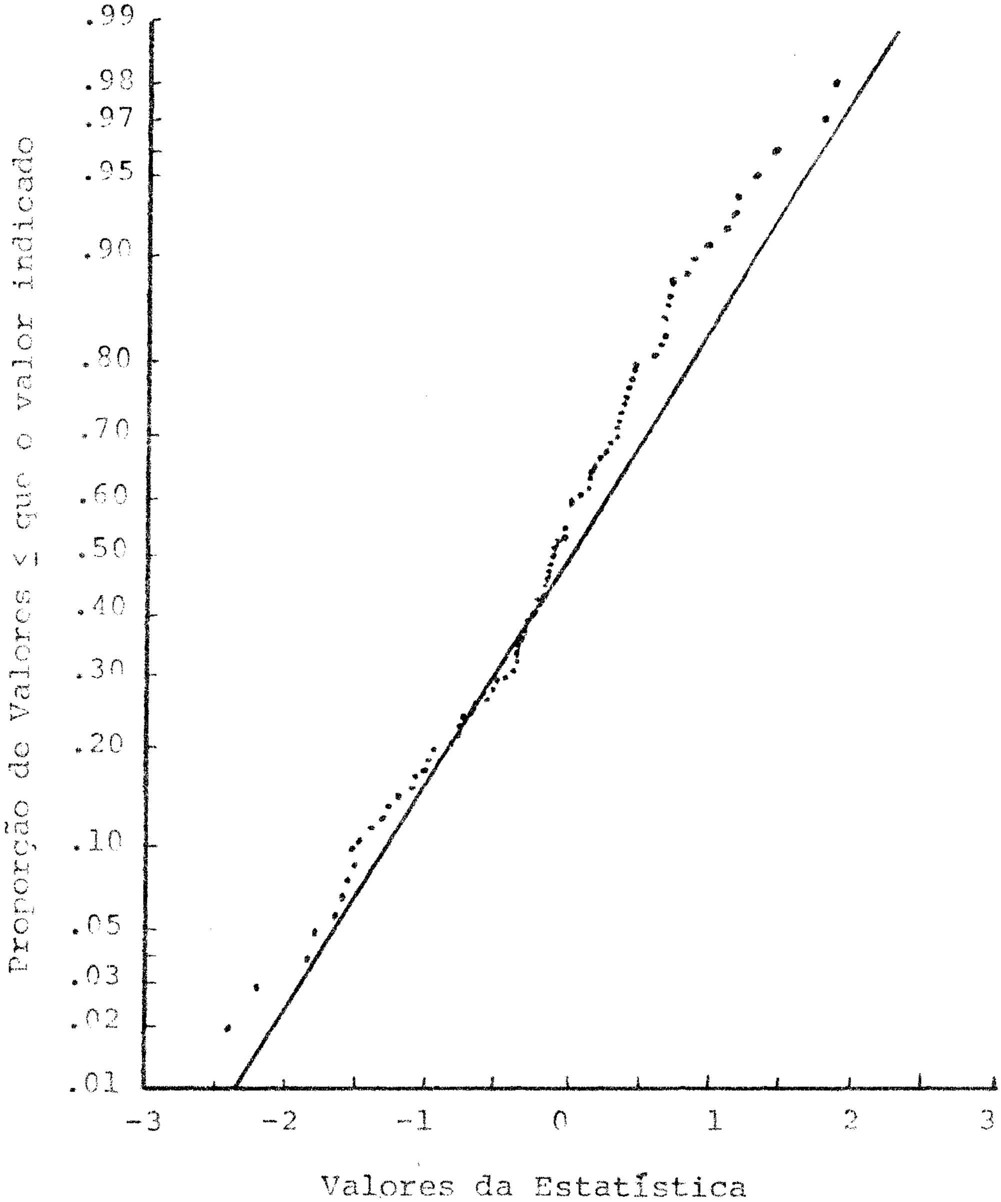

Proporção de Valores ≤ que o valor indicado
.99
.98
.97
.95
.90
.80
.70
.60
.50
.40
.30
.20
.10
.05
.03
.02
.01
-3
-2
-1
0
1
2
3
Valores da Estatística

FIGURA B



.99
.98
.97
.95
.90
.80
.70
.60
.50
.40
.30
.20
.10
.05
.03
.02
.01
-3
-2
-1
0
1
2
3
Proporção de Valores ≤ que o valor indicado
Valores da Estatística

FIGURA C

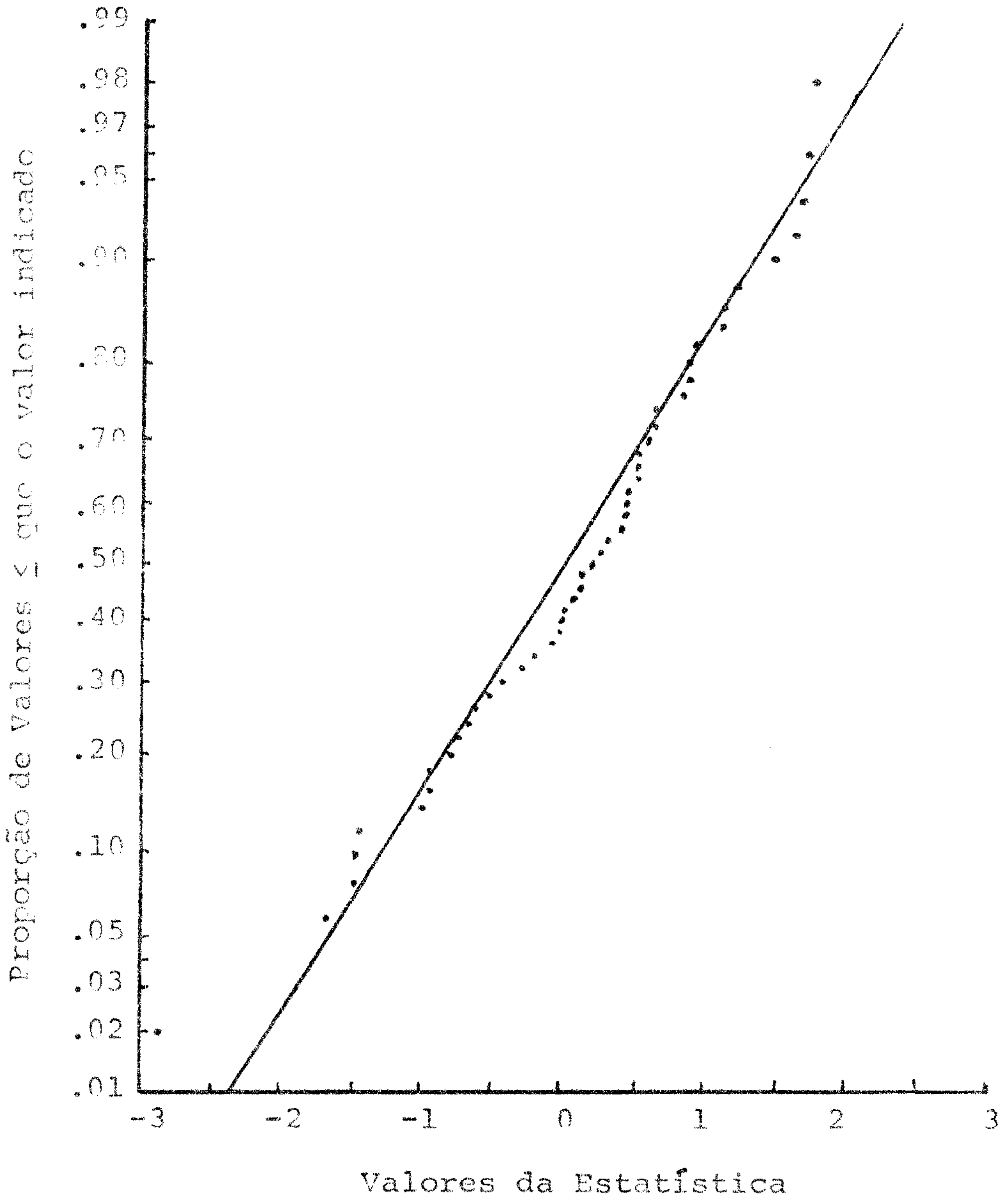
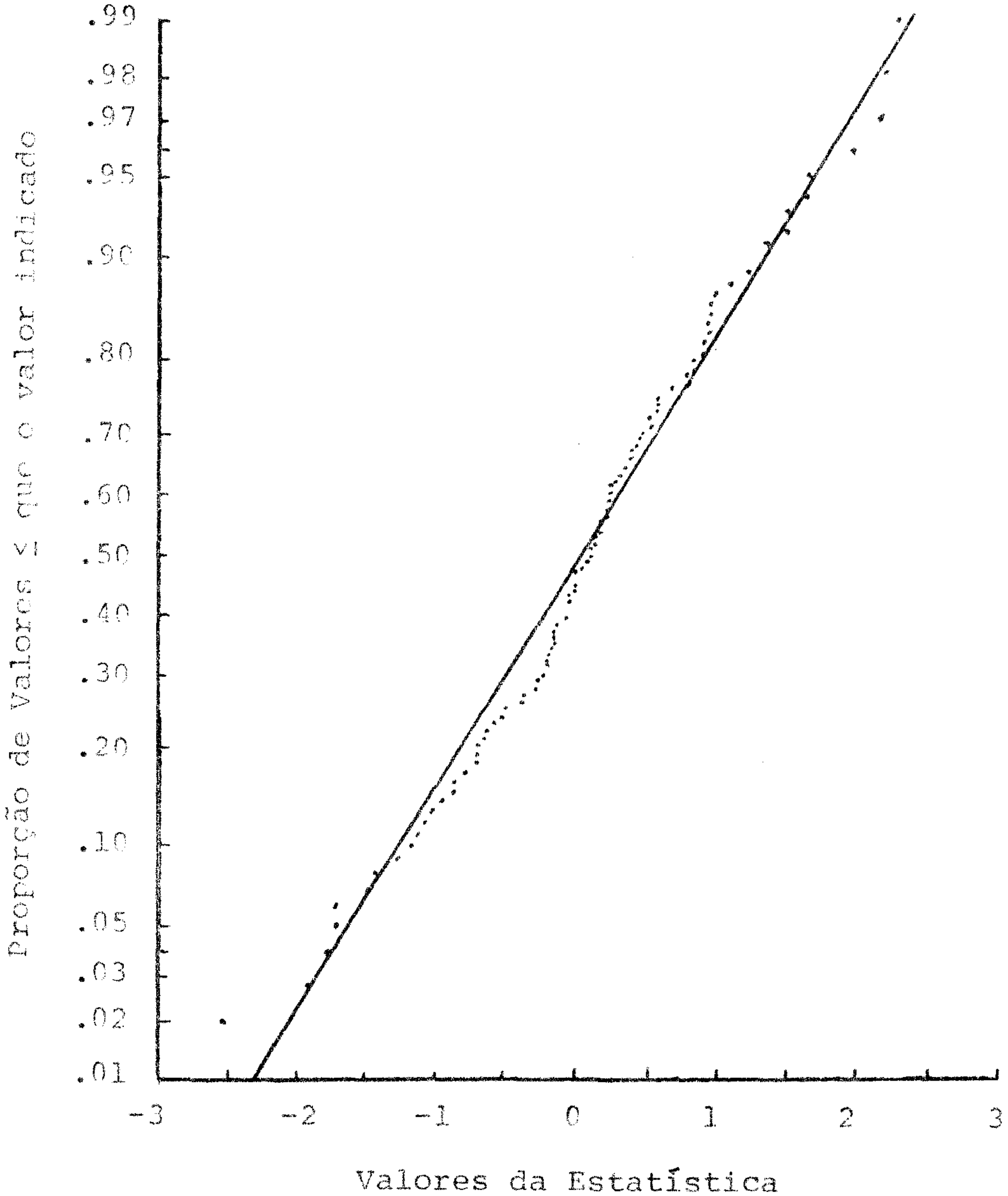

Proporção de Valores ≤ que o valor indicado
.99
.98
.97
.95
.90
.80
.70
.60
.50
.40
.30
.20
.10
.05
.03
.02
.01
-3
-2
-1
0
1
2
3
Valores da Estatística

FIGURA D



.99
.98
.97
.95
.90
.80
.70
.60
.50
.40
.30
.20
.10
.05
.03
.02
.01
-3
-2
-1
0
1
2
3
Proporção de Valores ≤ que o valor indicado
Valores da Estatística

FIGURA E

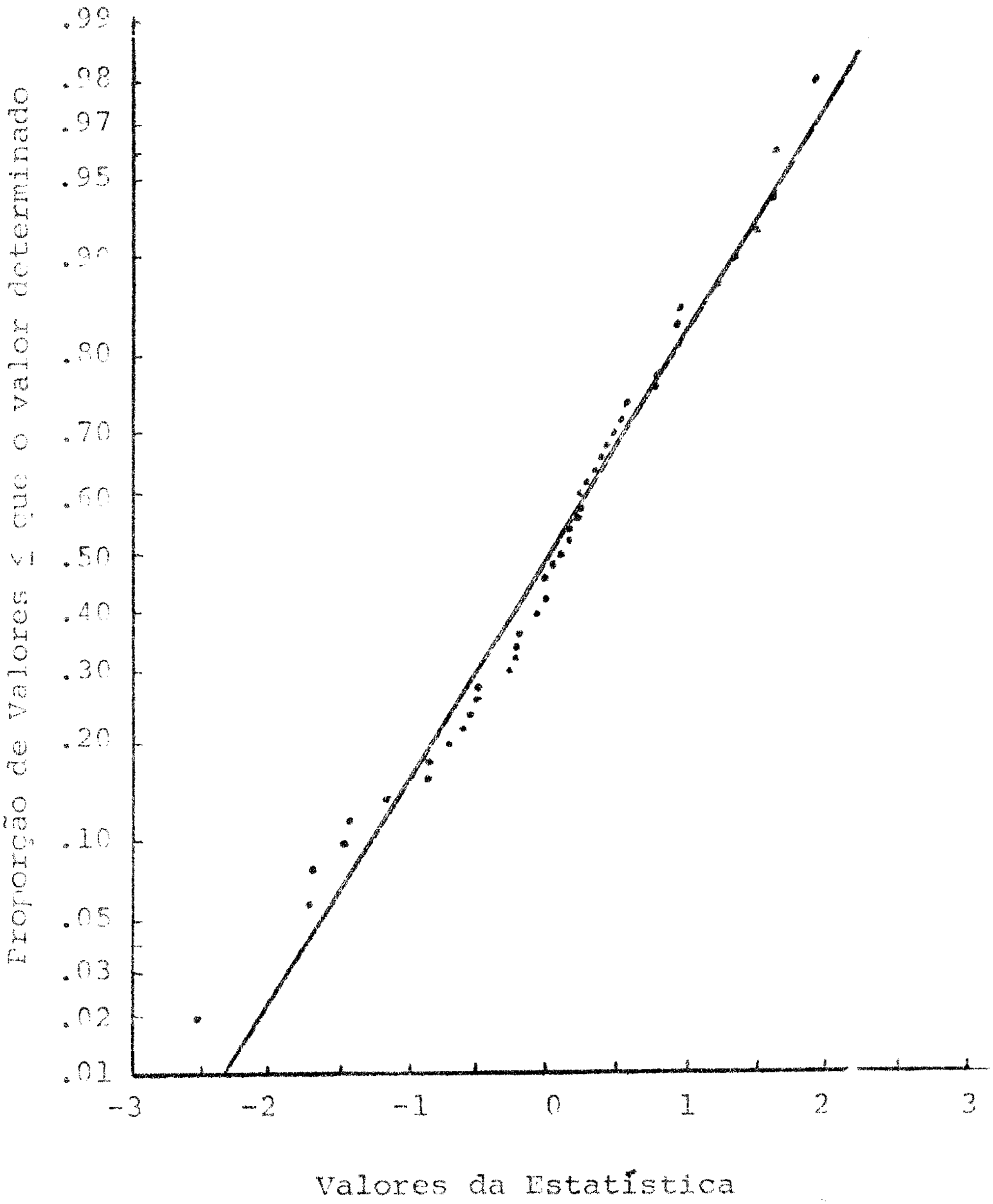

.99
.98
.97
.95
.90
.80
.70
.60
.50
.40
.30
.20
.10
.05
.03
.02
.01
Proporção de Valores ≤ que o valor determinado
-3
-2
-1
0
1
2
3
Valores da Estatística

FIGURA F

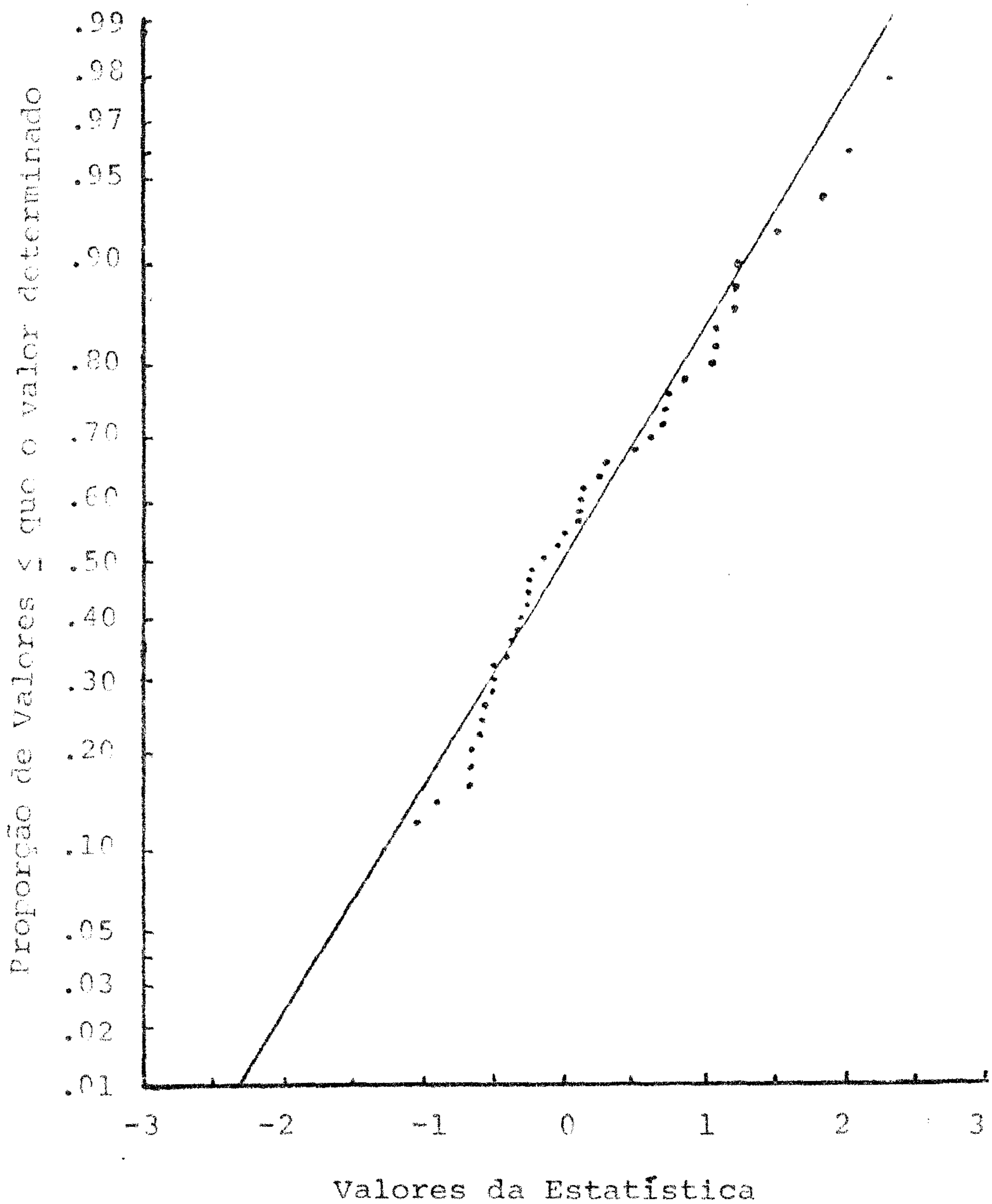

Proporção de Valores ≤ que o valor determinado
.99
.98
.97
.95
.90
.80
.70
.60
.50
.40
.30
.20
.10
.05
.03
.02
.01
-3
-2
-1
0
1
2
3
Valores da Estatística

FIGURA G

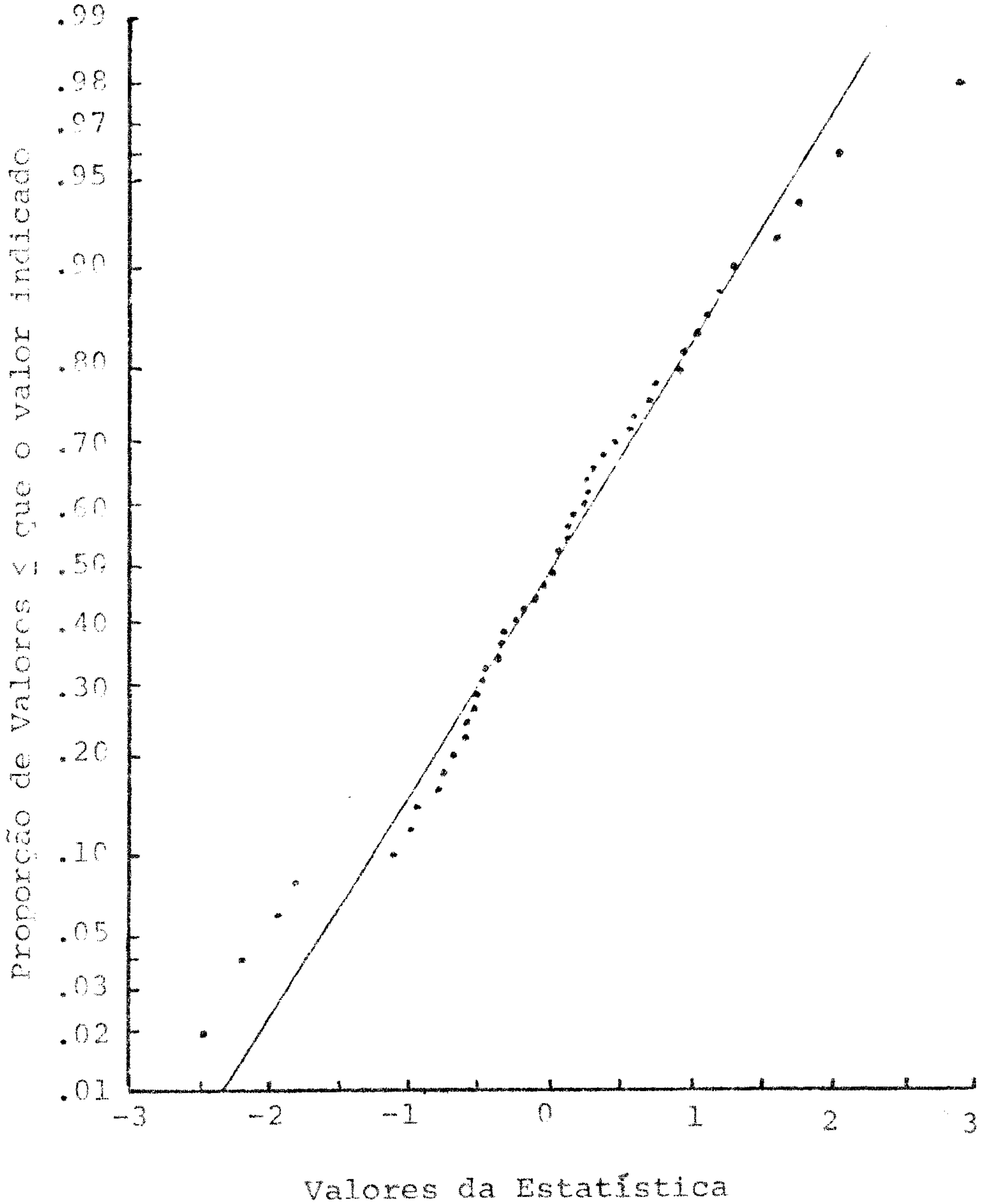

.99
.98
.97
.95
.90
.80
.70
.60
.50
.40
.30
.20
.10
.05
.03
.02
.01
-3
-2
-1
0
1
2
3
Proporção de Valores ≤ que o valor indicado
Valores da Estatística

FIGURA H

2.8 - Problemas Amostrais

No uso da teoria assintótica, um problema aparece no cálculo da variância, quando ocorre empate nos $N_{i.s}$. Por exemplo, para a amostra de tamanho 50 na tabela,

	C_1	C_2	C_3	
L_1	9	0	2	11
L_2	2	10*	4	16
L_3	1	1	14*	16
L_4	1	0	6	7
	13	11	26	50

temos

$$\hat{\lambda}(L/C) = \frac{33 - 16}{50 - 16} = .5 \ .$$

Podemos tomar $P_{(m).} = P_{2.}$ ou $P_{3.}$. No primeiro $\Sigma^* P_{(m)j} = 10/50$ e no segundo $\Sigma^* P_{(m)j} = 14/50$. Teremos então para cada escolha, variâncias diferentes, que resultarão diferentes intervalos de confiança para $\lambda(L/C)$. Para o exemplo dado acima, os intervalos obtidos através de uma distribuição normal para um coeficiente de confiança de .95, são respectivamente:

$$(.2805, .7195)$$

e

$$(.3132, .6868) \ .$$

Este problema não tem, ainda, uma solução teórica, mas para aplicações práticas podemos tomar a variância como sendo a média das possíveis variâncias assintóticas ou o máximo delas. Esta última solução tem a vantagem de garantir um limite inferior para o coeficiente de confiança (ou superior para o nível de significância) e por este motivo, será usada em nossas futuras aplicações, quando necessário. Uma terceira solução seria a aleatorização, esta porém nos parece ter apenas um apelo teórico.

Outro problema que pode ocorrer em amostras reais, mesmo que nossas suposições populacionais sejam verdadeiras, é que $P_{(m).} = 1$. Isto significa que todas as observações estão na mesma linha e que não é possível estimar $\lambda(L/C)$ pontualmente. Para este caso podemos considerar $0 \leq \lambda(L/C) \leq 1$ e não rejeitar a hipótese nula sobre $\lambda(L/C)$ qualquer que ele seja. Esta é apenas uma maneira de dizermos, que a amostra não deu nenhuma informação sobre $\lambda(L/C)$; ou que a amostra ou politomia L, precisa ser redefinida para um prosseguimento do estudo.

2.9 - Amostragem Multinomial em cada coluna

Ao invés de retirarmos uma amostra com reposição de todas as rs celas, pode ser necessário ou desejável que fixemos um tamanho de amostra para cada coluna, ou seja, uma amostra regida pelo produto de s distribuições multinomiais independentes.

Com os totais de coluna fixados, os N_{ij}'s estão agora sujeitos a s restrições. $\sum_i N_{ij} = n_{.j}$, além de $\sum_i \sum_j N_{ij} = n$. Para cada j e qualquer $n_{.j}$, $n_{.j}\left[(N_{.j}/n_{.j}) - (\rho_{ij}/\rho_{.j})\right]$ tem média zero, variância $(\rho_{ij}/\rho_{.j})\ [1 - (\rho_{ij}/\rho_{.j})]$ e covariância $-(\rho_{ij}/\rho_{.j})(\rho_{i'j}/\rho_{.j})$, para $i \neq i'$. Quando $n_{.j} \to \infty$, a distribuição assintótica conjunta para todo i é multinormal com médias zero e mesma estrutura de variância e covariância do caso finito. A distribuição assintótica conjunta para todo (i, j) segue da independência entre as colunas, mas devemos ainda assumir que o limite da razão $n_{.j}/n$ existe e é diferente de zero ou um quando $n \to \infty$.

Para este procedimento amostral não é possível, sem outros conhecimentos, estimar os ρ_{ij}'s dada a independência na distribuição amostral das probabilidades condicionais $(\rho_{ij}/\rho_{.j})$. Assim podemos, por exemplo, assumir conhecidas as probabilidades marginais $\rho_{.j}$'s. Quando os $\rho_{.j}$'s são conhecidos, escolhemos, para um dado tamanho total da amostra n, os $n_{.j} = n\,\rho_{.j}$ (*). Um estimador natural de $\rho_{ij}/\rho_{.j}$ é $N_{ij}/n_{.j}$, se $n_{.j} = n\,\rho_{.j}$, temos N_{ij}/n como o estimador de ρ_{ij}. Se não conhecemos os $\rho_{.j}$'s, mas temos os $n_{.j}$'s proporcionais, o mesmo critério pode ser adotado.

A medida $\lambda(L/C)$ pode ser estimada da mesma maneira que em 2.6, mas a estrutura probabilística do estimador é diferente dada a independência entre as colunas. Sob as mesmas suposições de 2.5 e assumindo que os $n_{.j}$ são proporcionais às probabilidades marginais, podemos mostrar de maneira análoga a 2.6, que $n\,(\hat{\lambda}(L/C) - \lambda(L/C)$ é, também nesse caso, assintoticamente nor

(*) $n\,\rho_{.j}$ não é em geral um inteiro, mas para nossos objetivos assintóticos este fato não será importante.

mal, com média zero e variância

$$[1/(1-\rho_{(m).})^2]\ \{(1-\rho_{(m).})^2 \sum_j \rho_{(m)j}\ [1-\rho_{(m)j}\ /\ \rho_{.j}] +$$

$$+ (1 - \sum_j \rho_{(m)j})^2 \sum \rho'_{(m)j}\ [1-(\rho'_{(m)j}/\rho_{(m).})] +$$

$$- 2(1-\rho_{(m).})(1-\sum_j \rho_{(m)j})\ [\Sigma^*\rho_{(m)j} - \Sigma(\rho_{(m)j}\ \rho'_{(m)j}/\rho_{.j})]\}$$

onde $\rho'_{(m)j}$ denota o ρ_{ij} da coluna j, para o qual $\rho_{i.}$ é máximo e Σ^* tem o mesmo significado anterior. Essa variância é zero se e somente se $\lambda(L/C) = 0$ ou 1. Podemos estimá-la pelo seu análogo amostral e dividir $n(\hat{\lambda}(L/C) - \lambda(L/C))$ pela raiz quadrada desse estimador de forma que, a expressão resultante tem distribuição - normal padrão.

2.10 - Comentários Finais

Goodman e Kruskal (1963), fazem um estudo sobre a comportamento assintótico de algumas das medidas estudadas no capítulo 1. Resultados similares, mas complexos, são obtidos para as outras medidas derivadas de $\lambda(L/C)$. Com exceção da medida de associação múltipla, as expressões para as variâncias assintóticas são de difícil dedução. No entanto, estes mesmos autores em 1972 mostram um método mais simplificado de obtenção destas variâncias.

Problemas como os do não cumprimento da unicidade dos máximos e o tamanho das amostras não proporcionais as probabilidades marginais - para o caso de amostragem independentes -, permanecem sem solução.

CAPÍTULO 3

SELEÇÃO DE VARIÁVEIS

3.1 - Preliminares

Como foi discutido brevemente na introdução, o método a ser proposto considera o caso de uma população, classificada de acordo com p+1 politomias $Y, X^{(1)}, X^{(2)}, \ldots, X^{(P)}$. A variável Y, aquela sobre a qual estamos interessados em fazer algum tipo de estudo, é chamada variável dependente ou resposta e as X's, chamadas independentes ou suportes, são aquelas que podem - a princípio - explicar ou estimar Y de alguma forma. Nosso objetivo será selecionar quais as variáveis X's, que melhor podem ser usadas para este objetivo.

Nós não pressuporemos a existência de uma relação funcional entre Y e os X's; esta relação, caso haja interesse, - pode ser encontrada e o modelo ajustado após a seleção das variáveis. Uma situação, onde o método poderá ser aplicado, seria por exemplo, na realização de um estudo piloto, com o objetivo de selecionar quais as variáveis que devem ser futuramente amostradas para estimar Y.

Consideraremos para a aplicação do método: uma amostra regida por uma distribuição multinomial sobre o cruzamento de todas as politomias em estudo e que as pressuposições necessárias, para a aplicação da teoria assintótica, são satisfeitas. Para a situação onde temos amostragens multinomiais independentes em -

cada classe, o método também poderá ser aplicado com as modificações necessárias - estudadas no capítulo anterior - nos resultados assintóticos.

Na escolha de uma medida de associação para a aplicação do método, levamos em consideração, que as politomias X e mesmo Y, são dos mais variados tipos. Outro aspecto importante foi que o tipo de associação, que estamos interessados em detectar, é o de previsão. A escolha de $\lambda(L/C)$ surgiu então naturalmente, com a vantagem adicional de uma fácil interpretabilidade.

A metodologia a ser proposta é análoga, em sua estratégia, ao método de regressão "forward", ou seja, a cada passo podemos ter ou não a inclusão de novas variáveis.

3.2 - O Método de Seleção

Consideremos uma amostra das politomias $Y, X^{(1)}, X^{(2)}, \ldots, X^{(P)}$. Nosso primeiro passo será selecionar a politomia X, com maior grau de associação com Y. Para tanto, podemos construir todas as classificações cruzadas $Y \times X^{(\ell)}$ $\ell=1,2,\ldots,P$ e calcular $\hat{\lambda}(Y/X^{(\ell)})$. A politomia X, que fornecer o maior $\hat{\lambda}(Y/X^{(\ell)})$ e portanto a maior associação com Y, será selecionada.

Após o primeiro passo, temos selecionado uma politomia $X^{(\ell_1)}$. Podemos então, construir novas politomias através da reenumeração dos p-1 pares $(X^{(\ell_1)}, X^{(\ell)})$, para $\ell \neq \ell_1$, e denotá-las por $X^{(\ell_1,\ell)}$. O segundo passo será então, análogo ao primeiro para as politomias $X^{(\ell_1,\ell)}$ e selecionará o par $(X^{(\ell_1)}, X^{(\ell)})$ mais associado com Y. Teremos aí escolhido a segunda variável

$X^{(\ell_2)}$ e podemos prosseguir agregando e aplicando o mesmo princípio de escolha, para $X^{(\ell_1)}, X^{(\ell_2)}, X^{(\ell)} = X^{(\ell_1, \ell_2, \ell)}, \ell \neq \ell_1, \ell_2$; $(X^{(\ell_1)}, X^{(\ell_2)}, X^{(\ell_3)}, X^{(\ell)}) = X^{(\ell_1, \ell_2, \ell_3, \ell)}$, $\ell \neq \ell_1, \ell_2, \ell_3$; e assim sucessivamente. A medida que estamos usando a partir do segundo passo, $\lambda(Y/X^{(\ell_1, \ell_2, \ldots, \ell_k, \ell)})$, pode e deve ser interpretada como a medida de associação múltipla definida em 1.2 ,c sendo sobre este ângulo denotada por $\lambda(Y/X^{(\ell_1)}, X^{(\ell_2)}, \ldots, X^{(\ell_k)}, X^{(\ell)})$.

Ao agregarmos uma nova politomia, o fato importante é o não decrescimento da medida de asociação. Para mostrarmos esta propriedade, consideremos as politomias A, B, C e (B, C), sendo a última obtida pela agregação de B e C. Seja:

ρ_{ij} = a probabilidade da cela (A_i, B_j).

ρ_{ijk} = a probabilidade da cela (A_i, B_j, C_k).

$$\lambda(A/B) = \frac{\sum_j \rho_{(m)j} - \rho_{(m).}}{1 - \rho_{(m).}}$$

$$\lambda(A/B,C) = \lambda(A/(B,C)) = \frac{\sum_j \sum_k \rho_{(m)jk} - \rho_{(m)..}}{1 - \rho_{(m)..}}$$

Como $\rho_{ij} = \sum_k \rho_{ijk}$, segue que

$$\rho_{(m).} = \max_i \rho_{i.} = \max_i \sum_j \rho_{ij} = \max_i \sum_j \sum_k \rho_{ijk} = \max_i \rho_{i..} = \rho_{(m)..}$$

Por outro lado,

$$\rho_{ijk} \leq \max_i \rho_{ijk} = \rho_{(m)jk} \qquad \forall\ (i,\ j,\ k)$$

$$\Rightarrow \rho_{ij} = \sum_k \rho_{ijk} \leq \sum_k \rho_{(m)jk} \qquad \forall\ (i,\ j)$$

$$\Rightarrow \rho_{(m)j} = \max_i \rho_{ij} \leq \sum_k \rho_{(m)jk} \qquad \forall\ j$$

$$\Rightarrow \sum_j \rho_{(m)j} \leq \sum_j \sum_k \rho_{(m)jk} \quad .$$

Usando estes dois resultados temos que

$$\frac{\sum_j \rho_{(m)j} - \rho_{(m).}}{1 - \rho_{(m).}} \leq \frac{\sum_j \sum_k \rho_{(m)jk} - \rho_{(m)..}}{1 - \rho_{(m)..}}$$

ou

$$\lambda(A/B) \leq \lambda(A/B,C) = \lambda(A/(B,C)) \quad .$$

Esta propriedade pode ser interpretada, considerando que a agregação de duas politomias produz uma partição da população mais refinada. Com isto, o número de celas vazias ou com pequenas probabilidades, tendem a aumentar, crescendo consequentemente o grau de associação.

Para exemplificar este aspecto consideremos uma população de 20 indivíduos classificados de acordo com três politomias Y, $X^{(1)}$, $X^{(2)}$ com 3, 4 e 5 classes respectivamente. Consideremos também a politomia obtida pela agregação de $X^{(1)}$ e $X^{(2)}$ e as seguintes classificações cruzadas $Y \times X^{(1)}$, $Y \times X^{(2)}$ e

$Y \times (X^{(1)}, X^{(2)})$.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Y :	2,	2,	1,	2,	3,	1,	3,	2,	2,	3,	2,	2,	1,	3,	3,	1,	2,	1,	2,	3
X_1 :	2,	3,	1,	1,	1,	1,	4,	2,	4,	1,	2,	4,	2,	4,	4,	1,	1,	3,	4,	3
X_2 :	2,	4,	3,	5,	2,	4,	2,	4,	2,	1,	2,	4,	2,	4,	4,	1,	1,	3,	4,	3
$(X_1\ X_2)$:	7,	9,	3,	10,	12,	4,	12,	13,	7,	14,	7,	9,	2,	14,	14,	1,	6,	3,	9,	13

$Y/X^{(1)}$	1	2	3	4	
1	3	1	1	0	5
2	2	3	1	3	9
3	2	0	0	4	6
	7	4	2	7	20

$Y/X^{(2)}$	1	2	3	4	5	
1	1	1	2	1	0	5
2	1	3	0	3	2	9
3	1	2	1	2	0	6
	3	6	3	6	2	20

$Y \backslash X^{(1)}X^{(2)}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	1	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
2	0	0	0	0	0	1	3	0	3	1	0	0	1	0	0	0	0	0	0	0	9
3	0	0	0	0	0	0	0	0	0	0	0	2	1	3	0	0	0	0	0	0	6
	1	1	2	1	0	1	3	0	3	1	0	2	2	3	0	0	0	0	0	0	20

$$\lambda(Y/X^{(1)}) = \frac{(3 + 3 + 1 + 4) - 9}{20 - 9} = \frac{11 - 9}{11} = \frac{2}{11}$$

$$\lambda(Y/X^{(2)}) = \frac{(1 + 3 + 2 + 3 + 2) - 9}{20 - 9} = \frac{11 - 3}{11} = \frac{2}{11}$$

$$\lambda(Y/X^{(1)},X^{(2)}) = \frac{(1+1+2+1+0+1+3+0+3+1+0+2+1+3+0+0+0+0+0+0) - 9}{20 - 9}$$

$$= \frac{18 - 9}{11} = \frac{9}{11}$$

Na classificação Y x $(X^{(1)}, X^{(2)})$ devemos observar que nada perderíamos se ao agregarmos as politomias $X^{(1)}$ e $X^{(2)}$ - considerássemos somente as classes com frequência diferente de zero. Isto é, no exemplo acima podemos desprezar as classes 5, 8, 11, 15, 16, 17, 18, 19, 20 e considerar a politomia $(X^{(1)}, X^{(2)})$ como tendo somente 11 classes. Este processo será por nós utilizado nas aplicações práticas.

Podemos agora usar o não decrescimento da medida de associação, para estabelecermos uma regra de parada no processo de seleção. Esta parada, no entanto, poderá ocorrer de maneiras distintas. A primeira forma ocorrerá, se um desejado nível de associação λ_o estabelecido a priori, for atingido estatisticamente. Usando a convergência assintótica, a cada passo podemos calcular o intervalo de confiança unilateral para um dado coeficiente $1 - \alpha$; se este intervalo contiver λ_o, o processo para e as variáveis selecionadas serão aquelas agregadas até este passo. Isto é o mesmo que realizar um teste, a um nível de significância α, da hipótese nula $\lambda(L/C) = \lambda_o$ contra $\lambda(L/C) < \lambda_o$.

Uma outra forma de parar o processo acontece, se a agregação de uma nova variável não produz um acréscimo em $\hat{\lambda}(L/C)$, isto é, não há redução na probabilidade do erro de previsão. Em assim sendo, selecionaremos as variáveis agregadas até o passo anterior. A terceira forma é aquela em que a associação máxima - $\hat{\lambda}(L/C) = 1$ - é atingida, ou seja, as variáveis selecionadas até este passo podem fazer uma previsão "exata". A última, e menos desejável, ocorre quando todas as variáveis já estão agregadas.

Embora não seja por nós considerada uma regra de parada, a ocorrência de $\hat{\lambda}(L/C)$ indeterminado ou zero provocará um truncamente no processo. Como vimos anteriormente, $\hat{\lambda}(L/C)$ é indeterminado, se e somente se, toda a amostra está concentrada em uma única classe de politomia dependente. Consideraremos esse fato como uma má definição da politomia ou um problema amostral: tamanho da amostra insuficiente ou amostragem deficiente. Por outro lado $\hat{\lambda}(L/C) = 0$, se e somente se, os máximos de cada classe da politomia independente estão em uma mesma classe da politomia dependente. Se isto ocorre devido a independência estatística, nada temos a selecionar em termos de previsão - as variáveis foram mal escolhidas - e caso contrário os mesmos comentários feitos para indeterminação são válidos. No entanto o truncamento do processo nestas situações irá ocorrer, se for o caso, já no primeiro passo.

O grau de associação, fixado para a primeira regra de parada, deve ser menor que um. Primeiro porque a convergência para a distribuição normal, só é válida se $0 < \lambda(L/C) < 1$; segundo $\hat{\lambda}(L/C) = 1$, se $\lambda(L/C) = 1$. Assim sendo, se associação total é desejada, somente a terceira regra deve ser usada, pois as outras representam uma frustração deste objetivo. Ao nosso ver, em ter mos práticos, a obtenção de $\lambda(L/C) = 1$ não requer um método de seleção no sentido proposto e por outro lado, não gostaríamos de pagar o preço necessário para termos tal grau de associação.

Uma variação no método deverá ocorrer, quando em algum passo tivermos duas ou mais agregações com a maior associação. Neste caso uma escolha pessoal, dentre as agregações que resultaram na maior associação, deve ser feita. Por exemplo, no primeiro passo se as politomias $X^{(i)}$ e $X^{(j)}$ têm o mesmo (e maior) grau de associação com Y podemos selecionar segundo algum critério uma delas. Caso o par $(X^{(i)}, X^{(j)})$ tenham conjuntamente um auto grau de asociação com Y, estas duas variáveis serão agregadas no próximo passo; se isto não ocorrer, é porque no segundo passo encontramos um outro par que fornece maior associação com Y. Neste ponto devemos lembrar que o método, não tem por objetivo selecionar o melhor conjunto de variáveis, mas reduzir o conjunto inicial de variáveis. Este problema também é encontrado nos métodos clássicos de regressão por passos onde nem sempre o melhor conjunto de variáveis é selecionado.

Uma última observação a ser feita é sobre a particularidade do primeiro passo no que se refere a: truncamentos no processo e a impossibilidade de aplicação da segunda regra de parada.

3.3 - Comentários Finais

Algumas críticas podem ser levantadas sobre o método proposto. Uma delas seria a utilização da teoria assintótica.

Este artifício é vastamente utilizado e plenamente aceito, inclusive nos métodos clássicos baseados em χ^2. No caso da medida utilizada, trabalhos de R.E. Greenwood e M.O. Glasgow em 1950 e R.E. Kozelka (1952 e 1956) sobre a distribuição do máximo em uma amostra com distribuição multinomial, justificam totalmente o uso da teoria assintótica.

Outra crítica, e essa realmente válida, diz respeito às fortes suposições populacionais de unicidade dos máximos. Podemos nos conformar com ela, se lembrarmos que este problema teórico, pode vir a ser resolvido em futuros estudos sobre convergência assintótica e pensarmos que na teoria de modelos lineares suposições para utilização da técnica por passos (stepwise) são - também bastante fortes.

CAPITULO 4

UM EXEMPLO DE APLICAÇÃO DO METODO

Para exemplificar o método foi escolhida uma amostra de 132 crianças em idade de alfabetização, nas quais foi aplicado o teste ABC, sendo então medida sua pontuação e também um conjunto de variáveis sócio-econômico biológicas que a princípio podem estar associadas com o resultado do teste ABC.

Sendo nosso objetivo apenas a exemplificação do método e não o estudo do problema em si, um estudo preliminar nos levou a escolher um subconjunto de dez variáveis que, individualmente, apresentaram o maior grau de associação de (λ) com a variável de interesse - pontos no teste ABC.

O subconjunto de variáveis escolhidas foi o seguinte:

i) Tempo que frequentou pré-primário (X_1)

ii) Se tem lugar para estudar (X_2)

iii) Esporte que pratica (X_3)

iv) Com quem dorme (X_4)

v) Tipo de alimentação (X_5)

vi) Se viaja (X_6)

vii) Se é super protegida (X_7)

viii) Localização da residência (X_8)

ix) Se assiste discussão dos pais (X_9)

x) Se os pais são separados (X_{10})

Estas variáveis acima - variáveis independentes - foram classificadas, respectivamente com os seguintes números de classes: 6, 2, 6, 6, 5, 3, 3, 3, 3 e 2. A pontuação obtida no teste ABC - variável dependente (Y) - foi dividida em 4 classes de forma que se obtivesse uma distribuição marginal para Y o mais próximo possível da uniforme.

Na seleção das variáveis independentes, os limites de confiança para o nível de associação foram calculados para um nível de significância de 0.05 e foram comparados com o nível de associação "desejado" de 0.9.

Com os parâmetros acima os resultados obtidos foram os seguintes:

*UM METODO DE SELECAO DE VARIAVEIS POLITOMICAS

*VARIAVEIS ASSOCIADAS COM O TESTE ABC

NUMERO DE VARIAVEIS INDEPENDENTES: 10
NUMERO DE INDIVIDUOS DA POPULACAO: 132
NIVEL DE ASSOCIACAO DESEJADO: 0.900
(1-A).CENTIL DA NORMAL PADRAO: 1.640

VARIAVEL DEPENDENTE:

PONTOS NO TESTE ABC   (4 CLASSES)

VARIAVEIS INDEPENDENTES:

FREQUENTOU PRE-PRIM.	(6 CLASSES)
TEM LUGAR P/ ESTUDAR	(2 CLASSES)
ESPORTE QUE PRATICA	(6 CLASSES)
COM QUEM DORME	(6 CLASSES)
TIPO DE ALIMENTACAO	(5 CLASSES)
VIAJA	(3 CLASSES)
SUPERPROTEGIDA	(3 CLASSES)
LOCALIZ. RESIDENCIA	(3 CLASSES)
DISCUSSOES DOS PAIS	(3 CLASSES)
SEPARACAO DOS PAIS	(2 CLASSES)

*UM METODO DE SELECAO DE VARIAVEIS POLITOMICAS

*VARIAVEIS ASSOCIADAS COM O TESTE ABC

PASSO	VARIAVEL SELECIONADA	LAMBDA	LIMSUP
1	ESPORTE QUE PRATICA	.14286	.26977
2	COM QUEM DORME	.31169	.55247
3	TIPO DE ALIMENTACAO	.54545	.75564
4	SUPERPROTEGIDA	.77922	.93424

O GRAU DE ASSOCIACAO DESEJADO FOI ATINGIDO.

*UM METODO DE SELECAO DE VARIAVEIS POLITOMICAS

*VARIAVEIS ASSOCIADAS COM O TESTE ABC

VARIAVEIS INDEPENDENTES SELECIONADAS:

ESPORTE QUE PRATICA

COM QUEM DORME

TIPO DE ALIMENTACAO

SUPERPROTEGIDA

Foram selecionadas 4 variáveis com um grau de associação final estimado de 0.77922. O limite superior de confiança após a seleção da quarta variável atingiu 0.93424, maior portanto que o grau de associação escolhido para a comparação, o que provocou a parada no processo de seleção.

Podemos observar pelo grau de associação da primeira variável selecionada que individualmente as variáveis não apresentam altos graus de associação com a variável dependente. O grau de associação final, 0.77922 , apesar de estar ainda muito afastado do desejado nos parece razoável para o nosso objetivo - selecionar o menor número de variáveis possível.

APENDICE 1

RESULTADOS BASICOS DE CONVERGENCIA

Teorema A1.

"Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes e identicamente distribuídas com média finita μ, então

$\bar{X}_n = (X_1 + X_2 + \ldots + X_n)/\, n \longrightarrow \mu$"

Referência | 4 |

Teorema A2.

"Se $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$ então:

i) $aX_n + bY_n \xrightarrow{P} aX + bY$ (a,b constantes)

ii) $X_nY_n \xrightarrow{P} XY$

iii) $X_n/Y_n \xrightarrow{P} X/Y$ $(P(Y_n \neq 0) = P(Y \neq 0) = 1)$"

Referência | 4 |

Teorema A3.

"Se $X_n \xrightarrow{P} X$ e $g : R \longrightarrow R$ é uma função contínua então:

$$g(X_n) \xrightarrow{P} g(X)"$$

Referência | 4 |

Teorema A4.

"Se $X_n \xrightarrow{p} X$ então $X_n \xrightarrow{D} X$"

Referência |4|

Teorema A5.

"Seja $\{P_n\}$ uma sequência de distribuições de probabilidade e V um evento tal que $\lim_{n\to\infty} P_n(V)$, denotado por $P(V)$, existe. Seja W um outro evento tal que $\lim_{n\to\infty} P_n(W)=1$. Então $\lim_{n\to\infty} P_n\,(V \cap W) = P(V)$"

Referência |15|

Teorema A6.

"Seja $\underset{\sim}{X}_n$ um vetor aleatório com distribuição multinomial com parâmetros n e $\underset{\sim}{\rho}' = (\rho_1, \rho_2, \ldots, \rho_n)$. Seja $U_n = \quad n\,(\underset{\sim}{X}_n / n - \underset{\sim}{\rho})$. Então $U_n \xrightarrow{d} N(0,\, D_\rho - \rho\rho')$ onde $D_\rho = \mathrm{diag}\,(\rho_1, \rho_2, \ldots, \rho_n)$"

Referência |3|

APENDICE 2

PROGRAMA E INTRUCÕES

A2.1 Introdução

Para aplicação do método, programas de computador encontram-se disponíveis, em fita magnética (MA54), no Centro de Computação da UNICAMP. A linguagem de computador usada foi FORTRAN IV; exceto por uma das sub-rotinas que também se encontra disponível em ASSEMBLER. Esta sub-rotina em ASSEMBLER deve ser preferida - com o objetivo de ganho no tempo de execução - sempre que o sistema usado for o DEC-10.

A2.2 Algorítmo Simplificado

1) Entra com

a) N = tamanho da amostra

b) $(Y, X^{(1)}, X^{(2)}, \ldots, X^{(P)})$ = matriz de dados

c) L_o = nível de associação desejado

d) $t_c = (1 - \alpha)$ centil da distribuição normal padrão

e) PRIOR = vetor com a ordem de prioridade das variáveis independentes para o caso de empate.

2) Calcula:

"$MTY = \max_{i} TY(i)$", onde $TY(i)$ é o número de indivíduos na classe $Y(i)$.

3) Verifica se LAMBDA é indeterminado:

"Se $MTY = N$, vai para 25"

4) Para $\ell = 1, 2, \ldots, p$:

a) "Constrói a tabela $Y \times X^{(\ell)}$"

b) "Calcula a soma dos máximos de cada coluna"

5) Calcula

"$MSMC = \max_{\ell} SMC(\ell)$"

6) Verifica se LAMBDA é zero:

"Se $MSMC = MTY$, vai para 26"

7) Seleciona $X^{(\ell_1)}, X^{(\ell_2)}, \ldots, X^{(\ell_k)}$ tal que:

"$SMC(\ell_i) = MSMC$"

8) Escolhe $X^{(\ell^*)}$ a variável com maior prioridade entre $X^{(\ell_1)}, \ldots, X^{(\ell_k)}$.

9) Faz:

$$"Z = X^{(\ell^*)}"$$

10) Verifica se a associação máxima foi atingida:

"Se MSMC = N, vai para 27"

11) Para a tabela Y x Z

a) "Calcula $\hat{\lambda}(Y/Z)$"

b) "Calcula $\hat{Var}(\hat{\lambda})$"

c) "Calcula o limite superior de confiança (LIMSUP) para λ"

12) Verifica se o grau de associação "desejado" foi atingido:

"Se LIMSUP > λ_o, vai para 28"

13) Verifica se todas as variáveis já foram selecionadas

"Se $Z = (X^{(\ell_1)},\ldots,X^{(\ell_p)})$, vai para 30"

14) Para as variáveis $X^{(\ell)}$ ainda não selecionadas:

a) "Faz a agregação $Z(\ell) = (Z, X^{(\ell)})$"

b) "Constrói as tabelas $Y \times Z(\ell)$"

c) "Calcula $SMC(\ell)$"

15) Calcula:

"$MSMCA = \max_{\ell} SMC(\ell)$"

16) Seleciona $X^{(\ell_1)}, X^{(\ell_2)}, \ldots, X^{(\ell_k)}$ tal que:

"$MSMCA = SMC(\ell_i)$"

17) Escolhe:

"$X^{(\ell^*)}$ a variável de maior prioridade entre $X^{(\ell_1)}, X^{(\ell_2)}, \ldots, X^{(\ell_k)}$"

18) Verifica se a inclusão da nova variável aumenta o grau de associação:

"Se MSMC = MSMCA, vai para 29"

19) Faz:

"MSMC = MSMCA"

20) Faz:

"$Z = Z(\ell^*)$"

21) Verifica se a associação máxima foi atingida:

"Se MSMC = N, vai para 27"

22) Calcula:

a) "$\hat{\lambda}(Y/Z)$"

b) "$\hat{V}ar\ (\lambda)$

c) "LIMSUP para λ"

23) Verifica se o grau de associação "desejado" foi atingido:

"Se LIMSUP > λ_o, vai para 28"

24) Volta para 13

25) O método não pode ser aplicado pois a medida de associação é indeterminada. FIM

26) Não existe evidência de associação entre Y e $X^{(1)}, X^{(2)}, \ldots, X^{(p)}$. FIM

27) A associação máxima (= 1) foi atingida. As variáveis selecionadas são aquelas que estão agregadas em Z. FIM

28) O grau de associação "desejado" foi atingido. As variáveis selecionadas são aquelas que estão -

agregadas em Z. FIM

29) A inclusão de novas variáveis não aumenta o grau de associação. Ficamos com as variáveis selecionadas até o passo anterior que já estão agregadas em Z. FIM

30) Todas as variáveis foram selecionadas. FIM

A.2.3 Listagens do Programa Principal e das Sub-rotinas

```
C          T E S E 4 . F 4 0    -    UM METODO DE SELECAO DE VARIAVEIS
C                                    MULTINOMIAIS
C
C                                    UNICAMP, 15/03/79
C
C
C
      INTEGER  Y(200),X(200,20),PRIOR(20),P,NX(20),
     1   NOMEY(5),NOMEX(20,5),TITULO(10),
     2   IY(15),VMIY(15),VS(20),VMS(20),
     3   Z(200),ZA(200),ZB(200),TABS(15,150),
     4   SMC,AREA(15,150),MCOL(15)
C
      REAL  L0,LAMBDA,LIMSUP
C
      COMMON /ALL/   N,KOUT
      COMMON /FVAR/  MIY,DENOMS,UMIY,MIY
C
C
      DATA  MAXIND/200/ MAXVAR/20/,MAXCLA/15/,MAXAGA/150/,
     1      KIN/20/,KOUT/3/,YABRE/'Y('/,FECHA/')'/
C
C
C
C        LE O ARQUIVO DE DADOS
C
C        LE TITULO, NUM. VARIAVEIS, NUM. INDIVIDUOS, NIVEL DE ASSOCIA-
C        CAO DESEJADO, E (1-ALFHA) CENTIL DA DISTRIBUICAO NORMAL PADRAO
      READ(KIN,1000,END=900) TITULO,P,N,L0,T0
 1000 FORMAT(10A4,/,I2,I3,2F5.3)
      IF(P.LT.1.OR.P.GT.MAXVAR) GO TO 800
      IF(N.LT.1.OR.N.GT.MAXIND) GO TO 801
      IF(L0.LE.0..OR.L0.GE.1.) GO TO 802
C
C        LE PRIORIDADE DE ESCOLHA DAS VARIAVEIS INDEPENDENTES
C        (EM CASO DE EMPATE)
      READ(KIN,1001,END=901) PRIOR
 1001 FORMAT(20I2)
      DO 100 I=1,P
       IP=PRIOR(I)
       IF(IP.LT.1.OR.IP.GT.P) GO TO 803
       IF(I.EQ.1) GO TO 100
       II=I-1
       DO 100 LOOP=1,II
        IF(PRIOR(LOOP).EQ.IP) GO TO 803
  100 CONTINUE
C
C        LE NUMERO DE CLASSES E O NOME DA VARIAVEL DEPENDENTE Y, E O MES-
C        MO PARA AS VARIAVEIS INDEPENDENTES X.
      READ(KIN,1002,END=902) NY,NOMEY
 1002 FORMAT(I2,5A4)
      IF(NY.LT.1.OR.NY.GT.MAXCLA) GO TO 804
      READ(KIN,1002,END=902) (NX(J),(NOMEX(J,K),K=1,5),J=1,P)
```

```
      DO 101 J=1,P
       IF(NX(J).LT.1.OR.NX(J).GT.MAXCL) GO TO 805
101   CONTINUE
C
C       LE, PARA CADA INDIVIDUO, SUAS CLASSES EM RELACAO 'AS VARIAVEIS
C       DEPENDENTE E INDEPENDENTES.
      DO 102 I=1,N
       READ(KIN,1003,END=902) Y(I),(X(I,J),J=1,P)
1003   FORMAT(2012)
       IF(Y(I).LT.1.OR.Y(I).GT.NY) GO TO 806
       DO 102 J=1,P
        IF(X(I,J).LT.1.OR.X(I,J).GT.NX(J)) GO TO 807
102   CONTINUE
C
C       VERIFICA SE NAO EXISTEM MAIS DADOS
      READ(KIN,1000,END=1) DUMMY
      GO TO 904
C
C       IMPRESSAO DOS DADOS GERAIS DESSA EXECUCAO
1     WRITE(KOUT,1004) TITULO,P,N,LO,TO,
     1 NOMEY,NY,(NOMEX(PRIOR(J),K),K=1,5),NX(PRIOR(J)),J=1,P)
1004  FORMAT('1'UM METODO DE SELECAO DE VARIAVEIS POLITOMICAS',
     1 //,1X,'0',10A4,
     2 ///,1X,'NUMERO DE VARIAVEIS INDEPENDENTES: ',I4,
     3   /,1X,'NUMERO DE INDIVIDUOS DA POPULACAO: ',I4,
     4   /,1X,'NIVEL DE ASSOCIACAO DESEJADO: ',F6.3,
     5   /,1X,'(1-@) CENTIL DA NORMAL PADRAO: ',F6.3,
     6  //,1X,'VARIAVEL DEPENDENTE: ',
     7  //,3X,5A4,2X,'(',I2,' CLASSES)',
     8  //,1X,'VARIAVEIS INDEPENDENTES:',
     9  //,(3X,5A4,2X,'(',I2,' CLASSES)'))
C
C
C       INICIALIZA VARIAVEIS
C
      DO 103 K=1,NY
       TY(K)=0
103   CONTINUE
C
C       CALCULA O NUMERO DE INDIVIDUOS EM CADA CLASSE DE Y
      DO 104 I=1,N
       K=Y(I)
       TY(K)=TY(K)+1
104   CONTINUE
C
C       OBTEM O MAXIMO DE TY
      MTY=0
      DO 105 K=1,NY
       IF(TY(K).GT.MTY) MTY=TY(K)
105   CONTINUE
C
C       VERIFICA SE A MEDIDA E' INDETERMINADA, OU SEJA, O METODO NAO SE
C       APLICA
```

```
      IF(MTY.EQ.N) GO TO 9000
      DENOM=FLOAT(N-MTY)
      DENOM3=DENOM*DENOM*DENOM
C
C       OBTEM AS CLASSES EM QUE MTY OCORRE
      NMTY=0
      DO 106 K=1,NY
       IF(TY(K).NE.MTY) GO TO 106
       NMTY=NMTY+1
       VMTY(NMTY)=K
106   CONTINUE
C
C       PRIMEIRO PASSO
C
C       MONTA AS TABELAS DE CONTINGENCIA DE Y VERSUS X(J), E
C       SELECIONA A VARIAVEL J1 QUE FORNECE O MAIOR LAMBDA (Y/X(J1))
      MSMC=0
      DO 107 L=1,P
       J=PRIOR(L)
       CALL MONTAB(Y,NY,X(1,J),NX(J),AREA,SMC)
       IF(MSMC.GE.SMC) GO TO 107
       MSMC=SMC
       J1=J
       NELEM=NY*NX(J)
       CALL COPIA(AREA,NELEM,TABS)
107   CONTINUE
      NCOL=NX(J1)
C
C       VERIFICA SE EXISTE REDUCAO NA PROBABILIDADE DE ERRO, OU SEJA,
C       SE LAMBDA E' DIFERENTE DE ZERO.
      IF(MSMC.EQ.MTY) GO TO 9001
C
C       COLOCA J1 EM VS (VARIAVEIS JA' SELECIONADAS).
      NVS=1
      VS(1)=J1
C
C       VERIFICA SE A ASSOCIACAO MAXIMA FOI ATINGIDA, OU SEJA, LAMBDA
C       IGUAL A UM.
      IF(MSMC.EQ.N) GO TO 9002
C
C       CALCULA LAMBDA
      LAMBDA=FLOAT(MSMC-MTY)/DENOM
C
C       CALCULA A VARIANCIA DE LAMBDA
      VAR=VARIAN(TABS,NY,NX(J1),MSMC,MCOL)
C
C       VERIFICA SE A ASSOCIACAO DESEJADA FOI ATINGIDA
      LIMSUP=LAMBDA+TO*SQRT(VAR)
      IF(LIMSUP.GT.1.) LIMSUP=1.
C
C       MOSTRA O RESULTADO DO PRIMEIRO PASSO
      WRITE(KOUT,1005) TITULO,NVS,(NOMEX(J1,K),K=1,5),LAMBDA,LIMSUP
1005  FORMAT('1*UM  METODO  DE  SELECAO  DE  VARIAVEIS  POLITOMICAS',
```

```
          1  [illegible] 10A4)
          2  ///,1X,'PASSO      VARIAVEL SELECIONADA',
          3  '            LARGURA      LIMSUP',
          4  //,1X,I3,7X,5A4,4X,F7.5,4X,F7.5)
             IF(LIMSUP.GE.LO) GO TO 9003
      C
      C        APENAS A VARIAVEL J1 NAO FORNECE O GRAU DE ASSOCIACAO DESEJADO.
      C        O PROCESSO CONTINUA ...
      C
      C        COM EXCESSAO DE J1, COLOCA EM VNS (VARIAVEIS AINDA NAO SELECIO-
      C        NADAS) AS DEMAIS P-1 VARIAVEIS.
             NVNS=0
             DO 109 L=1,P
              J=PRIOR(L)
              IF(J.EQ.J1) GO TO 109
              NVNS=NVNS+1
              VNS(NVNS)=J
      109    CONTINUE
      C
      C        INICIALIZA Z (AGREGACAO DE TODAS AS VARIAVEIS ATE' ENTAO SELE-
      C        CIONADAS) COM OS ELEMENTOS DE J1.
             NZ=NX(J1)
             CALL COPIA(X(1,J1),N,Z)
      C
      C
      C        PASSOS SUBSEQUENTES
      C
      C        VERIFICA SE O NUMERO DE VARIAVEIS JA' FOI ESGOTADO
      2      IF(NVNS.EQ.0) GO TO 9004
      C
      C        AGREGA 'A Z AS VARIAVEIS AINDA NAO SELECIONADAS
             MSMCA=0
             DO 111 L=1,NVNS
              J=VNS(L)
              CALL AGREGA(Z,NZ,X(1,J),NX(J),ZA,NZA,MAXAGR,AREA)
              CALL MONTAB(Y,NY,ZA,NZA,AREA,SMC)
              IF(MSMCA.GE.SMC) GO TO 111
              MSMCA=SMC
              JS=J
              NZS=NZA
              NELEM=NY*NZS
              CALL COPIA(AREA,NELEM,TABS)
              CALL COPIA(ZA,N,ZS)
      111    CONTINUE
             NCOL=NZS
      C
      C        VERIFICA SE HOUVE GANHO
             IF(MSMCA.EQ.MSMC) GO TO 9005
      C
      C        HOUVE, PORTANTO JS ESTA' SELECIONADA.
             MSMC=MSMCA
      C
      C        ADICIONA JS 'A VS
```

```
      NVS=NVS+1
      VS(NVS)=JS
C
C        RETIRA JS DE VNS
      DO 113 L=1,NVNS
       IF(VNS(L).EQ.JS) GO TO 3
113   CONTINUE
3     DO 114 J=L,NVNS
       VNS(L)=VNS(L+1)
114   CONTINUE
      NVNS=NVNS-1
C
C        Z PASSA A SER A AGREGACAO QUE INCLUI A VARIAVEL JS
      NZ=NZS
      CALL COPIA(ZS,N,Z)
C
C        CONSIDERA AS REGRAS DE PARADA:
C        A ASSOCIACAO MAXIMA FOI ATINGIDA?
      IF(MSMC.EQ.N) GO TO 9006
C
C        CALCULA O NOVO LAMBDA E SUA VARIANCIA
      LAMBDA=FLOAT(MSMC-MTY)/DENOM
      VAR=VARIAN(TABS,NY,NX(JS),MSMC,MCOL)
C
C        A ASSOCIACAO DESEJADA FOI ATINGIDA?
      LIMSUP=LAMBDA+TO*SQRT(VAR)
      IF(LIMSUP.GT.1.) LIMSUP=1.
C
C        MOSTRA O RESULTADO DESSE PASSO
      WRITE(KOUT,1006) NVS,(NOMEX(JS,K),K=1,5),LAMBDA,LIMSUP
1006  FORMAT(/,1X,I3,7X,5A4,4X,F7.5,4X,F7.5)
      IF(LIMSUP.GE.LO) GO TO 9007
C
C        E, FINALMENTE, REPETE-SE O PROCESSO
      GO TO 2
C
C
C        TRATAMENTO DE ENTRADAS INVALIDAS
C
C        NUMERO DE VARIAVEIS INVALIDO
800   WRITE(KOUT,1080) P
1080  FORMAT('0ERRO NA ENTRADA:',
     1    /,' NUMERO DE VARIAVEIS INVALIDO (',I2,')',/)
      STOP
C
C        NUMERO DE INDIVIDUOS INVALIDO
801   WRITE(KOUT,1081) N
1081  FORMAT('0ERRO NA ENTRADA:',
     1 /,' NUMERO DE INDIVIDUOS INVALIDO (',I3,')',/)
      STOP
C
C        LO INVALIDO
802   WRITE(KOUT,1082) LO
```

```
1082  FORMAT('OERRO NA ENTRADA:',
     1  /,' LO INVALIDO (',F5.2,')',/)
      STOP
C
C        PRIORIDADE DE ESCOLHA DE VARIAVEL INVALIDA
803   WRITE(KOUT,1083) I
1083  FORMAT('OERRO NA ENTRADA:',
     1  /,  PRIORIDADE DE ESCOLHA DA ',I2,'-ESIMA VARIAVEL ',
     2  'INVALIDA',/)
      STOP
C
C        NUMERO DE CLASSES DE Y INVALIDO
804   WRITE(KOUT,1084) NY
1084  FORMAT('OERRO NA ENTRADA:',
     1  /,' NUMERO DE CLASSES DA VARIAVEL Y INVALIDO (',
     2  I2,')',/)
      STOP
C
C        NUMERO DE CLASSES DE X(J) INVALIDO
805   WRITE(KOUT,1085) J,NX(J)
1085  FORMAT('OERRO NA ENTRADA:',
     1  /,' NUMERO DE CLASSES DA VARIAVEL',I2,' INVALIDO (',
     2  I2,')',/)
      STOP
C
C        VALOR DE Y INVALIDO
806   WRITE(KOUT,1086) I,Y(I)
1086  FORMAT('OERRO NA ENTRADA:',
     1  /,' VALOR DA VARIAVEL Y PARA O INDIVIDUO',I4,
     2  ' INVALIDO (',I2,')',/)
      STOP
C
C        VALOR DE X INVALIDO
807   WRITE(KOUT,1087) J,I,X(I,J)
1087  FORMAT('OERRO NA ENTRADA:',
     1  /,' VALOR DA VARIAVEL',I3,' PARA O INDIVIDUO',I4,
     2  ' INVALIDO (',I2,')',/)
      STOP
C
C
C        NUMERO INCOMPATIVEL DE REGISTROS DE ENTRADA
C
C        EOF NO PRIMEIRO REGISTRO
900   WRITE(KOUT,1090)
1090  FORMAT('OERRO NA ENTRADA:',
     1     /,' ENCONTROU FIM-DE-ARQUIVO AO LER O PRIMEIRO REGISTRO',/)
      STOP
C
C        EOF NO SEGUNDO REGISTRO
901   WRITE(KOUT,1091)
1091  FORMAT('OERRO NA ENTRADA:',
     1     /,' ENCONTROU FIM-DE-ARQUIVO AO LER O SEGUNDO REGISTRO',/)
      STOP
```

```
      C
      C        EOF AO LER NUMERO DE CLASSES E NOME DAS VARIAVEIS
      902  WRITE(KOUT,1092)
      1092 FORMAT('0ERRO NA ENTRADA:',
          1     /,' ENCONTROU FIM-DE-ARQUIVO AO LER OS REGISTROS COM',
          2     /,' NUMERO DE CLASSES E NOME DAS VARIAVEIS',/)
           STOP
      C
      C        EOF AO LER OS DADOS DO I-ESIMO INDIVIDUO
      903  WRITE(KOUT,1093) I
      1093 FORMAT('0ERRO NA ENTRADA:',
          1     /,' ENCONTROU FIM-DE-ARQUIVO AO LER O REGISTRO CORRES-',
          2     /,' PONDENTE AO',I4,'-ESIMO INDIVIDUO',/)
           STOP
      C
      C        MAIS REGISTROS DO QUE O INFORMADO
      904  WRITE(KOUT,1094)
      1094 FORMAT('0ERRO NA ENTRADA:',
          1     /,' HA'' MAIS REGISTROS NO ARQUIVO DE DADOS',
          2     /,' DO QUE O DECLARADO',/)
           STOP
      C
      C
      C        AQUI SAO FEITAS AS SAIDAS DO SISTEMA
      C
      C
      C        O METODO NAO PODE SER APLICADO
      9000 WRITE(KOUT,1900)
      1900 FORMAT(/,'0O METODO NAO PODE SER APLICADO, POIS A MEDIDA DE ',
          1  /,' ASSOCIACAO E'' INDETERMINADA.',
          2  /,' A POLITOMIA Y FOI MAL DEFINIDA OU O TAMANHO DA ',
          3  /,' AMOSTRA E'' INSUFICIENTE.')
           STOP
      C
      C        NAO EXISTE EVIDENCIA DE ASSOCIACAO
      9001 WRITE(KOUT,1901)
      1901 FORMAT(/,'0NAO EXISTE EVIDENCIA DE ASSOCIACAO.')
           STOP
      C
      C        ASSOCIACAO MAXIMA FOI ATINGIDA COM UMA VARIAVEL
      9002 LAMBDA=1.
           WRITE(KOUT,1005) TITULO,NVS,(NOMEX(J1,K),K=1,5),LAMBDA
           WRITE(KOUT,1902)
      1902 FORMAT(/,'0O GRAU DE ASSOCIACAO MAXIMO FOI ATINGIDO COM ',
          1  /,' APENAS UMA VARIAVEL.')
           GO TO 9008
      C
      C        ASSOCIACAO DESEJADA FOI ATINGIDA COM UMA VARIAVEL
      9003 WRITE(KOUT,1903)
      1903 FORMAT(/,'0O GRAU DE ASSOCIACAO DESEJADO FOI ATINGIDO COM ',
          1  /,' APENAS UMA VARIAVEL.')
           GO TO 9008
           STOP
```

```
C
C         TODAS AS VARIAVEIS FORAM SELECIONADAS
9004  WRITE(KOUT,1904)
1904  FORMAT(/,'0TODAS AS VARIAVEIS JA'' FORAM SELECIONADAS.')
      GO TO 9008
C
C         NAO HA' AUMENTO NA ASSOCIACAO
9005  WRITE(KOUT,1905)
1905  FORMAT(/,'0A INCLUSAO DE NOVAS VARIAVEIS NAO AUMENTA O GRAU DE',
     1  ' ASSOCIACAO.')
      GO TO 9008
C
C         A ASSOCIACAO MAXIMA FOI ATINGIDA
9006  LAMBDA=1.
      WRITE(KOUT,1006) NVS,(NOMEX(JS,K),K=1,5),LAMBDA
      WRITE(KOUT,1906)
1906  FORMAT(/,'0O GRAU DE ASSOCIACAO MAXIMO FOI ATINGIDO.')
      GO TO 9008
C
C         A ASSOCIACAO DESEJADA FOI ATINGIDA
9007  WRITE(KOUT,1907)
1907  FORMAT(/,'0O GRAU DE ASSOCIACAO DESEJADO FOI ATINGIDO.')
C
C         IMPRESSAO DAS VARIAVEIS SELECIONADAS, TABELA E LAMBDA FINAIS
9008  WRITE(KOUT,1908) TITULO,((NOMEX(VS(J),K),K=1,5),J=1,NVS)
1908  FORMAT('1*UM  METODO  DE  SELECAO  DE  VARIAVEIS  POLITOMICAS',
     1   //,1X,'*',10A4,
     2  ///,1X,'VARIAVEIS INDEPENDENTES SELECIONADAS:',
     3    /,(/,3X,5A4))
      WRITE(KOUT,1909) NOMEY,(YABRE,I,FECHA,I=1,NY)
1909  FORMAT(//,1X,'TABELA DE CONTINGENCIA DE Z FINAL VERSUS Y:',
     1  //,3X,'Z : AGREGACAO DAS VARIAVEIS INDEPENDENTES ',
     2            'SELECIONADAS',
     3   /,3X,'Y : VARIAVEL DEPENDENTE - ',5A4,
     4  //,10X,15(2X,A2,I2,A1))
      CALL IMPTAB(TABS,NY,NCOL,14+NVS*2)
      WRITE(KOUT,1911) LAMBDA
1911  FORMAT(//,1X,'GRAU DE ASSOCIACAO FINAL:',F7.5,////)
      STOP
      END
```

```
      SUBROUTINE MONTAB(Y,NY,X,NX,TAB,SMC)
C
C       ROTINA PARA MONTAR A TABELA DE CONTINGENCIA Y VERSUS X.
C       ALEM DISSO, FORNECE A SOMA DOS MAXIMOS DAS COLUNAS.
C
C
      INTEGER Y(1),X(1),TAB(NY,NX),SMC
C
      COMMON /ALL/ N
C
      DO 100 I=1,NY
       DO 100 J=1,NX
        TAB(I,J)=0
100   CONTINUE
C
      DO 101 I=1,N
       KY=Y(I)
       KX=X(I)
       TAB(KY,KX)=TAB(KY,KX)+1
101   CONTINUE
C
      SMC=0
      DO 102 J=1,NX
       MAX=0
       DO 103 I=1,NY
        NIND=TAB(I,J)
        IF(NIND.GT.MAX) MAX=NIND
103    CONTINUE
       SMC=SMC+MAX
102   CONTINUE
C
      RETURN
      END
```

```
      FUNCTION VARIAN(TAB,NY,NX,SMC,MCOL)
C
C        CALCULA A VARIANCIA DO ESTIMADOR DE LAMBDA
C
C
      INTEGER TAB(NY,NX),MCOL(1),VMTY(1),SMC.SE,SEMIN
C
      COMMON /ALL/    N
      COMMON /PVAR/   MTY.DENOM3,NMTY,VMTY
C
C        GUARDA ONDE OCORREU O MAXIMO DE CADA COLUNA
      DO 100 J=1,NX
       MAX=0
       DO 101 I=1,NY
        NIND=TAB(I,J)
        IF(NIND.LE.MAX) GO TO 101
        MAX=NIND
        MCOL(J)=I
101    CONTINUE
100   CONTINUE
C
C        CALCULA "SE" MINIMO (QUE IMPLICA EM MAIOR VARIANCIA)
      SEMIN=N
      DO 102 LOOP=1,NMTY
       SE=0
       IVMTY=VMTY(LOOP)
       DO 103 J=1,NX
        I=MCOL(J)
        IF(I.EQ.IVMTY) SE=SE+TAB(I,J)
103    CONTINUE
       IF(SE.LT.SEMIN) SEMIN=SE
102   CONTINUE
C
C        E CALCULA-A
      VARIAN = FLOAT((N-SMC)*(SMC+MTY-2*SEMIN)) / DENOM3
      RETURN
      END
```

```
      SUBROUTINE AGREGA(Z,NZ,X,NX,ZA,NZA,MAXAGR,CLASSE)
C
C       ROTINA PARA FAZER A AGREGACAO DE "X" A "Z".
C
      INTEGER Z(1),X(1),ZA(1),CLASSE(MAXAGR,1)
C
      COMMON /ALL/ N,KOUT
C
C       VERIFICA QUAIS AS CLASSES DE Z E X QUE OCORREM CONJUNTAMENTE
      DO 100 I=1,NZ
       DO 100 J=1,NX
        CLASSE(I,J)=0
100   CONTINUE
C
      DO 101 I=1,N
       KZ=Z(I)
       KX=X(I)
       CLASSE(KZ,KX)=1
101   CONTINUE
C
      NCLASS=0
      DO 102 I=1,NZ
       DO 102 J=1,NX
        IF(CLASSE(I,J).EQ.0) GO TO 102
        IF(NCLASS.EQ.MAXAGR) GO TO 900
        NCLASS=NCLASS+1
        CLASSE(I,J)=NCLASS
102   CONTINUE
C
      NZA=NCLASS
C
C       E CLASSIFICA OS INDIVIDUOS SEGUNDO A AGREGACAO
      DO 103 I=1,N
       KZ=Z(I)
       KX=X(I)
       ZA(I)=CLASSE(KZ,KX)
103   CONTINUE
C
      RETURN
C
C       OVERFLOW NO NUMERO DE CLASSES DA AGREGACAO
900   WRITE(KOUT,1000)
1000  FORMAT('0ERRO NO PROCESSAMENTO:',
     1    /,' O NUMERO MAXIMO DE CLASSES DA AGREGACAO FOI',
     2      ' ULTRAPASSADO.')
      STOP
      END
```

```
      SUBROUTINE IMPTAL(TAB,NY,NZ,LINIMP)
C
C       ROTINA PARA IMPRIMIR A TABELA DE CONTINGENCIA Y VERSUS Z.
C
C
      INTEGER TAB(NY,NZ)
C
      COMMON /ALL/         N,KOUT
C
      NLIN=LINIMP
      DO 100 LOOP=1,NZ
       IF(NLIN.LT.59) GO TO 1
       WRITE(KOUT,1000)
1000   FORMAT(1H1)
       NLIN=0
1      WRITE(KOUT,1001) LOOP,(TAB(I,LOOP),I=1,NY)
1001   FORMAT(3X,'Z(',I3,')',10I7)
       NLIN=NLIN+1
100   CONTINUE
C
      RETURN
      END
```

```
                                        TITLE   COPIA   - ROTINA PARA TRANSFERIR BLOCOS DE MEMORIA
                                        SUBTTL  ADALBERTO F SINDO  -  FSEADE, 16/03/79
                                        XLIST
                                        LIST
                                ;
                                ;       C O P I A        ROTINA PARA COPIAR UMA AREA DE MEMORIA EM OU-
                                ;                        TRA, UTILIZANDO A INSTRUCAO 'BLT'.
                                ;
                                ;       CHAMADA:         CALL COPIA(ORIG,NPAL,DEST)
                                ;
                                ;       ONDE:            ORIG - ENDERECO DA AREA QUE SERA TRANSFERIDA
                                ;                        NPAL - NUMERO DE PALAVRAS QUE SERAO TRANSFERIDAS
                                ;                        DEST - ENDERECO DA AREA DESTINO.
                                ;
                                        ENTRY   COPIA
000000'  254 00 0 00 000003'    COPIA:  JRST    .+3             ;ENTRADA DE F10
000001'  201 00 0 16 000003             MOVEI   10,NUMARG(ARG)  ;F40, PEGA ENDERCO DE RETORNO
000002'  261 17 0 00 000000             PUSH    P,10            ;EMPILHA ENDERECO DE RETORNO.
000003'  261 17 0 00 000002             PUSH    P,T2            ;SALVA T2
000004'  201 01 1 16 000002             MOVEI   T1,@DEST(ARG)   ;PEGA ENDERECO DE DESTINO
000005'  200 02 0 00 000001             MOVE    T2,T1           ;INICIALIZA ENDERECO DE FINAL DE BLT
000006'  505 01 1 16 000000             HRLI    T1,@ORIG(ARG)   ;COMPLETA BLT-POINTER COM END. DE ORIGEM
000007'  270 02 1 16 000001             ADD     T2,@NPAL(ARG)   ;OBTEM ENDERECO FINAL DA TRANSFERENCIA
000010'  251 01 0 02 777777             BLT     T1,-1(T2)       ;E TRANSFERE AS NPAL PALAVRAS.
000011'  262 17 0 00 000002             POP     P,T2            ;RESTAURA T2
000012'  263 17 0 00 000000             POPJ    P,              ;E RETORNA.
                                        END
```

```
C
C        SUBROUTINE COPIA(FONTE,NELEM,DEST)
C
C        ESTA E' A VERSAO EM FORTRAN-IV DA ROTINA COPIA:
C        RETIRE OS INDICADORES DE COMENTARIO PARA EXECU-
C        TAR O PROGRAMA FORA DO DEC-10.
C
C        INTEGER FONTE(1),DEST(1)
C        DO 100 LOOP=1,NELEM
C         DEST(LOOP)=FONTE(LOOP)
C100     CONTINUE
C        RETURN
C        END
```

A2.4 Instruções para uso no DEC-10

O programa principal encontra-se dimensionado com os seguintes parâmetros:

i) tamanho da amostra (N), menor ou igual a 200 (MAXIND)

ii) número de variáveis independentes, menor ou igual a 15 (MAXVAR)

iii) número de classes por variável, menor ou igual a 15 (MAXCLA)

iv) número de classes em uma agregação menor ou igual a 150 (MAXAGR).

Para uso de diferentes parâmetros, modificações são apenas nos comandos INTERGER (linhas 8, 9, 10, 11, 12), DATA (linha 2) nos seguintes dimensionamentos:

```
INTERGER Y("N"),  X("N", "P"),  PRIOR ("P"),..., NX("P")
      1  ..., NOMEX ("P", "?"),...
      2  TY("MAXCLA"), VMTY("MAXCLA"), VS("P"), VMS("P")
      3  Z("N"), ZA("N"), ZS("N"), TABS("MAXCLA", "?")
      4  ..., AREA ("MAXCLA"), MCOL("MAXCLA")
DATA MAXIND/"N"/, MAXVAR/"P"/, MAXCLA/"P"/, MAXAGR/"?"/
```

e nos comandos FORMAT 1001 and 1003 (linhas 21 e 62). Não são necessárias modificações nas sub-rotinas.

. O arquivo de dados; com o nome FOR20.DAT e com as classes de cada variável codificadas por 01, 02,...,"MAXCLA", deve ser gravado no seguinte formato:

REGISTRO	FORMATO	DESCRIÇÃO
1	A10	o título
2	I2,I3,2F5.3	número de variáveis, número de variáveis, nível de associação desejado e o $(1 - \alpha)$ centil da distribuição normal padrão
3	20I2	ordem de escolha das variáveis independentes para o caso de empate
4	I2,5A4	número de classes da variável dependente Y e seu rótulo
5	I2,5A4	número de classes da variável $X^{(1)}$ e seu rótulo
6	I2,5A4	número de classes da variável $X^{(2)}$ e seu rótulo

....

. . . .

5+p	I2,5A4	número de classes da variável $X^{(P)}$ e seu rótulo
5+p+1	21I2	$Y_1, X_1^{(1)}, X_1^{(2)}, \ldots, X_1^{(P)}$
5+p+2	21I2	$Y_2, X_2^{(1)}, X_2^{(2)}, \ldots, X_2^{(P)}$
. . . . . .		
5+p+n	21I2	$Y_n, X_n^{(1)}, X_n^{(2)}, \ldots, X_n^{(P)}$

. Com o programa principal suas sub-rotinas e o arquivo FOR20.DAT gravados em disco, use o comando padrão para execução de programas em FORTRAN, EX, seguido dos nomes TESE4 e COPIA, isto é,

. EX TESE4, COPIA

e uma saída, com formato planejado para 72 colunas (listagens podem ser pedidas para terminal (LA30) ou impressora), e a saída será obtida em um arquivo de nome ??????. LPT

REFERÊNCIA BIBLIOGRÁFICA

[1] Allen, D. (1971). Mean Square Error of Prediction as a Criterion for Selecting Variables. Technometrics 13, 469 - 75.

[2] Anscombe, F.J. (1967). Topics in the Investigation of Least Squares. J. Roy. Statistics Soc., Series B, 29 , 1 - 29.

[3] Bishop, Y.M.M.; Fienberg, S.E.; Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice. The MIT Press, Massachusetts.

[4] Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

[5] Daniel, C. e Wood, F.S. (1971). Fitting Equations. Data: Computer Analysis of Multifactor Date Scientists and Enginners, J. Wiley, New York.

[6] Draper, N.R. e Smith, H. (1966). Applied Regression Analysis. John Wiley and Soons. Inc., New York,

[7] Fisher, R.A. (1948). Statistical Methods for Research.Workers, Thenth Edition, New York, Hefner Publ. Co..

[8] Furnival, G.M. e Wilson, R.W. (1974). Regression by Leaps and Bounds. Technometrics 16, 499 - 511.

[9] Gini, C. (1912). Variabilitã e Mutabilita, Contributo allo Studio delle Distribuzioni; Relazione Statische.Studi Economico. Universitá di Ceglicri.

[10] Goodman, L.A. (1963). On Methods for Comparing Contingency

[18] Guttman, L. (1941). An Outline of the Statistics Theory of Prediction. Supplementary Study B-1 in Horst. The Prediction of Personal Adjustment. Social Science Research , New York 48, 253 - 318.

[19] Kendall, M.G. (1948). <u>The Advanced Theory of Statistics</u>. London Charles Griffin and Co. Ltd.

[20] Kozelka, R.M. (1952). On Some Special Order Statistics from the Multinomial Distribution. Unpublished Ph.D. dissertation, Havard University.

[21] Kozelka, R.M. (1956). Approximate Upper Percentage Points for Extreme Values in Multinomial Sampling Annals of Mathematical Statistics 27, 507 - 12.

[22] Light, R.J. e Margolin, B.H. (1971). Analysis of Variance for Categorical Data. J. of the American Statistical Association 66, 534 - 44.

[23] Mallows, C.L. (1973). Some Comments on C_p Technometrics 15, 661 - 75.

[24] Pearson, K. (1904). On the Theory of Contingency and Normal Correlation. Draper's Co. Res. Mem. Biométrica - Ser. 1.

[25] Tukey, J.W. (1967). Discussion of Anscombe's Paper. J. Roy. Statistics. Soc., Series B, 29, 47 - 48.

[26] Yule, G.U. (1912). On the Methods of Mesuring Association Between Two Attributes. J. of Roy. Statistical Soc. 75, 579 - 642.

[27] Yule, G.V. e Kendall, M.G. (1950). <u>An Introduction to the Theory of Statistics</u>. London, Charles Griffin and Co. Ltd.

Tables. J. Roy. Statistics Soc., Series A, 126 , 94 - 108.

[11] Goodman, L.A. (1971). The Analysis of Multidimensional Contingency Tables. Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications. Technometrics 13, 33 - 61.

[12] Goodman, L.A. (1973). Guided and Unguided Methods for the Selection of Models for a Set of T Multidimensional Contingency Tables. J. of the Am. Statistical Association 68, 165 - 75.

[13] Goodman, L.A. e Kruskal, W.H. (1954). Mesures of Association for Cross Classification. J. of the American Statistical Association 49, 732 - 64.

[14] Goodman, L.A. e Kruskal, W.H. (1959).Measures of Association For Cross Classification II: further discussion and references. J. of the American Statistical Association 54, 123 - 63.

[15] Goodman, L.A. e Kruskal, W.H. (1963). Measures of Association for Cross Classification III: aproximate sampling theory. J. of the American Statistical Association 58, 310 - 64.

[16] Goodman, L.A. e Kruskal, W.H. (1972). Measures of Association for Cross Classification IV: simplification of asymptotic Variances. J. of the American Statistical Association 67, 415 - 21.

[17] Greenwood, R.E. e Glasgov, M.O. (1950). Distribution of Maximum and Minimum Frequencies in a Sample Drawn from a Multinomial Distribution. Annals of Mathematical Statistics 21, 416 - 24.