

ALGUMAS TÉCNICAS DE ANÁLISE

MULTIVARIADA DISCRETA

MASAHARU KUNIYOSHI

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ÁREA DE CONCENTRAÇÃO ESTATÍSTICA

ORIENTADOR, PROF. DR. ADOLPHO WALTER PIMAZONI CANTON

SÃO PAULO, NOVEMBRO DE 1982

ã minha mulher,
Zaira.

às minhas filhas
Natalie e Lia.

AGRADECIMENTOS

Queremos aqui registrar nossos sinceros agradecimentos a todos os que direta ou indiretamente colaboraram para a concretização deste trabalho, e em particular:

- ao Prof. Dr. Adolfo Walter P. Canton pela dedicação com que realizou a árdua tarefa de orientação;
- aos funcionários da Biblioteca do IME-USP pela inestimável colaboração;
- aos professores e colegas do IME pela atenção e incentivos;
- ã Dona Selma Lemos de Melo, secretária da C.P.G. do IME-USP pela atenção e solicitude;
- ã Yoshie Yamanaka pelo excelente trabalho de revisão;
- ã Regina Helena da Silva e Luzia do Carmo Namiki pelo paciente e meticoloso trabalho de datilografia.

SUMÁRIO

	Página
CAPÍTULO 1 -	
1.1 - INTRODUÇÃO	2
CAPÍTULO 2 - ANÁLISE DE RIDIT (AR)	
2.1 - INTRODUÇÃO.	4
2.2 - NOTAÇÃO, COMPUTAÇÃO DE "RIDIT" E MÉ- DIA "RIDIT".	5
2.3 - A MÉDIA "RIDIT COMO INSTRUMENTO PA- RA MEDIR RELAÇÃO ENTRE VARIÁVEIS . .	9
2.4 - APLICAÇÃO.	18
CAPÍTULO 3 - ANÁLISE DE REGRESSÃO	
3.1 - INTRODUÇÃO.	27
3.2 - ANÁLISE DE REGRESSÃO SIMPLES E CANÔ- NICA.	28
3.3 - COMPONENTES PRINCIPAIS E RELAÇÕES PRINCIPAIS	29
3.4 - ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA	30
3.5 - ANÁLISE MULTIVARIADA.	33
3.5.1 - Mínimos Quadrados Ordinários (MQO)	33
3.5.2 - Correlação Canônica (CC)	34
3.5.3 - Regressão Canônica (RC)	35
3.6 - DADOS CATEGÓRICOS.	35
3.6.1 - Transformação de Dados Qualitati- vos.	35
3.6.2 - Regressão Canônica com Dados Ca- tegorizados	36
3.7 - APLICAÇÃO	37

CAPÍTULO 4 - ANÁLISE DE CORRESPONDÊNCIA	
4.1 - INTRODUÇÃO	44
4.2 - DEFINIÇÕES E INTERPRETAÇÕES.	45
4.3 - MAXIMIZAÇÃO DE COEFICIENTES DE CORRE LAÇÃO.	47
4.4 - REPRESENTAÇÃO GRÁFICA DOS FATORES. .	52
4.5 - ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA (ACM).	55
4.6 - REGRAS DE INTERPRETAÇÃO.	60
4.7 - APLICAÇÃO.	61
CAPÍTULO 5 - CONCLUSÕES	
5.1 - INTRODUÇÃO.	69
5.2 - RESUMO DAS CONCLUSÕES DAS ANÁLISES .	69
APÊNDICE 1 - CARACTERIZAÇÃO MATEMÁTICA DE "RIDIT".	72
APÊNDICE 2 - AUTOVETORES DE B NA MÉTRICA DE A	78
APÊNDICE 3 - TABELA DE FREQUÊNCIA DE RENDA (Y) POR ESCOLARIDADE, IDADE E ATIVIDADE ($E \times I \times A$)	81
BIBLIOGRAFIA	82

CAPÍTULO 1

1.1 - INTRODUÇÃO

Nos dias de hoje, são muito comuns os levantamentos por meio de entrevistas e/ou questionários. Estes levantamentos são utilizados em pesquisas na área sociológica, econômica etc., devido à necessidade de se conhecer melhor o perfil do indivíduo ou entidade sujeitos à pesquisa. Fazemos entrevistas ou observações para coletar as informações dos mesmos, cujos dados são multivariados, muitas vezes de grande dimensão tornando difícil a sua interpretação. É inegável que, por trás dessa grande quantidade de dados, estão ocultas valiosas informações. Daí a necessidade de analisar melhor a estrutura dos dados de tal forma que se capte a informação latente que os dados contêm. Com esta finalidade foram desenvolvidas várias técnicas que viabilizam a interpretação dos dados.

O objetivo deste trabalho é apresentar algumas técnicas com finalidade exploratória e fazer algumas comparações entre elas, ressaltando que nenhuma tentativa de análise confirmatória é feita.

No capítulo 2 descrevemos a técnica de Análise de "ridit". Uma das limitações desta técnica é que a variável "ridit" deve ser variável ordinal. Apesar disso, para o que propomos, ela é de extrema importância, principalmente na análise de dados de pes

quisas epidemiológicas onde foi inicialmente desenvolvida.

A seguir, no capítulo 3, apresentamos técnicas similares: a de análise de regressão simples e a canônica. Esses métodos são derivados de métodos quantitativos. Também apresentamos uma comparação, quanto à qualidade, dos resultados obtidos através desses métodos.

No capítulo 4 apresentamos uma das técnicas exploratórias mais empregadas pelos pesquisadores europeus no campo de análise multivariada: a Análise de Correspondência. Esta análise é bastante flexível quanto aos dados que são sujeitos à análise, ou seja, dados resultantes de levantamento por meio de entrevista. Em geral esses dados são provenientes de um número muito grande de variáveis e de categorias. A associação entre as mesmas (variáveis) é captada ou mostrada "visualmente" por meio desta técnica.

Finalmente, no capítulo 5, apresentamos um resumo das conclusões dos resultados obtidos pelas diferentes técnicas exploratórias, bem como uma comparação entre essas técnicas.

Para análise confirmatória, que não foi objeto deste trabalho, sugerimos o uso dos modelos log-lineares e qui-quadrado, baseados na medida de associação.

Os dados, que serão usados para ilustração das várias técnicas exploratórias apresentadas neste trabalho, são provenientes de uma "Pesquisa Estrutural de Renda de Indústria de Manufaturados" de 1972, realizada na Alemanha (Israëls, 1981). As variáveis utilizadas são Renda Bruta Anual (Y), Escolaridade (E), Idade (I) e Atividade (A) e todas estas informações estão no apêndice 3.

A tabela 1-1 apresenta um resumo dos dados referidos.

Tabela 1.1 - Classificação e Distribuição de
Frequência das Variáveis

Variável	Notação	Classificação	Frequência
Renda (Y) u.m.	Y_1	menos de 12	9219
	Y_2	12 — 22	101646
	Y_3	22 — 32	80832
	Y_4	32 — 42	24966
	Y_5	igual ou mais de 42	15170
Escolaridade (E)	E_1	Baixa	42171
	E_2	Média	156283
	E_3	Alta	33379
Idade (I) em anos	I_1	21 — 30	56710
	I_2	30 — 40	63021
	I_3	40 — 50	62790
	I_4	50 — 65	49312
Atividade (A) Industriais	A_1	Têxtil	10993
	A_2	Vestuário	7563
	A_3	Outros	75195
	A_4	Metalurgia Básica	9459
	A_5	Metalurgia	100216
	A_6	Alimentação	28407

CAPÍTULO 2

ANÁLISE DE RIDIT (AR)

2.1 - INTRODUÇÃO

Apresentamos neste capítulo a AR, que é uma técnica de análise de dados multivariados, desenvolvida por Bross (1958).

É uma técnica muito simples de ser aplicada, com vantagem de ser tão eficiente quanto outras técnicas mais complexas.

O nome "ridits" foi escolhido por causa da analogia com "probits" e "logits".

Como outras famílias de "it", a família "ridit" representa um tipo de transformação. Enquanto "probits" são relativos à distribuição teórica (distribuição normal), ridits são relativos à distribuição empírica. Em outras palavras, "ridits" são baseados na distribuição observada de várias respostas para um conjunto especificado de elementos. As três primeiras letras significam "Relative to an Identified Distribution" (Relativo à distribuição identificada).

O método consiste em comparar duas distribuições de uma variável ordinal através de uma média "ridit". Uma das distribuições pertence ao grupo de referência chamada de distribuição i

dentificada. A outra distribuição de freqüência pertence a um grupo experimental. Escolhido o grupo de referência, temos a linha básica para a pesquisa. É, portanto, fundamental que os objetivos sejam descritos claramente, para uma boa escolha da linha básica.

É importante lembrar que o conjunto de referência deve ser suficientemente grande para assegurar a estabilidade de "ridits".

Por conveniência consideramos a variável ordinal, também chamado de variável "ridit", como uma variável contínua que é observada como uma variável discreta em nível ordinal.

A média "ridit" é definida como a probabilidade de um elemento do grupo de referência, escolhido aleatoriamente, ter "escore" menor do que um elemento escolhido aleatoriamente do grupo experimental.

O "ridit" é um número real entre 0 e 1, definido para cada classe de variável ordinal como probabilidade condicional, baseado na distribuição de freqüência do grupo de referência.

As caracterizações das propriedades matemáticas foram desenvolvidas recentemente por Dennis (1980), Brockett (1977) e são apresentadas no apêndice 1.

2.2 - NOTAÇÃO, COMPUTAÇÃO DE "RIDIT" E MÉDIA "RIDIT"

Primeiramente introduzimos notações para o desenvolvimento da técnica de "ridit".

X : Grupo de referência

Y : Grupo de experiência

$N(X)$: Número total de elementos no grupo X

$N(Y)$: Número total de elementos no grupo Y

- $N(X_i)$: Número total de elementos no grupo X_i
 $N(Y_i)$: Número total de elementos no grupo Y_i
 $R(X_i)$: Rídit da classe i da variável "rídit" com K classes

As distribuições de frequência do grupo de referência e experimental com respeito à variável ordinal são apresentadas na tabela 2.1.

Tabela 2.1 - Frequência de grupo de referência e experimental

Classe de var. rid. / Grupo	1	2	K	Total
X (referência)	$N(X_1)$	$N(X_2)$	$N(X_K)$	$N(X)$
Y (experimental)	$N(Y_1)$	$N(Y_2)$	$N(Y_K)$	$N(Y)$

O "rídit" de classe i é igual à probabilidade de um elemento x_1 , escolhido aleatoriamente de X , ter "escore" menor do que o elemento x_2 , dado que x_2 pertence a X_i , ou seja

$$R(X_i) = P(x_1 < x_2 | x_2 \in X_i) \quad (2.1)$$

Considerando a natureza contínua das variáveis, $R(X_i)$ pode ser estimado por

$$R(X_i) = \left[\sum_{j=1}^{i-1} N(X_j) + N(X_i)/2 \right] / N(X) \quad (2.2)$$

Para cada i , $R(X_i)$ é considerado como sendo a nova escala de valores da i -ésima classe da variável ordinal, que corresponde, aproximadamente, ao valor da função-distribuição de frequência acumulada daquela classe. Utilizando estas novas escalas de valores $R(X_i)$, podemos calcular a média da variável ordinal para o grupo experimental Y .

A média "ridit" de Y , dada a distribuição de frequência de X , é definida por:

$$R(Y/X) = \sum_{i=1}^K R(X_i) \cdot \frac{N(Y_i)}{N(Y)} \quad (2.3)$$

Supondo que

$$P(x < y \mid x \in X_i \text{ e } y \in Y_i) = 0.5$$

verifica-se que:

$$R(Y/X) = P(x < y) \quad (2.4)$$

Se considerarmos x e y variáveis discretas temos:

$$R(Y/X) = P(x < y) + 0.5 P(x = y) \quad (2.5)$$

$R(Y/X)$ é uma média ponderada de "ridits", sendo pesos as frequências relativas de Y sobre as da variável "ridit".

Vamos supor que X e Y são amostras de uma população. A interpretação probabilística de (2.4) e (2.5) mostra a relação entre a média "ridit" e a estatística U de Mann-Whitney (Jansen, 1980), considerando $m=N(X)$ e $n=N(Y)$, temos:

$$\widehat{R(Y/X)} = 1 - U/(m.n) \quad (2.6)$$

Desta forma, $\widehat{R(Y/X)}$ pode ser usada para verificar se a população tem a mesma distribuição com respeito à variável ordinal. Utilizando a propriedade de U para grandes valores de m e n , $\widehat{R(Y/X)}$ tem aproximadamente distribuição normal com valor esperado $E[\widehat{R(Y/X)}] = 0.5$ e variância dada por

$$\text{var } \widehat{R(Y/X)} = \frac{1}{(m n)^2} \text{var } U \quad (2.7)$$

onde

$$\text{var } U = \frac{m n}{(m+n)(m+n-1)} \left[\frac{(m+n)^3 - (m+n)}{12} - \sum_{i=1}^K \frac{(t_i^3 - t_i)}{12} \right]$$

t_i = número de elementos do grupo X e do grupo Y pertencentes à i -ésima classe da variável ordinal.

Consideremos agora, o caso em que Y é baseado numa amostra e X não, isto é, $R(X_i)$ fixos e característica estocástica de $\widehat{R(Y/X)}$ é devido à amostragem de Y . Temos então a distribuição de frequência $\widehat{R(Y/X)}$ com X fixos e Y como amostra de uma população P .

Introduzindo a notação p_i ($i=1,2,\dots,K$) como frequência relativa da variável ordinal da população P , a média ridit para a população P é dada por

$$R(P/X) = \sum_{i=1}^K R(X_i) \cdot p_i \quad (2.8)$$

A média "ridit" de Y , dada a distribuição de frequência de X , é uma variável aleatória.

$$\widehat{R(Y/X)} = \sum_{i=1}^K R(X_i) \hat{p}_i \quad (2.9)$$

que tem como média e variância

$$E[\widehat{R(Y/X)}] = R(P/X) \quad (2.10)$$

$$\begin{aligned} \text{var} |\widehat{R(Y/X)}| &= E \cdot \left\{ \sum_{i=1}^K R(X_i) \hat{p}_i - \sum_{i=1}^K R(X_i) p_i \right\}^2 \\ &= \frac{1}{N(Y)} \left\{ \sum_{i=1}^K R^2(X_i) p_i - \left[\sum_{i=1}^K R(X_i) \cdot p_i \right]^2 \right\} \quad (2.11) \end{aligned}$$

Para grandes amostras, utilizamos estes resultados para construir intervalo de confiança de $R(P/X)$.

2.3 - A MÉDIA "RIDIT" COMO INSTRUMENTO PARA MEDIR RELAÇÃO ENTRE VARIÁVEIS

Utilizando a análise de "ridit", podemos investigar a influência de variáveis explicativas na variável "ridit". Suponhamos que uma ou mais variáveis explicativas são medidas junto com a variável "ridit". Desta forma, o grupo experimental pode ser dividido em subgrupos correspondentes às classes de variáveis explicativas.

Vamos supor que a relação que queremos medir é a influência da variável S ou variável T , ambas sobre a distribuição de frequência da variável ordinal Z . As variáveis S e T podem ser nominais ou ordinais.

- NOTAÇÕES COMPLEMENTARES

S	consiste em p classes S_1, S_2, \dots, S_p
T	consiste em q classes T_1, T_2, \dots, T_q
Z	consiste em K classes Z_1, Z_2, \dots, Z_k
Y_{S_i}	é um subgrupo de Y , cujos elementos pertencem à i -ésima classe de S
Y_{T_j}	é um subgrupo de Y cujos elementos pertencem à j -ésima classe de T
$Y_{S_i \cap T_j}$	é um subgrupo de Y cujos elementos pertencem à classe $S_i \cap T_j$
$N(S_i)$	número de elementos que pertencem à classe S_i
$N(T_j)$	número de elementos que pertencem à classe T_j

Primeiramente, consideramos que o grupo experimental é parte de uma população P , que é estratificada pela variável S , isto é, estamos analisando somente as variáveis Z e S . Retiramos de cada estrato uma amostra de $N(S_i)$ elementos, $i=1,2,\dots,p$. Desta

forma temos p subgrupos experimentais disjuntos Y_{S_i} ($i=1,2,\dots,p$)
 onde $Y = \bigcup_{i=1}^p S_i$.

Definindo $P(Y_i)$ como a probabilidade de retirar um elemento de Y_{S_i} , e a média do subgrupo como $R(Y_{S_i}/X)$ temos que a média "ridit" para o grupo experimental Y é dado por

$$R(Y/X) = \sum_{i=1}^p R(Y_{S_i}/X) \cdot P(Y_{S_i}) \quad (2.12)$$

Se desejarmos obter amplitude de intervalos de confiança, aproximadamente iguais, basta selecionar o mesmo número de elementos por estrato S_i .

No gráfico da fig. 2.1 estão representadas as médias "ridits" e seus intervalos de confiança, baseados nos dados numéricos apresentados no capítulo anterior.

Vamos supor que os dados são a amostra de uma população estratificada pela variável Idade (S) e que o grupo de referência é formado pelo total de amostra, isto é $X = \bigcup_{i=1}^p S_i$. O intervalo de confiança ilustrado é arbitrário, pois o mesmo é "muito pequeno" para a escala utilizada.

Analisando o gráfico, podemos notar que as diferenças entre as médias "ridits" $R(Y_{S_i}/X)$ no caso $i=1,2,3,4$ e a média "ridit" da população é acentuada. Se a variável S é uma variável ordinal, podemos verificar se a relação entre as variáveis Z e S são lineares ou não.

Utilizando intervalo de confiança das médias "ridits", podemos analisar somente um subgrupo Y_{S_i} de cada vez. Se a variável Z é normalmente distribuída dentro de cada subgrupo Y_{S_i} , com variâncias iguais, podemos usar a análise de variância para verificar se há ou não diferenças entre as médias "ridits" $R(Y_{S_i}/X)$, $i=1,2,3,4$. Na maioria das vezes a suposição de normalidade não é

satisfeita. Neste caso, temos que usar o teste estatístico de Kruskal-Wallis (Kruskal, 1952) de análise de variância não paramétrica. O que estamos fazendo na verdade, é testar a hipótese:

H_0 : A variável Z tem para cada Y_{S_i} a mesma média

H_1 : A variável Z não tem para cada Y_{S_i} a mesma média.

Lembramos que consideramos escala "ríditi" para a variável Z .

Logo abaixo, temos a tabela 2.2 e as notações que são utilizados no teste de Kruskal-Wallis.

Tabela 2.2 - Distribuição de freqüência observada da variável Z em cada subgrupo Y_{S_i}

Var. Sub Z Grupos	1	2	K	Total
Y_{S_1}	$N(Y_{S_1}^{(1)})$	$N(Y_{S_1}^{(2)})$	$N(Y_{S_1}^{(K)})$	N_1
Y_{S_2}	$N(Y_{S_2}^{(1)})$	$N(Y_{S_2}^{(2)})$	$N(Y_{S_2}^{(K)})$	N_1
⋮	⋮	⋮		⋮	⋮
⋮	⋮	⋮		⋮	⋮
⋮	⋮	⋮		⋮	⋮
Y_{S_p}	$N(Y_{S_p}^{(1)})$	$N(Y_{S_p}^{(2)})$	$N(Y_{S_p}^{(K)})$	N_1
Total	$N(Y^{(1)})$	$N(Y^{(2)})$	$N(Y^{(K)})$	pN_1

A média de postos do r -ésimo grupo de observações

empatadas \bar{e}

$$MQ_r = \frac{N(Y_r)+1}{2} + \sum_{j=1}^{r-1} N(Y_j) \quad (2.13)$$

Desta forma, a soma dos postos na i -ésima subamostra Y_{S_i} é dada por

$$Q_i = \sum_{r=1}^k MQ_r \cdot N(Y_r)$$

O teste de Kruskal - Wallis é dado por

$$\chi_r^2 = \left[\frac{12}{N(N+1)} \sum_{i=1}^k \frac{Q_i^2}{N_i} - 3(N+1) \right] / \left[1 - \frac{\sum_{j=1}^k T_j^2}{N^3 - N} \right] \quad (2.15)$$

com

$N = pN_1$ número total de observações em todas as subamostras combinadas

$Q_i =$ a soma de postos de observações na i -ésima subamostra

$N_i =$ número de observações na i -ésima subamostra, estabelecendo que $N_i = N_1$ para $i = 1, 2, \dots, p$

$T_j = t_j^3 - t_j$

$t_j =$ número de observações empatadas de Y no j -ésimo grupo de "escore" empatado, estabelecendo que

$$t_j = N(Y_j)$$

Sob a hipótese H_0 , a estatística de Kruskal-Wallis en contra-se tabulada. Para grandes valores de p e grandes subamostras, esta distribuição é aproximada pela distribuição qui-quadrada com $(p-1)$ graus de liberdade.

Considerando as variáveis S e T , a população é dividida em pq subpopulações $P_{S_i \cap T_j}$ disjuntas. Desta forma, fazemos amostragens independentes de pq subgrupos experimentais $Y_{S_i \cap T_j}$ disjuntas, sendo $Y = \bigcup_{i,j} Y_{S_i \cap T_j}$. A média "ridit" para os grupos experimentais é $R(Y_{S_i \cap T_j}/X)$ e a média "ridit" para o grupo experimental Y é

$$R(Y/X) = \sum_i \sum_j R(Y_{S_i \cap T_j}/X) P(Y_{S_i \cap T_j}/Y_{S_i}) P(Y_{S_i}) \quad (2.16)$$

onde

$P(Y_{S_i \cap T_j}/Y_{S_i})$: probabilidade para selecionar elementos de $Y_{S_i \cap T_j}$, dado que pertence a Y_{S_i} .

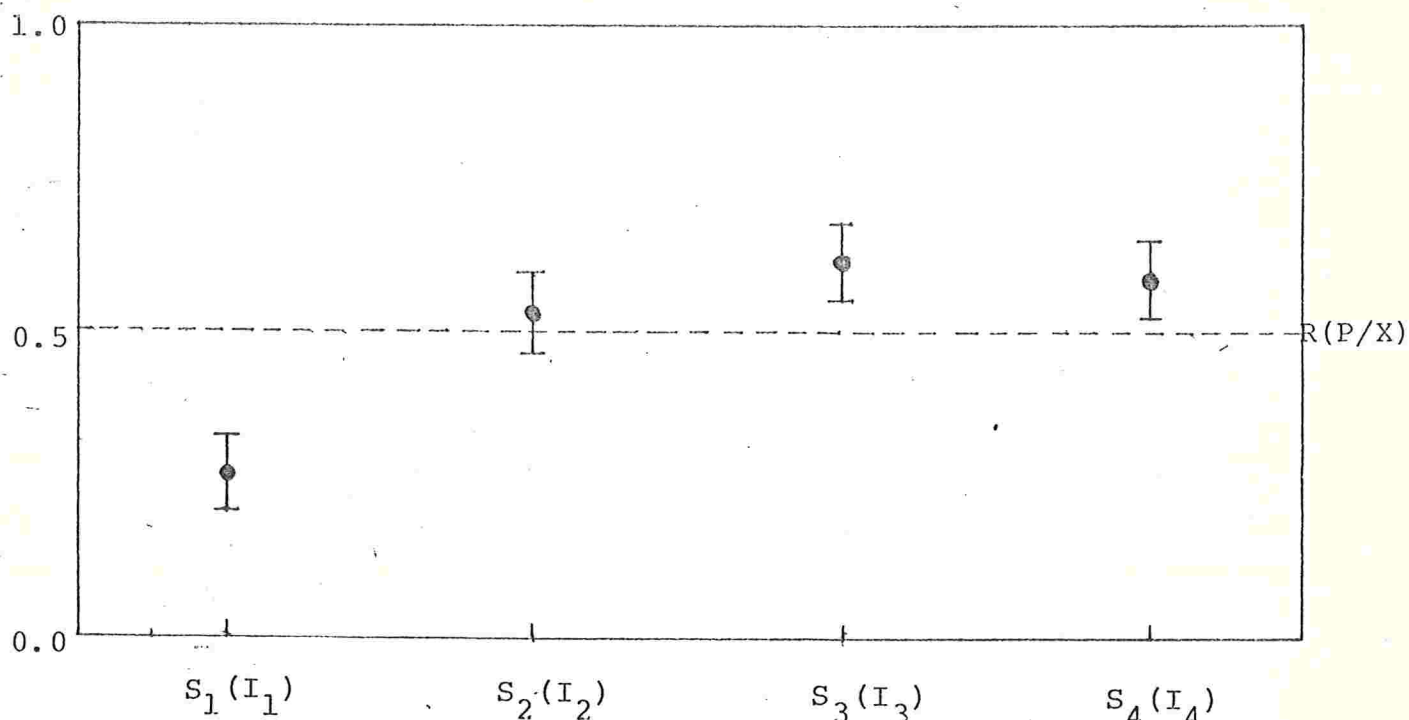
$P(Y_{S_i})$: probabilidade de selecionar um elemento de Y_{S_i} .

A influência de ambas as variáveis pode ser ilustrada graficamente na Fig. 2.2, baseada nos dados numéricos do capítulo 1. Vamos utilizar o mesmo critério adotado no exemplo anterior exceto na estratificação, que é feita pela variável $S \times T$ (Idade \times Escolaridade).

Como no caso anterior, o intervalo de confiança de média "ridit" para cada $Y_{S_i \cap T_j}$, comparada com a média "ridit" da população, mostra para quais valores de S e T a distribuição Z é significativamente diferente. (ver Fig. 2.2).

Podemos testar a influência de S_i e T_j por meio da análise de variância. Se a condição de normalidade não é satisfeita, podemos usar o modelo não paramétrico de análise de variância com dois fatores, denominado Teste de Friedman (Friedman, 1937).

Fig. 2.1 - Médias Ridit e seus intervalos de confiança para 4 subgrupos, definidas pela variável S (Idade).



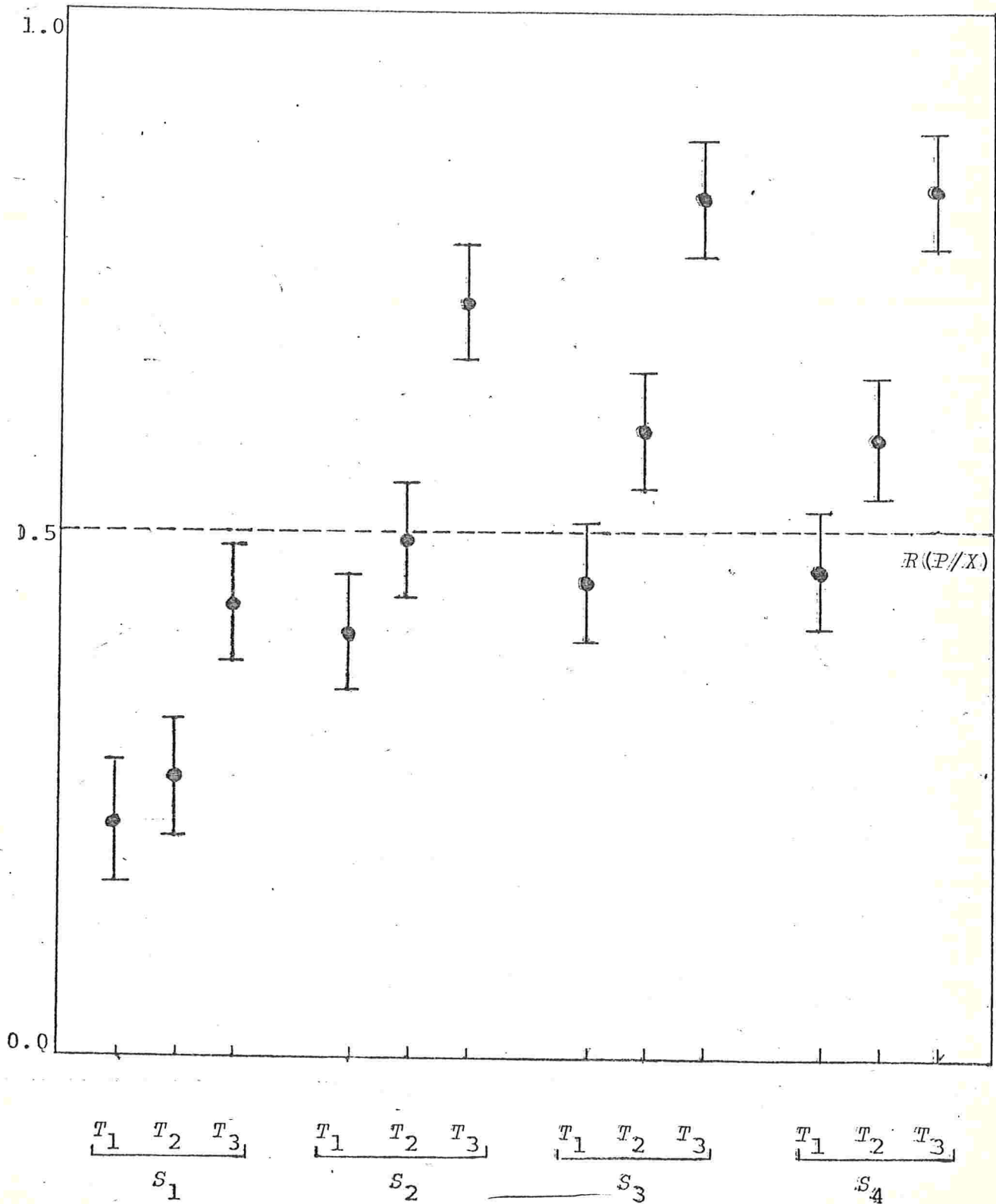
Obs.: $X = \bigcup_{i=1}^4 S_i$

Calculamos a média "ridit" para cada $Y_{S_i \cap T_j}$ e os resultados são colocados na tabela de dupla entrada com p linhas e q colunas.

Tabela 2.3 - Média "ridit" dos subgrupos experimentais definidos pelas variáveis S e T .

Variável T	1	2	...	q	
Variável S					
1	$R(Y_{S_i \cap T_j} / X)$				$R(Y_{S_1} / X)$
2					$R(Y_{S_2} / X)$
...					...
...					...
p					$R(Y_p / X)$
	$R(Y_{T_1} / X) R(Y_{T_2} / X) \dots R(Y_{T_q} / X)$				

Fig. 2.2 - Médias "Ridits" e intervalos de confiança para subgrupos $Y_{S_i} \cap T_j$ ($i=1,2,3$ e 4 e $j = 1,2$ e 3).



Suponhamos, então, que para cada casela da tabela (2.3), um elemento é selecionado aleatoriamente da subpopulação correspondentes, ou seja: para cada bloco (variável S) selecionamos um elemento para cada tratamento (variável T).

Vamos considerar o seguinte modelo estatístico baseado na Tabela 2.3.

$$y = \mu + \alpha_i + \beta_j + e_{ij} \quad \begin{array}{l} i=1,2,\dots,p \\ j=1,2,\dots,q \end{array} \quad (2.17)$$

onde

$$y = R(Y_{S_i \cap T_j} / X) ,$$

μ = é a média geral

α_i = é o efeito do i -ésimo bloco

β_j = é o efeito do j -ésimo tratamento

e_{ij} = variável erro com distribuições independentes

e com as restrições:

$$\sum_i \alpha_i = \sum_j \beta_j = 0$$

Baseados no modelo acima vamos testar:

H_0 : As populações (tratamento) dentro do bloco são idênticas

H_1 : Pelo menos um tratamento é diferente dos demais.

Atribuimos "posto" baseado nas médias "ridits" para cada linha (bloco) da Tabela (2.3) separadamente. Estes "postos" (1,2,3,...,q) são aleatoriamente ordenados para cada linha (bloco).

O teste estatístico χ_r^2 definido por Friedman é uma função da Soma de Postos nas colunas.

$$\chi_r^2 = \frac{12}{pq(q+1)} \left\{ \sum_{j=1}^q Q_j^2 - 3p(q+1) \right\} \quad (2.18)$$

onde:

p = número de linhas ,

q = número de colunas,

Q_j = soma de postos na j -ésima coluna.

Para pequenos valores de p e q , temos tabelas probabilidades exatas de χ_r^2 , e para grandes valores, a distribuição χ_r^2 aproxima-se da distribuição qui-quadrada com $(q-1)$ graus de liberdade.

O procedimento do teste de Friedman pode ser estendido sucessivamente para um número maior de variáveis.

Vamos incluir a terceira variável nominal, ou ordinal V , no grupo experimental, da mesma forma que foi feita anteriormente.

Dadas as variáveis S e T , podemos testar se os efeitos de V têm ou não influência na distribuição da variável "ridit" Z dos subgrupos experimentais definidos pelas variáveis S , T e V .

Logo abaixo, temos a tabela para testar a influência da variável V .

Tabela 2.4 - Médias "ridits" para subgrupos experimentais, definidos por S , T e V

Variável Variável	V_1 V_r	
$S_1 \cap T_1$		$R(Y_{S_1 \cap T_1} / X)$
$S_1 \cap T_2$	$R(Y_{S_i \cap T_j \cap V_k} / X)$	$R(Y_{S_1 \cap T_2} / X)$
\vdots		\vdots
$S_p \cap T_q$		$R(Y_{S_p \cap T_q} / X)$
	$R(Y_{V_1} / X)$ $R(Y_{V_r} / X)$	

O teste estatístico associado à tabela 2.4 é conduzido de maneira análoga ao descrito anteriormente.

2.4 - APLICAÇÃO

O exemplo numérico que vamos utilizar são os dados descritos no capítulo 1. O problema proposto se refere à análise de estrutura de um conjunto de dados de uma população. Para este caso, podemos escolher, como grupo de referência, o conjunto de todos os grupos experimentais. Desta forma, obtemos os "ridits" $R(X_i)$ ($i=1,2,3,4,5$) de cinco classes de renda do grupo de referência, cujo resultado está na tabela abaixo.

Tabela 2.5 - "Ridits" de grupos de referência

Renda (Y)	Classificação	freqüência	$R(X_i)$
Y_1	$0 \text{ --- } 1\mu$	9219	0.0199
Y_2	$1\mu \text{ --- } 2\mu$	101646	0.2590
Y_3	$2\mu \text{ --- } 3\mu$	80832	0.6525
Y_4	$3\mu \text{ --- } 4\mu$	24966	0.8807
Y_5	$4\mu \text{ --- } +\infty$	15170	0.9607

Obs.: $\mu = 12 \times 1\text{u.m}$

Em seguida apresentamos três tabelas de médias "ridits" referentes as variáveis Escolaridade (E), Idade (I) e Atividade (A), calculadas conforme a fórmula (2.3).

Tabela 2.6 - Média "Ridit" de Escolaridade

Escola- ridade	Classifi- cação	$P_{ij} = \frac{N(E_i \cap Y_j)}{N(E_i)}$					$R(E_i/X)$
		Y_1	Y_2	Y_3	Y_4	Y_5	
E_1	Baixa	0.06	0.59	0.28	0.04	0.03	0.4010
E_2	Média	0.04	0.46	0.36	0.09	0.04	0.4728
E_3	Alta	0	0.16	0.37	0.25	0.22	0.7158

Pela tabela acima, podemos dizer que a escolaridade está influenciando na renda, isto é, quanto maior o grau de escolaridade, maior é a renda. Destacamos uma nítida diferença entre a categoria E_3 e as categorias E_1 e E_2 .

Tabela 2.7 - Média "Ridit" de Idade

Idade (anos)	Classi- ficação	$P_{ij} = \frac{N(I_i \cap Y_j)}{N(Y_i)}$					$R(I_i/X)$
		Y_1	Y_2	Y_3	Y_4	Y_5	
I_1	18 — 30	0.15	0.72	0.12	0.01	0	0.2766
I_2	30 — 40	0.01	0.41	0.43	0.11	0.04	0.5225
I_3	40 — 50	0.01	0.28	0.44	0.16	0.11	0.6071
I_4	50 — 65	0.01	0.35	0.39	0.14	0.11	0.5750

O fator Idade, na Renda, apresenta uma característica interessante: à medida que a Idade aumenta melhora a Renda, exceto na última faixa de Idade, entre 50 e 65 anos (I_4), decrescendo um pouco em relação à categoria anterior (I_3). A categoria de Renda Mais Jovem (I_1) encontra-se isolada das demais.

Tabela 2.8 - Médias "Ridits" de Atividade

Atividade	Classificação	$P_i = \frac{N(A_i \cap Y_j)}{N(A_i)}$					$R(A_i/X)$
		Y_1	Y_2	Y_3	Y_4	Y_5	
A_1	Têxtil	0.01	0.34	0.42	0.15	0.08	0.5718
A_2	Vestuário	0.12	0.53	0.25	0.06	0.04	0.3943
A_3	Outros	0.04	0.40	0.35	0.13	0.08	0.5246
A_4	Metal.Básica	0.01	0.48	0.30	0.07	0.05	0.4321
A_5	Metalurgia	0.02	0.46	0.36	0.10	0.06	0.5005
A_6	Alimentação	0.05	0.47	0.32	0.10	0.06	0.4776

Podemos concluir que a categoria A_1 possui a melhor renda; A_2 é a categoria que apresenta renda mais baixa e A_5 é a atividade que se aproxima mais da média de renda da população.

Utilizando o mesmo procedimento das tabelas 2.6, 2.7 e 2.8, podemos calcular as médias "ridits" para subgrupos experimentais "menores".

Através das figuras, podemos fazer uma análise mais detalhada, além daquelas já feitas. (Figuras 2.2 a 2.8)

- A Renda relativa à Idade (I) se comporta da mesma forma para todas as Atividades (A). Notamos, no entanto, que as intensidades das variações de atividade para atividade são diferentes. Isto demonstra que existe uma pequena interação entre as variáveis Atividade (A) e Idade (I).
- A renda da Faixa de Idade entre 18 e 29 anos (I_1) é sempre bem abaixo da média da renda da respectiva Atividade (A) a que pertence.
- A variável Escolaridade (E) é uma variável ordinal em que as médias de Escolaridade Baixa (E_1) e Média (E_2) são sempre pró

ximas uma da outra, exceto na Atividade (A_2) para faixa de idade entre 40 e 49 anos (I_3) e 50 e 65 anos (I_4).

Para complementar a análise, podemos montar a seguinte tabela 2.9 mostrando a diferença de médias entre as categorias E_3 e E_1 dentro da categoria Idade (I) e Atividade (A).

Tabela 2.9 - Diferenças de Médias "ridits" entre E_3 e E_1 nas categorias $I \times A$

Atividade	I d a d e			
	I_1	I_2	I_3	I_4
A_1	0.1749	0.3198	0.3115	0.3707
A_2	0.1805	0.4100	0.2963	0.3051
A_3	0.2275	0.2969	0.3286	0.3576
A_4	0.3659	0.5353	0.4550	0.3179
A_5	0.1922	0.3155	0.3860	0.3587
A_6	0.1769	0.3156	0.4284	0.3523

A tabela 2.9 mostra que a influência da escolaridade na renda cresce à medida que aumenta a Idade, às vezes diminuindo um pouco na última Faixa de 50 e 65 anos (I_4). A Atividade Metalúrgica (A_4) é a que apresenta maior influência desta variável. (Escolaridade)

Através de observações, podemos destacar o seguinte:

- i) Os efeitos das variáveis Escolaridade (E), Idade (I) e Atividade (A) na variável renda (Y) não é igual à simples soma dos efeitos individuais das variáveis, devido à existência de interação entre elas.
- ii) A categoria Escolaridade Alta (E_3) e a categoria Mais Jovem (I_1) são individualmente, as que se destacam das demais em

relação à variável Renda (Y).

Dependendo do tipo da pesquisa, podemos construir tabelas, as mais variadas, permitindo uma análise mais completa da pesquisa.

L E G E N D A

(De Fig.2.3 a Fig.2.6)

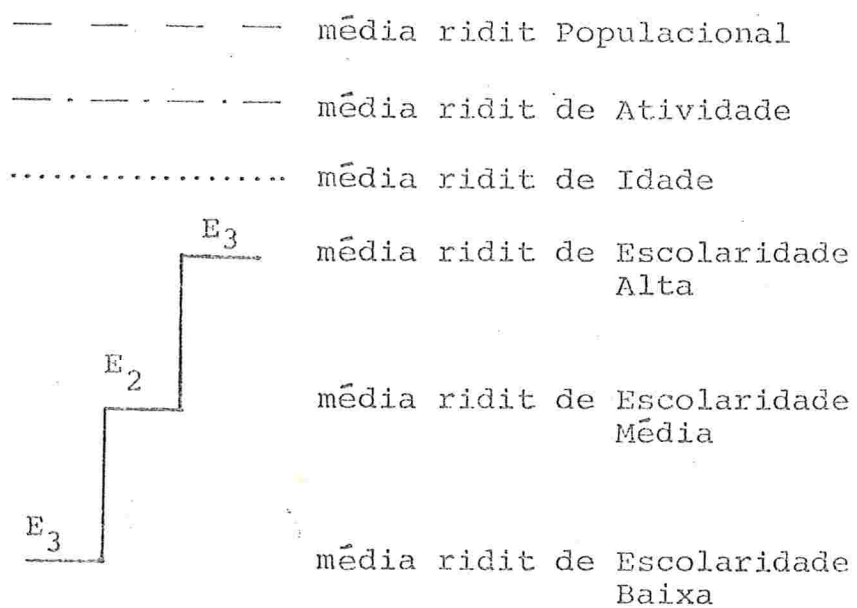


Fig. 2.3 - Médias Ridits de Atividade Têxtil

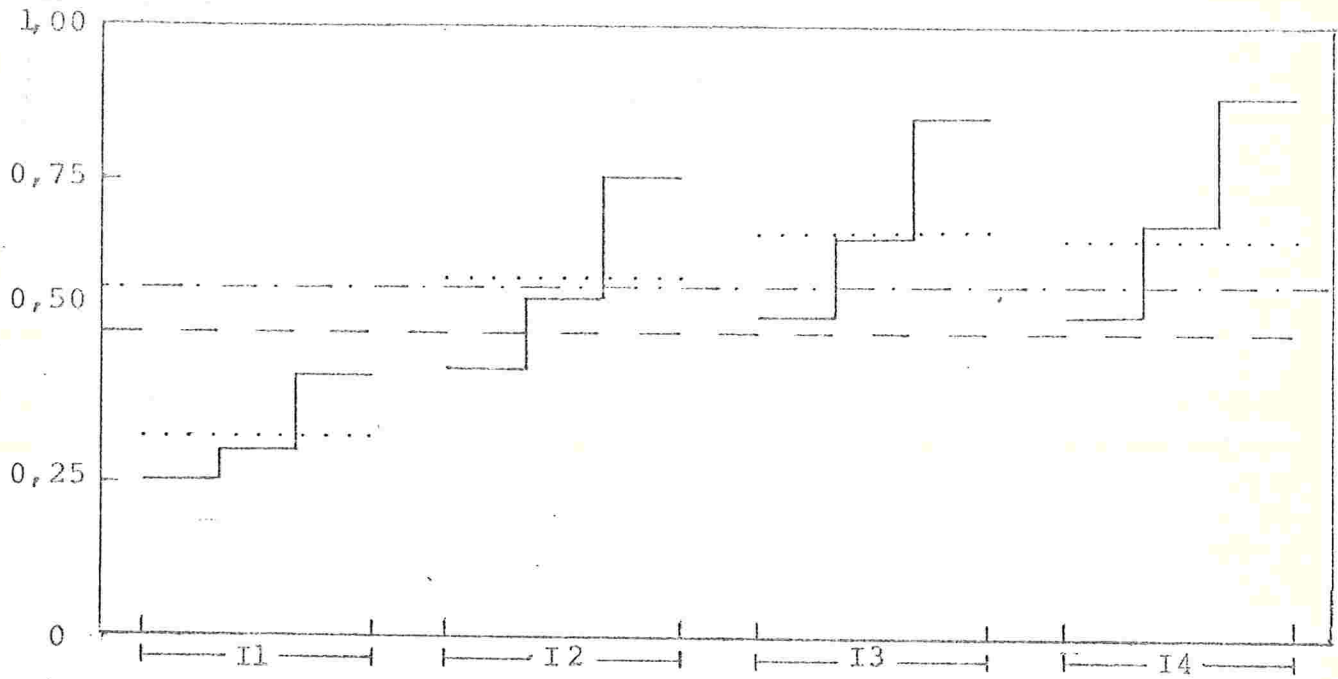


Fig. 2.4 - Médias Ridits de Atividade Vestuário

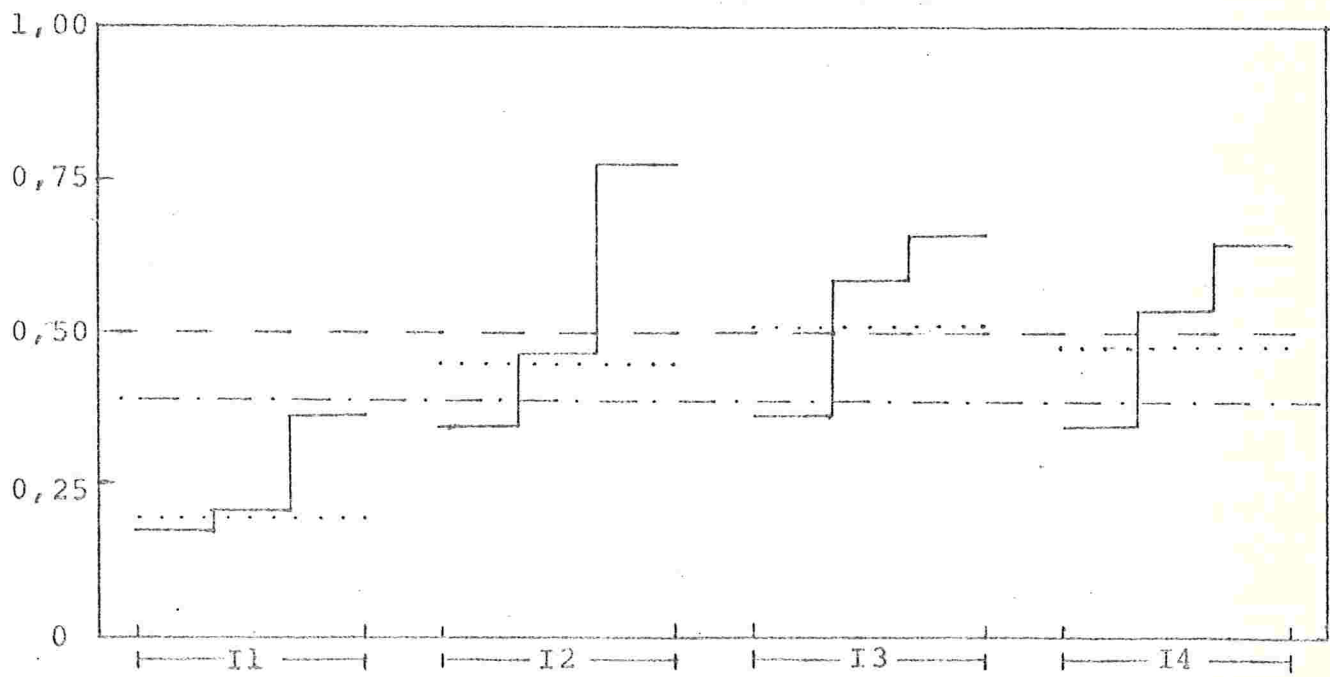


Fig. 2.5 - Médias Ridits de Atividade Outros

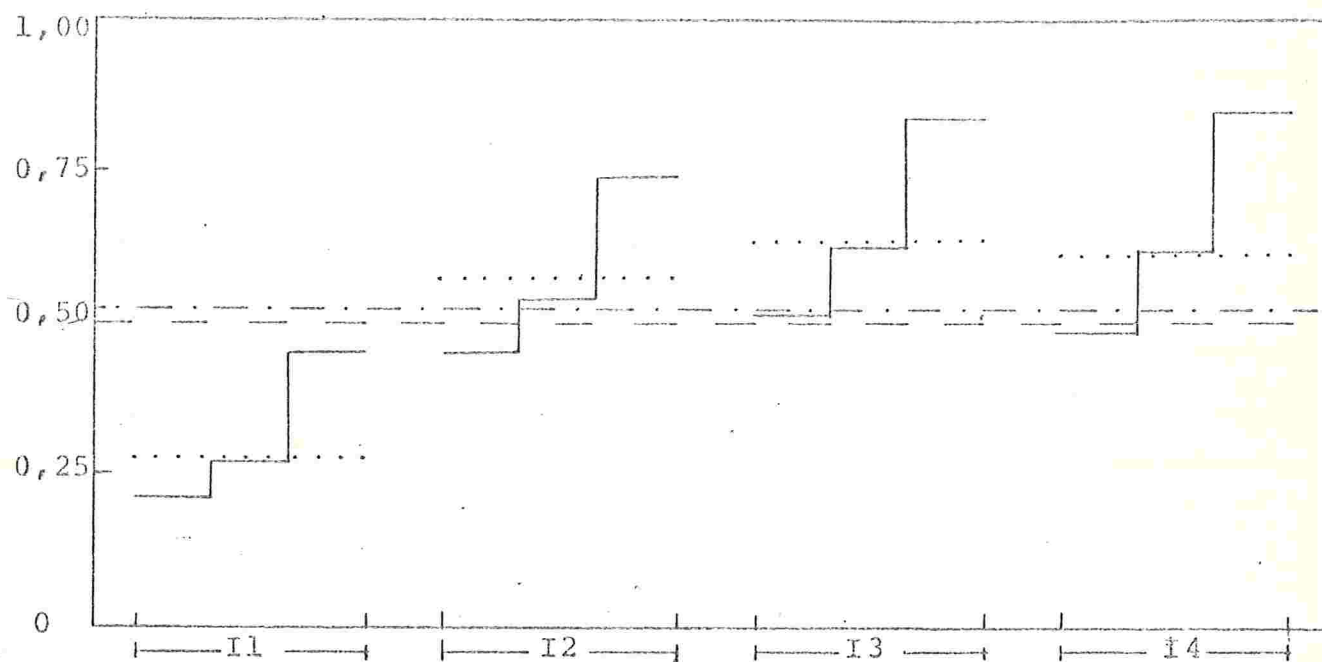


Fig. 2.6 - Médias Ridits de Atividade Metalurgia Básica

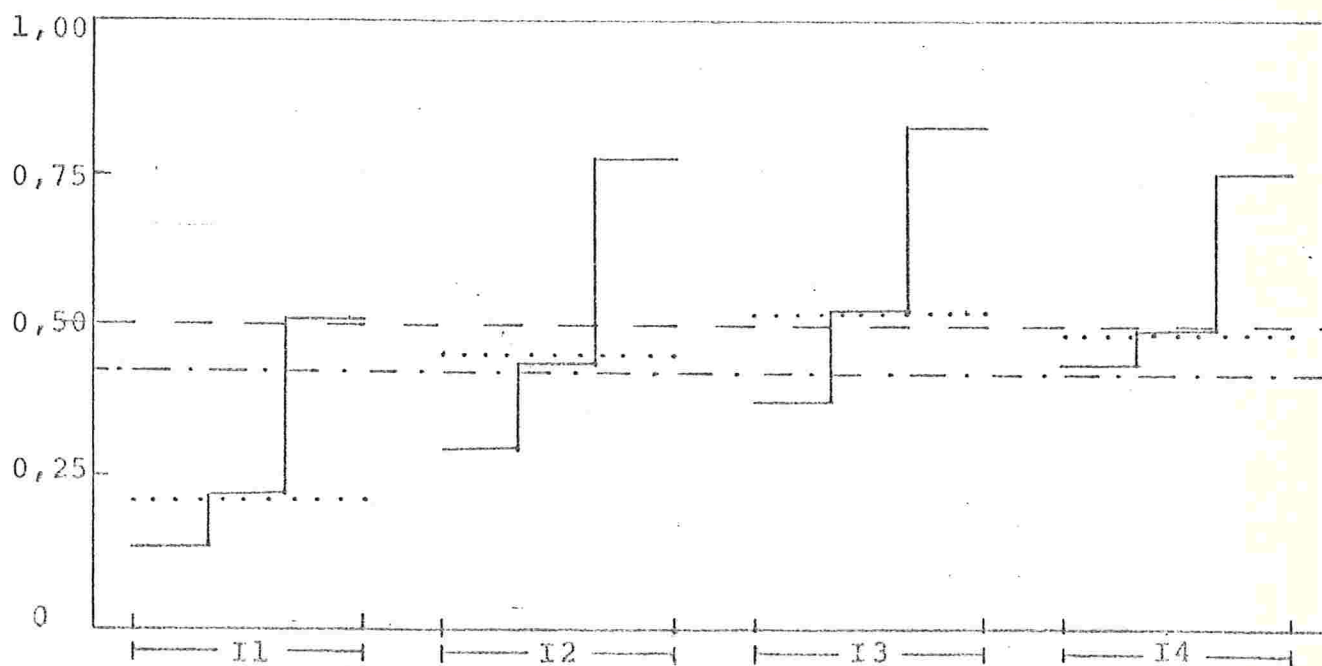


Fig. 2.7 - Médias Ridits de Atividade Metalurgia

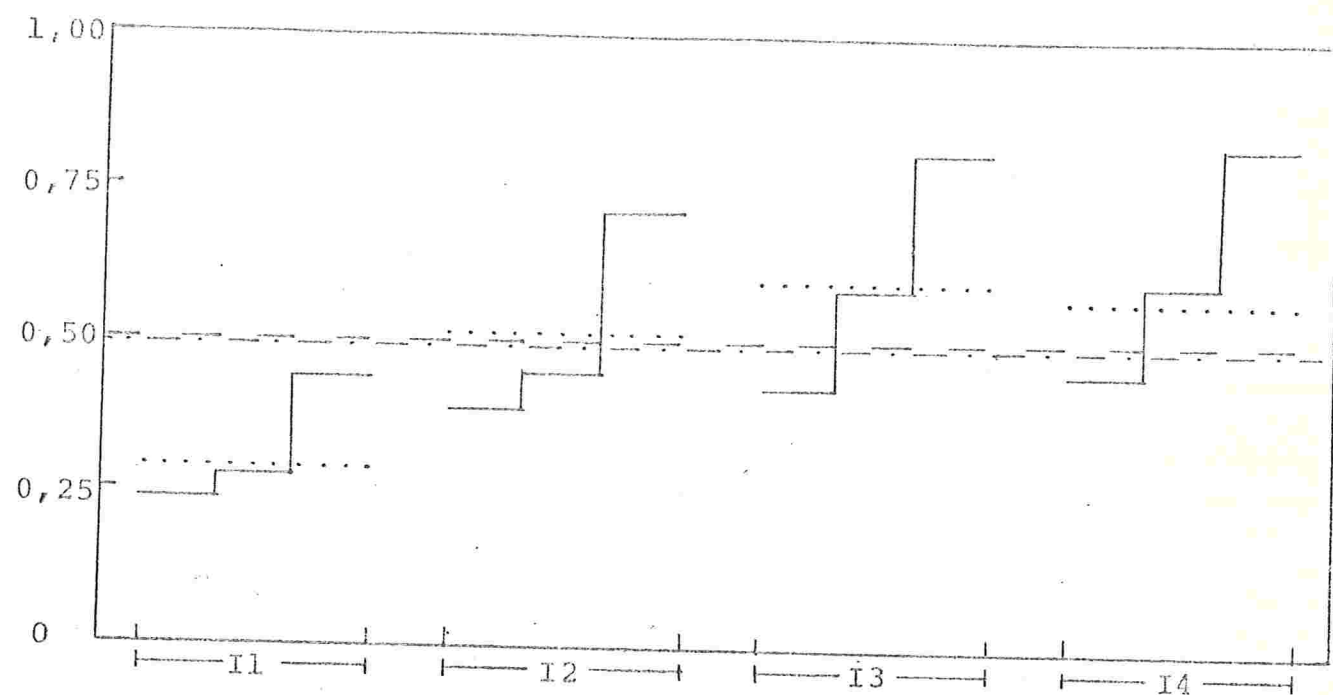
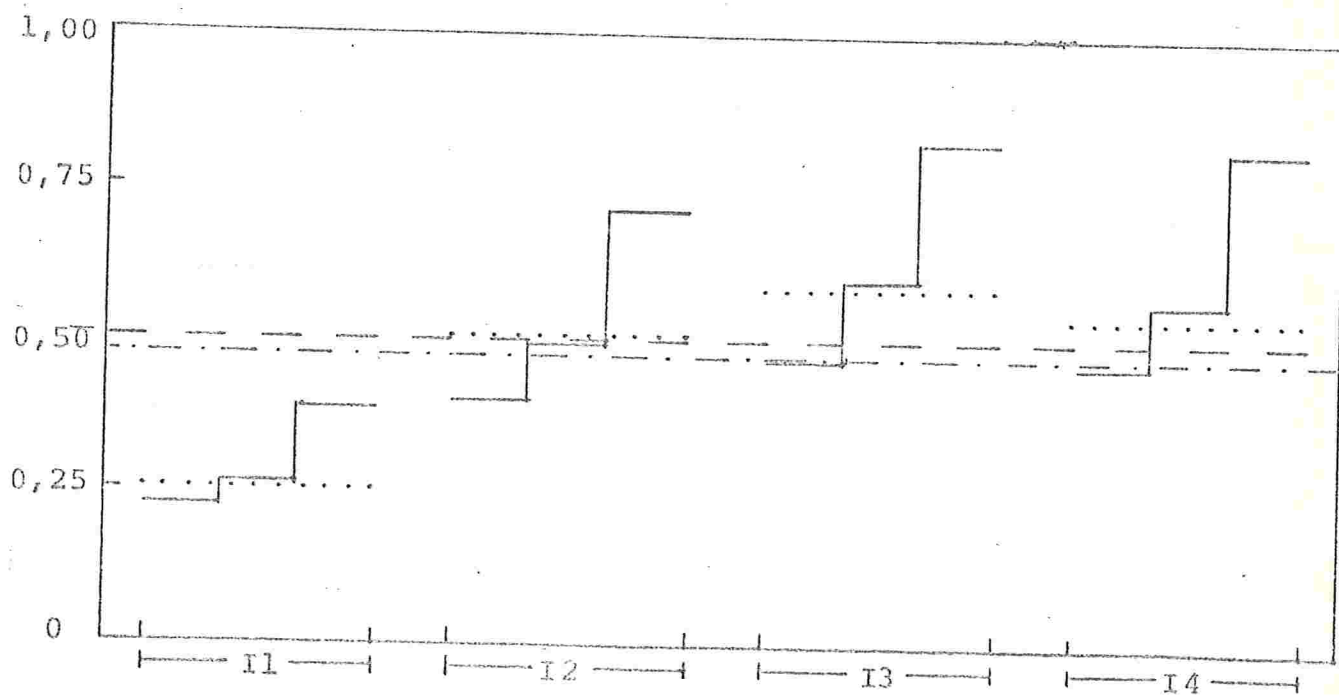


Fig. 2.8 - Médias Ridits de Atividade Alimentação



CAPÍTULO 3

ANÁLISE DE REGRESSÃO

3.1 - INTRODUÇÃO

Na maioria das pesquisas, lidamos com abundância de variáveis qualitativas que frequentemente dificultam a análise de dados. A grande maioria das técnicas estatísticas desenvolvidas está relacionada com dados quantitativos.

O que apresentamos neste capítulo é o aproveitamento de uma das técnicas mais conhecidas, a análise de regressão. Para a aplicação desta técnica efetuamos uma transformação nos dados qualitativos.

As variáveis explicativas x_j e a variável dependente y são tratadas como variáveis binárias 0 e 1.

Apresentamos, a seguir, dois métodos de análise de regressão bem como comparações entre eles:

- i) Análise de Regressão Simples (ARS).
- ii) Análise de Regressão Canônica (ARC).

3.2 - ANÁLISE DE REGRESSÃO SIMPLES E CANÔNICA

A equação de regressão simples (RS) de y em x_1, x_2, \dots, x_p pode ser escrita da seguinte forma:

$$Y_i = \sum_{j=1}^p \beta_j x_{ij} + e_{ij} \quad \begin{array}{l} i = 1, 2, \dots, n \\ j = 1, 2, \dots, p \end{array} \quad (3.1)$$

onde

β_j são parâmetros,

e_{ij} erro

ou na forma matricial

$$\underline{Y} = \underline{X} \underline{\beta} + \underline{E} \quad (3.2)$$

$n \times 1 \quad n \times p \quad p \times 1 \quad n \times 1$

Temos um vetor \underline{y} cujos elementos podem assumir valores conhecidos w_1, w_2, \dots, w_k correspondentes a K categorias. Na matriz \underline{X} as colunas são "dummy" e determinam as classes de variáveis explicativas às quais o indivíduo pertence. Os β 's são coeficientes de regressão e \underline{E} , o vetor de erros.

Consideremos w_i um elemento do vetor $\underline{W} (k \times 1)$ e $\underline{Z} (n \times k)$, uma matriz de incidência cujas linhas possuem exatamente um valor unitário e o resto, zeros. O valor 1 (um) indica a categoria em que um particular indivíduo pertence, portanto a equação (3.2) pode ser reescrita na forma:

$$\underline{Z} \underline{W} = \underline{X} \underline{\beta} + \underline{E} \quad (3.3)$$

Agora, consideremos w_1, w_2, \dots, w_k desconhecidos. O problema se resume em achar uma combinação linear $\underline{Z} \underline{W}$, das colunas de \underline{Z} , e uma combinação linear de $\underline{X} \underline{\beta}$, de tal forma que $\underline{X} \underline{\beta}$ seja um preditor "ótimo" de $\underline{Z} \underline{W}$. Portanto, estimamos $\underline{\beta}$ e \underline{W} maximizando R^2 (coeficiente de determinação).

Como em ARS e em ARC, a estimação de R^2 requer somente a determinação dos momentos da segunda ordem: $\underline{Y}^T \underline{Y}$, $\underline{Y}^T \underline{X}$ e $\underline{X}^T \underline{X}$. Esta técnica foi desenvolvida por Gifi (1980), usando o princípio de Mínimos Quadrados, e por Keller (1980), usando o princípio de Máxima Verossimilhança. Ambos chegaram ao mesmo resultado.

3.3 - COMPONENTES PRINCIPAIS E RELAÇÕES PRINCIPAIS

Consideremos n indivíduos sob observação, para cada um dos quais o vetor $\underline{y}(1 \times r)$ é medido. Vamos supor inicialmente que os dados são quantitativos.

O modelo proposto é

$$\begin{matrix} \underline{Y} \\ n \times r \end{matrix} = \begin{matrix} \underline{X} \\ n \times r \end{matrix} + \begin{matrix} \underline{U} \\ n \times r \end{matrix} \quad (3.4)$$

onde \underline{Y} é a matriz de observação de posto r , \underline{X} a matriz ajustada para os valores verdadeiros e \underline{U} é a matriz de erros com distribuição normal. Os elementos de \underline{X} são supostos não estocásticos e, por outro lado, os elementos de \underline{U} são estocásticos e independentes. Suponhamos que cada linha de \underline{U} tem esperança zero e matriz de covariância $\sigma^2 \underline{\Omega}(r \times r)$.

Estamos interessados nas relações lineares que possam existir entre as colunas de \underline{X} . Podemos, então, considerar duas relações:

a) Relações Principais (RP)

De acordo com esta especificação, existe uma matriz $\underline{B}(r \times p)$ ($p < r$, posto $(B) = p$) tal que

$$\underline{X} \underline{B} = \underline{0} \quad (3.5)$$

isto quer dizer que as classes de \underline{X} estão restritas a um subespaço de dimensão $(r-p)$.

b) Componentes Principais (CP)

De acordo com esta especificação, consideramos a estrutura na forma complementar e supomos a existência de uma matriz \underline{Q} ($n \times q$) e uma matriz \underline{A} ($r \times q$) ($q < r$, posto $(A) = q$), tal que

$$\underline{X} = \underline{Q}\underline{A}^T \quad (3.6)$$

isto é, as colunas são geradas por "expansão" de um espaço de dimensão menor.

Se as restrições em \underline{X} forem sempre $r = p+q$ então RP e CP são equivalentes.

3.4 - ESTIMACÃO DE MÁXIMA VEROSSIMILHANÇA

O que nos interessa particularmente é o modelo RP. Neste caso obteremos o Estimador de Máxima Verossimilhança (EMV) de \underline{B} .

Seja $\underline{\Omega}$ de posto s ($p < s < r$). Então podemos escrevê-la como:

$$\underline{\Omega} = \underline{F}\underline{D}\underline{F}^T \quad (3.7)$$

onde \underline{F} ($r \times s$) é uma matriz de autovetores de $\underline{\Omega}$, correspondente às raízes características (rc) não nulas. Estas rc são os elementos da matriz diagonal \underline{D} ($s \times s$).

Consideremos uma matriz $\underline{R} = [\underline{F}, \underline{G}]$ tal que

$$\underline{R}\underline{\Omega}\underline{R}^{-1} = [\underline{F}, \underline{G}] \underline{\Omega} \begin{bmatrix} \underline{F}^T \\ \underline{G}^T \end{bmatrix} = \begin{bmatrix} \underline{D} & \underline{O} \\ \underline{O} & \underline{O} \end{bmatrix} \quad (3.8)$$

onde \underline{G} ($r \times (r-s)$) é a matriz cujos elementos são autovetores correspondentes às rc nulas. Temos então

$$\underline{G}^T \underline{G} = \underline{I}_{r-s}, \quad \underline{F}^T \underline{G} = \underline{O} \quad \text{e} \quad \underline{F}^T \underline{F} = \underline{I}_s.$$

Como $\underline{Y}\underline{G} = \underline{X}\underline{G} + \underline{U}\underline{G}$ e $\text{cov}(\underline{U}\underline{G}) = \underline{0}$, verifica-se que $\underline{Y}\underline{G} = \underline{X}\underline{G}$ (com probabilidade 1).

Estimar \underline{B} pelo método de Máxima Verossimilhança significa minimizar

$$\text{tr} (\underline{Y}-\underline{X}) \underline{\Omega}^+ (\underline{Y}-\underline{X})^T \quad (3.9)$$

sujeita a $\underline{X}\underline{B} = \underline{0}$ e $\underline{Y}\underline{G} = \underline{X}\underline{G}$; $\underline{\Omega}^+$ é pseudo-inversa de $\underline{\Omega}$.

Sejam \underline{N}_1 ($p \times n$) e \underline{N}_2 ($(r-1) \times n$) as matrizes de multiplicadores de Lagrange. Assim temos:

$$L = \text{tr} (\underline{Y}-\underline{X}) \underline{\Omega}^+ (\underline{Y}-\underline{X})^T + 2\text{tr}(\underline{N}_1 \underline{X}\underline{B}) + 2\text{tr} \underline{N}_2 (\underline{Y}-\underline{X})\underline{G} \quad (3.10)$$

A condição de primeira ordem com relação a \underline{X} é dada por

$$\underline{X} \underline{\Omega}^+ = \underline{Y} \underline{\Omega}^+ - \underline{N}_1^T \underline{B}^T + \underline{N}_2^T \underline{G}^T \quad (3.11)$$

Sabemos que $\underline{\Omega}^+ = \underline{F} \underline{D}^{-1} \underline{F}^T$. Substituindo-o em (3.11) temos:

$$\underline{X} \underline{F} \underline{D}^{-1} \underline{F}^T = \underline{Y} \underline{F} \underline{D}^{-1} \underline{F}^T - \underline{N}_1^T \underline{B}^T + \underline{N}_2^T \underline{G}^T \quad (3.12)$$

Pós-multiplicando (3.12) por $\underline{F} \underline{D}$ resulta

$$\underline{X}\underline{F} = \underline{Y}\underline{F} - \underline{N}_1^T \underline{B}^T \underline{F} \underline{D} \quad (3.13)$$

Combinando este com $\underline{Y}\underline{G} = \underline{X}\underline{G}$

$$\underline{X}[\underline{F}, \underline{G}] = \underline{Y}[\underline{F}, \underline{G}] - [\underline{N}_1^T \underline{B}^T \underline{F} \underline{D}, \underline{0}] \quad (3.14)$$

ou

$$\underline{X}\underline{R} = \underline{Y}\underline{R} - [\underline{N}_1^T \underline{B}^T \underline{F} \underline{D}, \underline{0}] \quad (3.15)$$

Pós-multiplicando por \tilde{R}^{-1}

$$\begin{aligned}\tilde{X} &= \tilde{Y} - [\tilde{N}_1^T \tilde{B}^T \tilde{F} \tilde{D}, \tilde{O}] \begin{bmatrix} \tilde{F}^T \\ \tilde{G}^T \end{bmatrix} \\ &= \tilde{Y} - \tilde{N}_1^T \tilde{B}^T \tilde{\Omega}\end{aligned}\quad (3.16)$$

Assim:

$$\tilde{X}\tilde{B} = \tilde{Y}\tilde{B} - \tilde{N}_1^T \tilde{B}^T \tilde{\Omega}\tilde{B}\quad (3.17)$$

Como $\tilde{X}\tilde{B} = \tilde{O}$ segue que

$$\tilde{N}_1^T = \tilde{Y}\tilde{B} (\tilde{B}^T \tilde{\Omega}\tilde{B})^{-1}\quad (3.18)$$

e

$$\tilde{Y} - \tilde{X} = \tilde{Y}\tilde{B} (\tilde{B}^T \tilde{\Omega}\tilde{B})^{-1} \tilde{B}^T \tilde{Y}\quad (3.19)$$

Substituindo (3.19) em (3.9) ou (3.10) resulta

$$\text{tr } \tilde{Y}\tilde{B} (\tilde{B}^T \tilde{\Omega}\tilde{B}) \tilde{B}^T \tilde{Y}^T\quad (3.20)$$

como a expressão a ser minimizada. Diferenciando-a em relação a \tilde{B} resulta (Vetter, 1970):

$$(\tilde{I} - \tilde{\Omega}\tilde{B} (\tilde{B}^T \tilde{\Omega}\tilde{B})^{-1} \tilde{B}^T) \tilde{Y}\tilde{Y}^T \tilde{B} (\tilde{B}^T \tilde{\Omega}\tilde{B})^{-1} = \tilde{O}\quad (3.21)$$

ou

$$\tilde{Y}\tilde{Y}^T \tilde{B} = \tilde{\Omega}\tilde{B}\tilde{C}\quad (3.22)$$

onde

$$\tilde{C} = (\tilde{B}^T \tilde{\Omega}\tilde{B})^{-1} \tilde{B}^T \tilde{Y}^T \tilde{Y}\tilde{B}\quad (3.23)$$

Uma solução, com diagonal $\tilde{\Delta}$, é

$$\tilde{Y}^T \tilde{Y}\tilde{B}_{\tilde{O}} = \tilde{\Omega}\tilde{B}_{\tilde{O}} \tilde{\Delta}\quad (3.24)$$

a qual significa que B_0 é o conjunto de autovetores de $\tilde{Y}^T \tilde{Y}$ na métrica de $\tilde{\Omega}$.

Como

$$\text{tr } \tilde{Y} B_0 (B_0^T \tilde{\Omega} B_0)^{-1} B_0^T \tilde{Y}^T = \text{tr } \tilde{\Lambda} \quad (3.25)$$

temos que considerar autovetores correspondentes à menor r de $\tilde{Y}^T \tilde{Y}$ na métrica de $\tilde{\Omega}$.

3.5 - ANÁLISE MULTIVARIADA

A regressão canônica é derivada das técnicas de Correlação Canônica e Mínimos Quadrados Ordinários. Para melhor compreensão, a seguir apresentaremos de forma sucinta estas três técnicas.

3.5.1 - Mínimos Quadrados Ordinários (MQO)

Mínimos Quadrados Ordinários derivam do modelo RP. Consideramos $p=1$ e $\tilde{\Omega} = \sigma^2 (e_i \cdot e_i^T)$, sendo e_i i -ésimo vetor unitário.

Sem perda de generalidade, consideramos $i=1$ e normalizamos $\tilde{B} = \underline{\beta} = (-1, \beta_1)^T$, e estabelecemos $\tilde{Y} = [y_1, Y_1]$, $\tilde{X} = [x_1, X_1]$ e $\tilde{U} = (u_1, U_1) = (u_1, 0)$ onde

$$\tilde{Y} = [y_1, y_2, \dots, y_r] \quad , \quad \tilde{X} = [x_1, x_2, \dots, x_r] \quad ,$$

$$\tilde{Y}_1 = [y_2, y_3, \dots, y_r] \quad \text{e} \quad \tilde{X}_1 = [x_2, x_3, \dots, x_r]$$

De $\tilde{X} \tilde{B} = \tilde{0}$ temos:

$$\tilde{X} \tilde{B} = -x_1 + \beta_1 X_1 = 0 \quad (3.26)$$

ou

$$\underline{x}_1 = \underline{X}_1 \underline{\beta}$$

Logo,

$$\underline{y}_1 = \underline{x}_1 + \underline{u}_1 = \underline{X}_1 \underline{\beta}_1 + \underline{u}_1 \quad (3.27)$$

Tomando $i = 1, 2, 3 \dots r$ sucessivamente, definimos MQO para cada coluna de \underline{X} . Notemos que a escolha de elementos não nulos em $\underline{\Omega}$ determina quais variáveis são exógenas e quais variáveis são endógenas.

3.5.2 - Correlação Canônica (CC)

Suponhamos que haja dois conjuntos de variáveis:

$$Y = \begin{bmatrix} Y_1 & Y_2 \\ r & r_1 \quad r_2 \end{bmatrix} \quad (3.28)$$

Os erros dentro de cada conjunto são correlacionados mas, entre os conjuntos, são não correlacionados. Além disso, as variâncias dos erros são supostamente proporcionais às variâncias amostrais, para que obtenhamos

$$\underline{\Omega} = \sigma^2 \begin{bmatrix} \underline{Y}_1^T \underline{Y}_1 & 0 \\ 0 & \underline{Y}_2^T \underline{Y}_2 \end{bmatrix} \quad (3.29)$$

Resolvendo o problema correspondente a RP para $\underline{\Omega}$ especificada em (3.29), podemos mostrar que o método é equivalente ao método de maximização de Correlação Canônica (CC)* entre dois conjuntos. Em contraste com RP, podemos aplicar em (CC)* duas, em vez de uma, normalizações nas estimativas dos coeficientes: uma correspondente ao subvetor \underline{y}_1 e outra, para o subvetor \underline{y}_2 . Esta é consequência de uma pequena modificação na definição.

3.5.3 - Regressão Canônica (RC)

A RC é intermediária entre MQO e CC. Há dois conjuntos de variáveis, como em CC, mas um deles é considerado não estocástico como em MQO. A RC é definida como um modelo RP na qual

$$\underline{\Omega} = \sigma^2 \begin{bmatrix} Y_1^T Y_1 & 0 \\ 0 & 0 \end{bmatrix} \quad (3.30)$$

Portanto, RC é um modelo de regressão com $r_1 > 1$ variáveis endôgenas a que são atribuídas escalas de valores de tal forma que elas mostrem máxima correlação entre as variáveis exôgenas. Podemos mostrar que a solução RC é igual à solução CC para uma normalização diferente nos coeficientes de subvetor correspondente a Y_2 .

3.6 - DADOS CATEGÓRICOS

3.6.1 - Transformação de Dados Qualitativos

Seja Z_i matriz de incidência, correspondente a i -ésima variável qualitativa, com m_i categorias. Seja w_i o vetor de escala de valores de ordem m_i da i -ésima variável qualitativa. Logo o vetor y_i ($n \times 1$):

$$y_i = Z_i w_i \quad i = 1, 2, \dots, r \quad (3.32)$$

contém escores quantitativos de n indivíduos. Vamos supor que temos r variáveis qualitativas. Então,

$$\underline{Y} = [y_1, y_2, \dots, y_r] \quad (3.32)$$

$$\underline{W} = \begin{bmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \dots & \\ 0 & & & w_r \end{bmatrix} \quad (3.33)$$

onde

$$W_i = \begin{bmatrix} W_{i1} \\ W_{i2} \\ \vdots \\ W_{im_i} \end{bmatrix} \quad (3.34)$$

Logo,

$$\underset{n \times r}{Y} = \begin{bmatrix} Z_1 W_1 & Z_2 W_2 & \dots & Z_r W_r \end{bmatrix} = ZW$$

Assim transformamos dados qualitativos Z em dados quantitativos Y , e podemos utilizar o método desenvolvido em (3.5). A única diferença é a presença de uma matriz de escala de valores W . Assim temos:

$$ZW = X + U \quad (3.37)$$

com

$$XB = 0$$

3.6.2 - Regressão Canônica com Dados Categorizados

Na seção anterior, mostramos que a solução do modelo RP é a solução β estimada pela minimização de $(\beta^T Y^T Y \beta) / (\beta^T \Omega \beta)$ para $p=1$. Assim para $Y = ZW$

$$\min \left\{ \frac{\beta^T W^T Z^T Z W \beta}{\beta^T \Omega \beta} \right\} \quad (3.38)$$

onde Ω e W são desconhecidas. No caso de RC Ω é especificado como

$$\Omega = \sigma^2 W^T \begin{bmatrix} Z_1^T & Z_1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix} W \quad (3.39)$$

$$e \quad \underline{\underline{\beta}}^T \underline{\underline{\Omega}} \underline{\underline{\beta}} = \sigma^2 \underline{\underline{\beta}}^T \underline{\underline{W}}^T \begin{bmatrix} \underline{\underline{z}}_1^T & \underline{\underline{z}}_1 & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{0}} \end{bmatrix} \underline{\underline{W}} \underline{\underline{B}} \quad (3.40)$$

Finalmente substituindo (3.40) em (3.38) temos uma nova expressão

$$\min \left\{ \frac{\underline{\underline{\gamma}}^T \begin{bmatrix} \underline{\underline{z}}_1^T \underline{\underline{z}}_1 & \underline{\underline{z}}_2^T \underline{\underline{z}}_1 \\ \underline{\underline{z}}_1^T \underline{\underline{z}}_2 & \underline{\underline{z}}_2^T \underline{\underline{z}}_2 \end{bmatrix} \underline{\underline{\gamma}}}{\underline{\underline{\gamma}}^T \begin{bmatrix} \underline{\underline{z}}_1^T \underline{\underline{z}}_1 & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{0}} \end{bmatrix} \underline{\underline{\gamma}}} \right\} \quad (3.41)$$

onde

$$\underline{\underline{\gamma}} = \underline{\underline{W}} \underline{\underline{B}}$$

A solução da expressão (3.41) consiste na determinação do autovetor correspondente à menor raiz de $\underline{\underline{\gamma}}^T \underline{\underline{\gamma}}$ na métrica de $\underline{\underline{\Omega}}$.

Este modelo pode ser interpretado como uma relação que explica um conjunto de variáveis qualitativas dependentes a partir do conjunto de variáveis qualitativas independentes.

3.7 - APLICAÇÃO

Nesta seção será feita a comparação de duas técnicas, através dos parâmetros estimados e dos coeficientes de determinação R^2 , utilizando os dados descritos no capítulo 1.

Para que seja possível estimar $\underline{\underline{\beta}}$ sem ambigüidade, é necessário que as colunas de $\underline{\underline{X}}$ sejam linearmente independentes. Devido a este motivo, deixamos fora uma categoria de cada variável (a variável "produto" como $E \times A$ é considerada como variável simples) e

"intercept" é inserido na equação de regressão. Assim as primeiras colunas de X consistem de "uns" (1).

Como variáveis exploratórias, as seguintes combinações de E , I e A são consideradas para análise:

- i) variáveis: E, I ou A (def. no capítulo 1).
- ii) combinações das variáveis sem interação: $E+I$, $E+A$, $A+I$ ou $E+I+A$.
- iii) combinação das variáveis com interação: $E \times I$, $E \times A$, $A \times I$ ou $A \times E \times I$.

Quando fazemos uma regressão na variável "produto" $E \times A \times I$, em vez de usar todas as 72 combinações de categorias como "dummy", podemos usar todas as 72 combinações constituídas pelos seguintes termos:

- i) marginais univariadas de E , A e I ($2+5+3=10$ termos),
- ii) marginais bivariadas E , A e I ($10+6+15=31$ termos);
- iii) interação de segunda ordem ($2 \times 5 \times 3=30$ interações),
- iv) mais o "intercept" completam as 72 combinações.

Apresentamos, primeiramente, a tabela de R^2 e os graus de liberdade de regressão de algumas combinações de parâmetros de algumas variáveis utilizando ARS e ARC.

Os graus de liberdade (gl) estão entre parênteses.

- a) Para ARS (p-1) gl
- b) Para ARS $[(K-2)+(p-1)] = (K+p-3)$ gl.

Tabela 3.1 - Coeficientes de Determinação e gl com Regressão Simples e Canônica de Y em E , A e I .

Sem Interação			Com Interação		
Variáveis Independentes	Regressão Simples	Regressão Canônica	Variáveis Independentes	Regressão Simples	Regressão Canônica
A	0,014 (5)	0,019 (8)	-	-	-
E	0,131 (2)	0,136 (5)	-	-	-
I	0,204 (4)	0,255 (6)	-	-	-
$A+E$	0,140 (7)	0,144 (10)	$A \times E$	0,141 (17)	0,144 (20)
$A+I$	0,216 (8)	0,270 (11)	$A \times I$	0,219 (23)	0,274 (26)
$E+I$	0,336 (5)	0,355 (8)	$E \times I$	0,345 (11)	0,358 (14)
$A+E+I$	0,344 (10)	0,364 (13)	$A \times E \times I$	0,355 (71)	0,371 (74)

Analisando a tabela acima, podemos notar que a variável Idade (I) é a que possui maior poder de explicação, seguida da variável Escolaridade (E) e, por último, a variável Atividade (A) que tem pouco poder de explicação.

Combinando as duas variáveis simples, o melhor par são as variáveis $E \times I$.

As interações entre as variáveis não são significativas, portanto, a sua contribuição na renda é mínima.

Podemos destacar que R^2 é sempre melhor para Regressão Canônica (RC) do que Regressão Simples (RS).

Para dar continuidade às comparações entre as duas técnicas, vamos padronizar o vetor de escalas $\underline{w} = (w_1, w_2, \dots, w_k)^T$. Temos então:

$$E(w) = \sum_{i=1}^k w_i p_i \quad (3.42)$$

$$\sigma^2(w) = \sum_{i=1}^k w_i^2 p_i - \left(\sum_{i=1}^k w_i p_i \right)^2 \quad (3.43)$$

onde

$$p_i = \frac{N(X_i)}{N(X)} \quad (\text{cfm notação do cap.2})$$

$$i = 1, 2, \dots, k$$

A escala de valores padronizados será denotada como

$$w_i^p = \frac{w_i - E(w)}{\sigma(w)} \quad (3.44)$$

Sendo $\underline{w} = (1, 2, 3, 4, 5)^T$ para RS, temos

$$\underline{w}^p = (-1,825; -0,764; 0,296; 1,357; 2,418)^T$$

Assim podemos montar uma tabela de escalas de renda correspondente aos principais modelos.

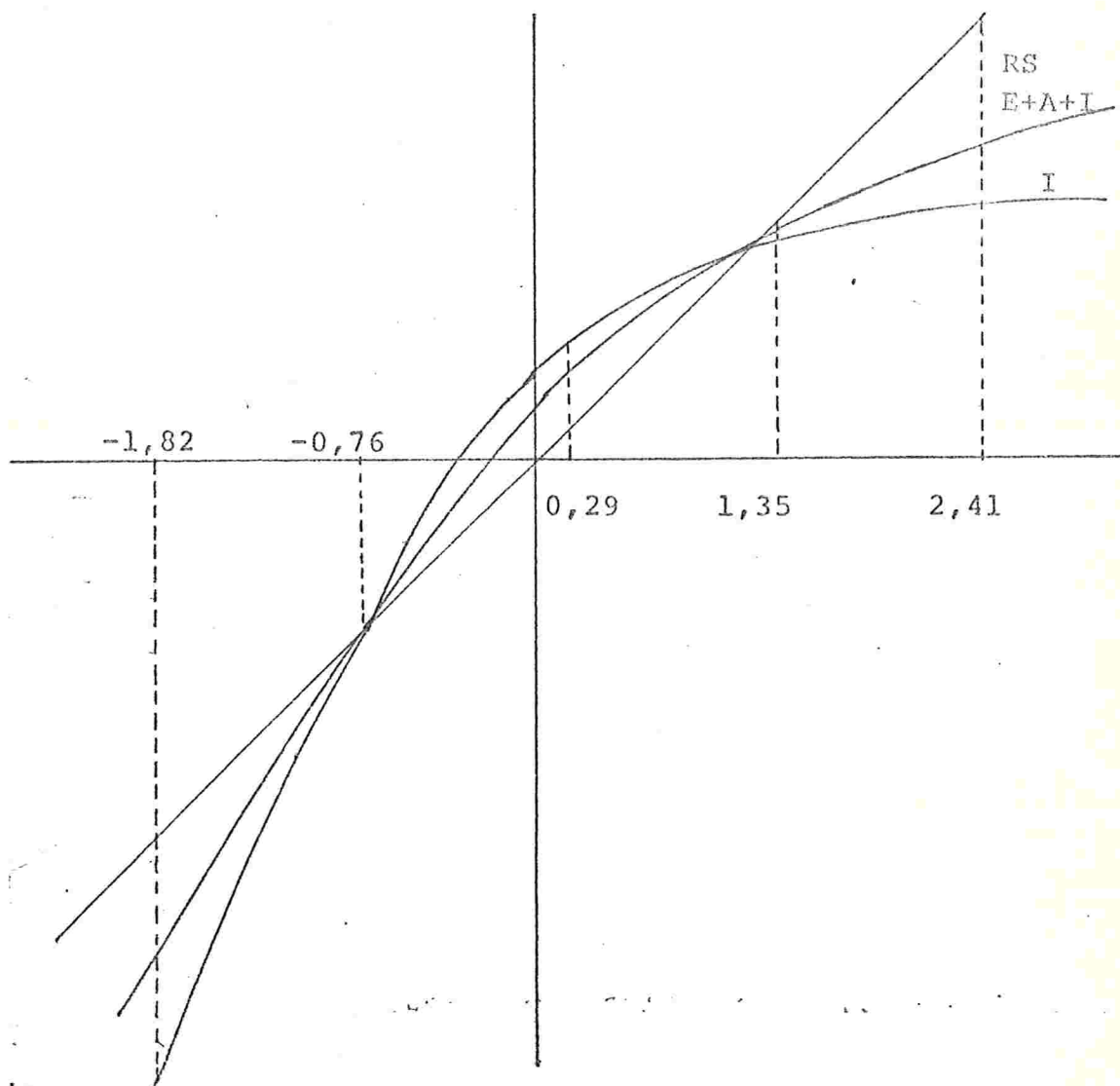
Tabela 3.2 - Escala de Renda

Classe de Renda	Regressão Simples	R e g r e s s ã o C a n ô n i c a				
		Sem Interação			Com Interação	
		A+E+I	E+I	I	A×E×I	E×I
Y ₁	-1,825	-2,440	-2,413	-2,951	-2,371	-2,305
Y ₂	-0,764	-0,797	-0,800	-0,756	-0,801	-0,809
Y ₃	0,296	0,555	0,555	0,736	0,537	0,533
Y ₄	1,357	1,287	1,280	1,049	1,304	1,302
Y ₅	2,418	1,744	1,765	1,208	1,797	1,835

Para ARC, as escalas correspondentes são todas parecidas com o modelo A+E+I com exceção do modelo I. Podemos fazer o gráfico dos três modelos: regressão simples, regressão canônica mo

delo I e modelo $A+E+I$.

Fig. 3.1 - Comparação de Escalas de Renda para RS e RC



Existe uma grande diferença nas escalas de valores utilizadas na RC e RS.

Na escala de valores de RC, a diferença entre w_1 e w_2 é muito grande. No entanto à medida que a faixa de renda aumenta diminuem as diferenças sucessivas.

Para aumentar R^2 , principalmente na RS, talvez deva ser tentada uma partição mais fina e adequada na faixa de renda.

Apresentamos, agora, a tabela de coeficientes de re-

gressão de A , E e I , sem interação e "intercept".

Os coeficientes de regressão apresentados são coeficientes "reduzidos", os quais são determinados de tal maneira que, para cada variável, a média de escore para todos os indivíduos seja igual a zero.

Tabela 3.3 - Coeficientes de Regressão Reduzidos para Regressão sem "intercept".

Variável Independente	Categoria	Reg. Simples $A+E+I$	Regressão Canônica		
			$A+E+I$	$E+I$	I
Atividade (A)	A_1	0,143	0,176	-	-
	A_2	-0,174	-0,243	-	-
	A_3	0,092	0,081	-	-
	A_4	-0,217	-0,258	-	-
	A_5	-0,049	-0,031	-	-
	A_6	-0,007	-0,021	-	-
Escolaridade (E)	E_1	-0,423	-0,425	-0,437	-
	E_2	-0,061	-0,040	-0,040	-
	E_3	0,819	0,726	0,741	-
Idade (I)	I_1	-0,746	-0,834	-0,833	-0,873
	I_2	-0,004	0,057	0,055	0,158
	I_3	0,374	0,394	0,395	0,399
	I_4	0,386	0,384	0,386	0,295

Não há diferença significativa nos parâmetros estimados através de vários modelos.

O coeficiente positivo pode ser interpretado como in

fluência positiva da categoria correspondente na renda, quando comparada com a média populacional. Assim como o coeficiente negativa, pode ser interpretado como influência negativa da categoria correspondente na renda.

- i) A categoria Escolaridade Alta (E_3) está bem distante das categorias Escolaridade Média (E_2) e Baixa (E_1) com influência positiva na renda.
- ii) Por outro lado, a categoria Idade Mais Jovem (I_1) está muito deslocada em relação às demais categorias com contribuição negativa na renda.
- iii) A variável Atividade é uma variável nominal, pois não há uma ordenação natural. A atividade Têxtil (A_1) é o que apresenta a melhor renda, enquanto que a que apresenta a mais baixa renda é a atividade Metalurgia Básica (A_4).

CAPÍTULO 4

ANÁLISE DE CORRESPONDÊNCIA

4.1 - INTRODUÇÃO

A Análise de Correspondência (AC) é uma técnica de visualização de tabelas de dupla entrada. Isto ocorre pela atribuição de uma escala de valores para cada uma das linhas e colunas da tabela.

Utilizando AC, não só conseguimos descrever o grau de associação entre as linhas e colunas, mas também o tipo de associação.

Esta técnica permite separar uma tabela em fatores, cada um destes fatores representando um aspecto do tipo de associação. Além disso, os fatores podem ser representados graficamente e os fatos, que estão ocultos na tabela, tornam-se claramente visíveis e de fácil interpretação.

O objetivo geral desta técnica é descrever a estrutura da tabela de uma forma exploratória.

Uma descrição minuciosa desta técnica pode ser encontrada nos trabalhos de Benzecri (1976) , Lebart (1977) e Sikkell (1980).

4.2 - DEFINIÇÕES E INTERPRETAÇÕES

Seja A uma tabela de freqüência de dupla entrada definida por:

$$A = \{a_{ij} | a_{ij} \geq 0, i=1,2,3,\dots,m; j=1,2,3,\dots,n\} \quad (4.1)$$

Sem perda de generalidade, suponhamos que $m \leq n$.

Sejam I_1, I_2, \dots, I_m as linhas e J_1, J_2, \dots, J_m as colunas de A . Desta forma a_{ij} é o número de elementos da linha I_i e da coluna J_j . Os totais marginais podem ser escritos da seguinte forma:

$$a_{i+} = \sum_{j=1}^n a_{ij} \quad (4.2)$$

$$a_{+j} = \sum_{i=1}^m a_{ij} \quad (4.3)$$

Suponhamos que para todo i , $a_{i+} > 0$ e para todo j , $a_{+j} > 0$ e que o total geral é dado por:

$$a_{++} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} \quad (4.4)$$

Como estamos interessados na freqüência relativa, definimos a matriz F com elementos $f_{ij} = a_{ij}/a_{++}$. Desta forma f_{ij} pode ser interpretado como a estimativa de probabilidade de um objeto cair nas categorias I_i e J_j simultaneamente. As distribuições marginais são:

$$f_{i+} = \sum_{j=1}^n f_{ij} \quad (4.5)$$

$$f_{+j} = \sum_{i=1}^m f_{ij} \quad (4.6)$$

e, naturalmente, vale a relação

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = 1 \quad (4.7)$$

Consideremos as matrizes diagonais de probabilidades marginais

$$\tilde{R} = \begin{bmatrix} f_{1+} & & & 0 \\ & f_{2+} & & \\ & & \ddots & \\ & & & f_{m+} \\ 0 & & & & 0 \end{bmatrix} \quad (4.8)$$

$$\tilde{S} = \begin{bmatrix} f_{+1} & & & 0 \\ & f_{+2} & & \\ & & \ddots & \\ & & & f_{+n} \\ 0 & & & & 0 \end{bmatrix} \quad (4.9)$$

Consideremos, agora, a probabilidade de um objeto, caindo na categoria I_i , cair na categoria J_j . Isto é, f_{ij}/f_{i+} ou na forma matricial $(\tilde{R}^{-1} \tilde{F})_{ij}$.

Para formalizar de forma adequada a interpretação que queremos mostrar, consideremos a seguinte matriz:

$$\tilde{M} = \begin{array}{c} I_1 \\ I_2 \\ I_3 \\ \vdots \\ I_m \\ J_1 \\ J_2 \\ \vdots \\ J_n \end{array} \begin{array}{c} I_1 \quad I_2 \quad \dots \quad I_m \quad J_1 \quad J_2 \quad \dots \quad J_n \\ \left[\begin{array}{cc|cc} & & & \\ & & & \\ & 0 & & \tilde{R}^{-1} \tilde{F} \\ \hline & & & \\ (\tilde{F} \tilde{S}^{-1})^T & & & 0 \end{array} \right] \end{array} \quad (4.10)$$

onde $(\tilde{F} \tilde{S}^{-1})^T$ é a transposta de $\tilde{F} \tilde{S}^{-1}$,

Analisando a matriz \tilde{M} podemos notar: a matriz é quadrada e a soma de qualquer linha é igual à unidade. Está claro que podemos considerá-la como uma matriz de Markov.

As categorias $I_1, I_2, \dots, I_m, J_1, J_2, \dots, J_n$ podem ser consideradas como possíveis condições da cadeia de Markov. Esta cadeia de Markov tem período maior ou igual a 2 desde que J_1, J_2, \dots, J_n podem ser alcançadas em somente um passo de I_1, I_2, \dots, I_m e vice-versa. Temos

$$M^2 = \begin{bmatrix} \tilde{R}^{-1} \tilde{F} (\tilde{F} \tilde{S}^{-1})^T & 0 \\ 0 & (\tilde{R} \tilde{S}^{-1})^T \tilde{R}^{-1} \tilde{F} \end{bmatrix} \quad (4.11)$$

matriz de Markov redutível, onde as submatrizes $\tilde{R}^{-1} \tilde{F} (\tilde{F} \tilde{S}^{-1})^T$ e $(\tilde{F} \tilde{S}^{-1})^T \tilde{R}^{-1} \tilde{F}$ também são matrizes de Markov. Ou em outras palavras: i, j -ésimo elemento da matriz $\tilde{R}^{-1} \tilde{F} (\tilde{F} \tilde{S}^{-1})^T$ pode ser considerado como probabilidade de alcançar I_i , a partir de I_j em dois passos.

4.3 - MAXIMIZAÇÃO DO COEFICIENTE DE CORRELAÇÃO

Agora, estamos à procura de um par de variáveis aleatórias que tenha distribuição bivariada $\{f_{ij} | i=1, \dots, m; j=1, 2, \dots, n\}$, que tenha a máxima correlação.

Podemos considerar uma distribuição bivariada I e J , onde os valores de I são as linhas I_1, I_2, \dots, I_m e os valores de J são as colunas J_1, J_2, \dots, J_n .

Definimos as variáveis aleatórias G e H com: $G = g_r$ se I assume o valor I_r e $H = h_s$, se J assume o valor J_s . Desta forma podemos interpretar g_r como escala de valores atribuída para a categoria I_r ($r=1, 2, \dots, m$) e h_s , uma escala de valores atribuída para

categoria J_s ($s=1, 2, \dots, n$), respectivamente. Estes valores são denotados na forma vetorial por \underline{g} e \underline{h} .

O objetivo da AC é o seguinte problema: Maximizar

$$\lambda(G, H) = \frac{E(GH) - E(G)E(H)}{\sigma(H) \cdot \sigma(G)} \quad (4.12)$$

onde "E" significa esperança matemática e " σ^2 " significa variância. O coeficiente de correlação λ é invariante mesmo sob multiplicação e/ou adição de uma constante nas variáveis (G e H); portanto, este problema é equivalente à maximização de

$$\psi(\underline{g}, \underline{h}) = \underline{g} \cdot \underline{F} \cdot \underline{h} \quad (4.13)$$

com as restrições

$$\sum_i f_{i+} g_i = \sum_j f_{+j} h_j = 0, \quad (4.14)$$

$$\sum_i f_{i+} g_i^2 = \sum_j f_{+j} h_j^2 = 1. \quad (4.15)$$

Quando exigimos que as esperanças G e H sejam iguais a zero, λ pode ser interpretada como coeficiente de correlação entre \underline{g} e \underline{h} .

A maximização da função ψ conduz ao problema de extração de raízes características. Existem $m-1$ pares de vetores localmente ótimos $(\underline{g}^{(2)}, \underline{h}^{(2)})$, $(\underline{g}^{(3)}, \underline{h}^{(3)})$, ..., $(\underline{g}^{(m)}, \underline{h}^{(m)})$. Um par $(\underline{g}^{(r)}, \underline{h}^{(r)})$ corresponde a um par de variáveis aleatórias G_r e H_r que têm coeficiente de correlação λ_r .

Assumimos que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. Notamos que a maior raiz característica é $\lambda_1 = 1$ cujos vetores correspondentes são iguais a $\underline{g}^{(1)} = (1, 1, \dots, 1)^T$ e $\underline{h}^{(1)} = (1, 1, \dots, 1)^T$ e portanto não satisfazem as condições (4.14). Em conseqüência, para esse caso, o coeficiente de correlação não é definido.

Para obtermos a solução desejada, escolhemos uma forma análoga à análise de Componentes Principais (Sikkel, 1980).

Vamos definir

$$g_i^* = g_i \sqrt{f_{i+}} \quad , \quad (4.16)$$

$$h_j^* = h_j \sqrt{f_{+j}} \quad , \quad (4.17)$$

$$f_{ij}^* = \frac{f_{ij}}{\sqrt{f_{i+} f_{+j}}} \quad , \quad (4.18)$$

Desta forma, a matriz de elementos $\{f_{ij}^*\}$ é denotada como

$$\tilde{F}^* = \tilde{R}^{-1/2} \tilde{F} \tilde{S}^{-1/2} \quad (4.19)$$

Agora o problema de AC pode ser reformulado:

Maximizar

$$\psi^* = \sum_i \sum_j f_{ij}^* g_i^* h_j^* \quad (4.20)$$

com as restrições

$$\sum_i g_i^* = 1 \quad , \quad (4.21)$$

$$\sum_j h_j^* = 1 \quad . \quad (4.22)$$

Consideremos a decomposição singular de \tilde{F}^*

$$\tilde{F}^* = \tilde{G}^* \Lambda \tilde{H}^* \quad (4.23)$$

onde

$$\tilde{G}^* = [g^{*(1)}, g^{*(2)}, \dots, g^{*(m)}] \quad , \quad (4.24)$$

$$\tilde{H}^* = [h^{*(1)}, h^{*(2)}, \dots, h^{*(m)}] \quad , \quad (4.25)$$

$$\tilde{\Lambda} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_m \\ & & & & 0 \end{bmatrix},$$

sendo

$$\tilde{G}^{*T} \tilde{G}^* = \tilde{I}_m \quad \text{e} \quad \tilde{H}^* \cdot \tilde{H}^{*T} = \tilde{I}_m.$$

As matrizes \tilde{G}^* e \tilde{H}^* têm propriedades de maximização da função Ψ^* .

Logo, $\tilde{x} = \tilde{g}^{*(r)}$ maximiza

$$f_r(\tilde{x}) = \tilde{x}^T \tilde{F}^* \tilde{F}^{*T} \tilde{x} \quad (4.27)$$

com as restrições

$$\sum_i \tilde{x}_i^2 = 1 \quad (4.28)$$

$$\sum_i \tilde{x}_i \tilde{g}_i^{*(s)} = 0 \quad (s=1, 2, \dots, r-1) \quad (4.29)$$

Temos então $\tilde{g}^{*(r)}$ como vetor característico (vc) de $\tilde{F}^* \tilde{F}^{*T}$ correspondente à raiz característica (rc) λ_r^2 .

Pós-multiplicando (4.23) por \tilde{H}^* obtemos

$$\tilde{F}^* \tilde{H}^{*T} = \tilde{G}^* \tilde{\Lambda} \quad (4.30)$$

Logo,

$$\tilde{F}^* \tilde{h}^{*(r)} = \lambda_r \tilde{g}^{*(r)} \quad (4.31)$$

Pré-multiplicando (4.31) por $\tilde{g}^{*(r)T}$ e usando (4.28) obtemos

$$\tilde{g}^{*(r)T} \tilde{F} \tilde{h}^{*(r)} = \sum_i \sum_j f_{ij} \tilde{g}_i^{(r)} \tilde{h}_j^{(r)} = \lambda_r \quad (4.32)$$

Naturalmente podemos trocar os papéis de $\tilde{g}^{*(r)}$ e $\tilde{h}^{*(r)}$.

Então temos:

$$\sum_j \tilde{h}_j^{*(r)} = 1, \quad (4.33)$$

$$\sum_j \tilde{h}_j^{*(r)} \tilde{h}_j^{*(s)} = 0 \quad (s=1,2,3,\dots,r-1) \quad (4.34)$$

Resolvendo o problema em termos de \tilde{F}^* , voltamos aos termos de \tilde{F} .

De (4.23) podemos escrever \tilde{F} como

$$\tilde{F} = \tilde{R}^{1/2} \tilde{G}^* \tilde{\Lambda} \tilde{H}^* \tilde{S}^{1/2} = \tilde{R} \tilde{G} \tilde{\Lambda} \tilde{H} \tilde{S} \quad (4.35)$$

onde

$$\tilde{G} = \tilde{R}^{-1/2} \tilde{G}^* \quad \text{e} \quad \tilde{H} = \tilde{H}^* \tilde{S}^{-1/2}.$$

Considerando $g_i^{(r)}$ como i -ésimo elemento do r -ésimo vetor coluna de \tilde{G} , e h_j como j -ésimo elemento do r -ésimo vetor coluna de \tilde{H}^T , podemos reformular as propriedades ortonormais de \tilde{G}^* e \tilde{H}^* . Temos então:

$$\sum_i f_{i+} g_i^{(r)} g_i^{(s)} = \delta_{rs}, \quad (4.36)$$

$$\sum_j f_{+j} h_j^{(r)} h_j^{(s)} = \delta_{rs} \quad \text{para } r=1,2,\dots,m; s=1,2,\dots,m \quad (4.37)$$

Obtemos, portanto, o equivalente de (4.31)

$$\lambda_r \tilde{g}^{(r)} = \tilde{R}^{-1} \tilde{F} \tilde{h}^{(r)} \quad (4.38)$$

e o correspondente

$$\lambda_r \tilde{h}^{(r)} = (\tilde{F} \tilde{S}^{-1})^T \tilde{g}^{(r)} \quad (4.39)$$

Combinando (4.38) e (4.39) encontramos $g^{(r)}$ e $h^{(r)}$ correspondentes à matriz de Markov $R^{-1} \tilde{F} (\tilde{F} \tilde{S}^{-1})^T$ e $h^{(r)}$ vc de $(\tilde{F} \tilde{S}^{-1})^T \tilde{R}^{-1} \tilde{F}$, ambas com raiz característica λ_r^2 .

4.4 - REPRESENTAÇÃO GRÁFICA DOS FATORES

Como dissemos anteriormente, a AC atribui para cada linha e cada coluna m escala de valores; para cada linha i , os valores $g_i^{(1)}, g_i^{(2)}, \dots, g_i^{(m)}$ e, para a coluna j , os valores $h_j^{(1)}, h_j^{(2)}, \dots, h_j^{(m)}$. Podemos dar peso para estes valores dependentes da escala de poder de explicação (λ).

Assim multiplicamos $g_i^{(r)}$ e $h_j^{(r)}$ por λ_r . Como a primeira escala de valores ($\lambda_1=1$) não fornece nenhuma informação, ela não será considerada. Desta forma, ficamos com os pontos R^{m-1} para cada linha e cada coluna:

$$(\lambda_2 g_i^{(2)}, \lambda_3 g_i^{(3)}, \dots, \lambda_m g_i^{(m)}) \text{ e } (\lambda_2 h_j^{(2)}, \lambda_3 h_j^{(3)}, \dots, \lambda_m h_j^{(m)}).$$

Vamos interpretar estas coordenadas para linhas.

As coordenadas descrevem a configuração de m pontos. Esta configuração pode ser interpretada como translação e rotação de uma outra configuração, onde a linha i têm as seguintes coordenadas:

$$\left(\frac{f_{i1}}{f_{i+} \sqrt{f_{+1}}}, \frac{f_{i2}}{f_{i+} \sqrt{f_{+2}}}, \dots, \frac{f_{in}}{f_{i+} \sqrt{f_{+n}}} \right) \quad (4.40)$$

Esta configuração é transladada de tal forma que a distribuição marginal, correspondente à $(\sqrt{f_{+1}}, \sqrt{f_{+2}}, \dots, \sqrt{f_{+n}})$, coincide com a origem. Em seguida, a rotação é efetuada de tal forma que a configuração é descrita, tão bem quanto possível, inicialmente pela 2ª coordenada, em seguida pela 3ª coordenada e assim por diante.

Um dos conceitos básicos na AC é o que se refere à "distância". É uma medida de dissimilaridade entre as linhas (colunas)

i_u e i_v (j_u, j_v).

Pelo procedimento de translação e rotação descrito, a distância entre pontos mantêm-se invariante. De acordo com as coordenadas (4.40), a distância é dada por

$$d(i_u, i_v) = \sum_{k=2}^m \lambda_k^2 (g_u^{(k)} - g_v^{(k)})^2 = \sqrt{\frac{\sum_{j=1}^n \left(\frac{f_{uj}}{f_{u+}} - \frac{f_{vj}}{f_{v+}} \right)^2}{f_{+j}}} \quad (4.41)$$

pois

$$d^2(i_u, i_v) = \sum_{k=2}^m (\lambda_k g_u^{(k)} - \lambda_k g_v^{(k)})^2 =$$

$$(4.38) \sum_{k=2}^m \sum_j \left(\frac{f_{uj}}{f_{u+}} h_j^{(k)} - \sum_{j'} \frac{f_{vj'}}{f_{v+}} h_{j'}^{(k)} \right)$$

$$= \sum_j \sum_{j'} \left(\frac{f_{uj}}{f_{u+}} - \frac{f_{vj}}{f_{v+}} \right) \left(\frac{f_{uj'}}{f_{u+}} - \frac{f_{vj'}}{f_{v+}} \right) \sum_{k=2}^m h_j^{(k)} h_{j'}^{(k)}$$

$$= \sum_j \sum_{j'} \left| \frac{\left(\frac{f_{uj}}{f_{u+}} - \frac{f_{vj}}{f_{v+}} \right) \left(\frac{f_{uj'}}{f_{u+}} - \frac{f_{vj'}}{f_{v+}} \right)}{\sqrt{f_{+j} \cdot f_{+j'}}} \right| \sum_{k=2}^m h_j^{*(k)} h_{j'}^{*(k)}$$

(4.42)

Em (4.17) H^* é uma base ortonormal de autovetores. Então

$$\sum_{k=2}^m h_j^{*(k)} h_{j'}^{*(k)} = \delta_{jj'}$$

e portanto (4.42) se reduz a (4.41).

Da mesma forma podemos repetir a interpretação de coordenadas para as colunas. Notamos, também, que a distância entre as colunas j_u e j_v é dada por:

$$d(j_u, j_v) = \sqrt{\sum_{k=2}^m \lambda_k^2 (h_i^{(k)} - h_v^{(k)})^2} = \sqrt{\frac{\sum_{i=1}^m \left(\frac{f_{iu}}{f_{i+}} - \frac{f_{iv}}{f_{i+}} \right)^2}{f_{i+}^2}} \quad (4.43)$$

Quanto maior a distância entre as linhas (colunas), maior é a distribuição dentro das linhas (colunas). Quando a fração marginal é pequena, tal diferença contribui com maior peso do que quando a fração marginal é maior. Se a distância da linha (coluna) à origem é grande, a distribuição dentro da linha (coluna) desvia mais da distribuição marginal.

Estabelecida a relação entre as linhas e a relação entre as colunas, consideremos a relação entre linhas e colunas.

A forma escalar de (4.35) é dada por

$$f_{ij} = f_{i+} f_{+j} \sum_{k=1}^m \lambda_k g_i^{(k)} h_j^{(k)} \quad (4.44)$$

ou

$$\frac{f_{ij}}{f_{i+} f_{+j}} - 1 = \sum_{k=2}^m (\lambda_k)^{-1} (\lambda_k g_i^{(k)}) (\lambda_k h_j^{(k)}) \quad (4.45)$$

Se o lado esquerdo da relação (4.45) é menor que zero temos $f_{ij} < f_{i+} f_{+j}$; portanto, i e j são dependentes negativamente, se o lado esquerdo é maior que zero, a linha i e a coluna j são dependentes positivamente.

O lado direito da relação (4.45) pode ser identificado como um produto interno, mas não na métrica Euclidiana original, pois os termos têm peso extra $1/\lambda_k$. Isto faz com que a interpretação exata seja um pouco difícil, mas as propriedades principais são claras.

Quando a linha i e a coluna j ficam quase na mesma direção da origem, os termos do lado direito de (4.45) são todos positivos e i e j são dependentes positivamente. Por outro lado, se i e j estão na direção oposta da origem, i e j são dependentes negativamente. Quanto maior a distância da origem maior é a dependência.

O primeiro par de fatores ($g^{(1)}$ e $h^{(1)}$) não contribui com nenhuma informação, por isso começamos a contagem a partir do segundo fator, que é o primeiro fator não-trivial.

Definimos

$$\rho_k = \lambda_{k+1} \quad (k=1, 2, \dots, m-1) \quad (4.46)$$

Assim ρ_1 é a correlação entre $g^{(2)}$ e $h^{(2)}$.

O grau que o fator k explica a completa associação da tabela é expressa por τ_k .

$$\tau_k = \frac{\rho_k^2}{\rho_1^2 + \rho_2^2 + \dots + \rho_{m-1}^2} \quad (k=1, 2, \dots, m-1) \quad (4.47)$$

4.5 - ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA (ACM)

Na ACM temos a seguinte estrutura de \tilde{F}

$$\tilde{F} = \begin{bmatrix} \tilde{F}_{11} & \tilde{F}_{12} \dots \tilde{F}_{1v} \\ \tilde{F}_{21} & \tilde{F}_{22} \dots \tilde{F}_{2v} \\ \vdots & \vdots \quad \quad \quad \vdots \\ \tilde{F}_{u1} & \tilde{F}_{u2} \quad \quad \tilde{F}_{uv} \end{bmatrix} \quad (4.48)$$

As linhas de \tilde{F} correspondem a u variáveis categóricas, e da mesma forma, as colunas de \tilde{F} correspondem a v variáveis categóricas.

Aqui, \tilde{F}_{rs} é uma tabela bimarginal de variáveis r e s ($r=1,2,\dots,u$; $s=1,2,3,\dots,v$).

Nesta situação, cada unidade da população é contada uv vezes em \tilde{F} . Isto faz com que a interpretação torna-se um pouco mais complicada do que no caso anterior.

A AC pode ser aplicada para \tilde{F} como se fosse tabela de dupla entrada simples, portanto, todos os procedimentos e teorias desenvolvidos anteriormente são válidos.

Isto pode ser justificado pelo seguinte argumento: a distância descrita em (4.41) e (4.43) é ainda uma medida de dissimilidade significativa para as linhas e as colunas de "supermatriz" \tilde{F} .

Podemos interpretar a distância entre i_u e i_v como média de distância entre a distribuição condicional em relação a v categorias de variáveis que formam a coluna.

Para simplificar, consideremos o caso $u=2$ e $v=1$ e definamos \tilde{F}_1 e \tilde{F}_2 como sendo \tilde{F}_{11} e \tilde{F}_{21} .

$$\tilde{F} = \begin{bmatrix} \tilde{F}_1 \\ \tilde{F}_2 \end{bmatrix} \quad (4.49)$$

Neste caso \tilde{F} consiste em duas Tabelas simples \tilde{F}_1 e \tilde{F}_2 .

Temos três variáveis categóricas: I , J e K cujas categorias são $\{I_1, I_2, \dots, I_{m_1}\}$, $\{J_1, J_2, \dots, J_{m_2}\}$ e $\{K_1, K_2, \dots, K_n\}$ respectivamente.

Consideremos três vetores aleatórios "dummy": X , Y e Z

$$X_r = 1 \quad \text{se } I = I_r \quad (r=1, 2, \dots, m_1)$$

$$Y_s = 1 \quad \text{se } J = J_s \quad (s=1, 2, \dots, m_2)$$

$$Z_t = 1 \quad \text{se } K = K_t \quad (t=1, 2, \dots, n)$$

$$X_r = Y_s = Z_t = 0 \quad \text{caso contrário.}$$

Temos, então:

a) $F_{\sim 1}$ uma tabela de dupla entrada $I \times K$

b) $F_{\sim 2}$ uma tabela de dupla entrada $J \times K$

A fração populacional, onde $X_r=1$, $Y_s=1$ e $Z_t=1$, é f_{rst} . Denotando as marginais por "+" vemos que F_{\sim} tem a seguinte estrutura:

$$F_{\sim} = \begin{bmatrix} f_{1+1} & f_{1+2} & \cdot & \cdot & \cdot & \cdot & \cdot & f_{1+n} \\ f_{2+1} & f_{2+2} & \cdot & \cdot & \cdot & \cdot & \cdot & f_{2+n} \\ \cdot & \cdot & & & & & & \cdot \\ \cdot & \cdot & & & & & & \cdot \\ f_{m_1+1} & f_{m_2+2} & & & & & & f_{m_1+n} \\ f_{+11} & f_{+12} & \cdot & \cdot & \cdot & \cdot & \cdot & f_{+1n} \\ f_{+21} & f_{+22} & \cdot & \cdot & \cdot & \cdot & \cdot & f_{+2n} \\ \cdot & \cdot & & & & & & \cdot \\ \cdot & \cdot & & & & & & \cdot \\ \cdot & \cdot & & & & & & \cdot \\ f_{+m_2 1} & f_{+m_2 2} & \cdot & \cdot & \cdot & \cdot & \cdot & f_{+m_2 n} \end{bmatrix} \quad (4.50)$$

Quando aplicamos AC em F_{\sim} , maximizamos

$$\psi(\underline{g}_1, \underline{g}_2, \underline{h}) = \underline{g}_1^T F_{\sim 1} \underline{h} + \underline{g}_2^T F_{\sim 2} \underline{h} \quad (4.51)$$

e com as restrições

$$\frac{1}{2} \left(\sum_i f_{i++} g_{1i} + \sum_j f_{+j+} g_{2j} \right) = \sum_k f_{++k} h_k = 0$$

$$\frac{1}{2} \left(\sum_i f_{i++} g_{1i}^2 + \sum_j f_{+j+} g_{2j}^2 \right) = \sum_k f_{++k} h_k^2 = 1$$

Proposição 1: Aplicação de AC para uma tabela composta $\begin{bmatrix} F \\ \sim 1 \\ F \\ \sim 2 \end{bmatrix}$ com variáveis G_1 e G_2 , associadas com as linhas, e H , associada com a coluna, significa maximizar a seguinte estatística:

$$\lambda(G_1, G_2, H) = \frac{1}{\sqrt{2}} \frac{\text{cov} \left[\frac{G_1 + G_2}{2}, H \right]}{\sqrt{\left[\sigma^2 \left(\frac{G_1}{2} \right) + \sigma^2 \left(\frac{G_2}{2} \right) \right] \sigma^2(H)}} \quad (4.54)$$

De fato, a relação (4.52) pode ser interpretada como

$$E\left(\frac{G_1 + G_2}{2}\right) = E(H) = 0.$$

Demonstraremos que $E(G_1) = E(G_2) = 0$.

Para esta finalidade, maximizamos a função definida anteriormente (4.51) com as restrições (4.52) e (4.53).

Utilizando o multiplicador de Lagrange, temos a seguinte função:

$$\begin{aligned} \psi = & g_1^T F_1 \tilde{h} + g_2^T F_2 \tilde{h} + \mu_1 E\left(\frac{G_1 + G_2}{2}\right) + \mu_2 E(H) + \\ & + \mu_3 \left[E\left(\frac{G_1^2 + G_2^2}{2}\right) - 1 \right] + \mu_4 \left[E(H^2) - 1 \right]. \end{aligned} \quad (4.55)$$

Derivando-a em relação a g_{1i} , temos:

$$\frac{1}{2} \sum f_{i+k} h + \frac{\mu_1}{2} f_{i++} + \mu_3 f_{i++} g_{1i} = 0 \quad (4.56)$$

e somando-a em i , temos:

$$\frac{1}{2} E(H) + \frac{\mu_1}{2} + \mu_3 E(G_1) = 0 \quad (4.57)$$

Da mesma forma para g_{2i} , temos:

$$\frac{1}{2} E(H) + \frac{\mu_1}{2} + \mu_3 E(G_2) = 0, \quad (4.58)$$

portanto:

$$E(G_1) = E(G_2) \quad \text{ou} \quad \mu_3 = 0.$$

Multiplicando (4.56) por g_{1i} e somando em i temos:

$$g_{1i}^T f_{i+k} h + \frac{\mu_1}{2} E(G_1) + \mu_3 E(G_1^2) = 0 \quad (4.59)$$

Da mesma maneira para g_{2i} , temos:

$$g_{2i}^T f_{i+k} h + \frac{\mu_1}{2} E(G_2) + \mu_3 E(G_2^2) = 0 \quad (4.60)$$

Somando (4.59) e (4.60) e utilizando as restrições (4.52) e (4.53), temos:

$$\psi(g_1, g_2, h) + 2\mu_3 = 0.$$

Nos casos em que $\psi > 0$, temos $\mu_3 \neq 0$ e $E(G_1) = E(G_2)$. Assim podemos interpretar as relações (4.52) e (4.53) como média e variância, exatamente como no caso (4.14) e (4.15).

Proposição 2 - Aplicação de AC para tabela composta de variáveis G_1, G_2, \dots, G_u , associadas com as linhas, e H_1, H_2, \dots, H_v , associadas com as colunas, significa maximizar a seguinte estatística:

$$\lambda(G_1, G_2, \dots, G_u, H_1, H_2, \dots, H_v) = \frac{1}{uv} \frac{\text{cov}(\bar{G}, \bar{H})}{\sqrt{\sigma^2(G), \sigma^2(H)}} \quad (4.62)$$

onde

$$\bar{G} = \frac{1}{u} \sum_{i=1}^u G_i, \quad \bar{H} = \frac{1}{v} \sum_{j=1}^v H_j,$$

$$\overline{\sigma^2(G)} = \frac{1}{u^2} \sum_{i=1}^u \sigma^2(G_i) \quad \text{e} \quad \overline{\sigma^2(H)} = \frac{1}{v^2} \sum_{j=1}^v \sigma^2(H_j)$$

A prova desta proposição é análoga à da proposição 1.

4.6 - REGRAS DE INTERPRETAÇÃO

Algumas das regras básicas para interpretar os resultados de AC são:

- i) A distância entre duas linhas (colunas) indica a dissimilaridade das distribuições dentro destas linhas (colunas). A distância entre linhas (colunas) e a origem indica dissimilaridade da distribuição dentro da linha (coluna) e a distribuição marginal correspondente.
- ii) Quando uma linha e uma coluna estão na mesma direção da origem, elas são positivamente dependentes. Quando ficam na direção oposta em relação à origem, elas são negativamente dependentes. Quanto mais distante da origem, maior é a dependência.
- iii) A intensidade da dependência, que é descrita pelo K-ésimo fator, é dada por ρ_k , que é a correlação entre K-ésima coordenada das linhas e K-ésima coordenada das colunas (K-ésimo fator não-trivial).
- iv) A medida de aproximação da configuração do K-ésimo fator é expressa por τ_k .

4.7 - APLICAÇÃO

Com finalidade ilustrativa, vamos tentar explorar, de forma conveniente, vários modelos para avaliar a eficiência desta técnica, utilizando os dados descritos anteriormente.

Analisando os resultados de vários modelos, Fig. 4.2 a Fig. 4.6, podemos denominar de Renda o primeiro eixo. Este explica cerca de 80% ($\tau_1 = 0.80$) das informações contidas nos dados.

Como o nosso objetivo é estudar a influência das variáveis Escolaridade, Idade e Atividade na variável R e n d a , análise ficará restrita no máximo aos dois primeiros eixos.

No modelo $(A+E+I) \times Y$ cada indivíduo é contado três vezes, pois a sua classificação é feita baseada nas três variáveis (E, I e A) independentemente. Neste modelo, detectamos qual das categorias destas variáveis corresponde a uma determinada categoria de Renda. Nota-se que a categoria de Escolaridade Alta corresponde à categoria de Renda máxima ($Y_5 = RE5$) e, por outro lado, a categoria de Idade Mais Jovem (I_1) está associada à categoria de Renda Mais Baixa ($Y_1 = RE1$). Quanto melhor o nível de Escolaridade, maior é a renda e, da mesma forma, quanto maior a Idade, melhor a Renda, exceto para última categoria de Idade (I_4) cuja renda é um pouco inferior em relação à categoria anterior (I_3). Notamos que a atividade que tem a melhor Renda (Y_5) é a atividade Têxtil (A_1) e a atividade Vestuário (A_2) possui Renda mais baixa. Os valores atribuídos para cada uma das variáveis são os valores das coordenadas do autovetor correspondente ao primeiro eixo e estão na tabela 4.1.

Este procedimento pode ser repetido satisfatoriamente para os demais modelos.

Vamos verificar a interação de segunda ordem, baseada nos modelos correspondentes a Fig. 4.3 a 4.5 ($E \times A$, $I \times A$ e $E \times I$). De acordo com a Fig. 4.3 podemos notar a existência de três conglomeradores correspondentes às categorias de Escolaridade: Alta, Média e Baixa. O conglomerado de Escolaridade Alta encontra-se bem

Tabela 4.1 - Escala de Valores das Variáveis

Renda = Y	Atividade = A	Escolaridade = E	Idade = I
$Y_1 = -0.88$	$A_1 = 0.26$	$E_1 = -0.36$	$I_1 = -0.85$
$Y_2 = -0.28$	$A_2 = -0.44$	$E_2 = -0.07$	$I_2 = 0.12$
$Y_3 = 0.20$	$A_3 = 0.08$	$E_3 = 0.77$	$I_3 = 0.41$
$Y_4 = 0.45$	$A_4 = -0.30$		$I_4 = 0.31$
$Y_5 = 0.60$	$A_5 = 0$		
	$A_6 = -0.10$		

isolado em relação aos outros. Se compararmos a ordem do nível de Renda de atividades dentro dos conglomerados, notamos uma ligeira alternância para algumas atividades. Isto demonstra a existência de uma pequena interação entre as variáveis Escolaridade e Atividade ($E \times A$) que influem na Renda.

Na Fig. 4.4, podemos separar, com alguma dificuldade, quatro conglomerados correspondentes à variável Idade (I_1, I_2, I_3 e I_4). O conglomerado correspondente à Idade Mais Jovem (I_1) encontra-se isolado do restante. Aqui, a interação das variáveis Idade e Atividade, influenciando na Renda, parece ser um pouco maior do que na figura anterior.

De acordo com Fig. 4.5, o fato do indivíduo ser jovem, mesmo tendo uma boa escolaridade, coloca-o na faixa de Renda relativamente baixa, portanto, se basearmos nas configurações da Fig. 4.2 e Fig. 4.5, notamos que a interação de ambas as variáveis é maior do que nos modelos anteriores.

Pela Fig. 4.6 notamos uma certa semelhança na configuração com a Fig. 4.5. Verifica-se que a inclusão da variável Atividade explica pouco ou influi pouco na forma como as faixas de Renda estão distribuídas. Destacamos as Atividades Têxtil (A_1) e Vestuário (A_2) que têm maior influência na variável Renda.

Fig. 4.1 - Análise de Correspondência na forma $(A+E+I) \times Y$
(O primeiro fator).

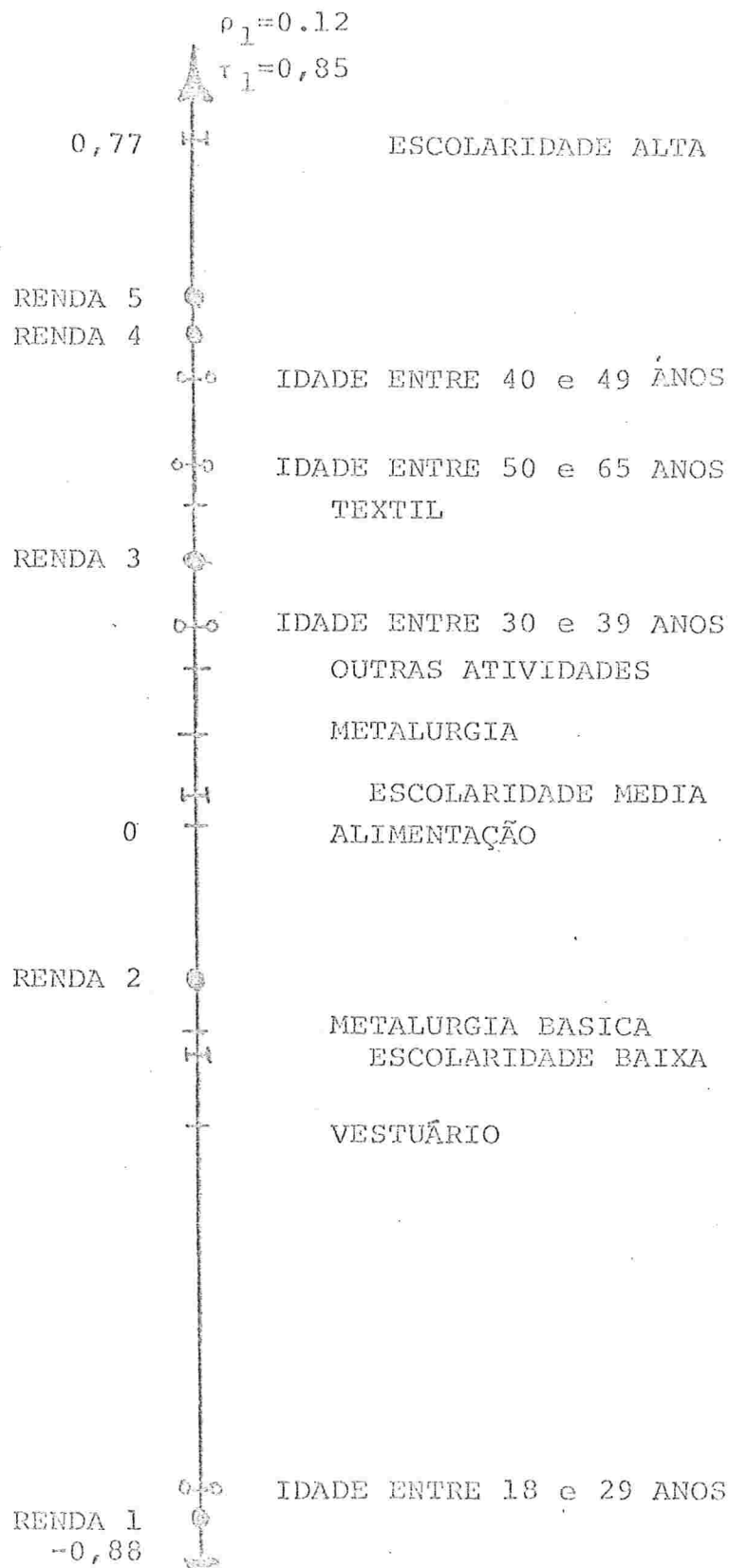
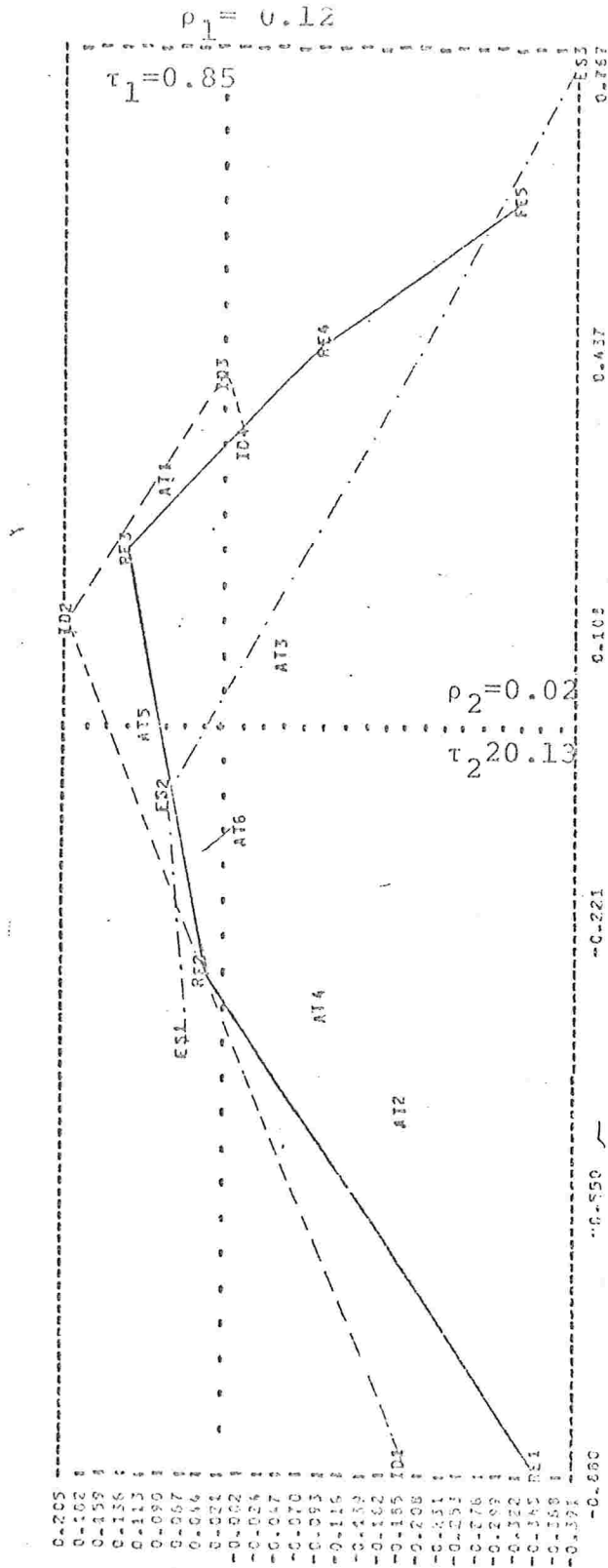
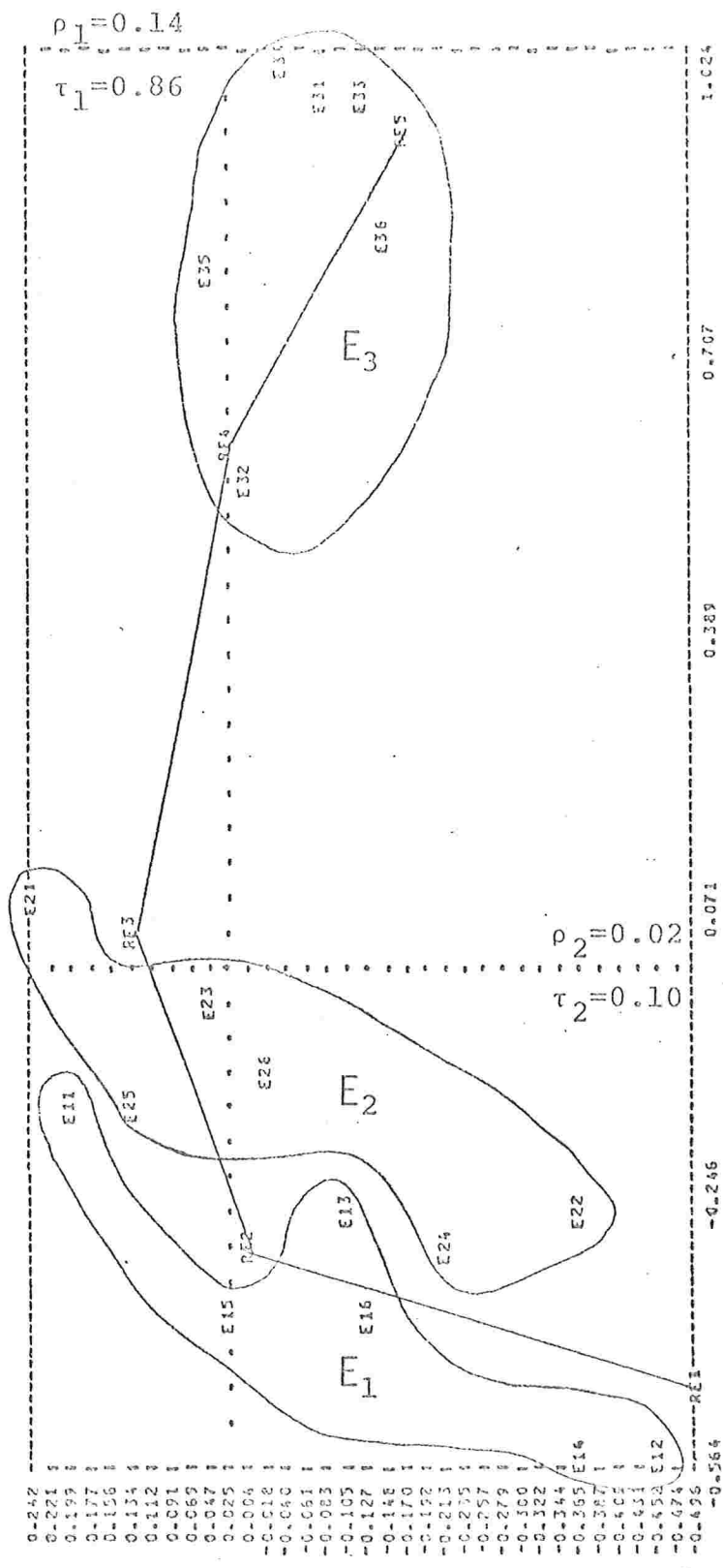


Fig. 4.2 - Análise de Correspondência na Forma: $(A+E+I) \times Y$



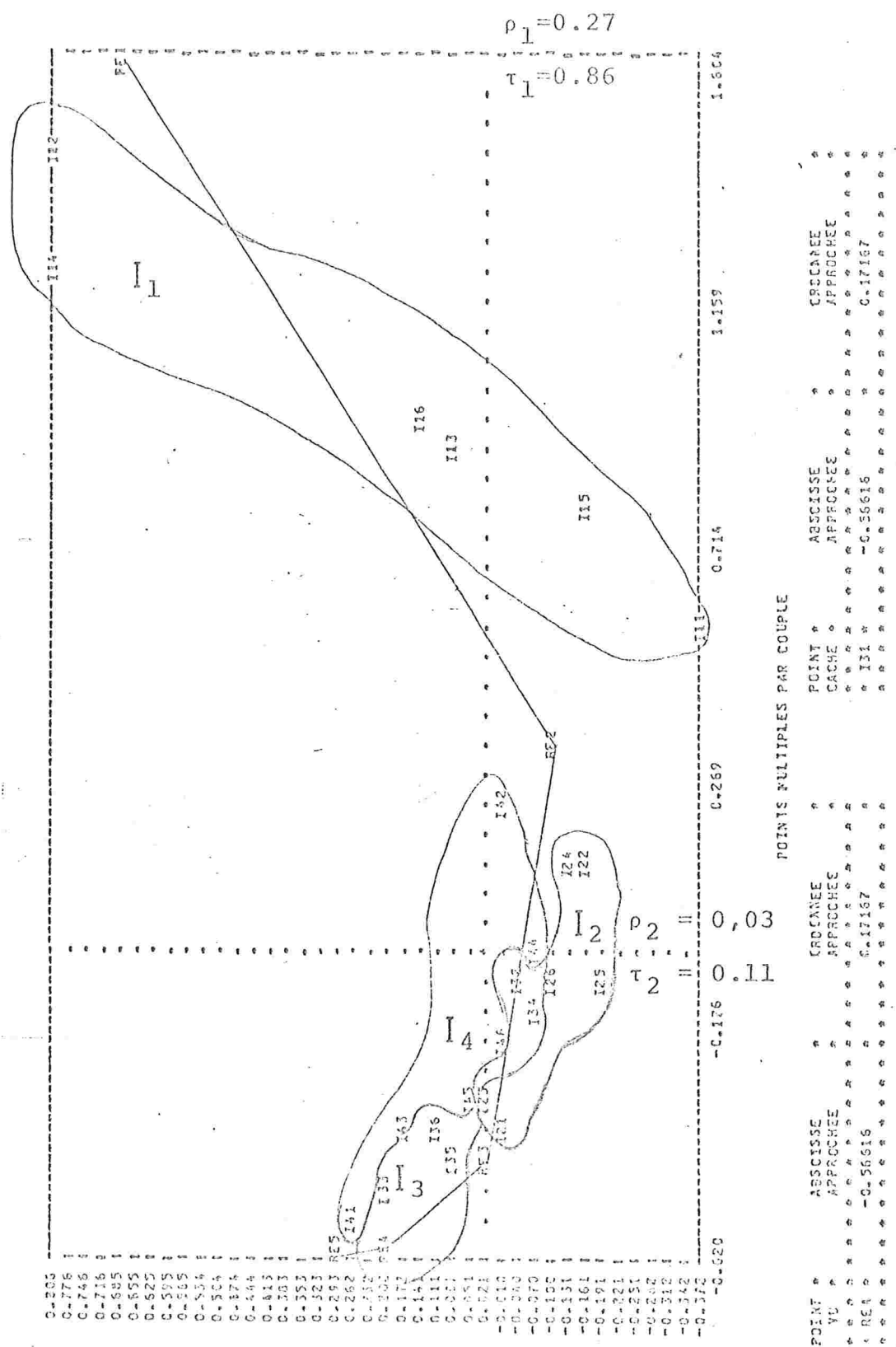
$AT_i = A_i ; ES_i = E_i ; ID_i = I_i ; RE_i = Y_i$

Fig. 4.3 - Análise de Correspondência na Forma: $(E \times A) \times Y$

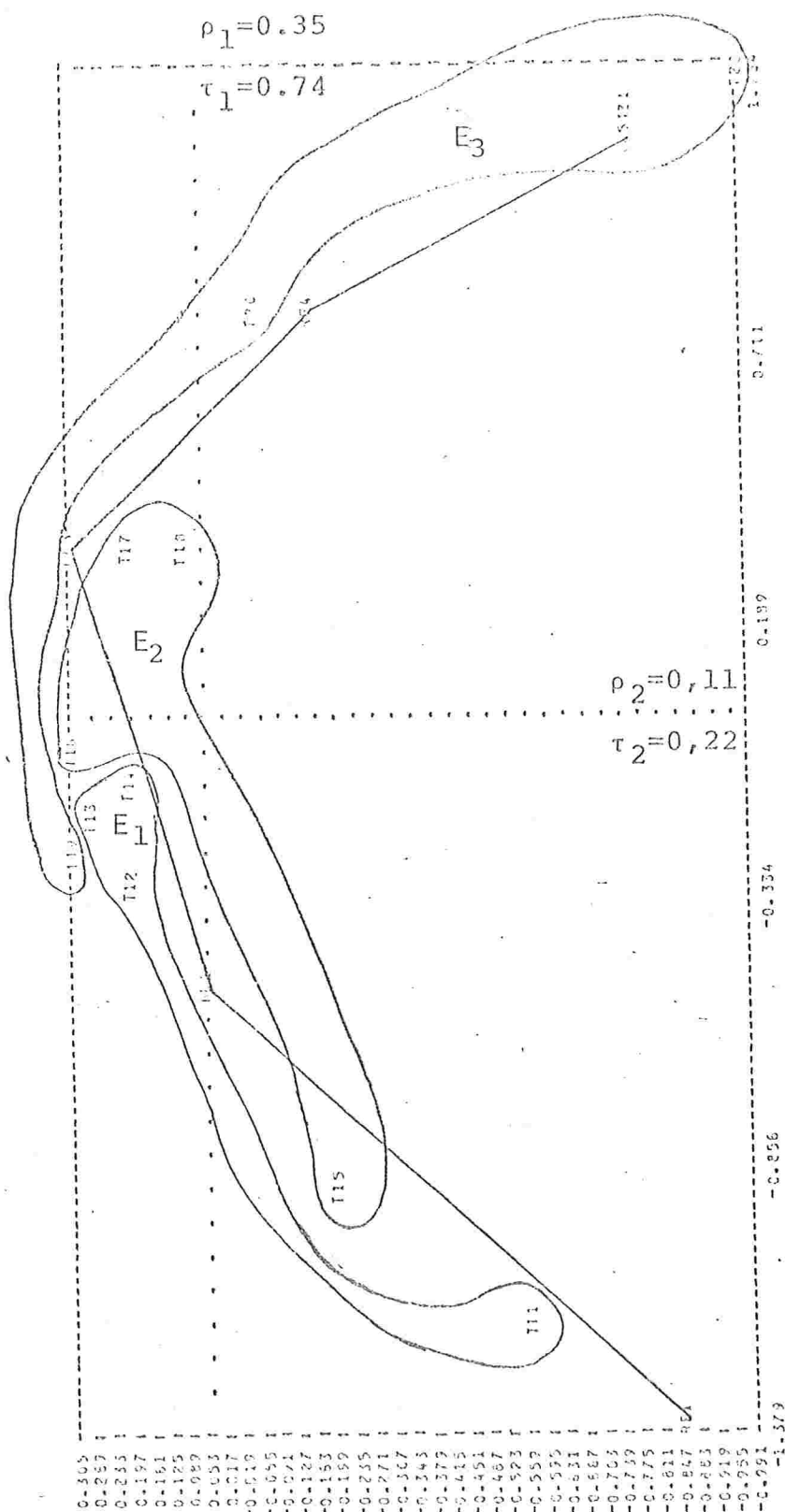


$$E_{ij} = \begin{cases} i(1, 2 \text{ e } 3) = \text{Escolaridade}(E_1, E_2 \text{ e } E_3) \\ j(1, 2, 3, 4, 5 \text{ e } 6) = \text{Atividade}(A_1, A_2, A_3, A_4, A_5 \text{ e } A_6) \end{cases}$$

Fig. 4.4 - Análise de Correspondência na Forma: $(I \times A) \times Y$

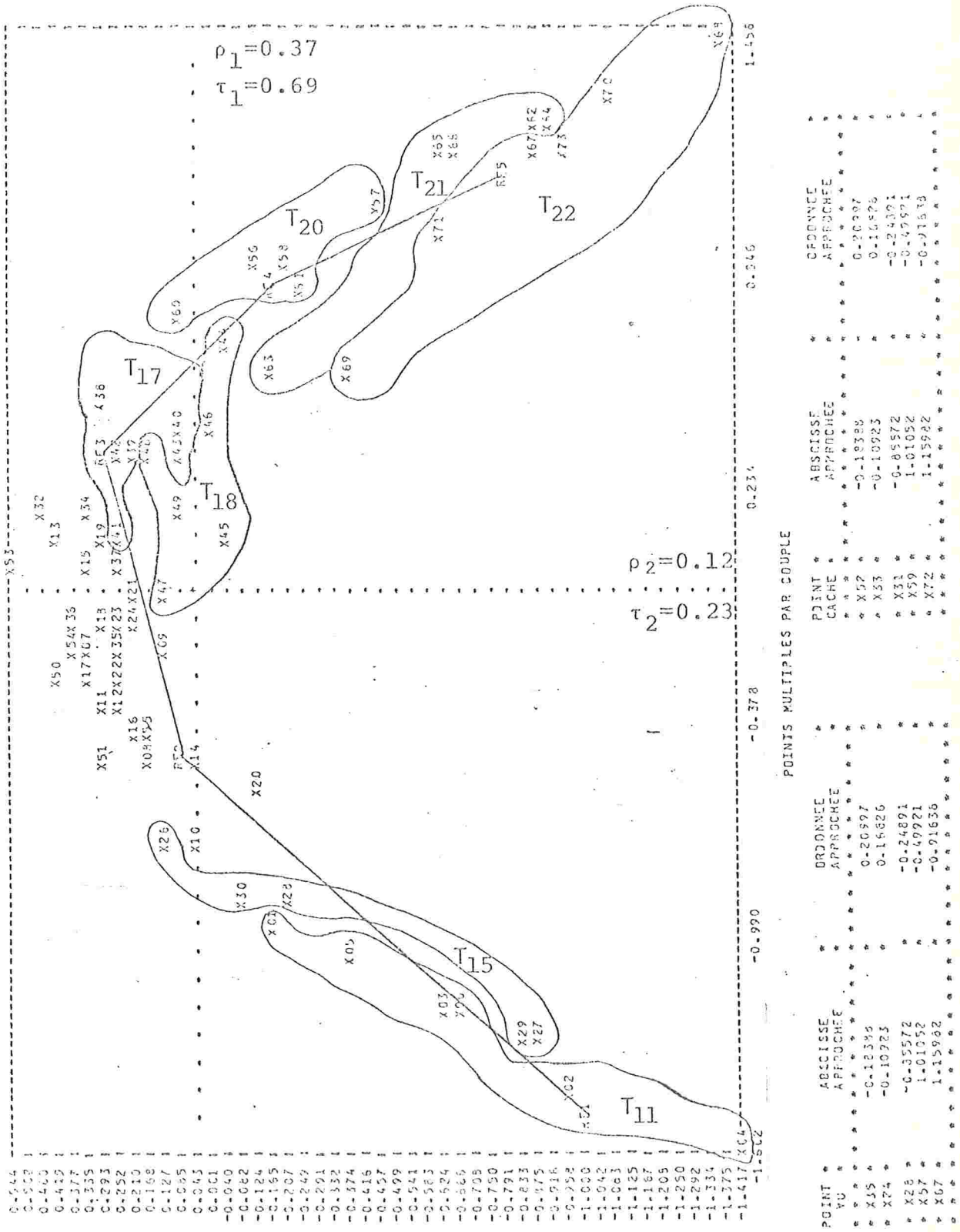


$$I_{ij} = \begin{cases} i(1,2,3 \text{ e } 4) = \text{Idade } (I_1, I_2, I_3 \text{ e } I_4) \\ j(1,2,3,4,5 \text{ e } 6) = \text{Atividade } (A_1, A_2, A_3, A_4, A_5 \text{ e } A_6) \end{cases}$$

Fig. 4.5 - Análise de Correspondência na Forma: $(E \times I) \times Y$ 

onde $\begin{cases} \ell = 4(E_i - 1) + I_j + 10 \\ E_i = 1, 2 \text{ e } 3; I_j = 1, 2, 3 \text{ e } 4 \end{cases}$

Fig. 4.6 - Análise de Correspondência na forma: $(A \times E \times I) \times Y$



X_{ℓ} onde $\begin{cases} \ell = 24 \times (E_i - 1) + 6(I_j - 1) + A_K \\ E_i = 1, 2 \text{ e } 3; I_j = 1, 2, 3 \text{ e } 4; A_K = 1, 2, 3, 4, 5, 6. \end{cases}$

CAPÍTULO 5

CONCLUSÕES

5.1 - INTRODUÇÃO

Neste capítulo, fazemos as comparações da utilidade e "performance" entre as técnicas apresentadas. Na verdade, este tipo de comparações baseado somente num tipo de pesquisa pode tornar arbitrárias as conclusões. Como é o caso da Análise de Rídit que pode ser utilizada para outras finalidades. Como exemplo, citamos a atribuição de pesos para as respostas de questionários de uma entrevista (Dennis, 1980). No caso da Análise de Correspondência e Análise de Regressão, podemos ter uma ou mais variáveis dependentes (respostas) e ainda existe a possibilidade da variável resposta não ser ordinal.

5.2 - RESUMO DAS CONCLUSÕES DAS ANÁLISES

- i) Quanto melhor a Escolaridade (E) maior é a Renda.
- ii) Uma nítida diferença de Renda entre as categorias de Escolaridade Alta (E_3) e as categorias de Escolaridades Média e Baixa (E_1, E_2), exceto para as categorias de Idades acima de 40 anos (I_3+I_4) da atividade Vestuário (A_2).

- iii) A categoria de Idade entre 40 e 49 anos (I_3) é a que apresenta melhor Renda, caindo um pouco na categoria seguinte (I_4).
- iv) A categoria Mais Jovem (I_1) é a que possui menor Renda e a que permanece isolada das demais categorias.
- v) A Renda da categoria Mais Jovem (I_1) é sempre bem abaixo da média de Renda da respectiva atividade a que pertence.
- vi) A influência da Escolaridade na renda cresce à medida que cresce a faixa de Idade, às vezes decrescendo um pouco na última categoria (I_4).
- vii) As categorias Escolaridade Alta (E_3) e Idade Mais Jovem (I_1) são, individualmente, as categorias que se destacam das demais em relação à variável renda.
- viii) A atividade Têxtil (A_1) possui a melhor Renda e a atividade Vestuário (A_2) apresenta a menor Renda.
- ix) Existe interação entre as variáveis.
- x) As variáveis influentes da Renda são, pela ordem de importância: Idade, Escolaridade e Atividade.

Os resultados básicos obtidos foram os mesmos para os três métodos. A diferença fundamental entre os três métodos é o grau de dificuldade na interpretação para alguns dos casos estudados.

A vantagem da Análise de Rídit e da Análise de Correspondência é a orientação visual que facilita a análise dos dados através de gráfico. No caso de Análise de Regressão, podemos quantificar e medir a importância de cada uma das variáveis e as respectivas categorias, inclusive a interação entre as variáveis.

Se utilizarmos convenientemente e explorarmos todo o potencial que as três técnicas oferecem, podemos efetuar excelente análise, pois podemos dizer de modo geral que os três métodos apresentados se complementam.

Para finalizar, citamos algumas técnicas de análise multivariadas discretas na análise de dados multivariados discretos:

Análise Discriminante Discreta, Análise de Conglomerados, Modelos Log-Linear, Uso de χ^2 para Medida de Associação, Análise Fatorial e Componentes Principais.

APÊNDICE 1

1. CARACTERIZAÇÃO MATEMÁTICA DE "RIDIT"

Como já vimos no Capítulo 1 "ridit" é um método de atribuir valores para dados categóricos ordenados.

Denotamos p_i ($i=1, 2, 3, \dots, k$) como proporção empírica de N elementos de grupo de referência X , que caem no i -ésima categoria. Na forma vetorial, temos:

$$\underline{p} = (p_1, p_2, \dots, p_k) \quad \text{onde } p_i \geq 0 \quad \text{e} \quad \sum p_i = 1$$

O ridit de i -ésimo categoria pode ser expresso da seguinte forma:

$$R(X_i) = h(i, \underline{p}) + 0,5. \quad (1)$$

onde

$$h(i, \underline{p}) = c \left(\sum_{j < i} p_j - \sum_{j > i} p_j \right) \quad \text{e} \quad c = 0,5$$

A função $h(i, \underline{p})$ será denominada função atribuição e, antes de apresentar os teoremas, vamos enunciar os postulados que serão utilizados para esta finalidade.

Postulado 1

Se existe uma resposta, então não há nenhuma possível ordenação e não podemos distinguir uma resposta padrão.

$$\boxed{h_1(1,1) = 0} \quad (2)$$

Postulado 2

Para duas categorias de respostas, se invertermos a distribuição empírica, os valores atribuídos devem ser invertidos com sinais trocados, mantendo-se a ordenação de categoria

$$\boxed{0 \leq h_2(2,p,1-p) = -h_2(1,1-p,p)} \quad (3)$$

Postulado 3 (propriedade de ramificação)

Suponhamos que há mais de duas respostas ($K > 2$), e que remos combinar duas respostas de categorias adjacentes. Assumimos que a categoria de resposta mantém o mesmo valor atribuído, e o valor dado para a nova categoria combinada é a média ponderada dos valores dados nas respostas das categorias originais.

Simbolicamente, isto quer dizer que, se i e $(i+1)$ ésimas categorias são combinadas, $q(q_1, q_2 \dots q_{k-1})$ é a nova distribuição empírica.

$$(q_t = p_t \text{ para } t < i \quad q_i = p_i + p_{i+1}, \quad q_t = p_{t+1} \text{ } t > i) ,$$

portanto temos:

$$h_{k-1}(i,q) = \frac{p_i}{p_i + p_{i+1}} h_k(i,p) + \frac{p_{i+1}}{p_i + p_{i+1}} h_k(i+1,p)$$

$$h_{k-1}(t, q) = h_k(t, p) \quad t \leq i-1 \quad (4)$$

$$h_{k-1}(t, q) = h_k(t+1, p) \quad t \geq i+1$$

Postulado 4

Se existem duas categorias, $h_2(2, p, 1-p) - h_2(1, p, 1-p)$ é não decrescente à medida que p cresce.

Esta suposição reflete na característica desejada: os valores atribuídos para as duas categorias não devem se aproximar na medida em que a diferença de proporções de respondentes cresce.

Baseados nestes postulados, podemos enunciar os seguintes teoremas de caracterizações.

Teorema 1: Uma função atribuição h_k satisfaz os Postulados de 1 a 4 se, e somente se, $h_k(i, p) = c \left(\sum_{j < i} p_j - \sum_{j > i} p_j \right)$ onde c é uma constante arbitrária.

Prova: Definimos $f(p) = h_2(2, p, 1-p)$, $0 \leq p \leq 1$.

Dos postulados 1, 2 e 3 temos:

$$0 = h_1(1, p+1-p) = p f(1-p) - (1-p) f(p) \quad (5)$$

Seja $p = (1+t)/2$ e definimos a função ϕ como

$$\phi(t) = \phi(2p-1) = (2/(1+t)) f((1+t)/2) = (1/p) f(p) \quad (6)$$

e da mesma forma temos:

$$\phi(-t) = (1/(1-p)) f(1-p) \quad (7)$$

De (5) concluímos que ϕ é uma função par, portanto,

$$f(p) = p \vartheta(|2p-1|) \quad (8)$$

Consideremos $x = \sum_{j < i} p_j$ e $y = \sum_{j > i} p_j$. Aplicando o postulado 3 temos:

$$h_k(i, p) = h_3(2, x, p_i, y) \quad p_i > 0 \quad (9)$$

Prosseguindo temos:

$$h_3(3, x, p_i, y) = h_2(2, x+p_i, y) = f(x+p_i) =$$

$$(8) \\ = (x+p_i) \vartheta(|2(x+p_i)-1|) = (x+p_i) \vartheta(|x+p_i-y|) \quad (10)$$

e

$$h_3(1, x, p_i, y) = h_2(1, x, p_i+y) = -h_2(2, p_i+y, x) = \\ (2) \\ = -f(y+p_i) = -(y+p_i) \vartheta(|y+p_i-x|) \quad (11)$$

Dos postulados 1 e 3 temos:

$$0 = xh_3(1, x, p_i, y) + p_i h_3(2, x, p_i, y) + yh_3(3, x, p_i, y) \quad (12)$$

ou

$$h_3(2, x, p_i, y) = \frac{-xh_3(1, x, p_i, y) - yh_3(3, x, p_i, y)}{p_i} \quad (12)$$

Utilizando (10) e (11) com (9) em (12) temos:

$$h_k(i, p) = \frac{x(1-x) \vartheta(|y+p_i-x|) - y(1-y) \vartheta(|x+p_i-y|)}{p_i} \quad (13)$$

Agora, como $\vartheta(t) = \vartheta(|2p-1|)$ é uma função par para $p = \frac{1}{2}$ segue que $\vartheta(|2p-1|) = h_2(2, p, 1-p) - h_2(1, p, 1-p)$ não pode satisfazer o postulado 4 a menos que ϑ seja constante. Chamamos esta constante de c e substituindo-a em (13) obtemos

$$h_k(i, p) = \frac{x(1-x)c - y(1-y)c}{p_i} = c \left[\frac{x(y+p_i) - y(x+p_i)}{p_i} \right] =$$

$$= c(x-y)$$

ou seja:

$$h_k(i, p) = c \left[\sum_{j < i} p_j - \sum_{j > i} p_j \right], \quad (14)$$

cqd.

Teorema 2: Seja h função real satisfazendo estas condições:

- i) h é contínuo
- ii) $h(x, y) = \lambda h(x, y + (1-\lambda)c) + (1-\lambda)h(x + \lambda c, y)$

$$\forall x, y, c \text{ e } \lambda$$

Então: $h(x, y) = a + b(x-y)$ para algum a e b

Prova: Seja $f \in C^1$. Diferenciando (ii) em relação a λ resulta

$$0 = h(x, y + (1-\lambda)c) - h(x + \lambda c, y) - \lambda c h_y(x, y + (1-\lambda)c)$$

$$+ (1-\lambda)c h_x(x + \lambda c, y), \quad (15)$$

onde h_x e h_y são derivadas parciais de h em relação a x e y respectivamente.

Fazendo $\lambda=1$ em (15) encontramos

$$-h_y(x, y) = \frac{(h(x+c, y) - h(x, y))}{c} \quad (16)$$

Desta forma

$$-h_y(x, y) = h_x(x, y) \quad (17)$$

ou
$$\frac{d}{ds} h(x+s, y+s) \equiv 0 \quad (18)$$

Integrando de $-y$ a 0 resulta:

$$h(x, y) = \int_{-y}^0 h(x+s, y+s) ds = h(x-y, 0) = \phi(x-y) \quad (19)$$

Substituindo $\phi(x-y)$ em (15) e fazendo $\lambda=0$, $x=y$ encontramos:

$$\phi(-c) = -c\phi'(0) + \phi(0) \quad (20)$$

e, por outro lado, para $\lambda=1$ e $x=y$ encontramos:

$$\phi(c) = c\phi'(0) + \phi(0) \quad (21)$$

De (20) e (21) podemos concluir que

$$\phi(z) = a+bz \quad (a \text{ e } b \text{ constantes}) \quad (22)$$

Assim, chega-se à conclusão quando $h \in C^1$.

Suponhamos, agora, que h é somente contínua e seja

$$h_n(x, y) = \iint h(x+s, y+t) W_{1/n}(s, t) ds dt \quad (23)$$

onde $W_r \in C^\infty$ e $W_r(x, y) = 0$ se

$$x^2 + y^2 > r^2 \quad \iint W_r(x, y) dx dy = 1$$

Utilizando a hipótese (ii) do Teorema 2 em h de (23) temos h_n satisfazendo (ii) e é C^1 . Logo,

$$h_n(x, y) = a_n + b_n(x-y).$$

Desde que $h_n \rightarrow h \Rightarrow h(x, y) = a+b(x-y)$

APÊNDICE 2

1. AUTOVETORES DE B NA MÉTRICA DE A

Vamos considerar as matrizes \underline{A} e \underline{B} . Consideremos as raízes de $|\underline{A} - \lambda \underline{B}| = 0$, que são raízes características (rc) de \underline{A} na métrica \underline{B} , e seus vetores satisfazendo a equação:

$$\underline{A}\underline{x} = \lambda \underline{B}\underline{x} \quad (1)$$

Se \underline{B} é não-singular, a equação (1) é reduzida ao problema algébrico de autovetores.

$$\underline{C}\underline{x} = \lambda \underline{x} \quad (2)$$

sendo

$$\underline{C} = \underline{B}^{-1} \underline{A}$$

Definição: \underline{B}^S é a "Strong Spectral Inverse" (SEI) de \underline{B}

$$\underline{B} = \underline{P}^{-1} \underline{J} \underline{P}$$

$$\underline{B}^S = \underline{P}^{-1} \underline{J}^+ \underline{P}$$

onde

\underline{J}^+ é pseudo-inversa da matriz de Jordan \underline{J} .

Se B é singular podemos expressá-la da seguinte forma:

$$\underline{B} = \underline{P}^{-1} \underline{J} \underline{P}$$

onde \underline{J} é a forma canônica de \underline{B} .

A matriz $B^S = \underline{P}^{-1} \underline{J}^+ \underline{P}$ é a inversa espectral de \underline{B} .

Lema: Se \underline{x} é a solução da equação (1) correspondente a λ , então $\underline{y} = \underline{P}\underline{x}$, onde $\underline{B} = \underline{P}^{-1} \underline{J} \underline{P}$ com λ satisfazendo

$$\underline{P} \underline{A} \underline{P}^{-1} \underline{y} = \lambda \underline{P} \underline{B} \underline{P}^{-1} \underline{y} \quad (3)$$

e, reciprocamente, se λ e \underline{y} representam a solução de (3), então λ e $\underline{x} = \underline{P}^{-1} \underline{y}$ satisfazem (1).

Prova: Pré-multiplicando (1) por \underline{P} e inserindo $\underline{P}^{-1} \underline{P}$, obtemos $\underline{P} \underline{A} \underline{P}^{-1} (\underline{P}\underline{x}) = \lambda \underline{P} \underline{B} \underline{P}^{-1} (\underline{P}\underline{x})$ que estabelece a primeira parte. Reciprocamente, pré-multiplicando (3) por \underline{P}^{-1} obtemos:

$$\underline{A} (\underline{P}^{-1} \underline{y}) = \lambda \underline{B} (\underline{P}^{-1} \underline{y}) \quad \text{que completa a prova.}$$

Pré-multiplicando (3) por \underline{J}^+ e fazendo $\underline{C} = \underline{J}^+ \underline{P} \underline{A} \underline{P}^{-1}$ obtemos a equação $\underline{C}\underline{y} = \lambda \underline{J}^+ \underline{J} \underline{y}$, cuja solução será analisada.

Teorema: Qualquer solução $\underline{y} = \underline{P}\underline{x}$ da equação (3), sujeita à condição $\underline{B}^S \underline{B}\underline{x} = \underline{x}$ para alguma inversa espectral \underline{B}^S , é um autovetor da matriz $\underline{C} = \underline{J}^+ \underline{P} \underline{A} \underline{P}^{-1}$, sendo λ seu correspondente autovalor. Além disso, qualquer autovetor $\underline{y} = \underline{P}\underline{x}$ é uma solução de (3), sendo λ autovalor relativo, se verificarmos que $\underline{B} \underline{B}^S (\underline{A}\underline{x}) = \underline{A}\underline{x}$ para alguma SEI \underline{B}^S .

Prova: Pré-multiplicando (3) por \underline{J}^+ temos $\underline{J}^+ \underline{P} \underline{A} \underline{P}^{-1} \underline{y} = \lambda \underline{J}^+ \underline{J} \underline{y}$. Usando $\underline{B} = \underline{P}^{-1} \underline{J} \underline{P}$ e $\underline{B}^S = \underline{P}^{-1} \underline{J}^+ \underline{P}$ obtemos

$$\underline{J}^+ \underline{P} \underline{A} \underline{P}^{-1} \underline{y} = \lambda \underline{P} \underline{B}^S \underline{B} (\underline{P}^{-1} \underline{y}) = \lambda \underline{y} \quad (4)$$

pela condição $B^S Bx = x$. Para estabelecer a segunda parte, pré-multiplicamos a $J^+ P A P^{-1} y = \lambda y$ por J . Substituindo J e J^+ obtemos:

$$P B B^S A P^{-1} y = \lambda P B P^{-1} y \quad (5)$$

ou

$$P B B^S A x = \lambda P B P^{-1} y \quad (6)$$

Aplicando a condição no teorema, temos o resultado desejado.

Apêndice 3 - Distribuição de Frequência de Renda (Y) por Escolaridade
Idade e Atividade (E×I×A)

Distribuição de Frequência

E	I	A	Y1	Y2	Y3	Y4	Y5	Total
1	1	1	34	237	8	2	0	281
1	1	2	174	295	6	0	0	475
1	1	3	752	1917	120	16	8	2813
1	1	4	221	212	0	0	0	433
1	1	5	545	2081	197	0	0	2823
1	1	6	465	1052	112	12	0	1641
1	2	1	8	170	134	14	0	326
1	2	2	18	361	84	31	0	494
1	2	3	40	1207	647	94	144	2132
1	2	4	33	338	55	2	0	428
1	2	5	13	2289	929	106	37	3374
1	2	6	5	957	320	74	31	1387
1	3	1	0	242	349	36	22	649
1	3	2	29	410	104	15	20	578
1	3	3	10	1293	1345	262	94	3004
1	3	4	22	403	159	14	7	605
1	3	5	23	2836	1676	122	113	4770
1	3	6	28	930	739	115	50	1862
1	4	1	4	358	403	87	39	891
1	4	2	41	350	70	12	24	497
1	4	3	41	1818	1406	325	170	3760
1	4	4	29	492	326	35	21	903
1	4	5	20	3314	1889	388	238	5849
1	4	6	26	1173	770	117	110	2196
2	1	1	69	1236	210	4	2	1521
2	1	2	614	1014	115	5	4	1752
2	1	3	2111	10706	1510	112	30	14469
2	1	4	584	899	190	0	0	1673
2	1	5	1783	13838	1507	91	22	17241
2	1	6	786	3846	504	39	16	5191
2	2	1	0	567	1010	126	25	1728
2	2	2	44	826	580	97	61	1608
2	2	3	38	5055	6905	1541	442	13981
2	2	4	31	824	488	86	23	1452
2	2	5	30	9959	7756	781	240	18766
2	2	6	108	2179	2229	443	143	5102

Apêndice 3 - (continuação)

<i>E</i>	<i>I</i>	<i>A</i>	Y1	Y2	Y3	Y4	Y5	Total
2	3	1	0	356	1230	417	88	2091
2	3	2	12	246	517	106	76	957
2	3	3	38	3485	6296	2384	1283	13486
2	3	4	8	722	805	145	111	1791
2	3	5	56	5322	9538	2751	913	18580
2	3	6	32	1265	2000	731	399	4427
2	4	1	0	261	612	362	165	1400
2	4	2	14	372	285	102	97	870
2	4	3	79	2906	4363	1960	1118	10426
2	4	4	23	627	478	111	87	1326
2	4	5	5	3629	6014	1770	1058	12476
2	4	6	38	1483	1532	594	322	3969
3	1	1	1	256	155	12	0	424
3	1	2	0	62	17	1	1	81
3	1	3	55	1068	640	197	4	1964
3	1	4	0	40	65	0	0	105
3	1	5	26	1841	1180	168	9	3224
3	1	6	25	379	159	33	3	599
3	2	1	0	45	344	276	87	752
3	2	2	0	9	50	32	36	127
3	2	3	11	373	1781	1325	705	4195
3	2	4	0	21	99	92	65	277
3	2	5	1	506	3398	1456	643	6004
3	2	6	0	146	302	304	136	888
3	3	1	1	6	156	217	232	612
3	3	2	0	18	31	13	15	77
3	3	3	0	109	808	1091	1315	3323
3	3	4	0	6	82	114	77	279
3	3	5	12	127	1661	1483	1625	4908
3	3	6	2	34	203	249	303	791
3	4	1	0	4	37	103	174	318
3	4	2	0	16	8	15	8	47
3	4	3	0	76	339	460	767	1642
3	4	4	0	26	62	36	63	187
3	4	5	1	98	634	583	885	2201
3	4	6	0	23	99	68	164	354

B I B L I O G R A F I A

- ANDERSON, T.W. 1958. *An introduction to multivariate statistical analysis*. New York, John Wiley. 374p. (Wiley Series in Probability and Mathematical Statistics)
- BISHOP, Y.M.M.; FIENBERG, S.E.; HOLLAND, P.W. 1975. *Discrete multivariate analysis: theory and practice*. Cambridge, MIT Press. 557p.
- BALESTRA, P. 1976. *La dérivation matricielle*. Paris, Sirey.
- BOULLION, T.L. & ODELL, P.L. 1971. *Generalized inverse matrices*. New York, Wiley-Interscience. 103p.
- BROCKETT, P.L. & LEVINE, A. 1977. On a characterization of ridents. *Annals of Statistics*, 5(6):1245-1248.
- BROSS, I.D.J. 1958. How to use rident analysis. *Biometrics*, 14(1):18-38.
- COHEN, J. & COHEN, P. 1975. *Applied multiple regression: correlation analysis for the behavioral sciences*. Hillsdale, Lawrence Erlbaum. 490p.

- DARROCH, J.N. 1974. Multiplicative and additive interaction in contingency tables. *Biometrika*, 61(2):207-214.
- DENNIS, T.B.; PORE, M.D.; TERRELL, G.R. 1980. Some properties of the riddit transformation. In: MEETING OF THE AMERICAN STATISTICAL ASSOCIATION, Houston, 1980. *Proceedings of the Statistical Computing Section*. Washington, ASA. p.303-308.
- DRAPER, N.R. & SMITH, H. 1966. *Applied regression analysis*. New York, John Wiley. 407p. (Wiley Series in Probability and Mathematical Statistics)
- FRIEDMAN, M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675-701
- X GIFI, A. 1980. *Niet-lineaire multivariate analyse*. Leyden, University of Leyden - Department of Data Theory.
- GREEN, P.E. & CARROLL, J.D. 1978. *Analyzing multivariate data*. Hinsdale, Dryden Press. 519p.
- X ISRAËLS, A.Z. et alii. 1981. *Multivariate analysis methods for discrete variables*. Netherlands, Centraal Bureau voor de Statistiek. (Department for Statistical Methods, BPA:3374-81-M1)
- HILL, M.O. 1974. Correspondence analysis: a neglected multivariate method. *Applied Statistics*, 23(3):340-354.
- X JANSEN, M. 1980. *Ridik analysis*. Netherlands, Centraal Bureau voor de Statistiek. 13p. (Department for Statistical Methods BPA:1059-80-M1)
- X KELLER, W.J. & WANSBEEK, T. 1981. *Multivariate methods for quantitative and qualitative data*. Netherlands, Centraal Bureau voor de Statistiek. 18p. (Department for Statistical Methods, 4834-81-M1)

- KRUSKAL, W.B. & WALLIS, W.A. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583-621.
- LEEUW, J. de. 1982. Generalized eigenvalue problems with positive semi-definite matrices. *Psychometrika*, 47(1):87-93.
- LEHMANN, E.L. 1975. *Nonparametrics: statistical methods based on ranks*. San Francisco, Holden-Day. 457p. (Holden-Day Series in Probability and Statistics)
- LYNCH, G.W. 1978. A decision theoretic approach to rидits. *Communications in Statistics, Theory and Methods*, A7(6):607-614
- MANN, H.B. & WHITNEY, D.R. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18(1):50-60.
- MANTEL, N. 1979. Rидit analysis and related ranking procedures—use at your own risk. *American Journal of Epidemiology*, 109(1):25-29.
- MCDONALD, R.P.; TORII, Y.; NISHISATO, S. 1979. Some results on proper eigenvalues and eigenvectors with applications to scaling. *Psychometrika*, 44(2):211-227.
- MORRISON, D.F. 1976. *Multivariate statistical methods*. 2.ed. New York, McGraw-Hill. 415p. (McGraw-Hill Series in Probability and Statistics)
- PESSOA, D. 1977. *Estatística não-paramétrica*. Rio de Janeiro, IMPA, 195p. Trab. apres. ao 11º Colóquio Brasileiro de Matemática, Poços de Caldas, 1977.
- SELVIN, S. 1977. A further note on the interpretation of Rидit analysis. *American Journal of Epidemiology*, 105(1):16-20.

- X SIKKEL, D. 1980. *An application of correspondence analysis to leisure time activities*. Netherlands, Centraal Bureau voor de Statistiek - Department of Statistical Methods. 40p. (Statistical Studies, 25)
- SIKKEL, D. 1981. *The relationship between canonical correlations and correspondence analysis*. Netherlands, Centraal Bureau voor de Statistiek. 12p. (Department for Statistical Methods, BPA:8696-81-M1)
- TIMM, N.H. 1975. *Multivariate analysis with applications in Education and Psychology*. Monterey, Brooks-Cole. 689p.
- VETTER, W.J. 1970. Derivative operations on matrices. *IEEE Transactions on Automatic Control*, AC-15(2):241-244.
- WILKS, S.S. 1962. *Mathematical statistics*. New York, John Wiley. 644p.
- YOUNG, F.W. 1981. Quantitative analysis of qualitative data. *Psychometrika*, 46(4):357-388.
- YOUNG, F.W.; LEEUW, J. de.; TAKANE, Y. 1976. Regression with qualitative and quantitative variables: an alternating least squares methods with optimal scaling features. *Psychometrika*, 41(4):505-529.