**Análise de Preferência
” Conjoint Analysis”**

Rinaldo Artes

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. Wilton Oliveira Bussab

– SÃO PAULO, 1991 –

ERRATA

Substituir a tabela 1.1, página 3, pela tabela abaixo

estímulo	Winchester	RAM	MODEM	Avaliação
1	60	2	não	8
2	60	2	sim	6
3	60	4	não	7
4	60	4	sim	5
5	120	2	não	3
6	120	2	sim	4
7	120	4	não	1
8	120	4	sim	2

Na tabela 4.9, página 53, substituir o valor **1,750**, correspondente à coluna $\beta^{(0)}$ da linha RAM, por **0,750**

Substituir as colunas r e $y_i^{(0)}$ da tabela 4.10, página 53, pelas colunas abaixo

r	$y(\beta_{pd}^{(0)})$
3,5	2,416
1,5	2,013
2,5	1,208
0,5	0,805
-1,5	-0,805
-0,5	-1,208
-3,5	-2,013
-2,5	-2,416

Página	linha	onde se lê	leia-se
3	10	$\mathbf{X}\beta_s + \epsilon$	$\mathbf{X}\beta_s$
14	4	entrevistados	entrevistado
14	20	preferido	estímulo preferido em cada par
19	5	inconveniência	inconveniência
21	3	intrínsecos	intrínsecos
48	19	β_i^*	β_i
49	5	$\sum_{i=1}^n$	$\sum_{j=1}^n$
49	10	$\sum_i$	$\sum_j$
52	25	$y_i^{(0)}$	$y(\beta_{pd}^{(0)})$
54	1	$y_{pd}^{(1)}$	$y(\beta_{pd}^{(1)})$
54	8	0.5%	0.1%
55	18	postos r ,	postos,
55	19	atribu'	atribuídos
55	20	idos	
56	16	impõe-se seguinte	impõe-se a seguinte
58	penúltima	sempre existirá	sob certas condições prova-se a existência
60	7	$\mathbf{R}$	$\mathbf{r}$
60	8	R	r
63	2	(1973a	(1973a)
69	7	vetor	utilidade
72	1	$\sum_{j=1}^q \sum_{l=1}^{n_j} r_s(\mathbf{x}_l^j) \frac{1}{n}$	$\sum_{j=1}^q \sum_{l=1}^{n_j} r_{jl} P(r_s(\mathbf{x}_l^j) = r_{jl})$
72	1	$\frac{1}{n} \sum_{j=1}^q \sum_{l=1}^{n_j} r_s(\mathbf{x}_l^j)$	$\frac{1}{n} \sum_{j=1}^q \sum_{l=1}^{n_j} r_{jl}$
72	4	$r_s(\mathbf{x}_l^j)$	r_{jl}
72	11 e 13	$\frac{1}{n(n+1)}$	$\frac{1}{n(n-1)}$
74	6	r_{sj}	r_{js}
75	7	r_{sj}	r_{js}
78	10	desejava-se	deseja-se
79	9 e 10	,isto é, $r_j \in W_k$	.
86	10	)	)'
87	9	=	$dt =$
87	10	$\int_{+\infty}^0$	$\int_{-\infty}^0$
89	4	$\frac{\epsilon^{x_1'^\beta}}{\sum_{l \in C_1} \epsilon^{x_l'^\beta}} =$	$\frac{\epsilon^{x_1'^\beta}}{\sum_{l \in C_1} \epsilon^{x_l'^\beta}} x_1' =$
90	1	acrescentar: ver definição de r_{jls} abaixo.	
92	18	assintoticamente	assintoticamente
97	14	6/9 x 3/9	(6 x 3)/9
104		ignorar a última linha da tabela 6.4	
105	2	0.623	0.627
105	11 e 12	ignorar a partir de "Calculou-se também ..."	
120	9	afeta	afetou

Página	linha	onde se lê	leia-se
131	última	de outro	dos outros
135	antepenúltima	referir ao	referir tanto ao
142	13	$\frac{\sum_i g_{ik}^2}{\sqrt{\sum_i \beta_i^{(k-1)^2}}}$	$\sqrt{\frac{\sum_i g_{ik}^2}{\sum_i \beta_i^{(k-1)^2}}}$
143	10	valor	método
145	4	$\frac{\partial y_j}{\partial \beta_i}$	$\frac{\partial y_j(\beta)}{\partial \beta_i}$
149	3	acrescentar ao final da linha: desde que exista pelo menos um $A_i, i \in P$, tal que $A_i > 0$. ou que pelo menos um $\beta_i, i \in P$, esteja livre de sinal (não esteja incluso na restrição 4.3.10)	
149	4	acrescentar: a) Considere que β_p não tenha sido incluído em 4.3.10.)	
149	9	$\frac{h}{ A_p }$	$\frac{h}{A_p}$
149	10	$-\frac{ha_{jk_1}}{ A_i }$	$-\frac{ha_{jk_1}}{A_i}$
149	12	acrescentar: b) Suponha $A_p > 0$ e tome a solução acima	
152	5	D_p	D_i
153	6	Snivasan	Srinivasan
154	4	e.	.
161	última	$2\epsilon'X(X'X)^{-1}X'\beta^*$	$2\epsilon'X(X'X)X'\beta^* + 2\beta'X(X'X)\beta^*$
162	2	$\beta^{*'}X'X\beta^*$	$\beta^{*'}X'X\beta^* - 2\beta'X'X\beta^*$
162	3	$)\epsilon'X(X'X)$	$) + \epsilon'X(X'X)^{-1}$

À Amélia

e aos

meus pais

AGRADECIMENTOS

Não poderia deixar de registrar meus agradecimentos às pessoas que participaram direta, ou indiretamente, da elaboração desta monografia, em especial,

ao Prof. Dr. **Wilton Oliveira Bussab**, pela paciência e atenção com que sempre me atendeu e pelo privilégio que me concedeu ao concordar em me orientar neste trabalho;

ao Prof. Dr. **Adolpho Walter P. Canton**, pelo incentivo dado durante o desenvolvimento deste tema;

à Profa. **Lisbeth K. Cordani**, pela minuciosa leitura dos originais e pelas sugestões dadas;

ao Prof. **Dario Nery**, à Profa. **Márcia D. Branco** e aos demais colegas do curso de mestrado, pelo importante apoio e trocas de experiências durante toda a etapa de elaboração deste trabalho.

Resumo

Esta monografia discute as etapas de uma análise de preferência ("conjoint analysis"), enfatizando o estudo dos métodos de estimação. São apresentados os métodos: PREFMAP, MONANOVA, LINMAP, PROBIT, PROBIT ordinal, LOGIT, LOGIT condicional, híbrido e bayesiano.

Inicialmente, são discutidos aspectos ligados ao planejamento de uma análise de preferência, tais como: determinação dos atributos relevantes, escolha da forma de avaliação, planejamento perfil completo e planejamento trade-off.

A seguir, são introduzidos os modelos usuais de preferência: vetor, ponto-ideal (simples, ponderado e geral) e utilidade.

Após o estudo dos mencionados métodos de estimação, são apresentados estudos empíricos extraídos da literatura, relacionados á análise de preferência. Complementa-se com um estudo original de comparação empírica dos métodos MONANOVA, LINMAP, LOGIT condicional, híbrido, bayesiano e mínimos quadrados.

Abstract

This monograph discuss the steps of a conjoint analysis, specially those related to the estimation methods. It is presented the methods: PREFMAP, MONANOVA, LINMAP, PROBIT, ordinal PROBIT, LOGIT, conditional LOGIT, hybrid and bayesian.

Aspects related to the design of a conjoint analysis are treated in the beginning, in special: relevants attributes determination, selection of the response variable, full profile and trade-off designs.

The vector, (simple, weighted and general) ideal-point and utility preference models are introduced.

After the presentation of the mentioned preference models, empirical studies papers related to conjoint analysis are discussed. Finally, an original empirical comparison of the estimation methods: MONANOVA, LINMAP, conditional LOGIT, hybrid, bayesian and least squares is reported.

Apresentação

Esta dissertação tem por objetivo apresentar as etapas da análise de preferência ("conjoint analysis"), ou seja, a análise de experimentos com variáveis respostas que expressam preferências individuais.

Sem a pretensão de ser um trabalho definitivo, procurou-se comparar alguns métodos de estimação com o intuito de sugerir guias de orientação para a adequada utilização de cada um deles. A ênfase maior do trabalho recai sobre o estudo de variáveis respostas que não seguem uma escala intervalar.

Não se adotou a tradução literal do termo em inglês (análise conjunta) para evitar confusões quanto ao teor deste trabalho, já que o termo "conjunta" é utilizado em outras áreas da estatística. No capítulo 1 são explicitadas as razões que levaram à adoção do termo em inglês "conjoint analysis". Nesse mesmo capítulo, é feita uma introdução às técnicas de análise de preferência, um breve histórico e uma formulação matemática para o problema.

No capítulo 2 são discutidos aspectos ligados ao planejamento e coleta de dados.

São introduzidos, no capítulo 3, os modelos clássicos de preferência (vetor, ponto-ideal e utilidade).

Os capítulos 4, 5 e 6 referem-se ao estudo de diversos métodos de estimação: PREFMAP, MONANOVA e LINMAP no 4; PROBIT e LOGIT no 5 e os métodos híbrido e bayesiano no 6.

No capítulo 7 resume-se diversos estudos empíricos de comparação extraídos da literatura e apresenta-se um estudo elaborado para esta dissertação.

Além desses 7 capítulos, utilizou-se uma série de apêndices, alguns deles para complementar aspectos teóricos ou aplicados de algum capítulo, ou mesmo apresentar pré-requisitos necessários à compreensão de certos temas. Nesta linha, destacam-se: o apêndice 1, onde se revê a classificação de variáveis segundo a escala de medida; o apêndice 2, onde é apresentado o problema de confundimento em planejamentos fracionários; o 3, onde se faz uma recapitulação da regressão isotônica e o 8, onde se desenvolve o estimador James-Stein.

Estabeleceu-se as seguintes convenções:

- expressões matemáticas em **negrito** referem-se a vetores ou matrizes;
- as tabelas foram numeradas no formato XX.YY, onde XX refere-se ao número do capítulo e YY ao número da tabela no capítulo; assim a tabela 2.1 é a primeira do capítulo 2, já a tabela A3.2 é a segunda do apêndice 3;
- para a numeração das fórmulas utilizou-se o formato XX.YY.ZZ, onde XX é o número do capítulo, YY o número da subseção (quando existir) e ZZ é o número da fórmula; assim a fórmula A4.2.3 é a terceira da segunda seção do apêndice 4.

ÍNDICE

1. Análise de Preferência	1
1.1. Introdução	1
1.2. Uma formulação matemática de AP e notação	2
1.3. Exemplo	3
1.4. Histórico	5
1.5. Aplicações	7
1.6. Etapas de uma AP	8
1.7. Proposta da dissertação	9
2. Coleta de Dados	10
2.1. Escolha dos atributos	10
2.2. Variável resposta	12
2.2.1. Ordenamento através de postos	13
2.2.2. Avaliação através de escalas	14
2.2.3. Comparação pareada	14
2.2.4. Outros tipos de avaliação	14
2.3. Apresentação dos estímulos	15
2.4. Planejamento para perfil completo	19
3. Modelos de Preferência	21
3.1. Modelo de preferência VETOR	21
3.2. Modelos PONTO-IDEAL	23
3.2.1. Modelo PONTO-IDEAL SIMPLES	23
3.2.2. Modelo PONTO-IDEAL PONDERADO	25
3.2.3. Modelo PONTO-IDEAL GERAL	26
3.2.4. Comentários sobre os modelos ponto-ideal	28
3.3. Modelo UTILIDADE	29
3.4. Hierarquia dos modelos	30
4. Métodos de Estimção	32
4.1. PREFMAP - PREference MAPping	32
4.1.1. Aplicação ao modelo vetor	33
4.1.2. Aplicação aos modelos ponto-ideal	34
4.1.2.1. Modelo ponto-ideal simples	34
4.1.2.2. Modelo ponto-ideal ponderado	35
4.1.2.3. Modelo ponto-ideal geral	36
4.1.3. Exemplo	40
4.1.4. Avaliação em escala de medida não métrica	45
4.2. MONANOVA - MONotonic ANalysis Of VAriance	48
4.2.1. Exemplo	52

4.3. LINMAP - LINear MAPping	55
4.3.1. Aplicado ao modelo vetor (utilidade)	56
4.3.1.1. Exemplo	59
4.3.2. Aplicado ao modelo ponto-ideal ponderado	60
4.3.2.1. Exemplo	63
4.3.3. Aplicado ao modelo ponto-ideal geral	64
4.3.3.1. Exemplo	65
4.3.4. Comentários	68
4.4. Significância dos atributos	70
4.4.1. Procedimento 1 - Para atributos que não sofrem restrições de ordem a priori	71
4.4.2. Procedimento 2 - Para atributos que sofrem restrições de ordem a priori	75
5. Modelos Para Dados Categorizados	77
5.1. Variável resposta dicotômica	77
5.2. Variável resposta ordinal	78
5.2.1. Exemplo	82
5.3. LOGIT condicional	84
5.3.1. Estimação a nível amostral	85
5.3.2. Estimação de modelos individuais	90
5.3.3. Exemplo	90
5.4. Comentários	92
6. Métodos Especiais	93
6.1. Método híbrido	93
6.1.1. O Modelo híbrido	96
6.1.1.1. Fase auto-explicativa	96
6.1.1.2. Fase conjunta	96
6.1.1.3. O Modelo	97
6.1.2. Modificações no processo	98
6.1.2.1. Nível de agregação	98
6.1.2.2. Avaliação	99
6.1.3. Estimação dos parâmetros	101
6.1.4. Exemplo	102
6.1.4.1. Importância dos atributos	105
6.2. Método bayesiano	106
6.2.1. Obtenção de	107
6.2.2. Comentário sobre a utilização do modelo bayesiano	109
7. Estudos Empíricos	110
7.1. Resultados extraídos da literatura	110
7.1.1. Coleta de dados	111
7.1.2. Má especificação do modelo	113
7.1.3. Comparação dos métodos de estimação	115

7.2. Estudo Empírico	117
7.2.1. Análise preliminar	120
7.2.2. Comparação da porcentagem de acerto da previsão	123
7.2.3. Comparação dos coeficientes de correlação de Spearman	127
7.3. Comentários	131
7.4. Conclusões e futuros estudos	132

APÊNDICES

1. Níveis de Mensuração	133
2. Planejamentos Fracionários	135
3. Regressão Isotônica	137
4. Método "Steepest Descent" para MONANOVA	140
5. Teoremas Relativos ao Método LINMAP	147
6. LINMAP - Modelos Ponto-Ideal	151
7. Estimador James-Stein	155
8. Resumo dos Estudos Empíricos da Seção 7.1.	163
9. Método de JOHNSON	170
10. Estudo Empírico	172
11. Tabelas Referentes ao Capítulo 7	178

BIBLIOGRAFIA	182
---------------------------	------------

CAPÍTULO 1

Análise de Preferência - "Conjoint Analysis"

1.1. Introdução

Entende-se por análise de preferência ("conjoint analysis"), AP, processos que permitam obter e analisar, através da estimação de modelos, experimentos cujas variáveis respostas expressam preferências individuais.

Deseja-se, em geral, estimar a importância de fatores pré estabelecidos na formação da preferência do entrevistado por diferentes versões de um produto, serviço ou situação hipotética. Considere, por exemplo, um caso onde se quer conhecer a importância dos fatores: capacidade do winchester (60 ou 120 Mb), memória RAM (2 ou 4 Mb) e existência de MODEM (sim ou não), no processo de escolha de um micro computador pessoal. As técnicas de **AP** exigem que diferentes formulações do produto sejam estabelecidas através de combinações dos níveis dos fatores de interesse (estímulos ou perfis), para posterior avaliação dos entrevistados (juízes).

No exemplo acima, capacidade do winchester, memória RAM e MODEM são os atributos do produto e, por exemplo, um micro com 60 Mb de winchester, 2 Mb de memória

RAM e presença de MODEM, é um possível estímulo (perfil) a ser avaliado.

Em geral, a estimação de modelos dá-se a nível individual, ou seja, para cada elemento da amostra (juiz) haverá a estimação de um modelo.

1.2. Uma formulação matemática de AP e notação

Seja p o número de fatores (atributos) envolvidos no experimento, cada um com l_k níveis. Considerados todos os atributos, seriam possíveis $\prod_{k=1}^p l_k$ combinações de níveis. Cada combinação recebe o nome de estímulo (perfil). Convém ressaltar que, nesta monografia, será denominado estímulo qualquer combinação de níveis de 2 ou mais atributos, não sendo necessário que níveis de todos os fatores envolvidos estejam presentes em todos os perfis¹. Cada indivíduo será submetido a n_s , $s = 1, 2, \dots, S$, estímulos.

Considere $\mathbf{X}_s^* = (\mathbf{x}_1^* \ \mathbf{x}_2^* \dots \mathbf{x}_{n_s}^*)' = (x_{jks}^*)$, matriz de dimensão n_s por p , onde x_{jks}^* é o nível do atributo k , presente no estímulo j avaliado pelo indivíduo s ; em outras palavras, cada linha de $\mathbf{X}_s^*$ representa um estímulo. Em geral, $\mathbf{X}_s^* = \mathbf{X}^*$ para $s = 1, 2, \dots, S$, ou seja, todos os indivíduos recebem os mesmos estímulos.

Considere também, $\mathbf{r}_s = (r_{js})$ de dimensão n_s por 1, onde r_{js} é a avaliação do estímulo j realizada pelo indivíduo s ; em outras palavras, expressa a preferência do indivíduo s pelo perfil j .

O objetivo de uma AP é, a partir de $\mathbf{r}_s$, $\mathbf{X}_s^*$ e de um modelo $\mathbf{f}_s$, estimar os parâmetros do modelo e encontrar uma variável latente $\mathbf{y}_s$, função de $\mathbf{r}_s$, $\mathbf{y}_s = (y_{js}) = \mathbf{g}_s(\mathbf{r}_s)$ (vetor de dimensões n_s por 1) sendo $\mathbf{y}_s = \mathbf{f}_s(\mathbf{X}_s^*) + \boldsymbol{\epsilon}_s$, onde $\boldsymbol{\epsilon}_s$ é um vetor n_s por 1 de erro aleatório do indivíduo s .

Como em geral r_{js} segue uma escala ordinal ou nominal, a função $\mathbf{g}_s$ tem a função de atenuar a rigidez da escala da variável resposta.

A função $\mathbf{f}_s$ representa a lei que rege o processo de escolha do entrevistado (modelo de preferência ou regra de composição); em geral, adota-se o modelo linear:

$$\mathbf{f}_s(\mathbf{X}_s^*) = \mathbf{X}\boldsymbol{\beta}_s \quad \text{onde,} \quad \mathbf{X} = (\mathbf{x}_1 \ \mathbf{x}_2 \dots \mathbf{x}_{n_s})' = (x_{jk})$$

¹ ver seção 2.3. para maiores detalhes

1.3. Exemplo

Volte-se ao exemplo do micro computador. A tabela 1.1 representa 8 estímulos formados a partir dos atributos winchester, memória RAM e MODEM, avaliados por um indivíduo através da atribuição de postos (de 1 a 8), sendo que quanto menor o posto, maior a preferência pelo estímulo. As colunas correspondentes aos atributos permitem formar a matrix X_s^* , enquanto que a coluna da avaliação é o vetor r_s .

TABELA 1.1. Avaliação dos estímulos do exemplo

estímulo	Winchester	RAM	MODEM	Avaliação
1	60	2	não	8
2	60	4 ²	sim	6
3	60	2 ⁴	não	7
4	60	4	sim	5
5	120	2	não	3
6	120	4 ²	sim	4
7	120	2 ⁴	não	1
8	120	4	sim	2

Inspirando-se em modelos aditivos de análise de variância, considere o seguinte modelo de preferência:

$$f_s(X_s^*) = X_s \beta_s + \epsilon_s \quad \text{onde:}$$

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_{1s} \\ \beta_{2s} \\ \beta_{3s} \end{pmatrix}$$

β_{1s} reflete a importância do nível "winchester de 60 Mb" para o sujeito s , recebendo a denominação de utilidade do nível; do mesmo modo, $-\beta_{1s}$, β_{2s} , $-\beta_{2s}$, β_{3s} e $-\beta_{3s}$ são,

respectivamente, as utilidades dos níveis "winchester de 120 Mb", "2 Mb de memória RAM", "4 Mb de memória RAM", "ausência de MODEM" e "presença de MODEM". Nesse contexto, $\beta_{1s} + \beta_{2s} - \beta_{3s}$ reflete a preferência do juiz s pelo estímulo: (60Mb de winchester, RAM de 2Mb, existência de MODEM), sendo chamado de utilidade total desse estímulo. Nesse contexto, denomina-se utilidade de um nível à sua contribuição na formação da utilidade total dos estímulos.

Utilizando o método de estimação MONANOVA², chega-se que:

$$y_s = g_s(r_s) = (1,737 \quad 1,732 \quad 1,732 \quad 1,727 \quad -1,728 \quad -1,728 \quad -1,736 \quad -1,736)'$$

$$\hat{\beta}_s = (1,732 \quad 0,004 \quad 0,001)'$$

Assim, ao nível inexistência de MODEM associa-se o valor 0,001; já ao nível existência, associa-se -0,001. Diz-se que 0,001 é a utilidade do nível inexistência de MODEM e -0,001 é a utilidade de existência.

Ao analisar o vetor $\hat{\beta}_s$, percebe-se que o fator mais importante é a capacidade do winchester, seguido da memória RAM e MODEM.

A preferência por um micro que possui 60 Mb de winchester, memória RAM de 2 Mb e MODEM (micro A) é dada, para o modelo de preferência adotado, pela soma das utilidades de cada um dos níveis acima, ou seja:

$$y(60\text{Mb}, 2\text{Mb}, \text{MODEM}) = 1,732 + 0,004 - 0,001 = 1,735$$

O valor 1,735 é a utilidade total do estímulo.

A utilidade total de um micro que possui as características: 60 Mb de winchester, 4 Mb de memória RAM sem MODEM (micro B) é 1,729; de onde se infere que entre os micros A e B, o preferido pelo entrevistado seria o B (lembre que a atribuição de postos foi realizada em ordem decrescente de preferência; assim, quanto menor o valor numérico da utilidade, maior a preferência pelo estímulo).

² ver capítulo 4.2

Ilustra-se desse modo, um dos pontos positivos da **AP**, que é a possibilidade de simular a preferência dos entrevistados (juízes) por diferentes produtos.

Outra informação que pode ser extraída do modelo dado refere-se à quantificação da (o) importância (peso) relativa (o) de cada fator. Green & Tull(1978) sugerem que a medida da importância de cada atributo seja diretamente proporcional à amplitude de sua função utilidade. Desta forma:

$$\text{Atributo A: amplitude} = 1,732 - (-1,732) = 3,464$$

$$\text{Atributo B: amplitude} = 0,004 - (-0,004) = 0,008$$

$$\text{Atributo C: amplitude} = 0,001 - (-0,001) = 0,002$$

$$\text{Amplitude total} = 3,464 + 0,008 + 0,002 = 3,474$$

Padronizando esses valores de modo que a soma dos pesos relativos seja 1, chega-se a:

$$\text{Importância relativa de A} = \frac{3,464}{3,474} = 99,7\%$$

$$\text{Importância relativa de B} = \frac{0,008}{3,474} = 0,2\%$$

$$\text{Importância relativa de C} = \frac{0,002}{3,474} = 0,1\%$$

1.4. Histórico

A origem da **AP** está ligada ao artigo de Luce & Tukey(1964). Nesse trabalho, estudou-se o efeito conjunto de dois ou mais atributos na ordenação de estímulos. Basicamente, estabeleceu-se as condições para a existência de escalas intervalares, associadas aos níveis dos fatores³, de modo que, quando combinadas através de um modelo de preferência, fossem reproduzidos, da maneira mais fiel possível, os postos originais. Essa teoria foi denominada medida conjunta ("conjoint measurement"⁴).

Outros pesquisadores seguiram estudando aspectos teóricos da medida conjunta (Krantz(1964), Tversky(1967), Barron(1977)).

³ na verdade, quanto maior o número de níveis, mais próximo de uma escala intervalar se estará

⁴ esse termo derivou o título em inglês "conjoint analysis"

Em 1965, Kruskal desenvolve a técnica MONANOVA (Kruskal(1965)), que nada mais é do que uma transformação para eliminar o efeito de interação em dados provenientes de planejamentos fatoriais. Essa técnica foi adotada como algoritmo de estimação de modelos de preferência. Durante a década de 60, outros algoritmos foram desenvolvidos (Carroll(1969), Young(1969)).

Em 1971, Green & Rao publicam o primeiro artigo sobre medida conjunta aplicado a "marketing" (Green & Rao(1971)). Esse artigo enfatiza os aspectos relativos à estimação de um modelo de preferência.

Observou-se, na prática, que psicometristas e pesquisadores de "marketing" simplesmente assumiam que as condições para existência das escalas intervalares eram válidas e se preocupavam apenas com aspectos numéricos do problema (Green & Srinivasan(1978)). Assim, fez-se necessário distinguir as duas posturas: aos trabalhos desenvolvidos na linha de Luce & Tukey, deu-se o nome de medida conjunta axiomática e aos outros de medida conjunta numérica.

A partir do artigo de Green & Rao, cresce o interesse pelo tema. As técnicas de estimação PREFMAP (Carroll(1972)) e LINMAP (Srinivasan & Shocker(1973 a e b)) e o método de coleta de dados "trade-off" (Jonhson(1974)), são dessa época.

O ano de 1978 foi bastante importante para a divulgação da medida conjunta, com a publicação da quarta edição do livro Green & Tull(1978), onde foi dedicada metade de um capítulo ao tema, e do artigo Green & Srinivasan(1978), onde se fez uma excelente recapitulação de todo o desenvolvimento da medida conjunta, até aquela data. Nesse artigo, os autores afirmam: "*... parece útil a adoção do nome "conjoint analysis" [AP], para cobrir modelos e técnicas que enfatizem a transformação de respostas subjetivas em parâmetros estimados*".

Nesse espírito mais abrangente (o escolhido no desenvolvimento dessa dissertação), técnicas como LOGIT e PROBIT (Aldrich & Nelson(1985) e McFadden(1974)) foram incluídas em AP.

Uma série de pesquisadores, sobretudo os ligados à área de "marketing", não se interessando pela linha de pesquisa original (medida conjunta axiomática), preocuparam-se somente com a estimação de modelos. Todavia, ainda é possível encontrar artigos recentes sobre medida conjunta axiomática (Michel(1988)).

Na década de 80, novas técnicas são sugeridas, como os métodos híbrido (Green, Goldberg & Montemayor(1981)) e bayesiano (Cattin, Gelfand & Danes (1983)).

Mais recentemente, percebe-se que a tendência de desenvolvimento da técnica é a criação de programas computacionais que permitam a coleta de dados e a análise de uma AP. Em alguns desses programas, o entrevistado interage com a máquina, que solicita os dados necessários para a estimação de modelos; entre eles cita-se o "Adaptive Conjoint Analysis" (Johnson(1987)), que foi analisado em Green, Krieger & Agarwal(1991).

Técnicas baseadas na teoria da integração de informação ("information integration theory"), Anderson(1970) e Louviere(1988)) não fazem parte do âmbito desta dissertação.

A popularidade de técnicas de AP foi constatada por Cattin & Wittink(1982) e Wittink & Cattin(1989) em levantamentos realizados em empresas de pesquisa de mercado sobre a utilização desses procedimentos.

1.5. Aplicações

Nesta seção, algumas possíveis aplicações de AP serão comentadas. Não se deseja aqui esgotar o tema, quer-se somente chamar a atenção para as potencialidades da técnica, através de alguns poucos exemplos.

Uma AP produz como resultado final uma estimativa, em geral individual, de um determinado modelo de preferência. Esse resultado, em grande parte das vezes, é utilizado em análises posteriores, onde, duas se destacam: segmentação de mercado e simulação da preferência por diferentes formulações de um produto (Green & Tull(1978), Green & Srinivasan(1978), Louviere(1988), Kamakura(1988)).

A segmentação do mercado pode ser realizada através de uma análise de agrupamentos (Bussab et al(1990)) aplicada aos parâmetros estimados, ou sobre o peso relativo dos atributos. Os grupos assim formados (segmentos) podem ser diferenciados através de outras variáveis coletadas por ocasião da pesquisa (sócio-econômicas, informações sobre marcas que utiliza, etc).

Uma das grandes vantagens de uma AP é que, através dos modelos individualmente estimados, pode-se simular a preferência dos juízes por diferentes formulações de um produto, expressas através da variação dos níveis dos atributos. Desse modo, a informação

que se tem sobre o comportamento do mercado é muito mais rica do que ao se estimar um único modelo para toda amostra. No capítulo 4 de Louvire(1988) são feitas sugestões sobre a melhor maneira de se utilizar as informações simuladas.

A seguir são apresentados alguns exemplos de como AP pode ser utilizada na prática.

- Uma nova versão de um produto será desenvolvida por uma empresa. Utilizando os modelos estimados, a empresa poderá calcular a nível individual, a utilidade total de uma série de possíveis versões para um grupo de consumidores; a partir daí, decidir quais características o produto deverá conter.
- Um produto será lançado no mercado. Através de simulações, é possível prever como será sua aceitação frente aos concorrentes.
- AP pode auxiliar na definição de uma estratégia de marketing. Suponha que um produto seja vendido no mercado por várias empresas. Consumidores de cada uma dessas marcas podem ser submetidos a uma AP, permitindo, desse modo, verificar a importância que os consumidores de cada marca dão a uma série de atributos pré-estabelecidos.
- Mahajan, Green & Goldberg(1982) apresentam um modo de medir a relação entre preço e demanda de um produto através de uma AP.

1.6. Etapas de uma AP

Green & Srinivasan(1978) sistematizaram as etapas de uma AP em 6 passos:

Passo 1. Seleção de um modelo de preferência - estabelecidos os atributos do produto, deve-se determinar como esses fatores se combinam na explicação da preferência do entrevistado, ou seja, qual é o modelo f_s que deve ser estimado. No capítulo 3, são discutidos os modelos de preferência clássicos (vetor, ponto-ideal e utilidade).

Passo 2. Seleção do método de coleta de dados - define-se como os níveis dos atributos serão combinados na formação dos estímulos. Na tabela 1.1, por exemplo, cada estímulo foi formado com níveis de todos os atributos envolvidos no problema (perfil-completo); no capítulo 2 é feita uma discussão sobre esse tema.

Passo 3. Construção do conjunto de estímulos a serem avaliados - este passo está ligado

ao tipo de planejamento que será adotado na formação dos estímulos, no caso de perfil completo ⁵, pode-se ter: planejamento fatorial completo, fatorial fracionário ou blocos incompletos balanceados. Esse assunto será abordado no capítulo 2.

Passo 4. Apresentação dos estímulos aos entrevistados - os estímulos podem ser descritos verbalmente, ou até mesmo através de desenhos.

Passo 5. Escolha da escala de medida para a variável dependente - trata-se da definição de como os estímulos serão avaliados (r_s). Algumas opções são apresentadas no capítulo 2.

Passo 6. Escolha do método de estimação - nos capítulos 4, 5 e 6 são apresentados alguns possíveis métodos de estimação para um modelo de preferência.

1.7. Proposta da dissertação

Deseja-se, com essa monografia, apresentar um resumo das etapas descritas na seção anterior, de modo a formar um texto de consulta para novos usuários de AP.

Visando assegurar ao leitor uma visão ampla dos passos de uma AP, tenta-se encontrar um meio termo entre o número de temas abordados e a profundidade em que cada um é estudado. Para aqueles que desejam um detalhamento maior de alguma técnica específica, é apresentada uma vasta bibliografia.

Optou-se priorizar o estudo dos métodos de estimação de modelos de preferência. Comparações entre esses métodos e também entre outros aspectos, ligados ao planejamento de uma AP, são encontradas em abundância na literatura. São apresentadas, no capítulo 7, as conclusões de 13 estudos empíricos encontrados em publicações científicas. Um estudo original, visando comparar alguns métodos de estimação, também é apresentado nesse capítulo.

⁵ onde os estímulos são formados através da variação de níveis de todos os atributos

CAPÍTULO 2

Coleta de Dados

Neste capítulo serão discutidos aspectos ligados ao planejamento de uma análise de preferência. Serão feitos alguns comentários sobre as variáveis respostas mais utilizadas, a apresentação dos estímulos e alguns tópicos sobre o planejamento propriamente dito.

2.1. Escolha dos atributos

Considere o seguinte exemplo: uma loja de computadores gostaria de saber a importância que os seus consumidores atribuem a diversas características de um micro computador, no instante da compra.

A fim de planejar um questionário que permita a utilização de uma técnica de análise de preferência, deve-se escolher fatores inerentes ao produto que interfiram na preferência do consumidor. A união desses fatores definirá tipos hipotéticos de micro computadores. Nesse exemplo são utilizados cinco atributos: capacidade do winchester (40, 60 ou 120 Mb), memória RAM (1, 2 ou 4Mb), clock (8, 16 ou 20 MHz), existência de MOUSE (sim ou não) e existência de MODEM (sim ou não).

Ao escolher os atributos deve-se ter o cuidado de selecionar os que realmente sejam importantes, pois, como será visto no decorrer deste capítulo, um número alto de fatores pode dificultar a avaliação do entrevistado, prejudicando a credibilidade dos resultados.

Louviere(1988) apresenta um exemplo de como determinar os atributos relevantes. É proposta a realização de pesquisas qualitativas, onde uma série de produtos deve ser apresentada aos entrevistados, que responderão perguntas do tipo:

- (1) Quais produtos desta série você compra ou gostaria de comprar?
- (2) Quais produtos desta série você não compra ou não gostaria de comprar?
- (3) Você afirmou que compraria estes produtos (retomar a resposta de (1)). O que há neles que os fazem atraentes a você?
- (4) Você afirmou que não compraria estes produtos (retomar a resposta de (2)). O que há neles que não os fazem atraentes a você?
- (5) Pensando nos produtos que você gostaria de adquirir, que mudanças deveriam os fabricantes ou vendedores fazer para que você perdesse o interesse em sua aquisição?
- (6) Pensando nos produtos que você não gostaria de adquirir, que mudanças deveriam os fabricantes ou vendedores fazer que você passasse a se interessar em sua aquisição?

Após a coleta destas informações, analistas devem agrupar as respostas obtidas em categorias mutuamente exclusivas e, por exemplo, analisando a sua distribuição de frequências, adotar algum critério para a escolha dos atributos que serão utilizados em uma AP.

Alpert(1971) discute e compara alguns métodos de identificação de atributos relevantes.

Cattin & Wittink(1982), em um estudo que visava conhecer a utilização de técnicas de AP por empresas americanas de pesquisa de mercado, verificaram que, em algumas situações práticas, foram identificados até 50 fatores em potencial. Contudo, para tornar factível a aplicação de técnicas descritas nesta dissertação, este número foi reduzido para algo em torno de 3 a 15 sendo, na maioria das vezes, usados de 6 a 7 atributos.

Definidos os fatores é necessário estabelecer os seus níveis. A fim de simplificar o

planejamento, não é indicado um número muito alto de níveis. Deve-se ter a preocupação em escolher apenas os que são de importância para a situação prática. Quanto menor o número de níveis, mais simples será o planejamento. Em geral, na definição dos atributos já se conhece quais os níveis de interesse. Em todo caso, em Louviere(1988), encontra-se um procedimento para a escolha dos níveis.

Green & Wind(1975) apresentam uma análise onde os níveis de um dos atributos são representados através de desenhos. Assim, se fosse de interesse conhecer a preferência do consumidor pelo "design" de um produto, poderia ser adicionado um atributo cujos níveis seriam os possíveis projetos.

Para o exemplo foram estabelecidos os seguintes níveis:

$$\begin{array}{lll} \text{winchester} - \begin{cases} 40 \text{ Mb} \\ 60 \text{ Mb} \\ 120 \text{ Mb} \end{cases} & \text{RAM} - \begin{cases} 1 \text{ Mb} \\ 2 \text{ Mb} \\ 4 \text{ Mb} \end{cases} \\ \\ \text{clock} - \begin{cases} 8 \text{ MHz} \\ 16 \text{ MHz} \\ 20 \text{ MHz} \end{cases} & \text{MOUSE} - \begin{cases} \text{sim} \\ \text{não} \end{cases} & \text{MODEM} - \begin{cases} \text{sim} \\ \text{não} \end{cases} \end{array}$$

2.2. Variável resposta

TABELA 2.1. Tipos de avaliações utilizadas em aplicações comerciais de AP, em %, segundo o ano

Avaliação	1971-1980	1981-1985
Escala	34	49
Ordenamento	45	36
Escolha Pareada	11	9
Outras	10	6

fonte: Cattin & Wittink(1989)

Um dos itens analisados nas pesquisas de Cattin & Wittink(1982) e Wittink & Cattin(1989), foi a escolha da variável resposta, ou seja, a definição do procedimento a ser

utilizado pelos entrevistados em suas avaliações. Constatou-se que apenas três métodos foram responsáveis por mais de 90% das aplicações comerciais da técnica, no período que vai de 1971 até 1985 (tabela 2.1). São eles:

- a.) ordenamento através de postos,
- b.) avaliação através de escalas e
- c.) comparação pareada

2.2.1. Ordenamento através de postos

Esta avaliação consiste na ordenação dos estímulos segundo a preferência do juiz. A ordenação pode ser feita em ordem crescente ou decrescente sem prejuízo da análise.

Como exemplo, suponha que os seguintes estímulos sejam apresentados ao entrevistado:

- estímulo I - winchester de 40 Mb, 1 Mb de memória RAM, clock de 20 MHz, com MODEM e MOUSE;
- estímulo II - winchester de 60 Mb, 4 Mb de memória RAM, clock de 8 MHz, sem MODEM e sem MOUSE;
- estímulo III - winchester de 120 Mb, 2 Mb de memória RAM, clock de 10 MHz, sem MODEM e com MOUSE.

Suponha também, que o estímulo II seja o preferido, seguido pelo III e pelo I. Assim, haverá a seguinte atribuição de postos para os estímulos I, II e III, respectivamente, 3 1 2, se a atribuição for em ordem decrescente e 1 3 2 se for em ordem crescente de preferência.

Em geral, é fácil para o entrevistado entender o objetivo de um ordenamento via postos. Algum problema pode ocorrer quando o número de perfis a serem ordenados é muito alto. Nesses casos, deve-se contar com um previsível cansaço do entrevistado, que, a partir de um certo momento, poderá ter dificuldades na execução da tarefa.

2.2.2. Avaliação através de escalas

Os estímulos são avaliados, um a um, a partir de uma escala de preferência que pode ser discreta ou contínua. Por exemplo, ao se utilizar uma escala de 5 pontos é apresentado ao entrevistado números variando de 1 a 5; ao número 1 pode-se associar uma frase do tipo: "nenhuma preferência" e, em contrapartida, ao número 5 associar a frase: "altamente preferido". Assim, pede-se ao respondente que atribua, para cada estímulo, um número de 1 a 5, de tal modo que quanto maior for o número, maior seja a preferência pelo estímulo:

nenhuma preferência 1 2 3 4 5 altamente preferido

Uma escala contínua pode ser obtida ligando as duas frases com um traço e pedindo que o juiz assinale, nessa linha, a posição que mais concorde com sua preferência, seguindo o critério de proximidade das frases (quanto mais próximo de uma frase, maior a concordância com ela). A avaliação seria dada através da distância de uma das frases à marca do entrevistado:

nenhuma preferência —————— altamente preferido

Uma situação semelhante à anterior é obtida classificando cada perfil segundo alguma escala ordinal (exemplo: ótimo, bom, regular, ruim ou péssimo).

2.2.3. Comparação pareada

Neste método, os estímulos são apresentados aos pares e o entrevistado deve escolher o ^{estímulo} preferido ^{em cada par} (Srinivasan & Shocker(1973a), Madansky (1980)).

Um método alternativo de avaliação seria pedir que o juiz, além de escolher o estímulo preferido, indique também a intensidade da diferença entre os dois estímulos. Jackson & Shenker(1982) apresentam um exemplo onde se utiliza esse tipo de avaliação.

2.2.4. Outros tipos de avaliação

Apesar de menos utilizados, são encontrados na literatura alguns outros tipos de variáveis resposta.

Avaliações podem ser obtidas através da atribuição de notas a cada perfil, ou ainda perguntando-se ao entrevistado quanto ele pagaria por determinado produto. Este último

tipo de variável resposta segue uma escala métrica e a análise pode ser realizada através dos métodos convencionais de análise de variância ou análise de regressão.

Um outro tipo de variável resposta pode ser obtido através da classificação de cada estímulo em grupos, por exemplo: compraria/não compraria o produto, aceitaria/não aceitaria a situação descrita pelos perfis, ... (Rao & Winter(1978), Punj & Staelin(1978)).

2.3. Apresentação dos estímulos

Basicamente dois modos de se construir um estímulo são utilizadas em uma **análise de preferência**: perfil completo e matrizes "trade-off".

TABELA 2.2. Estímulos construídos através de um planejamento perfil-completo

estímulo	winchester	RAM	clock	MOUSE	MODEM
1	40	1	8	não	não
2	40	2	16	não	sim
3	40	4	20	sim	não
4	60	1	16	sim	sim
5	60	2	20	não	não
6	60	4	8	não	não
7	120	1	20	não	sim
8	120	2	8	sim	não
9	120	4	16	não	não
10	40	1	20	sim	não
11	40	2	8	não	sim
12	40	4	16	não	não
13	60	1	8	não	não
14	60	2	16	sim	não
15	60	4	20	não	sim
16	120	1	16	não	não
17	120	2	20	não	não
18	120	4	8	sim	sim

No perfil completo, os níveis de todos os atributos são combinados de modo a gerarem os estímulos. Esses estímulos são descrições do objeto de estudo (produtos, serviços,

etc). Ao entrevistado caberá avaliar estes perfis. A tabela 2.2 apresenta a formação de alguns possíveis estímulos associados ao problema proposto no início do capítulo. O número total de possíveis estímulos é dado pelo produto do número de níveis de cada atributo e será, neste caso, 108.

Neste tipo de enquête praticamente não há restrições quanto ao tipo de variável dependente a ser utilizada; em geral todas se aplicam.

Winchester RAM	40	60	120
1			
8			
16			

Winchester Clock	40	60	120
40			
60			
20			

Winchester MODEM	40	60	120
sim			
não			

MODEM MOUSE	sim	não
sim		
não		

MOUSE Winchester	sim	não
40		
60		
120		

Clock RAM	8	16	20
1			
2			
4			

MODEM RAM	sim	não
1		
2		
4		

MOUSE RAM	sim	não
1		
2		
4		

MODEM Clock	sim	não
8		
16		
20		

MOUSE Clock	sim	não
8		
16		
20		

FIGURA 2.1. Estímulos construídos através de matrizes Trade off

Ao se adotar como variável resposta o ordenamento através de postos, Green & Tull(1978) sugerem que cada estímulo seja escrito em um cartão e que, inicialmente, seja pedido ao entrevistado que os agrupe em três pilhas, a primeira com os perfis que lhe agradem, a segunda com aqueles que lhe são indiferentes e a última com aqueles que lhe desagradem. Este processo pode ser repetido até que não haja pilhas com um número excessivo de cartões. Em seguida, o respondente deverá ordenar os cartões de cada pilha segundo sua preferência. Note que, devido sua própria complexidade, este tipo de procedimento deve ser adotado preferencialmente em entrevistas pessoais.

A matriz "trade-off" (Johnson(1974)), também conhecida na literatura como avaliação dois-por-vez, difere em muito do perfil completo. Agora, os fatores serão dispostos dois a dois em tabelas e em cada uma, separadamente, as combinações de seus níveis serão ordenadas de acordo com a preferência do entrevistado (vide figura 2.1).

Um dos inconvenientes associados a este método é a possibilidade de haver uma certa perda em termos de realismo (Green & Srinivasan(1978), Johnson(1974)), primeiro porque ao se analisar apenas dois fatores por vez, estarão sendo ignoradas algumas possíveis interações envolvendo um número maior de atributos. Outra situação onde ocorre perda de realismo se dá quando existe uma relação natural forte entre dois (ou mais) fatores (correlação ambiental). Por exemplo, suponha que o objetivo de um estudo seja conhecer a preferência do consumidor por certo tipo de refrigerante. Assuma que entre os atributos de interesse estejam: o tamanho da embalagem, 290ml, 1000ml ou 1500ml e o preço do produto. É óbvio que existe uma relação natural entre os dois atributos, quanto maior a embalagem, mais caro é o produto. Espera-se, no entanto, que o consumidor prefira pagar menos por mais bebida. Ao ordenar os níveis de preço e quantidade, produtos inviáveis poderão ser os de maior preferência. A conclusão que se tira é que quando a correlação ambiental é muito alta, torna-se mais razoável a adoção de um esquema perfil completo, onde se tem um controle maior da factibilidade do perfil e onde, além disso, todos os atributos são avaliados simultaneamente (Green & Srinivasan(1978)).

Outro problema intrínseco ao método "trade-off" ocorre quando o número de tabelas a serem analisadas é muito alto. Nessa situação existe a possibilidade do entrevistado adotar um padrão de resposta, sem levar em conta os atributos envolvidos; por exemplo, ele pode começar a atribuir postos mais altos às posições de um canto e ir decrescendo conforme dele se afaste, em qualquer tabela que analise.

Foi constatada uma tendência do entrevistado considerar apenas as variações de

um atributo no ordenamento das matrizes, ignorando o segundo fator (Green & Tull(1978)). Este tipo de problema pode também acontecer com o método perfil completo. Quando o número de atributos é alto, existe a possibilidade do entrevistado ignorar as variações daqueles de menor importância.

As matrizes "trade-off", todavia, são de mais fácil aplicação, possibilitando inclusive sua utilização em pesquisas pelo correio.

Nos estudos de Cattin & Wittink(1982) e Wittink & Cattin(1989), nota-se um ligeiro aumento na popularidade do perfil completo e uma acentuada diminuição no uso das matrizes "trade off", quando se compara o período 1971-1980 com o período 1981-1985, tabela 2.3. Segundo os autores isto se deve, em parte, ao artificialismo do método "trade off".

Analisando a tabela 2.3, nota-se que alguns métodos não foram mencionados. São eles:

- **combinação** - trata-se de construir estímulos utilizando para alguns o método perfil completo e, para outros, as matrizes "trade off".
- **comparações pareadas** - a razão pela qual este tema não foi discutido é que tal procedimento consiste, simplesmente, na comparação de perfis completos dois a dois. Em termos da construção dos estímulos, será idêntico ao já descrito para perfil completo. Em David(1988) são discutidos alguns planejamentos que permitem a redução do número de pares a serem avaliados pelo entrevistado; em Jackson & Shenker(1982), é descrita a aplicação de um planejamento cíclico na redução do número de pares.

TABELA 2.3. Tipos de construções de estímulos utilizados em aplicações comerciais de ap, em %, segundo o ano

Tipo de construção	1971-1980	1981-1985
Perfil Completo	56	61
Matrizes Trade off	27	6
Combinação	14	10
Comparação pareada	-	10
Outros	3	13

fonte: Wittink & Cattin(1989)

2.4. Planejamento para perfil completo

Mesmo definido o método de construção dos estímulos, alguns problemas quanto à sua aplicabilidade continuam existindo. Em geral, a **análise de preferência** tem por objetivo a estimação de modelos individuais. Na prática, isto obriga que o número de perfis a serem avaliados por um único sujeito seja muito alto. A inconveniência resultante é que a confiabilidade das respostas tende a diminuir, uma vez que, como já foi salientado, deve-se contar com um eventual cansaço do entrevistado. A fim de evitar esta ocorrência, surge a necessidade da redução do número de perfis.

Green(1974) discute, sucintamente, alguns tópicos de planejamento que têm por objetivo facilitar a fase de coleta de dados. Ele destaca que a diminuição no número de estímulos a serem apresentados depende basicamente da complexidade do modelo de preferência. Admitindo a inexistência de alguns efeitos de interação, pode-se abrir mão da avaliação de alguns perfis através da adoção de planejamentos fracionários (Box, Hunter & Hunter(1978), Addelman(1962)).

Um exemplo de aplicação de um planejamento fracionário pode ser encontrado na tabela 2.2. Para que um modelo com todos os termos pudesse ser estimado, seria necessária a avaliação de 108 estímulos. Ignorando-se os termos de interação foi possível reduzir este número para 18.

O problema de confundimento em planejamentos fracionários é brevemente discutido no apêndice 2.

A adoção de um planejamento fracionário pode ser incapaz de reduzir consideravelmente o número de estímulos, quando um dos fatores possui um número muito alto de níveis. Nesta situação, Green(1974) sugere que se adote um método de estimação em múltiplos estágios. Em um primeiro momento seria pedido ao juiz que avaliasse apenas os níveis deste fator, por exemplo, atribuindo nota 10 ao de maior preferência, suponha nível A1, nota 0 ao de menor preferência, suponha nível A2 e, aos outros, notas intermediárias, o que acabaria formando uma escala de preferência para este atributo. A seguir, os perfis seriam construídos utilizando-se apenas os níveis A1 e A2 deste fator. O modelo seria estimado para estes dois níveis. Assim, estariam determinadas as utilidades de A1 e A2, que seriam utilizadas como limites de uma nova escala de preferência. O passo final seria obter as utilidades dos níveis intermediários; isto se dá fazendo a mudança da escala 0-10 para a nova escala.

Moore & Holbrook(1990) discutem planejamentos alternativos quando se tem uma alta correlação ambiental entre os atributos.

Um alto número de atributos também pode prejudicar a avaliação de um estímulo. Nesse caso, é interessante estudar a possibilidade da adoção de um planejamento em blocos incompletos ao acaso.

CAPÍTULO 3

Modelos de Preferência

Como já foi discutido durante a introdução, há dois ajustes intrínsecos a uma AP. O primeiro trata-se da determinação da função $y_s = g_s(R_s)$ e o segundo corresponde à modelagem da variável transformada $y_s = f_s(X_s) + \epsilon_s$. A função f , conhecida por modelo de preferência ou regra de composição, tem por finalidade exprimir como os atributos se combinam na explicação da preferência do entrevistado por um determinado estímulo.

O objetivo deste tópico é apresentar os modelos de preferência mais utilizados em uma AP. Será suposto, para os modelos vetor e ponto-ideal, que os atributos estejam em uma escala de medida intervalar ou razão.

Fica subentendida a existência de um termo aditivo de erro aleatório, em todos os modelos a serem apresentados. A critério do analista, pode-se supor a existência de um termo aditivo constante, a ser estimado, em cada um dos modelos a serem introduzidos.

3.1. Modelo de preferência VETOR

Trata-se do modelo usual de regressão. Sua utilização presuppõe que a função de

preferência do estímulo, y_s , aumente, ou diminua, linearmente de acordo com as variações dos atributos supostos quantitativos. Algebricamente é dado por:

$$y_s = X\beta_s \Rightarrow y_{js} = x'_j\beta_s \Rightarrow y_{js} = \sum_{k=1}^p \beta_{ks}x_{jk}. \quad (3.1.1)$$

Os parâmetros β_{ks} representam a importância (peso) de um determinado atributo para o indivíduo s .

Carroll(1972) apresenta um outro modo de visualizar este modelo. Considere os estímulos dispostos graficamente, em um espaço onde cada eixo representa um atributo: espaço de multiatributos. A idéia básica dessa interpretação é imaginar, nesse espaço, a existência de vetores, cada um associado a um indivíduo, que indicam a direção do crescimento de sua preferência. Dessa forma, a ordem de preferência dos estímulos para um determinado sujeito seria dada pela projeção dos estímulos em seu correspondente vetor.

Figura 3.1. Representação gráfica do modelo vetor, em um plano formado por dois atributos (X_1 e X_2)

Na figura 3.1, é representada uma situação com apenas dois atributos, 1 e 2. As letras maiúsculas representam a localização dos estímulos, dada pela matriz de planejamento. Os números romanos indicam os vetores de dois sujeitos. Ao projetar os estímulos nestes vetores, observa-se que a ordem de preferência para o primeiro indivíduo é dada por ECABD e, para o segundo, por DEABC.

Algebricamente esta interpretação pode ser descrita como:

$$y_{js} = a_s \mathbf{v}_s' \mathbf{x}_j \quad , \text{ onde:} \quad (3.1.2)$$

a_s é uma constante desconhecida.

$\mathbf{v}_s$ é um vetor ($p \times 1$) de norma 1 sobre o qual serão projetados os estímulos do indivíduo s .

Definindo $\beta_s = a_s \mathbf{v}_s$, vem que $y_{js} = \mathbf{x}_j' \beta_s$, que é a equação do modelo vetor anteriormente apresentada.

Como foi salientado no início deste tópico, ao utilizar este modelo acredita-se que a preferência por um estímulo simplesmente cresça (decresça) conforme se alterem os valores dos atributos. Em termos práticos, isto nem sempre ocorre. Em muitos casos um modelo monotônico não é suficientemente abrangente para descrever a preferência de um indivíduo. Surge, então, a necessidade de modelos mais complexos.

3.2. Modelos PONTO - IDEAL

Os modelos ponto-ideal, também conhecidos por "unfolding" (Carroll(1972)), ou euclidiano (Davison(1983)), partem do princípio de que para cada indivíduo existe um estímulo que é o de preferência máxima (ponto-ideal) e que, quanto mais parecido com ele o produto for, maior será sua preferência (Green & Carmone(1970)).

Considerando os estímulos como pontos em um espaço formado pelos atributos, esses modelos ajustam a distância entre o ponto-ideal, a ser estimado, e os estímulos, cujas coordenadas são definidas pela matriz de planejamento. Quanto menor for essa distância, maior será a preferência pelo estímulo.

Os modelos ponto-ideal são hierarquicamente classificados como simples, ponderado e geral.

3.2.1. Modelo PONTO - IDEAL SIMPLES

O modelo ponto-ideal simples, ou euclidiano simples, é dado por:

$$y_{js} = \beta_s(x_j - \Pi_s)'(x_j - \Pi_s) = \sum_{k=1}^p \beta_s(x_{jk} - \pi_{ks})^2 \quad , \text{onde:} \quad (3.2.1)$$

$\Pi_s = (\pi_{ks})$ vetor ($p \times 1$) é o ponto-ideal, a ser estimado, referente ao indivíduo s . Cada coordenada desse vetor corresponde a um fator;

β_s é o peso (importância) do conjunto de atributos para o indivíduo s .

Esse é o modelo ponto-ideal mais elementar. O peso, β_s , não é estabelecido para cada atributo em separado, mas para todo o conjunto de fatores envolvidos no estudo. Parte-se do princípio de que todos os atributos são igualmente importantes. A única mudança que pode ocorrer é no peso estimado para cada indivíduo.

Figura 3.2. Comportamento da preferência por um determinado atributo, segundo um modelo ponto-ideal

Figura 3.3. Linhas de isopreferência de um modelo ponto-ideal simples, com dois atributos (X_1 e X_2), para dois indivíduos (I e II)

Pode-se visualizar o que acontece neste modelo através das figuras 3.2 e 3.3.

Na figura 3.2, está representado o comportamento do modelo de preferência de um modelo ponto-ideal simples em relação a um atributo, quanto menor for y_j , maior será a

preferência pelo nível.

A figura 3.3 representa a projeção de dois modelos ponto-ideal simples, com dois atributos, no plano definido pelos fatores X_1 e X_2 . Os círculos são interpretados como linhas de isopreferência e os centros são os ponto-ideais de dois modelos. Para o modelo I a ordem de preferência é dada por ABDCE e para o II ECADB.

3.2.2. Modelo PONTO - IDEAL PONDERADO

Analiticamente, este modelo é dado por:

$$y_{js} = (x_j - \Pi_s)' \beta_s^m (x_j - \Pi_s) = \sum_{k=1}^p \beta_{ks} (x_{jk} - \pi_{ks})^2 \quad \text{onde:} \quad (3.2.2)$$

$\beta_s^m = \text{diag}(\beta_{1s}, \beta_{2s}, \dots, \beta_{ps})$ é a matriz diagonal com os pesos dos atributos.

A diferença básica deste modelo em relação ao anterior é a possibilidade dos atributos terem diferentes pesos.

Figura 3.4 Linhas de isopreferência de um modelo ponto-ideal ponderado, com 2 atributos (X_1 e X_2), para dois indivíduos (I e II)

Figura 3.5. Linhas de isopreferência de um modelo ponto-ideal geral, para dois indivíduos (I e II), com 2 atributos (X_1 e X_2).

Vê-se, na figura 3.4, como ficam as linhas de isopreferência para dois indivíduos em um modelo com dois atributos. Agora, estas linhas serão elipsóides cujo centro continua sendo o ponto ideal. Ao analisar esta figura, percebe-se que para o indivíduo I o atributo X_2 é o mais importante enquanto que para o indivíduo II X_1 é o mais importante. A ordem de preferência dos estímulos para o sujeito I é ADBCE e para o sujeito II é ECABD.

Muitos autores consideram o modelo ponto-ideal ponderado como sendo o modelo ponto-ideal, ignorando as outras versões (Green & Srinivasan(1978)).

3.2.3. Modelo PONTO - IDEAL GERAL

Este modelo é análogo ao ponto-ideal ponderado, com a particularidade de permitir a existência de interação entre os atributos. A interação é traduzida como uma mudança de direção nos eixos da elipse, como é exemplificado na figura 3.5. Nesse gráfico, tem-se que para o indivíduo I, a ordem de preferência dos estímulos é dada por ACBDE e para o sujeito II é ECADB.

Basicamente, o modelo é dado por:

$$y_{js} = (\mathbf{x}_j - \mathbf{\Pi})' \mathbf{A} (\mathbf{x}_j - \mathbf{\Pi}) \quad \text{onde, } \mathbf{A} \text{ é uma matriz de transformação.}$$

Encontra-se na literatura, duas possibilidades para a escolha da matriz $\mathbf{A}$: Davison(1983) e Carroll(1972), que serão doravante denominadas versões I e II, respectivamente.

Na versão I, a seguinte transformação é proposta:

$$\mathbf{x}_{js}^* = \mathbf{T}_s \boldsymbol{\beta}_s^{m*} \mathbf{x}_{js} \quad \text{e} \quad \mathbf{\Pi}_s^* = \mathbf{T}_s \boldsymbol{\beta}_s^{m*} \mathbf{\Pi}_s \quad \text{onde:}$$

$\mathbf{T}_s$ é uma matrix de dimensão p por p de transformação linear e $\boldsymbol{\beta}^{m*} = \text{diag}(\beta_1^*, \beta_2^*, \dots, \beta_p^*)$.

Parte-se do modelo seguinte, que é a fórmula da distância entre o estímulo j e o ponto ideal no novo sistema de coordenadas:

$$y_{js} = (\mathbf{x}_{js}^* - \mathbf{\Pi}_s^*)' (\mathbf{x}_{js}^* - \mathbf{\Pi}_s^*).$$

Desenvolvendo a expressão anterior tem-se

$$y_{js} = (\mathbf{x}_j - \Pi_s)' \beta_s^{m*} \mathbf{T}_s' \mathbf{T}_s \beta_s^{m*} (\mathbf{x}_j - \Pi_s).$$

Definindo $\Lambda_s^* = \mathbf{T}_s' \mathbf{T}_s$ e

$$y_{js} = (\mathbf{x}_j - \Pi_s)' \beta_s^{m*} \Lambda_s^* \beta_s^{m*} (\mathbf{x}_j - \Pi_s),$$

é possível demonstrar que $\beta_s^{m*} \Lambda_s^* \beta_s^{m*} = (\beta_{is}^* \beta_{js}^* \lambda_{ijs}^*)$, o que resulta em

$$y_{js} = \sum_{i=1}^p \beta_{ks}^{*2} \lambda_{kks}^* (x_{jk} - \pi_{ks})^2 + \sum_{i=1}^p \sum_{l \neq k} \beta_{ks}^* \beta_{ls}^* \lambda_{kls}^* (x_{jk} - \pi_{ks})(x_{jl} - \pi_{ls})$$

Fazendo

$$\beta_{ks} = \beta_{ks}^{*2} \lambda_{kks}^* \quad \text{e que} \quad \lambda_{kls} = \frac{\lambda_{kls}^*}{\lambda_{kks}^* \lambda_{lls}^*},$$

o modelo passa a ser

$$y_{js} = \sum_{k=1}^p \beta_{ks} (x_{jk} - \pi_{ks})^2 + \sum_{k=1}^p \sum_{l \neq k} \beta_{ks}^{\frac{1}{2}} \beta_{ls}^{\frac{1}{2}} \lambda_{kls} (x_{jk} - \pi_{ks})(x_{jl} - \pi_{ls}), \quad (3.2.3)$$

onde β_{ks} seria a importância de cada fator e o termo λ_{kls} a interação entre os atributos k e l .

A versão II assume que tanto $\mathbf{x}_j$ como Π_s são operacionados por uma matriz de transformação ortogonal $\mathbf{T}_s^*$ e assim as distâncias entre os estímulos e o ponto ideal transformados, ponderadas pela matriz β^m , são calculadas. Desta forma, tem-se

$$\mathbf{x}_{js}^* = \mathbf{T}_s^* \mathbf{x}_j = (x_{jks}^*) \quad \text{e} \quad \Pi_s^* = \mathbf{T}_s^* \Pi_s = (\pi_{ks}^*),$$

de onde vem que

$$y_{js} = (\mathbf{x}_j^* - \Pi_s)' \beta_s^m (\mathbf{x}_{js}^* - \Pi_s) = (\mathbf{x}_j - \Pi_s)' \mathbf{T}_s' \beta_s^m \mathbf{T}_s (x_j - \Pi_s).$$

$$\text{Tomando } \Lambda_s^* = \mathbf{T}_s' \beta_s^m \mathbf{T}_s = (\lambda_{ijs}^*),$$

$$y_{js} = (\mathbf{x}_j - \Pi_s)' \Lambda_s^* (\mathbf{x}_j - \Pi_s). \quad (3.2.4)$$

Na seção 4.1.2.3, é retomada a comparação entre as duas versões.

3.2.4. Comentários sobre os modelos ponto-ideal

Nem sempre a estimativa do suposto "ponto-ideal" pode ser interpretada como a localização do estímulo de maior preferência. Isto só ocorre se duas condições estiverem satisfeitas:

- a. a estimativa $\hat{\beta}_{ks}$ do peso correspondente ao atributo k seja positiva e,
- b. que, por ocasião do planejamento, se estabeleça que os estímulos sejam avaliados de modo que y_{js} cresça conforme diminua sua preferência.

Carroll(1972) discorre sobre a violação da condição "a". O autor sugere que quando for violada, $\hat{\pi}_{ks}$ seja interpretado como a localização do estímulo de menor preferência, denominando este ponto como anti-ideal. Desta forma, quanto mais próximo se estiver deste ponto menor será a preferência.

A violação da condição "b" pode ser evitada no momento do planejamento. Contudo se isto não tiver acontecido pode-se adotar o modelo ponto-ideal proposto por Green & Srinivasan(1978). Este modelo não invalida os comentários feitos ao longo do capítulo, e é dado por:

$$\tilde{y}_{js} = -y_{js}$$

A figura 3.6 é a representação gráfica de um modelo ponto-ideal (tendo $\tilde{y}$ na ordenada) com 2 atributos, sendo que ambos têm pesos positivos. A figura 3.7 ilustra o que acontece quando um dos pesos é negativo.

Figura 3.6. Gráfico de $-y_s$ para um modelo ponto-ideal ponderado com 2 atributos, sendo β_1 e β_2 positivos

fonte: Carroll(1972)

Figura 3.7. Gráfico de $-y_s$ para um modelo ponto-ideal ponderado com 2 atributos, sendo β_1 negativo e β_2 positivo

fonte: Carroll(1972)

3.3. Modelo UTILIDADE

Este é o modelo mais genérico de todos, uma vez que não é imposta nenhuma forma pré definida ao comportamento de y_{js} . Tem-se

$$y_{js} = \sum_{k=1}^p f_{ks}(x_{jk}) \quad \text{onde:}$$

$f_{ks}(\cdot)$ é uma função que descreve a importância de cada nível do fator k na formação da preferência do estímulo.

O modelo pode ser escrito de uma maneira mais atrativa considerando $\mathbf{X} = (x_{jks})$, uma matriz n por q , onde q é soma do número de níveis de cada atributo e

$$x_{jkl} = \begin{cases} 1 & \text{se o nível } l \text{ do atributo } k \text{ aparece no estímulo } j \\ 0 & \text{se não aparecer.} \end{cases}$$

Assim sendo, pode-se escrever y_{js} de maneira análoga ao modelo vetor:

$$y_{js} = \sum_{k=1}^p \sum_{l=1}^{n_k} \beta_{kls} x_{jkl}$$

onde n_k é o número total de níveis do atributo k .

O parâmetro β_{kls} é denominado como a utilidade do nível l do atributo k para o indivíduo s e mede a importância desse nível para o entrevistado.

Vale ressaltar que esse é o único modelo que se aplica diretamente a variáveis qualitativas. Trata-se de um modelo de análise de variância.

3.4. Hierarquia dos modelos

Pode-se mostrar que os modelos apresentados nas seções 3.1, 3.2 e 3.3 pertencem a uma mesma família e que os mais complexos englobam os mais simples.

Note que no modelo utilidade não se faz restrições ao comportamento de $f_{ks}(\cdot)$; logo pode-se encarar os modelos apresentados em 3.1 e 3.2 como casos particulares deste.

O modelo ponto-ideal geral, em qualquer das versões, reduz-se automaticamente ao modelo ponto-ideal ponderado se a transformação T_s for a matriz identidade.

Do mesmo modo, no modelo ponto-ideal ponderado, se todos os atributos forem igualmente importantes, chega-se ao modelo ponto-ideal simples.

A passagem do ponto-ideal ponderado para vetor exige um pouco mais de reflexão sobre o significado do ponto ideal. Como foi dito anteriormente, quanto mais próximo um estímulo estiver do ideal, menor será o valor de sua função preferência. No modelo vetor a preferência simplesmente cresce, ou decresce, linearmente. Isto pode ser obtido a partir do ponto-ideal ponderado, tomando os pontos ideais tendendo a $\pm\infty$.

A passagem do ponto-ideal simples para o modelo vetor fica mais clara ao se considerar a expressão 4.1.2.

Esta ordem é mais claramente exposta no capítulo que trata do método de estimação PREFMAP (seção 4.1.).

CAPÍTULO 4

Métodos de Estimação

Este capítulo apresenta os principais métodos de estimação de um modelo de AP. Entenda por estimação do modelo, tanto a estimação de g_s ($y_s = g_s(r_s)$), quanto a de f_s ($y_s = f_s(X)$).

Parte-se da suposição de que os métodos serão utilizados na estimação de modelos individuais. A fim de facilitar a notação dos modelos, será omitido o índice s referente ao indivíduo.

4.1. PREFMAP - PREference MAPping

O método PREFMAP, Carroll(1972), utiliza técnicas derivadas da regressão linear múltipla na estimação dos efeitos dos atributos. Nesse contexto, os atributos seriam as variáveis independentes e a avaliação subjetiva do estímulo a dependente. Basicamente é estabelecida uma linearização dos modelos de preferência e a estimação desses modelos é feita pelo método dos mínimos quadrados.

Inicialmente será abordada uma versão métrica, onde a variável dependente segue

uma escala intervalar, ou razão e, posteriormente será introduzida a versão não métrica. Será assumido, nesta primeira fase, que $y = r$, ou seja, a função g é a identidade.

Este procedimento aplica-se à estimação de modelos de preferência vetor e ponto-ideal.

Da mesma forma que no capítulo de regras de composição, os termos de erro não serão explicitados. Subentenda-se que, para todos os modelos, exista um resíduo aditivo.

Supõe-se, para validade dos testes a serem apresentados, que as hipóteses usuais de uma análise de regressão (Draper & Smith(1981)) estejam satisfeitas.

4.1.1. Aplicação ao modelo VETOR

Como foi anteriormente visto, o modelo vetor é dado por

$$y_j = \sum_{k=1}^p \beta_k x_{jk}.$$

Como pode ser observado, trata-se do modelo de regressão linear múltipla. A estimação é efetuada através dos procedimentos usuais (mínimos quadrados) e o coeficiente de determinação, R_V^2 , pode ser utilizado como uma medida da qualidade do ajuste.

Eventualmente, pode-se querer estimar o vetor $\mathbf{v}$ da equação (3.1.2) ($y_j = \mathbf{a}\mathbf{v}'\mathbf{x}_j = \beta' \mathbf{x}_j$). Considerando $\hat{\beta}_k$ como a estimativa de mínimos quadrados de β_k , Carroll(1972) sugere, uma vez que o vetor $\mathbf{v}$ é de norma unitária, que suas coordenadas sejam estimadas por:

$$\hat{\mathbf{v}} = (\hat{v}_k) \quad \text{onde:} \quad \hat{v}_k = \frac{\hat{\beta}_k}{(\sum_{k=1}^p \hat{\beta}_k^2)^{\frac{1}{2}}}$$

4.1.2. Aplicação aos modelos ponto-ideal

4.1.2.1. Modelo ponto-ideal simples

Partindo da equação (3.2.1) tem-se

$$y_j = \sum_{k=1}^p \beta (x_{jk} - \pi_k)^2.$$

Trabalhando a equação acima chega-se a

$$y_j = \beta \sum_{k=1}^p x_{jk}^2 + \sum_{k=1}^p (-2\beta \pi_k x_{jk}) + \sum_{k=1}^p \beta \pi_k^2.$$

Fixe

$$c = \sum_{k=1}^p \beta \pi_k^2 \quad \text{e} \quad \alpha_k = -2\beta \pi_k. \quad (4.1.1)$$

Substituindo as expressões acima no modelo de preferência, chega-se a

$$y_j = \beta \sum_{k=1}^p x_{jk}^2 + \sum_{k=1}^p \alpha_k x_{jk} + c. \quad (4.1.2)$$

Os parâmetros a serem estimados no modelo (4.1.2) são as quantidades β , α_k e c . As estimativas devem ser obtidas através do método de mínimos quadrados.

Obtidas as estimativas dos parâmetros e utilizando (4.1.1), estima-se o ponto ideal por

$$\hat{\pi}_k = \frac{\hat{\alpha}_k}{-2\hat{\beta}}. \quad (4.1.3)$$

Se $\hat{\beta} = 0$, o modelo (4.1.2) será idêntico ao modelo vetor. Deste modo pode-se verificar se a introdução deste parâmetro trouxe alguma melhoria significativa ao modelo,

através de um teste F parcial (Montgomery & Peck(1982)). O teste verifica a hipótese (H_1) de haver melhoria significativa ao se passar de um modelo vetor para um modelo ponto-ideal simples. Sendo R_V^2 o coeficiente de determinação do modelo vetor e R_E^2 o coeficiente de determinação do modelo ponto-ideal simples, a estatística desse teste é dada por:

$$F = \frac{R_E^2 - R_V^2}{1 - R_E^2} (N - p - 2) \stackrel{H_0}{\sim} F(1, N - p - 2).$$

4.1.2.2. Modelo ponto-ideal ponderado

O modelo ponto-ideal ponderado é dado por:

$$y_j = \sum_{k=1}^p \beta_k (x_{jk} - \pi_k)^2$$

Desenvolvendo esta equação tem-se

$$y_j = \sum_{k=1}^p \beta_k x_{jk}^2 + \sum_{k=1}^p (-2\beta_k \pi_k) x_{jk} + \sum_{k=1}^p \beta_k \pi_k^2$$

Com o fim de facilitar a estimação do modelo

$$c = \sum_{k=1}^p \beta_k \pi_k^2 \quad \text{e} \quad \alpha_k = -2\beta_k \pi_k. \quad (4.1.4)$$

Deste modo, o modelo fica sendo

$$y_j = \sum_{k=1}^p \beta_k x_{jk}^2 + \sum_{k=1}^p \alpha_k x_{jk} + c. \quad (4.1.5)$$

Os estimadores de β_k , α_k e c são obtidos através da aplicação do método de mínimos quadrados. De modo análogo ao item anterior, estima-se os pontos ideais a partir de (4.1.4), obtendo-se

$$\hat{\pi}_k = \frac{\hat{\alpha}_k}{-2\hat{\beta}_k}. \quad (4.1.6)$$

O modelo (4.1.2) é um caso particular de (4.1.5): basta que $\beta_k = \beta$, $k = 1, 2, \dots, p$. Isto permite a utilização de testes F parciais para verificar se houve ganho significativo ao se passar do modelo ponto-ideal simples para o ponto-ideal ponderado, ou mesmo do vetor para o ponto-ideal ponderado. Denotando o coeficiente de determinação do modelo ponto-ideal ponderado por R_P^2 , tem-se então, sob a hipótese nula de não haver melhoria com a mudança de modelo:

a.. Ponto-ideal simples para ponderado

$$F = \frac{R_P^2 - R_E^2}{1 - R_P^2} \frac{N - 2p - 1}{p - 1} \stackrel{H_0}{\sim} F(p - 1, N - 2p - 1).$$

b.. Vetor para ponto-ideal ponderado

$$F = \frac{R_P^2 - R_V^2}{1 - R_P^2} \frac{N - 2p - 1}{p} \stackrel{H_0}{\sim} F(p, N - 2p - 1).$$

4.1.2.3. Modelo ponto-ideal geral

Inicialmente o modelo será estimado para a versão II. A I será abordada na sequência.

Da equação (3.2.4) vem

$$y_j = (\mathbf{x}_j - \Pi)' \Lambda' (\mathbf{x}_j - \Pi) = \mathbf{x}_j' \Lambda' \mathbf{x}_j - 2\Pi' \Lambda' \mathbf{x}_j + \Pi' \Lambda' \Pi.$$

Fazendo

$$\boldsymbol{\alpha}' = -2\Pi' \Lambda' = (\alpha'_j) \quad \text{e} \quad c = \Pi' \Lambda' \Pi,$$

vem

$$y_j = \mathbf{x}_j' \mathbf{\Lambda}^* \mathbf{x}_j + \mathbf{x}_j' \boldsymbol{\alpha}^* + c = \sum_{k=1}^p \sum_{l=1}^p \lambda_{kl}^* x_{jk} x_{jl} + \sum_{k=1}^p \alpha_k^* x_{jk} + c.$$

Uma vez que $\mathbf{\Lambda}^*$ é simétrica, vem, da equação acima que

$$y_j = \sum_{k=1}^p \lambda_{kk}^* x_{jk}^2 + \sum_{k=1}^p \alpha_k^* x_{jk} + \sum_{k=1}^p \sum_{l < k}^p 2\lambda_{kl}^* x_{jk} x_{jl} + c. \quad (4.1.7)$$

Como anteriormente, utiliza-se o procedimento de mínimos quadrados para estimar $\mathbf{\Lambda}^*$, $\boldsymbol{\alpha}^*$ e c . A estimativa do ponto ideal pode ser obtida através de:

$$\hat{\Pi} = -\frac{1}{2} \hat{\mathbf{\Lambda}}^{*-1} \hat{\boldsymbol{\alpha}}^*. \quad (4.1.8)$$

Por definição, $\mathbf{\Lambda}^* = \mathbf{T}^{*'} \boldsymbol{\beta}^m \mathbf{T}^{*'} (3.2.4)$. Logo, um modo de obter uma estimativa dessas matrizes seria através da decomposição espectral de $\hat{\mathbf{\Lambda}}^*$. $\hat{\mathbf{T}}^*$ seria o transposto da matriz de vetores característicos de $\hat{\mathbf{\Lambda}}^*$ e, $\hat{\boldsymbol{\beta}}^m$ seria a matriz diagonal formada pelas raízes características de $\hat{\mathbf{\Lambda}}^*$.

A estimação do modelo proposto na versão I será semelhante a esta. O modelo, nesse caso, é dado por

$$y_j = \sum_{k=1}^p \beta_k (x_{jk} - \pi_k)^2 + \sum_{k=1}^p \sum_{l \neq k}^p \beta_k^{\frac{1}{2}} \beta_l^{\frac{1}{2}} \lambda_{kl} (x_{jk} - \pi_k)(x_{jl} - \pi_l).$$

Desenvolvendo-se, chega-se a

$$\begin{aligned} y_j = & \sum_{k=1}^p \beta_k x_{jk}^2 + \sum_{k=1}^p (-2\beta_k x_{jk} \pi_k) + \sum_{k=1}^p \beta_k \pi_k^2 + \sum_{k=1}^p \sum_{l \neq k}^p \beta_k^{\frac{1}{2}} \beta_l^{\frac{1}{2}} \lambda_{kl} x_{jk} x_{jl} + \\ & + \sum_{k=1}^p \sum_{l \neq k}^p (-2\beta_k^{\frac{1}{2}} \beta_l^{\frac{1}{2}} \lambda_{kl} x_{jk} \pi_l) + \sum_{k=1}^p \sum_{l \neq k}^p \beta_k^{\frac{1}{2}} \beta_l^{\frac{1}{2}} \lambda_{kl} \pi_k \pi_l. \end{aligned}$$

Assumindo:

$$c = \sum_{k=1}^p \beta_k \pi_k^2 + \sum_{k=1}^p \sum_{l \neq k}^p \beta_k^{\frac{1}{2}} \beta_l^{\frac{1}{2}} \lambda_{kl} \pi_k \pi_l \quad ;$$

$$b_{kl} = 2\beta_k^{\frac{1}{2}} \beta_l^{\frac{1}{2}} \lambda_{kl} \quad \text{e} \quad (4.1.9)$$

$$\alpha_k = -2\beta_k \pi_k - 2 \sum_{k=1}^p \sum_{l \neq k}^p \beta_k^{\frac{1}{2}} \beta_l^{\frac{1}{2}} \lambda_{kl} \pi_k \quad \text{que equivale a:}$$

$$\boldsymbol{\alpha} = -2\mathbf{U}\boldsymbol{\Pi} \quad (4.1.10)$$

onde: $\boldsymbol{\alpha} = (\alpha_j)_{(px1)}$ e

$\mathbf{U} = (u_{ij})_{p \times p}$, sendo

$$u_{ij} = \begin{cases} \beta_i^{\frac{1}{2}} \beta_j^{\frac{1}{2}} \lambda_{ij} = \frac{b_{kl}}{2} & \text{se } i \neq j \\ \beta_i & \text{se } i = j, \end{cases} \quad (4.1.11)$$

vem

$$y_j = \sum_{k=1}^p \beta_k x_{jk}^2 + \sum_{k=1}^p \alpha_k x_{jk} + \sum_{k=1}^p \sum_{l < k}^p b_{kl} x_{jk} x_{jl} + c. \quad (4.1.12)$$

Utilizando o método de mínimos quadrados encontra-se as estimativas de β_k , α_k , b_{kl} e c .

De (4.1.9) e de (4.1.10) tira-se, respectivamente, que:

$$\hat{\lambda}_{kl} = \frac{\hat{b}_{kl}}{2\hat{\beta}_k^{\frac{1}{2}} \hat{\beta}_l^{\frac{1}{2}}} \quad (4.1.13)$$

$$\hat{\boldsymbol{\Pi}} = -\frac{1}{2} \hat{\mathbf{U}}^{-1} \hat{\boldsymbol{\alpha}}. \quad (4.1.14)$$

Considerando R_G^2 como sendo o coeficiente de determinação do modelo acima, tem-se que as estatísticas F para testar as hipóteses nulas de não haver melhoria significativa ao

se passar dos modelos ponto-ideal ponderado, ponto-ideal simples e vetor para o modelo ponto-ideal geral são, respectivamente:

$$F = \frac{R_G^2 - R_P^2}{1 - R_G^2} \frac{N - \frac{1}{2}p(p+3) - 1}{\frac{1}{2}p(p-1)} \stackrel{H_0}{\sim} F\left(\frac{1}{2}p(p-1), N - \frac{1}{2}p(p+3) - 1\right)$$

$$F = \frac{R_G^2 - R_E^2}{1 - R_G^2} \frac{N - \frac{1}{2}p(p+3) - 1}{\frac{1}{2}p(p+1) - 1} \stackrel{H_0}{\sim} F\left(\frac{1}{2}p(p+1) - 1, N - \frac{1}{2}p(p+3) - 1\right)$$

$$F = \frac{R_G^2 - R_V^2}{1 - R_G^2} \frac{N - \frac{1}{2}p(p+3) - 1}{\frac{1}{2}p(p+1)} \stackrel{H_0}{\sim} F\left(\frac{1}{2}p(p+1), N - \frac{1}{2}p(p+3) - 1\right).$$

Observe que as equações (4.1.7) e (4.1.12) são análogas, valendo as seguintes relações:

$$\hat{\lambda}_{kk}^* = \hat{\beta}_k \quad , \quad 2\hat{\lambda}_{kl}^* = \hat{b}_{kl} \quad , \quad \hat{\alpha}_k^* = \hat{\alpha}_k.$$

Note que:

$$\Lambda^* = (\lambda_{kl}^*)_{p \times p} = \begin{pmatrix} \beta_1 & \frac{b_{12}}{2} & \dots & \frac{b_{1p}}{2} \\ \frac{b_{21}}{2} & \beta_2 & \dots & \frac{b_{2p}}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{b_{p1}}{2} & \frac{b_{p2}}{2} & \dots & \beta_p \end{pmatrix} = \mathbf{U}.$$

Por estes fatos, conclue-se que (4.1.8) e (4.1.10) são idênticas, ou seja, os pontos ideais das duas versões serão os mesmos.

Diferenças começam a surgir quando se deseja estimar a matriz de transformação e os pesos β .

4.1.3. EXEMPLO

Como ilustração da técnica, retome o exemplo dos micros, capítulo 2, considerando apenas os atributos quantitativos (capacidade do winchester, memória RAM e clock). Suponha que a tabela 4.1 represente a avaliação, realizada em ordem decrescente de preferência, de um conjunto de estímulos realizada por um determinado indivíduo.

Inicialmente, será realizado o ajuste de todos os modelos aos dados e a seguir será feita a escolha do melhor modelo.

TABELA 4.1. Exemplo

estímulo	winchester	RAM	clock	avaliação
1	40	1	8	4,45
2	40	2	20	0,80
3	40	4	16	2,10
4	40	1	20	1,10
5	40	2	16	0,85
6	40	4	8	5,50
7	60	1	20	0,50
8	60	2	16	0,60
9	60	4	8	5,20
10	60	1	16	0,40
11	60	2	8	4,40
12	60	4	20	0,65
13	120	1	16	1,00
14	120	2	8	4,70
15	120	4	20	1,20
16	120	1	8	4,30
17	120	2	20	0,70
18	120	4	16	1,40

TABELA 4.2. Pesos relativos (importância) para o modelo VETOR

Atributo	estimativa	v	I_k (%)
Winchester	-0,049	-0,027	2,3
RAM	0,324	0,179	15,0
Clock	-1,782	-0,984	82,7

Nas aplicações dos modelos a esses dados, tomou-se o cuidado de padronizar as colunas winchester (W), RAM (R) e clock (C). Isso foi feito para que as estimativas β_k reflitam apenas a importância do atributo e não uma diferença de escala. As variáveis padronizadas serão representadas por W_p , M_p e C_p .

a. Estimação do modelo VETOR

Através de uma regressão linear múltipla, chegou-se à seguinte estimativa do modelo:

$$\hat{y}_j = 2.214 - 0.049W_p + 0.324M_p - 1.782C_p \quad R_V^2 = 0.9014$$

O sinal negativo de W_p e C_p indica que quanto maior o valor dessas variáveis, menor será y_j , sendo maior a preferência pelo estímulo. O inverso acontece com M_p .

Percebe-se, analisando o modelo, que o atributo mais importante para o entrevistado é o clock, seguido pela memória RAM e pela capacidade do winchester. A tabela 4.2 contém um índice que procura quantificar a importância relativa de cada atributo, bem como as coordenadas estimadas do vetor v . O índice, calculado para o atributo k , é dado por:

$$I_k = \frac{|\hat{\beta}_k|}{\sum_{i=1}^p |\hat{\beta}_i|}.$$

b. Estimação do modelo ponto-ideal simples

Aplicando mínimos quadrados ao modelo (4.1.2), chega-se à seguinte estimativa

$$\hat{y}_j = 0.781 - 0.309W_p + 0.136M_p - 1.594C_p + 0.506(W_p^2 + M_p^2 + C_p^2) \quad R_E^2 = 0.9631.$$

A partir de (4.1.3) chega-se às seguintes estimativas dos pontos ideais

$$\hat{\pi}_{W_p} = 0.3053 \quad \hat{\pi}_{M_p} = -0.1344 \quad \text{e} \quad \hat{\pi}_{C_p} = 1.5751.$$

Passando esses valores para a escala original, vem que:

$$\hat{\pi}_W = 84.0\text{Mb} \quad \hat{\pi}_M = 2.2\text{Mb} \quad \text{e} \quad \hat{\pi}_{C_p} = 22.8\text{MHz}.$$

Logo, o micro ideal, para o entrevistado, deveria ter um winchester de 84.0Mb, 2.2Mb de memória RAM e um clock de 22.8MHz.

A importância do conjunto de atributos, dada por $\hat{\beta}$, é de 0.506. Esse valor só adquire um significado maior quando comparado com o de outros indivíduos.

c. Estimação para o modelo ponto-ideal ponderado

Estimando-se o modelo (4.1.5), chega-se a:

$$\hat{y}_j = 0.799 - 0.283W_p + 0.266M_p - 1.452C_p + 0.455W_p^2 + 0.156M_p^2 + 0.888C_p^2 \quad R_p^2 = 0.9879.$$

As estimativas dos pontos ideais, obtidas através de (4.1.6), nas escalas padronizada e real são, respectivamente:

$$\hat{\pi}_{W_p} = 0.3110 \quad \hat{\pi}_{M_p} = -0.8526 \quad \text{e} \quad \hat{\pi}_{C_p} = 0.8176$$

$$\hat{\pi}_W = 84.2\text{Mb} \quad \hat{\pi}_M = 1.2\text{Mb} \quad \text{e} \quad \hat{\pi}_C = 18.9\text{MHz}.$$

Com base nesse modelo, o micro ideal possui um winchester de 84.2 Mb, uma memória RAM de 1.2 Mb e um clock de 18.9 MHz. Em relação ao modelo ponto-ideal simples, a maior diferença encontra-se na memória RAM ideal.

Na tabela 4.3 estão estimadas a importância de cada atributo. Nota-se que o clock foi, segundo o modelo, o fator predominante na escolha do micro.

TABELA 4.3. Pesos relativos para o modelo Ponto-Ideal Ponderado

Atributo	estimativa	I_k (%)
Winchester	-0,283	14,1
RAM	0,266	13,3
Clock	-1,452	72,6

d. Estimação do modelo ponto-ideal geral

Estimando o modelo (4.1.12), chega-se a

$$\hat{y}_j = 0.768 - 0.247W_p + 0.272M_p - 1.442C_p + 0.481W_p^2 + 0.158M_p^2 + 0.891C_p^2 - \\ -0.027W_pM_p - 0.020W_pC_p - 0.147M_pC_p \quad R^2 = 0.9934. \quad (4.1.15)$$

De (4.2.13) chega-se a

$$\hat{\lambda}_{WM} = -0.0489 \quad \hat{\lambda}_{WC} = -0.0153 \quad \hat{\lambda}_{MC} = -0.1959.$$

A partir de (4.1.11) tem-se que

$$\hat{U} = \begin{pmatrix} 0.481 & -0.0135 & -0.0100 \\ -0.0135 & 0.158 & -0.0735 \\ -0.0100 & -0.0735 & 0.891 \end{pmatrix}.$$

Além disso

$$\alpha = (-0.247 \quad 0.272 \quad -1.442).$$

De (4.1.14), os pontos ideais são:

$$\hat{\pi}_{W_p} = 0.2594 \Rightarrow \hat{\pi}_W = 82.4Mb \quad (4.1.16)$$

$$\hat{\pi}_{MP} = -0.4792 \Rightarrow \hat{\pi}_M = 1.7Mb \quad (4.1.17)$$

$$\hat{\pi}_{CP} = 0.7726 \Rightarrow \hat{\pi}_C = 18.6Mb. \quad (4.1.18)$$

Repare que esses valores não diferem em muito dos obtidos para o modelo ponto-ideal ponderado.

TABELA 4.4. Estatísticas F para os testes de melhoria

TESTE	F	graus de liberdade	nível descritivo
VETOR para Ponto-Ideal Simples	21,7	1 e 13	0,0004
Ponto-Ideal Simples para Ponderado	11,3	2 e 11	0,0022
Ponto-Ideal Ponderado para Geral	2,2	3 e 8	0,1658

TABELA 4.5. Estímulos ordenados

estímulo	winchester	RAM	clock	avaliação
1	40	1	8	15
2	40	2	20	6
3	40	4	16	12
4	40	1	20	9
5	40	2	16	7
6	40	4	8	18
7	60	1	20	2
8	60	2	16	3
9	60	4	8	17
10	60	1	16	1
11	60	2	8	14
12	60	4	20	4
13	120	1	16	8
14	120	2	8	16
15	120	4	20	10
16	120	1	8	13
17	120	2	20	5
18	120	4	16	11

e. Escolha do melhor modelo

Na tabela 4.4 estão dispostos os testes de melhoria para verificar a existência de ganho significativo ao se passar de um modelo de menor para um de maior complexidade. Observa-se que a única melhoria não significativa foi a que avaliava a passagem de um modelo ponto-ideal ponderado para um ponto-ideal geral. Desse modo, segundo esses testes, o modelo a ser adotado deve ser o ponto-ideal ponderado.

4.1.4. Avaliação em escala de medida não métrica

Até agora, o desenvolvimento da técnica exigiu que a avaliação r_j fosse medida em uma escala métrica. Será apresentado um procedimento que permite a estimação dos modelos anteriores, mesmo quando a escala de medida for não métrica, por exemplo, ordinal.

A idéia básica é encontrar para cada indivíduo uma transformação monotônica, $g(\cdot)$, que aplicada a r_j faça com que o modelo abaixo se ajuste via mínimos quadrados:

$$y_j = d_j$$

onde d_j é qualquer um dos modelos analisados neste tópico.

Algoritmo proposto

Passo 1. Estimar o modelo utilizando r_j como variável dependente.

Passo 2. Com o modelo estimado, calcular os valores ajustados para cada estímulo.

Passo 3. Aplicar o algoritmo de regressão isotônica, ver apêndice 3, aos valores ajustados pelo modelo, encontrando $y^{(1)}$.

Passo 4. Estime novamente o modelo, utilizando $y^{(1)}$ como variável dependente.

Passo 5. Repita os passos 2, 3 e 4 até que algum critério de convergência seja satisfeito.

Ao final do processo serão obtidas as estimativas apropriadas e os coeficientes de regressão para serem utilizados na estimação dos pontos ideais, matrizes de rotação e pesos.

Os testes F perdem sua validade, mas podem ser utilizados como indicadores da melhoria do ajuste ao se mudar de modelo.

TABELA 4.6. Processo de estimação

Estímulo	r	y_p	$y^{(1)}$	$y^{(2)}$	...	$y^{(88)}$	$y^{(89)}$
10	1	3,046	2,697	2,811	...	4,671	4,673
7	2	2,347	2,697	2,811	...	4,671	4,673
8	3	3,281	3,281	3,519	...	4,671	4,673
12	4	4,545	4,454	4,729	...	5,763	5,766
17	5	5,777	5,777	6,241	...	5,918	5,917
2	6	6,728	6,728	6,675	...	5,918	5,917
5	7	7,717	6,921	6,872	...	6,595	6,594
13	8	6,126	6,921	6,872	...	6,595	6,594
4	9	7,092	7,092	6,872	...	6,595	6,594
15	10	9,511	9,511	9,390	...	7,478	7,473
18	11	11,526	10,915	10,605	...	9,225	9,223
3	12	10,305	10,915	10,605	...	9,901	9,902
16	13	14,197	12,916	12,867	...	12,837	12,838
11	14	11,634	12,916	12,867	...	12,837	12,838
1	15	15,193	15,193	15,204	...	13,698	13,692
14	16	15,863	15,863	15,389	...	13,996	13,996
9	17	16,147	16,147	16,403	...	18,666	18,671
6	18	19,966	19,966	20,270	...	20,966	20,966

TABELA 4.7. Parâmetros estimados no decorrer do processo de estimação

Efeito	(0)	(1)	(2)	...	(88)	(89)
W	-1,766	-1,720	-1,6492	...	-0,968	-0,966
M	1,580	1,576	1,5401	...	1,337	1,337
C	-3,804	-3,698	-3,6062	...	-3,515	-3,515
W^2	4,357	4,071	3,8672	...	1,983	1,979
M^2	0,710	0,565	0,5236	...	0,901	0,901
C^2	1,709	1,811	1,9116	...	1,878	1,877
WM	0,526	0,371	0,2569	...	-0,106	-0,107
WC	-0,167	0,072	0,2369	...	0,203	0,202
MC	-0,600	-0,615	-0,6425	...	-1,109	-1,110
R^2	0,94107	0,99103	0,99476	...	0,99998	0,99998
$diffmax$		1,477	2,325	...	0,009	0,009

Como ilustração do processo iterativo, considere os dados da tabela 4.5, que representam uma possível ordenação por postos dos estímulos da tabela 4.1. Note que essa ordenação é coerente com os dados de 4.1.

Foi escolhido o ponto-ideal geral, como modelo a ser estimado. Como critério de parada do processo iterativo, calculou-se a diferença relativa máxima entre os parâmetros estimados em iterações sucessivas:

$$difmax = \max \left\{ \frac{\beta_j^{(k)} - \beta_j^{(k-1)}}{\beta_j^{(k-1)}} ; j = 1, 2, \dots, 9 \right\}.$$

Quando, por duas iterações sucessivas, essa diferença foi inferior a 0,01, interrompeu-se o processo de estimação.

A tabela 4.6, resume o processo de estimação do modelo. A coluna **r** representa os postos atribuídos aos estímulos. Utilizando esses valores como variável dependente, passo 1, ajustou-se o modelo (4.1.12). Através do modelo estimado, previu-se os valores de y para cada estímulo, coluna y_p , que foram submetidos a uma regressão isotônica, passo 3, encontrando-se $y^{(1)}$. A seguir, o modelo (4.1.12) foi novamente ajustado, só que tendo $y^{(1)}$ como variável resposta. Aos valores previstos pelo novo modelo aplicou-se os passos 3 e 4 e assim sucessivamente até o passo 89, onde satisfiz-se o critério de convergência empregado.

A tabela 4.7 contém as estimativas dos parâmetros para alguns passos do processo iterativo e seus correspondentes coeficientes de determinação.

O modelo estimado foi:

$$\hat{y}_j = 5,0072 - 0,9660W_p + 1,3371M_p - 3,5148C_p + 1,9787W_p^2 + 0,9010M_p^2 + 1,8774C_p^2 - \\ - 0,1068W_pM_p + 0,2020W_pC_p - 1,1099M_pC_p,$$

de onde vem que:

$$\hat{U} = \begin{pmatrix} 1,9787 & -0,0534 & 0,01010 \\ -0,0534 & 0,9010 & 0,5550 \\ 0,0100 & 0,5550 & 1,8774 \end{pmatrix}$$

$$\alpha' = (-0.9660 \quad 1,3371 \quad -3,5148).$$

Consequentemente, por (4.1.14), vem que:

$$\hat{\pi}_{W_p} = 0.1296 \quad \hat{\pi}_{M_p} = -1.5976 \quad \hat{\pi}_{C_p} = 1,4014$$

$$\hat{\pi}_W = 77,9Mb \quad \hat{\pi}_M = 0,3Mb \quad \hat{\pi}_C = 21,9MHz.$$

Se esses valores forem comparados com os obtidos em (4.1.16), (4.1.17) e (4.1.18), a única diferença sensível estará no atributo memória RAM. Essa diferença deve ser atribuída à perda de informação ao se passar de uma escala de medida intervalar para uma ordinal.

A qualidade do ajuste do modelo pode ser medida através do coeficiente de correlação de postos de Spearman, calculado entre os postos previstos pelo modelo e os observados. No caso acima, encontrou-se um coeficiente de 0,9814, o que indica um bom ajuste do modelo.

4.2. MONANOVA - MONotonic ANalysis Of VARIance

MONANOVA, análise de variância monotônica, Kruskal(1965), é um procedimento utilizado na estimação de modelos utilidade, com a avaliação dos estímulos sendo, em geral, realizada através de sua ordenação, postos. Consiste em encontrar, iterativamente, a transformação monotônica dos postos que melhor se ajuste a um modelo aditivo (modelo de análise de variância sem interações).

O modelo a ser estimado é

$$y_j = \sum_{i=1}^p \beta_i x_{ji} + \epsilon_j,$$

sujeito à restrição

$$r_j < r_{j'} \Rightarrow y_j \leq y_{j'}. \quad (4.2.1)$$

Fixado um valor para β , tem-se que

$$y_j(\beta) = \sum_{i=1}^p \beta_i x_{ji}.$$

O método de estimação indicado para o procedimento é o de mínimos quadrados. Deseja-se encontrar β e $y = g(r) = (y_j)$ que minimizem a expressão

$$\sum_{j=1}^n [y_j - y_j(\beta)]^2, \quad \text{sujeita à restrição (4.2.1).}$$

Observe que se $y_j = 0$, $j = 1, 2, \dots, n$ e $\beta = 0$, a soma de quadrados acima atingirá o seu valor mínimo. Logo torna-se necessário realizar modificações de modo a evitar essa solução trivial.

Kruskal(1965) sugere que seja minimizada a seguinte função

$$Str(g, \beta) = \sqrt{\frac{\sum_j [y_j - y_j(\beta)]^2}{\sum_j [y_j(\beta) - \bar{y}(\beta)]^2}} \quad (4.2.2)$$

sujeita à restrição (4.2.1), onde $\bar{y}(\beta)$ é a média dos $y_j(\beta)$.

A quantidade acima recebe o nome de "stress". O termo do denominador tem o papel de evitar a solução trivial e fazer com que g não seja uma simples transformação linear dos postos (decorrência do teorema A4.1).

O "stress" pode ser interpretado como uma medida da falta de ajuste do modelo. Valores altos do "stress" indicam que os dados não se ajustaram bem e, por outro lado, valores baixos indicam um ajuste satisfatório. Mais adiante serão feitas mais algumas considerações sobre este índice.

Fixado g e considerando $\hat{\beta}(g)$ como aquele que minimiza a expressão acima, será chamado de "stress" direto de g a quantidade $Str(g, \hat{\beta}(g))$, que passará a ser uma medida de falta de ajuste para um particular g .

Deste modo, o problema passa a ser o de encontrar $\hat{g}$ que minimize o "stress" direto. $Str(\hat{g}, \hat{\beta}(\hat{g}))$ será denominado "stress" monotônico dos dados. $\hat{g}$ e $\hat{\beta}(\hat{g})$ serão,

respectivamente, a transformação procurada e os parâmetros estimados do modelo.

Algebricamente, tem-se que

$$Str(\hat{g}, \hat{\beta}(\hat{g})) = \min_{\{y_j = g(r_j)\}} \min_{\beta} Str(g, \beta)$$

sujeito à restrição (4.2.1).

Kruskal(1965) sugere duas diferentes abordagens para o caso de se permitir empates: a primeira consiste em considerar a restrição $r_j \leq r_{j'} \Rightarrow y_j \leq y_{j'}$ (esta abordagem foi denominada primária); na outra abordagem, que recebeu o nome de secundária, aconselha-se acrescentar a seguinte restrição: $r_i = r_j \Rightarrow y_i = y_j$.

Para simplificar a minimização, aconselha-se fixar inicialmente os valores de β e, a partir daí, encontrar os valores ajustados y_j . Para cada β será encontrada uma diferente transformação g . O valor do "stress" desta "nova" variável será o seu "stress" direto.

Grande parte dos procedimentos iterativos utiliza o gradiente como indicador do sentido da modificação que deverá ser realizada em β , a fim de minimizar o "stress". Kruskal(1965) sugere o seguinte procedimento:

- a) Estipular um valor inicial para β .
- b) Encontrar $y_j(\beta)$.
- c) Encontrar y_j a partir de $y_j(\beta)$.
- d) Calcular o "stress".
- e) Calcular o gradiente.
- f) A partir do gradiente modificar β .
- g) Retomar o passo b) até que alguma regra de parada seja obedecida.

No passo c), deve-se estimar y_j , de modo que esta estimativa seja ótima em termos de minimizar o "stress", sujeito à restrição (4.2.1). Para isto aplica-se a técnica regressão isotônica¹, Barlow et al(1972), que é utilizada na estimação por mínimos quadrados de

¹ vide apêndice 3

funções sujeitas a restrições de ordem. Nesta etapa, dado que β é considerado fixo, os valores y_j que minimizam a função (4.2.2) sujeita a (4.2.1) correspondem aos estimadores de mínimos quadrados submetidos à mesma restrição, visto que o denominador de (4.2.2) é, neste momento, uma constante.

Originalmente, MONANOVA foi desenvolvida no contexto de análise de variância usual. Trata-se de uma técnica que tem por objetivo encontrar uma transformação da variável dependente que elimine, do modelo, os termos de interação. Sua aplicação limita-se a modelos aditivos e planejamentos fatoriais.

Para facilitar a interpretação do "stress", é interessante conhecer o seu campo de variação. Com este intuito prova-se o seguinte resultado:

RESULTADO $Str[g, \beta] \leq 1$

prova:

$$Str[g, \beta] = \sqrt{\frac{\sum_j [y_j - y_j(\beta)]^2}{\sum_j [y_j(\beta) - \bar{y}(\beta)]^2}}$$

Fixado qualquer β ; $\beta \in \mathbb{R}^p$, tem-se que $y_j = g(r_j)$ é a função que minimiza $\sum_{j=1}^n (h(r_j) - y_j(\beta))^2$, na classe das funções h isotônicas, sujeita à restrição de ordem.

Por outro lado, $\bar{y}$ é uma função isotônica de r . Logo,

$$\sum_{j=1}^n (y_j - y_j(\beta))^2 \leq \sum_{j=1}^n (\bar{y}(\beta) - y_j(\beta))^2 = \sum_{j=1}^n (y_j(\beta) - \bar{y}(\beta))^2$$

Aplicando este resultado na equação do "stress", vem que:

$$Str[g, \beta] \leq 1 \quad \blacksquare$$

Vê-se, então, que "stress" com valor próximo a 1 é indesejável, pois indica falta de ajuste dos dados ao modelo. Kruskal(1964a), em um contexto de escalonamento multi-dimensional, cria uma regra prática para julgar a qualidade de um ajuste, baseando-se no

"stress"; segundo este autor, valor de "stress" superior a 20% indica um péssimo ajuste, de 15 a 20% um ajuste ruim, de 10 a 15% um ajuste regular, de 5 a 10% um ajuste satisfatório e abaixo de 5% um excelente ajuste.

No apêndice 4, é descrita uma versão do método de minimização "steepest descent", proposta por Kruskal(1964b), bem como as derivadas do "stress" com respeito a β , para o cálculo do gradiente. Cabe ressaltar que não foi descartada a possibilidade de existência de mínimos locais.

4.2.1. Exemplo

Suponha o exemplo apresentado na seção 1.2. Para facilitar o processo computacional, os postos serão centrados (vide tabela 4.8).

O modelo a ser estimado é:

$$y_j = \sum_{i=1}^3 \beta_i x_{ji} \quad (4.2.3.)$$

sujeito à restrição (4.2.1).

O primeiro passo do processo iterativo é estipular um valor inicial para o vetor β , β^0 . Como o processo de estimação a ser utilizado resultará em estimadores de mínimos quadrados, uma opção razoável para se encontrar β^0 consiste em realizar uma regressão linear nos dados, tomando y como variável dependente. Assim feito, encontrou-se as estimativas expostas na tabela 4.9.

Objetivando simplificar o processo computacional, Kruskal(1964b) recomenda que os parâmetros estimados sejam padronizados, de modo que seu quadrado médio resulte em um ($\frac{\sum_{j=1}^p \beta_j^2}{p} = 1$). Note que isto não alterará o valor do "stress", uma vez que este não é afetado por transformações lineares. Os parâmetros padronizados estão representados por $\beta_{pd}^{(0)}$ na tabela 4.9.

Tendo um valor inicial para β , a próxima etapa consiste em utilizar o modelo de regressão estimado e calcular os valores $y_{(i)}^{(0)}$. Observe, tabela 4.10, que $y_{(i)}^{(0)}$ não obedece à restrição (4.2.1.). Efetua-se, então, uma regressão isotônica nesta variável, obtendo-se $y^{(0)}$, que será entendida como o valor inicial da "preferência".

O próximo passo consiste no cálculo do "stress".

Em seguida, considerando o "stress", como uma função de β , calcula-se o seu gradiente e, através de algum método numérico, por exemplo Kruskal(1964b), modifica-se o valor $\beta_{pd}^{(0)}$, obtendo-se $\beta_{pd}^{(1)}$.

TABELA 4.8. Dados iniciais para análise

estímulo	r	X		
1	3,5	1	1	1
2	1,5	1	1	-1
3	2,5	1	-1	1
4	0,5	1	-1	-1
5	-1,5	-1	1	1
6	-0,5	-1	1	-1
7	-3,5	-1	-1	1
8	-2,5	-1	-1	-1

TABELA 4.9. Parâmetros estimados

atributo	$\beta^{(0)}$	$\beta_{pd}^{(0)}$	$\beta_{pd}^{(1)}$	$\beta_{pd}^{(2)}$	$\beta_{pd}^{(3)}$	$\beta_{pd}^{(4)}$	$\beta_{pd}^{(5)}$	$\beta_{pd}^{(6)}$
Winch.	2,000	1,611	1,627	1,659	1,704	1,729	1,732	1,732
RAM	1,750	0,604	0,563	0,473	0,294	0,103	0,021	0,004
Clock	0,250	0,201	0,188	0,158	0,098	0,034	0,007	0,001

TABELA 4.10. Valores ajustados

r	$y_i^{(0)}$	$y^{(0)}$	$y^{(1)}$	$y^{(2)}$	$y^{(3)}$	$y^{(4)}$	$y^{(5)}$	$y^{(6)}$
3,5	2,416	2,416	2,378	2,290	2,096	1,866	1,760	1,737
2,5	1,208	1,611	1,627	1,659	1,704	1,729	1,732	1,732
1,5	2,013	1,611	1,627	1,659	1,704	1,729	1,732	1,732
0,5	0,805	0,805	0,877	1,027	1,312	1,591	1,703	1,727
-1,5	-0,805	-1,007	-1,065	-1,185	-1,410	-1,626	-1,711	-1,728
-0,5	-2,416	-1,007	-1,065	-1,185	-1,410	-1,626	-1,711	-1,728
-2,5	-1,208	-2,215	-2,190	-2,132	-1,998	-1,832	-1,753	-1,736
-3,5	-2,416	-2,215	-2,190	-2,132	-1,998	-1,832	-1,753	-1,736

A partir deste momento o processo entra em recorrência: tendo $\beta_{pd}^{(1)}$ estima-se $y_{pd}^{(1)}$ através de (4.2.3.); aplica-se regressão isotônica encontrando $y^{(1)}$; calcula-se o "stress" e o gradiente; modifica-se $\beta_{pd}^{(1)}$ e assim por diante, até que seja obedecida alguma regra de parada.

No exemplo apresentado, o procedimento convergiu na sexta iteração. A regra de parada baseou-se na diminuição do "stress". Desta forma, $\beta_{pd}^{(6)}$ seria o vetor de parâmetros estimado e $y^{(6)}$ a "função preferência" estimada.

O valor do "stress" de 0,5% leva a acreditar no sucesso da análise. Resta, agora, interpretar o resultado final.

O valor estimado da preferência para cada nível dos atributos recebe o nome de utilidade (U). Este valor será dado através do valor estimado do parâmetro correspondente a cada fator. Assim, tem-se:

$$U(A_2) = -U(A_1) = 1,732$$

$$U(B_2) = -U(B_1) = 0,004$$

$$U(C_2) = -U(C_1) = 0,001.$$

Repare que a preferência por um estímulo é dada através da soma das utilidades dos níveis dos atributos que o compõem; isto permite que o modelo (4.2.3.) seja reescrito como:

$$y_j = U(A_j) + U(B_j) + U(C_j).$$

A importância relativa de cada atributo foi calculada na seção 1.3.

4.3. LINMAP - LINear MAPping

Este procedimento foi desenvolvido para o caso em que os estímulos são apresentados aos pares e o juiz deve escolher o preferido em cada um. O método pode ser adaptado para qualquer modelo de preferência. Será apresentado o desenvolvimento para modelo vetor (que é idêntico ao do modelo utilidade), Srinivasan & Shocker(1973b), e para modelo ponto ideal ponderado, Srinivasan & Shocker(1973a).

A estimação do modelo é realizada através de programação linear (método SIMPLEX), Yoshida(1987). Desse modo, é necessário definir uma função objetivo (função que se pretende minimizar) e um conjunto de restrições lineares.

Seja $J = \{1, 2, \dots, n\}$, o conjunto dos estímulos descritos em termos de p atributos e $P = \{1, 2, \dots, p\}$, o conjunto desses atributos.

Por facilidade de notação a avaliação subjetiva é representada por um conjunto de pares ordenados

$$\Omega = \{(j, k) : \text{o estímulo } j \text{ é o preferido em uma comparação com o estímulo } k\}.$$

Considere $y_j = \sum_{i \in P} \beta_i x_{ji}$. Para todo par ordenado $(j, k) \in \Omega$ diz-se que uma solução é consistente se

$$y_j \geq y_k. \quad (4.3.1)$$

Caso fosse considerada uma avaliação via postos, a expressão acima corresponderia à restrição de ordem do procedimento MONANOVA, quando os postos são atribuídos aos estímulos em ordem crescente de preferência, já que, por (4.3.1), quanto maior for o valor de y_j , maior será a preferência pelo estímulo.

Ao se fazer uma avaliação pareada é possível que ocorram situações de intransitividade, ou seja (j, k) , (k, l) e (l, j) ; esses casos podem ser atribuídos ao fato de que muitas vezes a preferência por estímulos diferentes é muito parecida, o que dificulta o trabalho do entrevistado. O método a ser apresentado não é prejudicado por estas ocorrências.

Dada a localização dos estímulos e um conjunto de pares ordenados Ω , o objetivo da análise é determinar a solução $\beta^* = (\beta_1^* \ \beta_2^* \ \dots \ \beta_p^*)$, para a qual as condições (4.3.1)

sofram o mínimo de transgressões. Assim sendo, Srinivasan & Shocker(1973a) definem como função objetivo o índice de falta de ajuste abaixo

$$B = \sum_{(j,k) \in \Omega} \delta_{jk}(y_k - y_j) \quad \text{onde}$$

$$\delta_{jk} = \begin{cases} 0 & \text{se } y_j \geq y_k \\ 1 & \text{se } y_j < y_k. \end{cases}$$

Ao considerar o par (j,k) , se $y_j < y_k$, então $(y_k - y_j)$ pode ser interpretado como uma parcela do erro atribuído à solução.

De modo análogo, uma medida de qualidade de ajuste pode ser definida como

$$G = \sum_{(j,k) \in \Omega} \delta_{kj}(y_j - y_k).$$

Tanto B quanto G são maiores ou iguais a zero.

4.3.1. Aplicado ao modelo vetor (ou utilidade)

Como foi visto anteriormente, o modelo vetor pode ser escrito como

$$y_j = \sum_{i \in P} \beta_i x_{ji}. \quad (4.3.2)$$

A solução do problema será representada por β^* .

O programa linear pretende encontrar β que minimize B . Isso gera um inconveniente, pois para a solução $\beta^* = 0$, a função objetivo atinge o valor mínimo zero. Visando evitar a solução trivial, impõe-se seguinte restrição linear: $G - B = h$, onde h é um número positivo arbitrário. Posteriormente será discutido o fato da escolha de h não alterar, em termos relativos, a solução do programa linear.

Note que

$$G - B = \sum_{(j,k) \in \Omega} \delta_{kj}(y_j - y_k) - \sum_{(j,k) \in \Omega} \delta_{jk}(y_k - y_j) = \sum_{(j,k) \in \Omega} (y_j - y_k) ,$$

o que reduz a restrição anterior a

$$\sum_{(j,k) \in \Omega} (y_j - y_k) = h. \quad (4.3.3)$$

Substituindo-se (4.3.2), tem-se

$$y_j - y_k = \sum_{i \in P} \beta_i x_{ji} - \sum_{i \in P} \beta_i x_{ki} = \sum_{i \in P} \beta_i (x_{ji} - x_{ki}).$$

Considere agora as constantes

$$a_{jki} = x_{ji} - x_{ki} \quad \text{e} \quad A_i = \sum_{(j,k) \in \Omega} a_{jki} = \sum_{(j,k) \in \Omega} (x_{ji} - x_{ki}).$$

Assim, a restrição (4.3.3) fica sendo

$$\sum_{i \in P} \beta_i A_i = h. \quad (4.3.4)$$

Reescrevendo B com as constantes definidas acima, chega-se a

$$B = \sum_{(j,k) \in \Omega} \delta_{jk}(y_k - y_j) = - \sum_{(j,k) \in \Omega} \delta_{jk} \sum_{i \in P} \beta_i a_{jki} = \sum_{(j,k) \in \Omega} z_{jk} \quad (4.3.5)$$

onde, $z_{jk} = -\delta_{jk} \sum_{i \in P} \beta_i a_{jki}$.

Quando $(y_k - y_j)$ é maior que zero, tem-se que $\delta_{jk} = 1$ e, conseqüentemente, $z_{jk} = y_k - y_j$; caso contrário, $z_{jk} = 0$. Desse modo, uma segunda restrição linear pode ser escrita como

$$z_{jk} \geq y_k - y_j = - \sum_{i \in P} \beta_i a_{jki}.$$

Uma restrição que pode ser imposta ao modelo refere-se ao sinal de β_i . Cabe ao analista decidir se vale a pena forçar que β_i seja sempre positivo, ou se é melhor considerá-lo livre de sinal.

Partindo das restrições anteriores, chega-se ao problema de programação linear:

Dado o conjunto de constantes $\{a_{jki} \text{ e } A_i : (j, k) \in \Omega \text{ } i \in P\}$, encontrar β^* e z_{jk}^* , $(j, k) \in \Omega$, tal que

$$\text{a função } \{F = \sum_{(j,k) \in \Omega} z_{jk}\} \text{ seja mínima, sujeita a} \quad (4.3.6)$$

$$\sum_{i \in P} a_{jki} \beta_i + z_{jk} \geq 0 \quad \forall (j, k) \in \Omega \quad (4.3.7)$$

$$\sum_{i \in P} A_i \beta_i = h \quad (4.3.8)$$

$$z_{jk} \geq 0 \quad \forall (j, k) \in \Omega. \quad (4.3.9)$$

Às vezes, à critério do analista, acrescenta-se

$$\beta_i \geq 0 \quad \forall i \in P \quad (4.3.10)$$

Considere o expoente * como indicador da solução ótima do programa linear, ou seja, F^* e B^* são, respectivamente, os valores mínimos de F e B . Está demonstrado no apêndice 5 que:

- A solução do programa linear realmente minimiza B , ou seja, $F^* = B^*$ (teorema A5.1) e
- quando se utiliza o método SIMPLEX sempre existirá uma solução ótima finita (teorema A5.2).

4.3.1.1. Exemplo

Considere os dados da tabela 4.11, onde se descreve a avaliação de 4 estímulos definidos pela variação de 2 atributos, X_1 e X_2 . A avaliação dos perfis é dada através de postos, sendo a ordenação conduzida do mais para o menos preferido.

A primeira dificuldade encontrada na utilização do procedimento LINMAP a esses dados está exatamente no tipo de avaliação realizada, já que o método foi desenvolvido pensando-se em comparações pareadas. Um modo de contornar esse problema é utilizar a ordem definida pelos postos para formar os pares ordenados de Ω , assim, como o estímulo 2 é, dentre os quatro, o preferido, tem-se os seguintes pares ordenados: (2, 1), (2, 3) e (2, 4); o estímulo 1 recebeu o segundo menor posto, originando assim os seguintes pares: (1, 3) e (1, 4); por fim tem-se que o estímulo 4 recebeu posto menor que o 3 originando: (4, 3). Assim procedendo, o conjunto Ω fica sendo

$$\Omega = \{(2, 1); (2, 3); (2, 4); (1, 3); (1, 4); (4, 3)\}.$$

Para a resolução do programa linear é necessário o cálculo das constantes a_{jki} e A_i . Os resultados dessas operações estão apresentados na tabela 4.12. Com esses valores, o programa linear fica sendo

$$\text{minimizar } F = z_{21} + z_{23} + z_{24} + z_{13} + z_{14} + z_{43} \quad \text{sujeito a}$$

$$-\beta_1 + 4\beta_2 + z_{21} \geq 0 \quad 3\beta_1 - 2\beta_2 + z_{23} \geq 0$$

$$-2\beta_1 - 8\beta_2 + z_{24} \geq 0 \quad 4\beta_1 - 6\beta_2 + z_{13} \geq 0$$

$$-1\beta_1 - 12\beta_2 + z_{14} \geq 0 \quad 5\beta_1 + 6\beta_2 + z_{43} \geq 0$$

$$8\beta_1 - 18\beta_2 = 1 \quad z_{jk} \geq 0 \quad \forall (j, k) \in \Omega.$$

Usando-se o método SIMPLEX, chega-se a $\beta_1^* = 0.08$, $\beta_2^* = -0.02$, $z_{21}^* = 0.16$, e que os z_{jk}^* restantes são iguais a zero, sendo então, $F^* = 0.16$.

A tabela 4.13 apresenta os valores da variável y . Lembre que pela própria construção do programa, quanto maior for o valor de y_j , maior será a preferência pelo estímulo; contudo, os postos foram atribuídos em ordem decrescente de preferência, ou seja, quanto menor o posto, maior é a preferência pelo perfil. Para que haja uma correspondência entre a ordem de R e y , considere o modelo: $\tilde{y} = -y$. Comparando a coluna correspondente a $\tilde{y}$ com a coluna R , nota-se que há inversão de ordem entre os estímulos 2 e 1, isso faz com que $\delta_{21} = 1$. Utilizando a fórmula de B chega-se no valor 0.16, idêntico, como era esperado, ao valor de F^* .

4.3.2. Aplicado ao modelo ponto-ideal ponderado

Como foi visto em tópicos anteriores, o modelo ponto ideal é dado por

$$y_j = \sum_{i \in P} \beta_i (x_{ji} - \pi_i)^2 = \sum_{i \in P} d_{ji}^2 \quad j \in J.$$

d_{ji}^2 mede a distância entre o atributo i e seu respectivo ponto ideal, logo quanto menor for y_j maior será a preferência pelo j -ésimo estímulo.

Um programa linear análogo ao anterior foi desenvolvido para esses modelos.

Nesse caso, o objetivo é encontrar a solução ótima (β^*, Π^*) , $\Pi^* = (\pi_i^*) \quad i \in P$, que minimize B sujeito a uma série de restrições lineares, análogas às do modelo vetor.

O desenvolvimento do programa seguinte encontra-se no apêndice 6.

TABELA 4.11. Dados iniciais para análise

estímulo	atributos		r
	X_1	X_2	
1	6	3	2
2	5	7	1
3	2	9	4
4	7	15	3

TABELA 4.12. Constantes do programa linear do modelo vetor, relativas aos dados da tabela 4.11

Par Ordenado	atributos	
	X_1	X_2
a_{21i}	-1	4
a_{23i}	3	-2
a_{24i}	-2	-8
a_{13i}	4	-6
a_{14i}	-1	-12
a_{43i}	5	6
A_i	8	-18

TABELA 4.13. Exemplo de aplicação do método LINMAP ao modelo VETOR

estímulo	atributos		avaliação r	avaliação	
	X_1	X_2		y	$\tilde{y}$
1	6	3	2	0,420	-0,420
2	5	7	1	0,260	-0,260
3	2	9	4	-0,020	0,020
4	7	15	3	0,260	-0,260

Dado o conjunto de constantes $\{a'_{jki} = x_{ji}^2 - x_{ki}^2; A'_i = \sum_{(j,k) \in \Omega} a'_{jki} \quad b_{jki} = -2(x_{ji} - x_{ki}) \quad e \quad D_i = \sum_{(j,k) \in \Omega} b_{jki} : (j,k) \in \Omega \quad i \in P\}$, encontrar β^* e Π^* que

$$\text{minimize a função } F = \sum_{(j,k) \in \Omega} z_{jk}, \quad \text{sujeita a} \quad (4.3.11)$$

$$\sum_{i \in P} a'_{jki} \beta_i + \sum_{i \in P} b_{jki} v_i + z_{jk} \geq 0, \quad (j,k) \in \Omega \quad (4.3.12)$$

$$z_{jk} \geq 0, \quad (j,k) \in \Omega \quad (4.3.13)$$

$$\sum_{i \in P} A'_i \beta_i + \sum_{i \in P} D_i v_i = h \quad (4.3.14)$$

$$\beta_i \geq 0, \quad i \in P \quad \text{a critério do analista.} \quad (4.3.15)$$

Nesse programa, A'_i , D_i , a'_{jki} e b_{jki} são constantes e $v_i = \beta_i \pi_i$.

Demonstra-se que:

- $F^* = B^*$
- existe uma solução ótima finita ao se utilizar o método SIMPLEX.

Essas demonstrações foram suprimidas por serem semelhantes às do modelo vetor e por estarem dispostas de modo detalhado em Srinivasan & Shocker(1973a).

Existe uma correspondência entre as soluções (β, Π) e (β, V) , $V = (v_i)$. Para passar de uma para outra deve-se considerar os seguintes casos ($\forall i \in P$) :

- se $\beta_i \neq 0$ e v_i qualquer, então $\pi_i = \frac{v_i}{\beta_i}$;
- se $\beta_i = 0$ e $v_i = 0$, define-se $\pi_i = 0$;
- se $\beta_i = 0$ e $v_i \neq 0$, é o caso de um modelo degenerado onde $\pi_i = \pm\infty$, adotando o sinal de v_i . Neste caso, quanto maior (ou menor) for x_{ji} maior será a preferência, ou seja

se comportaria como se fosse um modelo vetor. Esta configuração não permite o cálculo de y_j . Visando contornar este problema, Srinivasan & Shocker(1973a) sugerem que seja considerado um modelo modificado, derivado diretamente da equação A6.1. Trata-se, na verdade, de um modelo misto, onde para aqueles atributos com ponto ideal infinito deve-se calcular o modelo vetor. Ter-se-ia assim

$$y_j = \sum_{i \in P_1} \beta_i (x_{ji} - \pi_i)^2 - 2 \sum_{i \in P_2} v_i x_{ji}, \quad (4.3.16)$$

onde $P_1 \subseteq P$, $\beta_i \neq 0$ ou $\beta_i = v_i = 0$ e $P_2 = P - P_1$.

4.3.2.1. Exemplo

Utilizando novamente os dados da tabela 4.11 é construído a seguir um programa linear para estimação de um modelo ponto-ideal ponderado. O conjunto Ω está descrito na seção 4.3.1.1.; com as informações nele contidas, pode-se proceder o cálculo das constantes do programa linear. Os valores dessas constantes encontram-se na tabela 4.14. Assim, o programa linear para esse problema é

minimizar $F = z_{21} + z_{23} + z_{24} + z_{13} + z_{14} + z_{34}$ sujeita a

$$-11\beta_1 + 40\beta_2 + 2v_1 - 8v_2 + z_{21} \geq 0$$

$$21\beta_1 - 32\beta_2 - 6v_1 + 4v_2 + z_{23} \geq 0$$

$$-24\beta_1 - 176\beta_2 + 4v_1 + 16v_2 + z_{24} \geq 0$$

$$32\beta_1 - 72\beta_2 - 8v_1 + 12v_2 + z_{13} \geq 0$$

$$-13\beta_1 - 216\beta_2 + 2v_1 + 24v_2 + z_{14} \geq 0$$

$$45\beta_1 + 144\beta_2 - 10v_1 - 12v_2 + z_{43} \geq 0$$

$$50\beta_1 - 312\beta_2 - 16v_1 + 36v_2 = 1$$

$$z_{jk} \geq 0 \quad (j, k) \in \Omega.$$

Quanto à solução do programa linear, tem-se que $z_{jk}^* = 0 \quad \forall (j, k) \in \Omega$, resultando $B^* = 0$. Este fato pode ser constatado da tabela 4.15 Quando do ajuste do modelo vetor a esses mesmos dados, obteve-se $B^* = 0.16$; a diminuição no valor desse índice deve-se a uma melhoria no ajuste ao se passar para o modelo ponto-ideal ponderado.

Quanto aos parâmetros do modelo tem-se: para o atributo X_1 , $\beta_1^* = 0.0103$ e $v_1^* = 0$, o que leva a um ponto ideal igual a zero; já para o atributo X_2 , $\beta_2^* = -0.0075$ e $v_2^* = -0.0518$, o que leva a um ponto ideal $\pi_2 = 6,9067$. Assim o modelo final passa a ser

$$y_j = 0.0103X_1^2 - 0.0075(X_2 - 6.9067)^2 = -\tilde{y}.$$

4.3.3. Aplicado ao modelo ponto-ideal geral

Como foi visto na equação (3.2.4), o modelo ponto ideal geral, em sua versão II é

$$y_j = (\mathbf{x}_j - \mathbf{\Pi})' \mathbf{\Lambda} (\mathbf{x}_j - \mathbf{\Pi}) \quad \text{onde: } \mathbf{\Lambda} = \mathbf{T} \mathbf{\beta}^m \mathbf{T}'.$$

Foi visto em (4.1.7) que

$$y_j = \sum_{i \in P} \lambda_{ii} x_{ji}^2 + \sum_{i \in P} \alpha_i x_{ji} + \sum_{i \in P} \sum_{l < i} 2\lambda_{il} x_{ji} x_{jl}.$$

A partir da equação anterior, encontra-se, no apêndice 6, o desenvolvimento do método LINMAP para esse modelo. O programa linear resultante cujo objetivo é encontrar a solução $(\mathbf{\Lambda}^*, \mathbf{\alpha}^*)$ que minimize o índice B , é dado por:

Dado o conjunto de constantes $\{a_{jki} = x_{ji} - x_{ki}, A_i = \sum_{(j,k) \in \Omega} a_{jki}, a'_{jki}, A'_i = \sum_{(j,k) \in \Omega} a'_{jki}, c_{jkil} = x_{ji}x_{jl} - x_{ki}x_{kl} \text{ e } E_{il} = \sum_{(j,k) \in \Omega} c_{jkil} : (j,k) \in \Omega, i \in P\}$, encontrar $\Lambda = (\lambda_{ij})$ e $\alpha = (\alpha_i)$, tal que minimiza a função

$$F = \sum_{(j,k) \in \Omega} z_{jk} \quad \text{sujeita a} \quad (4.3.17)$$

$$\sum_{i \in P} a'_{jki} \lambda_{ii} + \sum_{i \in P} a_{jki} \alpha_i + \sum_{i \in P} \sum_{l < i} c_{jkil} \lambda_{il} + z_{jk} \geq 0 \quad (j,k) \in \Omega \quad (4.3.18)$$

$$\sum_{i \in P} A'_i \lambda_{ii} + \sum_{i \in P} A_i \alpha_i + \sum_{i \in P} \sum_{l < i} E_{il} \lambda_{il} = h \quad (4.3.19)$$

$$z_{jk} \geq 0 \quad (j,k) \in \Omega. \quad (4.3.20)$$

Dada a semelhança entre as equações (4.1.7) e (4.1.12), o modelo ponto-ideal geral pode ser estimado em sua versão I, através do mesmo programa linear acrescido da restrição $\lambda_{ii} > 0$, uma vez que a raiz quadrada desse parâmetro deve ser extraída em (4.1.13).

De (4.1.8) tira-se que

$$\hat{\Pi} = -\frac{1}{2} \Lambda^{-1} \alpha$$

e, como $\Lambda = T' \beta T$, as matrizes T e β podem ser obtidas através da decomposição espectral de Λ , como discutido no capítulo 4.1.

4.3.3.1. Exemplo

A partir dos dados da tabela 4.11, calculou-se as constantes c_{jkil} e E_{il} , do programa linear acima; as outras já foram calculadas nos tópicos anteriores. Assim, para esse exemplo, tem-se o seguinte programa linear:

TABELA 4.14. Constantes do programa linear do modelo ponto-ideal ponderado, relativas aos dados da tabela 4.11

Par Ordenado	X_1		X_2	
	a_{jk1}	b_{jk1}	a_{jk2}	b_{jk2}
(2,1)	-11	2	40	-8
(2,3)	21	-6	-32	4
(2,4)	-24	4	-176	16
(1,3)	32	-8	-72	12
(1,4)	-13	2	-216	24
(4,3)	45	-10	144	-12
	$A_1 = 15 \quad D_1 = 15$		$A_2 = -312 \quad D_2 = 36$	

TABELA 4.15. Exemplo de aplicação do método LINMAP ao modelo Ponto Ideal Ponderado

estímulo	atributos		avaliação r	avaliação	
	X_1	X_2		y	$\tilde{y}$
1	6	3	2	0,256	-0,256
2	5	7	1	0,257	-0,257
3	2	9	4	0,008	-0,008
4	7	15	3	0,013	-0,013

TABELA 4.16. Exemplo de aplicação do método LINMAP ao modelo Ponto-Ideal Geral

estímulo	atributos		avaliação r	avaliação	
	X_1	X_2		y	$\tilde{y}$
1	6	3	2	0,369	-0,369
2	5	7	1	0,549	-0,549
3	2	9	4	0,035	-0,035
4	7	15	3	1,292	-1,292

minimizar a função $F = z_{21} + z_{23} + z_{24} + z_{13} + z_{14} + z_{43}$ sujeita a:

$$-11\lambda_{11} + 40\lambda_{22} - \alpha_1 + 4\alpha_2 + 17\lambda_{12} + z_{21} \geq 0$$

$$21\lambda_{11} - 32\lambda_{22} + 3\alpha_1 - 2\alpha_2 + 17\lambda_{12} + z_{23} \geq 0$$

$$-24\lambda_{11} - 176\lambda_{22} - 2\alpha_1 - 8\alpha_2 - 70\lambda_{12} + z_{24} \geq 0$$

$$32\lambda_{11} - 72\lambda_{22} + 4\alpha_1 - 6\alpha_2 + z_{13} \geq 0$$

$$-13\lambda_{11} - 216\lambda_{22} - \alpha_1 - 12\alpha_2 - 87\lambda_{12} + z_{14} \geq 0$$

$$45\lambda_{11} + 144\lambda_{22} + 5\alpha_1 + 6\alpha_2 + 87\lambda_{12} + z_{43} \geq 0$$

$$50\lambda_{11} - 312\lambda_{22} + 8\alpha_1 - 18\alpha_2 - 36\lambda_{12} = 1.$$

Como no modelo anterior, $z_{jk} = 0 \quad \forall (j, k) \in \Omega$, indicando um ajuste perfeito ($B^* = 0$). Já os parâmetros estimados são

$$\Lambda = \begin{pmatrix} 0.00070361 & 0.01061272 \\ 0.01061272 & -0.00431692 \end{pmatrix} \quad \alpha = 0 \Rightarrow \Pi = 0.$$

O modelo estimado fica sendo

$$y_j = 0.00070361x_1^2 - 0.00431692x_2^2 + 0.0212254x_1x_2.$$

Os valores previstos para esse modelo encontram-se na tabela 4.16.

4.3.4. Comentários

É fácil ver que se o primeiro membro de (4.3.8), (4.3.14) e (4.3.19) for igualado a um outro número q , tem-se que: $F^*(q) = \frac{q}{h} F^*(h)$, ou seja, $B^*(q) = \frac{q}{h} B^*(h)$ e além disso $\beta^*(q) = \frac{q}{h} \beta^*(h)$ e $V^*(q) = \frac{q}{h} V^*(h)$.

O parágrafo anterior garante que os pesos ótimos relativos serão os mesmos. Logo, por simplicidade, pode-se adotar $h = 1$, como foi feito nos exemplos anteriores, sem prejuízo dos resultados da análise.

Srinivasan & Shocker(1973a) sugerem a utilização de um novo índice, C , para medir o ajuste do modelo. Este coeficiente é sugerido devido às suas boas propriedades e é dado por

$$C = \frac{B}{h + B}.$$

Suas principais propriedades são:

a. Como se trata de uma transformação monotônica de B , qualquer solução (β, V) (ou β no caso de modelo vetor ou utilidade) que otimize B também otimizará C . Será, então, considerado o índice ótimo como sendo

$$C^* = \frac{B^*}{h + B^*}.$$

b. Ao igualar o primeiro membro de (4.7) a um valor q ao invés de h , C^* não se alterará.

prova:

$$C^*(q) = \frac{B^*(q)}{q + B^*(q)} = \frac{(\frac{q}{h})B^*(h)}{q + (\frac{q}{h})B^*(h)} = \frac{B^*}{h + B^*(h)} = C^*(h) \quad \blacksquare$$

c. $B^* \geq 0 \Rightarrow 0 \leq C^* < 1$. Essa propriedade permite, em termos de interpretação, fazer uma analogia entre esse índice e um índice de "stress".

d. Utilizando o fato de que $G - B = h$, tem-se

$$C^* = \frac{B^*}{h + B^*} = \frac{B^*}{G^* - B^* + B^*} = \frac{B^*}{G^*};$$

isso leva à afirmação de que C^* independe do número de pares em Ω , uma vez que este influenciaria tanto B^* quanto G^* .

Para os exemplos descritos nas seções 4.3.1.1., 4.3.2.1. e 4.3.3.1. tem-se que C^* é respectivamente 0.14, 0 e 0, indicando um ajuste perfeito nos dois últimos casos.

Mullet & Karson(1986) realizaram um trabalho de simulação visando conhecer o comportamento do índice C^* , quando utilizado na estimação de um modelo vetor, sob diferentes planejamentos fracionários (38 ao todo). Para cada planejamento foram geradas 2000 diferentes avaliações e para cada uma delas foi ajustado o modelo e calculado C^* . Isso feito, calculou-se para cada planejamento alguns percentis da distribuição de C^* , sua média e coeficiente de assimetria.

É possível encontrar mais de uma solução ótima para o programa. Nesses casos Srinivasan & Shocker(1973a) sugerem alguns métodos de escolha, entre eles, escolher aquela solução que violar o menor número de inequações.

Os pesos, β_i não refletem apenas a importância dos atributos, mas também suas unidades de medida. Para evitar esta dupla interferência sugere-se utilizar $x'_{ji} = \frac{x_{ji} - \bar{x}_i}{\sigma_i}$. Desta forma os pesos β'_i medirão apenas a importância da dimensão. Pode ser provado, como consequência do teorema A5.3, que, ao utilizarmos x'_{ji} , há uma solução ótima com $\beta'_i = \sigma_i^2 \beta_i$ e além disso para o modelo ponto-ideal ponderado $\pi'_i = \frac{\pi_i - \bar{x}_i}{\sigma_i}$.

Esta análise é desenvolvida a nível individual, os juízes terão diferentes pesos e pontos ideais para os atributos. Para que se possa comparar os pesos dos diferentes juízes, a fim de saber se o atributo tem ou não a mesma importância para todos, aconselha-se que os pesos sejam padronizados de modo que, por exemplo, sua soma seja sempre 1 para cada juiz.

Srinivasan & Shocker(1973a) apresentam também alguns métodos alternativos de ajuste. Um deles consiste em minimizar o número de violações da desigualdade $y_k \geq y_j$, $(j, k) \in \Omega$. Sugerem ainda um programa linear aplicado ao caso em que podemos não ter preferência por alguns pares.

4.4 Significância de atributos

Em Kohli(1988), encontram-se procedimentos para o cálculo da significância de atributos de um modelo utilidade, onde a avaliação dos estímulos é dada através de sua ordenação e onde se aplicou um planejamento perfil-completo.

Os procedimentos propostos baseiam-se em técnicas não paramétricas utilizando, sobretudo, resultados obtidos no desenvolvimento dos testes de Kruskal-Wallis e Friedman (Conover(1980)).

A preocupação com o estudo da significância se justifica pelos seguintes fatos:

a. Em muitas situações práticas, uma AP é seguida por uma análise de conglomerados nos parâmetros estimados para cada indivíduo. Dessa forma, saber quais atributos são significantes em cada agrupamento pode auxiliar a caracterização dos segmentos.

b. Outra utilidade da AP é o fato de permitir a realização de simulações no sentido de se prever a preferência dos juízes por estímulos que não são avaliados no planejamento. Assim, se é sabido quais atributos não são significantes em um grupo de indivíduos, estes fatores podem ser ignorados, o que, em alguns casos, resulta em uma economia de tempo de processamento. Kohli(1988) comenta que, para simular a preferência por todos os possíveis estímulos em um problema envolvendo 200 entrevistados, oito atributos e, em cada atributo, cinco níveis, um computador FPS, cuja velocidade é 10 vezes maior que a de um DEC-10 e o dobro de um VAX-3600, gasta em torno de uma hora de processamento e, caso dois atributos fossem retirados, este tempo se reduziria a dois minutos.

$$8 \text{ atributos} \Rightarrow 200 \times 5^8 = 78.125.000 \text{ simulações}$$

$$6 \text{ atributos} \Rightarrow 200 \times 5^6 = 3.125.000 \text{ simulações}$$

c. A informação sobre quais atributos foram, ou não, significantes para uma dada população, além de facilitar a análise dos resultados, auxilia o planejamento de futuros estudos de AP.

Para o prosseguimento deste estudo, os atributos são divididos em duas classes:

- atributos que não possuem uma ordem de preferência dos níveis definida a priori, sem restrição de ordem (exemplo "design" de um produto) e,

- atributos que possuem uma ordem de preferência dos níveis definida a priori, com restrição de ordem, como no caso do fator "preço" (preços menores tendem a ser preferidos)².

4.4.1 Procedimento 1 - Para atributos que não sofrem restrições de ordem a priori

Considere a seguinte notação:

n_j , $j = 1, 2, \dots, q$ número de vezes que o nível j aparece entre os estímulo;

$x_l^j, l = 1, 2, \dots, n_j$ estímulos nos quais o nível j do atributo aparece ;

$r_s(x_l^j)$ posto que o indivíduo s associou ao estímulo x_l^j e

$$r_{js} = \sum_{l=1}^{n_j} r_s(x_l^j).$$

Para o desenvolvimento do método, é suposto inicialmente que a preferência do indivíduo s não difere ao longo dos q níveis do atributo. Caso o atributo não seja significativo, é de se esperar que não exista uma associação aleatória entre os níveis do atributo e o ordenamento por postos dos estímulos. Sob esta hipótese, cada posto é atribuído a um estímulo com probabilidade $\frac{1}{n}$. Desta forma, verificam-se os seguintes resultados:

Lema. Havendo associação aleatória entre os níveis do atributo e o ordenamento por postos dos estímulos, tem-se que:

a-) $E[r_s(x_l^j)] = \frac{n+1}{2},$

b-) $\text{Var}[r_s(x_l^j)] = \frac{(n+1)(n-1)}{12},$

c-) $\text{Cov}[r_s(x_l^j), r_s(x_{l'}^j)] = -\frac{(n+1)}{12} \quad l \neq l'.$

prova:

a-) Sob a hipótese de associação aleatória tem-se que cada posto é atribuído a um perfil (estímulo) com probabilidade $\frac{1}{n}$, então:

² Srinivasan, Jain & Malhotra(1983), constataram que quando é possível, a priori, estabelecer restrições de ordem na preferência dos níveis de um atributo e essas restrições são impostas no procedimento de estimação do modelo, aumenta-se a capacidade preditiva do modelo estimado

$$E[r_s(\mathbf{x}_1^j)] = \sum_{j=1}^q \sum_{l=1}^{n_j} r_s(\mathbf{x}_1^j) \frac{1}{n} = \frac{1}{n} \sum_{j=1}^q \sum_{l=1}^{n_j} r_s(\mathbf{x}_1^j).$$

Como o juiz associa um único posto a cada estímulo, a soma dos postos ao longo destes perfis será:

$$\sum_{j=1}^q \sum_{l=1}^{n_j} r_s(\mathbf{x}_1^j) = \frac{n(n+1)}{2}.$$

Substituindo na equação anterior, chega-se a:

$$E[r_s(\mathbf{x}_1^j)] = \frac{(n+1)}{2} \quad \blacksquare$$

b-) A demonstração é suprimida por ser análoga à do item a-).

c-) Partindo-se da suposição de associação aleatória, a probabilidade de que os perfis $\mathbf{x}_1^j$ e $\mathbf{x}_1^{j'}$ sejam atribuídos aos postos $r_s(\mathbf{x}_1^j)$ e $r_s(\mathbf{x}_1^{j'})$, respectivamente, é $\frac{1}{n(n-1)}$. Desta forma:

$$\text{Cov}[r_s(\mathbf{x}_1^j); r_s(\mathbf{x}_1^{j'})] = E\{[r_s(\mathbf{x}_1^j) - E[r_s(\mathbf{x}_1^j)]] [r_s(\mathbf{x}_1^{j'}) - E[r_s(\mathbf{x}_1^{j'})]]\} =$$

$$\sum_{t=1}^n \sum_{u \neq t}^n \left\{t - \frac{n+1}{2}\right\} \left\{u - \frac{n+1}{2}\right\} \frac{1}{n(n-1)}.$$

Adicionando e subtraindo o termo correspondente a $t=u$, vem que:

$$\text{Cov}[r_s(\mathbf{x}_1^j); r_s(\mathbf{x}_1^{j'})] = \sum_{t=1}^n \sum_{u=1}^n \left\{t - \frac{n+1}{2}\right\} \left\{u - \frac{n+1}{2}\right\} \frac{1}{n(n+1)} -$$

$$- \sum_{t=1}^n \left\{t - \frac{(n+1)}{2}\right\}^2 \frac{1}{n(n-1)} =$$

$$= \left\{ \sum_{t=1}^n \left[t - \frac{n+1}{2} \right] \sum_{u=1}^n \left[u - \frac{n+1}{2} \right] \frac{1}{n(n-1)} - \sum_{t=1}^n \left\{ t - \frac{n+1}{2} \right\}^2 \frac{1}{n(n-1)} \right\}.$$

$\sum_{t=1}^n \left[t - \frac{n+1}{2} \right] = 0$ e, além disso

$$\sum_{t=1}^n \left[t - \frac{n+1}{2} \right]^2 \frac{1}{n(n-1)} = \frac{1}{n-1} \text{Var}[r_s(\mathbf{x}_1^j)] = \frac{n+1}{12}.$$

Logo

$$\text{Cov}[r_s(\mathbf{x}_1^j), r_s(\mathbf{x}_{l'}^j)] = -\frac{(n+1)}{12} \quad l \neq l' \quad \blacksquare$$

Teorema 4.4.1. Havendo associação aleatória entre os níveis do atributo e o ordenamento por postos dos estímulos, tem-se que:

$$a-) E[r_{js}] = \frac{n_j(n+1)}{2}$$

$$b-) \text{Var}[r_{js}] = \frac{n_j(n+1)(n-n_j)}{12}$$

prova:

$$a-) E[r_{js}] = E\left[\sum_{l=1}^{n_j} r_s(\mathbf{x}_l^j)\right] = \sum_{l=1}^{n_j} E[r_s(\mathbf{x}_l^j)] \stackrel{\text{Lema}}{=} \frac{n_j(n+1)}{2}.$$

$$\begin{aligned} b-) \text{Var}(r_{js}) &= \sum_{l=1}^{n_j} \text{Var}[r_s(\mathbf{x}_l^j)] + \sum_{\mathbf{x}_1^j} \sum_{\mathbf{x}_{l'}^j, \mathbf{x}_1^j \neq \mathbf{x}_{l'}^j} \text{Cov}[r_s(\mathbf{x}_1^j); r_s(\mathbf{x}_{l'}^j)] \stackrel{\text{Lema}}{=} \\ &= \frac{n_j(n+1)(n-n_j)}{12} \quad \blacksquare \end{aligned}$$

O teorema de Wald Wolfowitz, Fraser(1957), permite mostrar que a distribuição da soma de n pontos aleatórios, independentes e identicamente distribuídos, converge em distribuição para uma distribuição normal. Aplicando este resultado a r_{js} , tem-se que, conforme n_j cresce, sua distribuição se aproxima de uma normal, desde que haja associação aleatória entre os estímulos e a atribuição de postos. Nestas condições:

$$\frac{r_{js} - E[r_{js}]}{\sqrt{\text{Var}[r_{js}]}} \xrightarrow{\mathcal{D}} N(0, 1)$$

$$c_{js} = \frac{\{r_{js} - E[r_{js}]\}^2}{\text{Var}[r_{js}]} \xrightarrow{\mathcal{D}} \chi^2_{(1)}.$$

Kruskal(1952) provou que se c_{js} for multiplicado por $\frac{(n-n_j)}{n}$, resulta:

$$c_s = \sum_{j=1}^q \frac{(n - n_j)}{n} c_{js} \xrightarrow{\mathcal{D}} \chi^2_{(q-1)}.$$

Pode-se provar que:

$$c_s = \frac{12}{n(n+1)} \sum_{j=1}^q \left\{ r_{sj} - \frac{n_j(n+1)}{2} \right\}^2 \frac{1}{n_j},$$

que é a estatística do teste de Kruskal-Wallis utilizado para verificar a identidade de distribuições populacionais.

Quando se trabalha a nível individual, n_j não será, na prática, um valor grande; logo, é preciso conhecer o comportamento de c_s nestas condições.

Em um estudo de simulação, desenvolvido por Kruskal & Wallis(1952), concluiu-se que para pequenos valores de n_j e para níveis de significância inferiores a 0,10, a aproximação χ^2 fornece um teste conservativo. Este resultado foi confirmado através de outros estudos posteriores (Feir-Walsh & Toothaker(1974) e Gabriel & Lachanbruch(1969)).

É bom ressaltar a existência da alternativa do cálculo da distribuição exata de c_s , quando o resultado obtido pela aproximação não parecer razoável.

Até agora, o teste foi desenvolvido a nível individual, mas, na maioria das vezes é interessante conhecer a significância do atributo de uma maneira global, levando em conta o conjunto de entrevistados. Para isto, sugere-se a utilização da estatística:

$$c = \sum_{s=1}^S c_s$$

Como c é a soma de S variáveis assintoticamente qui-quadrado, sua distribuição também será assintoticamente qui-quadrado, com $S(p - 1)$ graus de liberdade.

4.4.2. Procedimento 2 - Para atributos que sofrem restrições de ordem a priori

Neste caso, considera-se a hipótese de que os níveis dos atributos sigam uma ordem a priori. Assim, quanto mais frequentemente a restrição de ordem é violada, maiores são as evidências de que o atributo não é significativo.

O teste proposto baseia-se em $r_j = \sum_{s=1}^S r_{sj}$, pois, se todos os juízes têm o mesmo ordenamento de preferência a priori, poderão diferir apenas no grau em que preferem um nível ao outro. Em um contexto de planejamento de experimentos, os respondentes serão encarados como blocos, através dos quais a significância dos atributos é testada.

RESULTADO. Sob a hipótese de não associação entre os níveis dos atributos e o ordenamento dos postos, os $r_s(x_i^j)$ são inteiros entre 1 e $n = \sum_{j=1}^p n_j$ aleatoriamente selecionados

$$1-) E(r_j) = E[\sum_{s=1}^S r_{js}] = SE[r_{js}] = \frac{Sn_j(n+1)}{2}$$

$$2-) Var(r_j) = \sum_{s=1}^S Var(r_{js}) = \frac{Sn_j(n+1)(n-n_j)}{12}.$$

Como r_s é a soma de Sn_j variáveis aleatórias, o teorema de Wald-Wolfowitz implica em que r_j tenha distribuição assintótica normal. A distribuição assintótica está ancorada no crescimento de Sn_j . Em estudos comerciais de AP isto acontece, já que grandes amostras são empregadas. Desta forma:

$$h_j = \frac{r_j - E(r_j)}{\sqrt{Var(r_j)}} \xrightarrow{\mathcal{D}} N(0, 1)$$

$$h_j^2 \xrightarrow{\mathcal{D}} \chi_{(1)}^2.$$

Wilks, em Friedman(1937), mostra que:

$$h = \sum_{j=1}^q \frac{h_j^2(q-1)}{q} \xrightarrow{\mathcal{D}} \chi_{(q-1)}^2,$$

$$h = \frac{12(q-1)}{Sq(n-1)} \sum_{j=1}^q \frac{[r_j - \frac{1}{2}Sn_j(n+1)]^2}{n_j(n-n_j)}.$$

Note que quando $n_j = 1$ o teste é idêntico ao teste de Friedman, Conover(1980).

CAPÍTULO 5

Modelos para Dados Categorizados

Os modelos para dados categorizados podem ser aplicados quando a avaliação é realizada através de escalas ordinais ou nominais. Tratam-se, na verdade, dos tradicionais modelos de regressão para variáveis qualitativas PROBIT e LOGIT. Dada a existência de farta literatura sobre o assunto, a ênfase maior deste capítulo será para versões menos usuais desses modelos: PROBIT para variáveis ordinais (McKelvey & Zavoina(1975)) e LOGIT condicional (McFadden(1974)).

Nos tópicos 5.1. e 5.2., será omitido dos modelos o índice referente ao indivíduo.

5.1. Variável resposta dicotômica

Esse é o caso mais comum de utilização dos modelos LOGIT e PROBIT, onde a variável resposta segue uma distribuição de Bernoulli. Num contexto de AP pode-se ter o seguinte exemplo de variável resposta:

$$r_j = \begin{cases} 0 & \text{se não compraria o produto descrito pelo estímulo } j \\ 1 & \text{se compraria o produto} \end{cases}$$

Um possível tratamento seria estudar o modelo:

$$y_j = P(r_j = 1 \mid \mathbf{X}) = f_j[\mathbf{X}].$$

As formas mais utilizadas para a função f_j são:

- $f_j[\mathbf{X}] = \Phi(\sum_{i=1}^p \beta_i x_{ji})$, onde Φ é a função distribuição acumulada da normal padrão.
- $f_j[\mathbf{X}] = \frac{\exp\{\sum_{i=1}^p \beta_i x_{ji}\}}{1 + \exp\{\sum_{i=1}^p \beta_i x_{ji}\}}$, que é a função logística (modelo LOGIT).

A estimação desses modelos é realizada através do método de máxima verossimilhança. Os modelos PROBIT e LOGIT foram exaustivamente estudados e há uma série de textos que abordam o tema, entre eles Aldrich & Nelson(1985).

Um exemplo de aplicação dessa metodologia em **AP** pode ser encontrado no artigo de Rao & Winter(1978). Nesse artigo, desejava-se estudar o interesse de um grupo de alunos na compra de uma máquina fotográfica. Os estímulos são descrições de câmeras e é perguntado se o entrevistado compraria, ou não, um aparelho com aquelas características. Os autores utilizam um modelo PROBIT para segmentar o mercado. Nesse exemplo são criados 16 estímulos compostos por 4 atributos com 2 níveis cada. São estimados 45 modelos, um para cada elemento da amostra, e aplica-se uma análise de agrupamentos aos coeficientes estimados, obtendo-se 4 conglomerados (segmentos). Assim, cada segmento é caracterizado conforme uma série de variáveis previamente obtidas.

5.2. Variável resposta ordinal

Este tópico diferencia-se do anterior pelo fato da variável dependente estar em uma escala ordinal. Para entender a aplicabilidade desta metodologia em **AP**, considere o caso em que juízes avaliam estímulos, classificando-os através do seguinte critério: gostou muito, indiferente ou não gostou; ou ainda quando se adota uma escala diferencial semântica discreta. Claramente tem-se uma variável resposta em escala ordinal. Apresenta-se agora o desenvolvimento de um modelo PROBIT aplicado a esse caso.

Considere r_j como sendo a avaliação do estímulo j , podendo assumir M diferentes categorias de resposta em uma escala ordinal ($W_1, W_2, \dots, W_M$).

Suponha também a existência de uma variável contínua, não observável, y_j , que

representa a preferência do juiz pelo estímulo j . A essa variável será aplicado o seguinte modelo linear

$$y = \mathbf{X}\beta + \epsilon \quad (5.2.1)$$

onde: $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$, vetor de parâmetros;

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

matriz de planejamento, com x_{ji} = valor do atributo i no estímulo j e $\epsilon = (\epsilon_j)$ vetor aleatório com $\epsilon \sim N_n(\mathbf{0}; \sigma^2 \mathbf{I})$.

A relação entre y_j e r_j é estabelecida, assumindo-se a existência de $M + 1$ números reais, $\xi_0, \xi_1, \dots, \xi_M$, com $\xi_0 = -\infty$ e $\xi_M = +\infty$; além disso, $\xi_0 \leq \xi_1 \leq \dots \leq \xi_M$, isto é, $r_j \in W_k$. O estímulo j é dito como pertencente à categoria de resposta k , se e só se $\xi_{k-1} < y_j \leq \xi_k$, ou seja:

$$r_j \in W_k \Leftrightarrow \xi_{k-1} < y_j \leq \xi_k \quad \text{para } 1 \leq k \leq M. \quad (5.2.2)$$

Aplicando (5.2.1) vem que:

$$\begin{aligned} \xi_{k-1} < y_j \leq \xi_k &\Leftrightarrow \xi_{k-1} \leq \sum_{i=0}^p \beta_i x_{ji} + \epsilon_j < \xi_k \Leftrightarrow \\ &\Leftrightarrow \frac{\xi_{k-1} - \sum_{i=0}^p \beta_i x_{ji}}{\sigma} < \frac{\epsilon_j}{\sigma} \leq \frac{\xi_k - \sum_{i=0}^p \beta_i x_{ji}}{\sigma}. \end{aligned} \quad (5.2.3)$$

Considere a seguinte notação: seja r_{jk} tal que

$$r_{jk} = \begin{cases} 1, & \text{se } r_j \in W_k \\ 0, & \text{caso contrário} \end{cases}; 1 \leq j \leq n \quad \text{e} \quad 1 \leq k \leq M.$$

Considere também a matriz

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nM} \end{pmatrix}.$$

Utilizando essa nova notação e a suposição de que ϵ segue uma distribuição normal multivariada, tem-se, a partir de (5.2.3) que:

$$P(r_{jk} = 1) = P(r_j \in W_k) = \Phi\left[\frac{\xi_k - \sum_{i=0}^p \beta_i x_{ji}}{\sigma}\right] - \Phi\left[\frac{\xi_{k-1} - \sum_{i=0}^p \beta_i x_{ji}}{\sigma}\right]. \quad (5.2.4)$$

O modelo (5.2.4) é não identificável, pois ao aplicar qualquer transformação de escala a y_j e também aos parâmetros $\xi_0, \xi_1, \dots, \xi_M$ o modelo resultante seria o mesmo acima descrito. Para contornar este problema, assume-se, sem perda de generalidade, que $\xi_1 = 0$ e $\sigma = 1$. Desta forma o modelo (5.2.4) torna-se

$$P(r_{jk} = 1) = P(r_j \in W_k) = \Phi\left[\xi_k - \sum_{i=0}^p \beta_i x_{ji}\right] - \Phi\left[\xi_{k-1} - \sum_{i=0}^p \beta_i x_{ji}\right] \quad (5.2.5)$$

O modelo estará estimado quando forem encontradas estimativas para os parâmetros $\xi_2, \dots, \xi_{M-1}, \beta_0, \dots, \beta_p$. ($Q = M+p-1$ parâmetros).

Considere a seguinte notação:

$$\Phi_{j,k} = \Phi\left(\xi_k - \sum_{i=0}^p \beta_i x_{ji}\right) \quad \text{para } 1 \leq k \leq M \quad \text{e } 1 \leq j \leq n.$$

Desta forma o modelo (5.2.5) pode ser escrito como:

$$P(r_{jk} = 1) = \Phi_{j,k} - \Phi_{j,k-1}. \quad (5.2.6)$$

Na estimação dos parâmetros do modelo será utilizado o método de máxima verossimilhança. A partir de (5.2.6) tem-se que a função de verossimilhança de $\mathbf{R}$ será dada por:

$$L = L(\mathbf{R}/\beta; \xi_2, \dots, \xi_{M-1}) = \prod_{j=1}^n \prod_{k=1}^M (\Phi_{j,k} - \Phi_{j,k-1})^{r_{jk}} \quad (5.2.7)$$

$$L^* = \log L = \sum_{j=1}^n \sum_{k=1}^M r_{jk} \log(\Phi_{j,k} - \Phi_{j,k-1}) \quad (5.2.8)$$

Para encontrar os estimadores deve-se encontrar $(\beta_0, \dots, \beta_k, \xi_2, \dots, \xi_{M-1})$ que maximize L^* , sujeito à restrição $\xi_1 \leq \xi_2 \leq \dots \leq \xi_{M-1}$. Em McKelvey & Zavoina(1975), são apresentadas as primeiras e segundas derivadas de $\log L$. Como as primeiras derivadas não formam um sistema linear, é necessário que a estimação seja feita através de algum procedimento iterativo (por exemplo, Newton Raphson). Souza(1986) discute e compara uma variedade de possíveis procedimentos.

McKelvey & Zavoina(1975) advertem sobre a possibilidade do processo iterativo levar a um mínimo local, ou até mesmo a não convergir, embora a experiência tenha mostrado que, em geral, convergência é atingida em um número pequeno de iterações que varia de 4 a 10, dependendo do número de parâmetros.

Ainda em McKelvey & Zavoina(1975), são apresentados alguns testes para o modelo, que se baseiam no fato dos estimadores obtidos serem de máxima verossimilhança. Estes testes são assintóticos, o que diminui o interesse pelo tópico, uma vez que ao se estimar modelos individuais apenas um número relativamente pequeno de estímulos é avaliado.

Uma medida da qualidade do ajuste foi desenvolvida baseando-se no coeficiente de determinação (R^2). A dificuldade em seu cálculo reside no fato da variável dependente, y_j , ser não observável, assim, algumas somas de quadrados utilizadas serão estimativas das reais.

Considere $\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_p)'$, o estimador de máxima verossimilhança de β e $\hat{y}_j = \sum_{i=0}^p \hat{\beta}_i X_{ji}$ e $\hat{e}_j = (y_j - \hat{y}_j)$ (desconhecido). Sabe-se, da teoria da regressão, que

$$\frac{\sum_{j=0}^n e_j^2}{n} \xrightarrow{P} \sigma^2.$$

Logo, uma possível estimativa para a soma de quadrados residual seria: $\hat{S}_R^2 = n\sigma^2 = n$, uma vez que fixou-se $\sigma^2 = 1$.

A soma de quadrados da regressão pode ser calculada da maneira usual: $S_E^2 = \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2$ onde $\bar{\hat{y}} = \sum_{j=1}^n \frac{\hat{y}_j}{n}$. Assim, estima-se a soma de quadrados total por

$$\hat{S}_T^2 = S_E^2 + \hat{S}_R^2.$$

Isto faz com que se tenha:

$$\hat{R}^2 = \frac{S_E^2}{\hat{S}_T^2} \xrightarrow{P} R^2.$$

A interpretação de $\hat{R}^2$ é equivalente á do coeficiente de determinação usual.

O modelo LOGIT, análogo a esse, pode ser obtido supondo que a distribuição acumulada de ϵ seja a logística.

5.2.1. Exemplo

Retome o exemplo dos micros, discutido no capítulo 4.1. Suponha que a avaliação dos estímulos tenha sido realizada através da escala diferencial semântica abaixo:

certamente compraria 1 2 3 certamente não compraria.

A tabela 5.1 fornece uma possível avaliação para o conjunto de estímulos apresentados na tabela 4.5¹.

O modelo de preferência a ser estimado é $y_j = \beta_0 + \beta_1 W_p + \beta_2 M_p + \beta_3 C_p + \epsilon_j$; onde W_p , M_p e C_p correspondem aos valores padronizados das colunas W, M e C. O modelo estimado é

$$\hat{y}_j = 0,958 + 0,095W_p + 0,445M_p - 1,747C_p \quad \hat{R}^2 = 0,7547.$$

¹ a relação entre as avaliações das tabelas 4.5 e 5.1 fica clara ao se notar que os estímulos que receberam postos de 1 a 7, de 8 a 11 e de 12 a 18, em 4.5, receberam os scores 1, 2 e 3, respectivamente, em 5.1

Analisando o modelo, percebe-se que o entrevistado hipotético prefere micros com alto valor de clock e com memória RAM e winchester pequenos. As importâncias relativas dos atributos W, M e C, são, respectivamente, 4,2 %, 19,5 % e 76,4%.

TABELA 5.1. Estímulos avaliados através de um escala diferencial semântica de três pontos

estímulo	winchester	RAM	clock	avaliação
1	40	1	8	3
2	40	2	20	1
3	40	4	16	3
4	40	1	20	2
5	40	2	16	1
6	40	4	8	3
7	60	1	20	1
8	60	2	16	1
9	60	4	8	3
10	60	1	16	1
11	60	2	8	3
12	60	4	20	1
13	120	1	16	2
14	120	2	8	3
15	120	4	20	2
16	120	1	8	3
17	120	2	20	1
18	120	4	16	2

TABELA 5.2. Parâmetros estimados para o modelo VETOR

parâmetro	estimativa	erro-padrão
β_0	0,9578	0,7154
β_1	0,0949	0,5435
β_2	0,4452	0,4877
β_3	-1,7466	0,8887
ξ_2	1,4064	0,9876

Outro parâmetro estimado foi $\hat{\xi}_2 = 1,40638$; isso leva ao estabelecimento da seguinte regra: se $y_j \leq 0$ então a avaliação prevista do estímulo j é 1, se $0 < y_j \leq 1,40638$, então a avaliação prevista para esse estímulo é 2 e, por fim, se $y_j > 1,40638$, prevê-se que

o estímulo j seria avaliado como 3.

Na tabela 5.2, estão dispostas as estimativas dos parâmetros e seus correspondentes erros-padrão².

TABELA 5.3. Probabilidades de classificação previstas pelo modelo vetor

estímulo	r_j	$\hat{r}_j$	$P(r_j = 1)$	$P(r_j = 2)$	$P(r_j = 3)$
10	1	1	0,4979	0,4215	0,0806
7	1	1	0,9124	0,0847	0,0029
8	1	2	0,3623	0,4918	0,1459
12	1	1	0,6236	0,3338	0,0426
17	1	1	0,8012	0,1867	0,0122
2	1	1	0,8561	0,1371	0,0068
5	1	2	0,3828	0,4833	0,1339
13	2	2	0,4332	0,4590	0,1078
4	2	1	0,9207	0,0769	0,0024
15	2	1	0,5604	0,3800	0,0596
18	2	2	0,1134	0,4649	0,4217
3	3	2	0,1606	0,5001	0,3392
16	3	3	0,0019	0,0670	0,9311
11	3	3	0,0011	0,0466	0,9523
1	3	3	0,0038	0,0989	0,8974
14	3	3	0,0006	0,0330	0,9664
9	3	3	0,0001	0,0090	0,9909
6	3	3	0,0001	0,0104	0,9895

Na tabela 5.3, encontram-se as probabilidades $P(r_j \in W_k)$, $j = 1, 2, \dots, 18$ e $k = 1, 2, 3$.

Convém lembrar que os modelos ponto-ideal também poderiam ser estimados, desde que se utilize as linearizações sugeridas no capítulo 4.1.

5.3. LOGIT condicional

Em um primeiro momento, será discutida a aplicação do método LOGIT condi-

² a seção 5.4 discute um procedimento para estimação do erro-padrão e também a montagem de testes de hipóteses para parâmetros estimados pelo método da máxima verossimilhança

cional ao caso em que se deseja estimar um único modelo para toda a amostra (estimação a nível amostral). Em seguida, será discutida a estimação de modelos individuais.

5.3.1. Estimação a nível amostral

Considere C_s o conjunto de estímulos apresentado ao indivíduo s . Para aplicação do método, a avaliação se dá através da escolha, pelo entrevistado, do estímulo preferido dentre os do conjunto C_s . Essa variável resposta pode ser representada por:

$$r_{js} = \begin{cases} 0 & \text{se o estímulo } j \text{ não for o escolhido em } C_s \\ 1 & \text{se o estímulo } j \text{ for o escolhido em } C_s \end{cases}.$$

O modelo de preferência utilizado é:

$$y_{js} = \sum_{i=1}^p \beta_i x_{ji} + \epsilon_{js} = V_j + \epsilon_{js} \quad \text{onde} \quad (5.3.1)$$

y_{js} é a utilidade do estímulo j para o indivíduo s e

ϵ_{js} é um termo de erro aleatório.

O método parte do princípio que o entrevistado preferirá o estímulo j^* se

$$y_{j^*s} > y_{js} \quad \forall j \neq j^*, j \in C_s.$$

Aplicando (5.3.1) a essa expressão chega-se a

$$\epsilon_{js} - \epsilon_{j^*s} < V_{j^*} - V_j \quad \forall j \in C_s, j \neq j^*.$$

Assim, a probabilidade de que o estímulo j seja o preferido no conjunto C_s , é dada por:

$$P_{j^*s} = P(\epsilon_{js} - \epsilon_{j^*s} < V_{j^*} - V_j \quad ; \forall j \in C_s, j \neq j^*) \quad (5.3.2)$$

A expressão (5.3.2) recebe a denominação de modelo de utilidade estocástico. Considerando $\phi_{\epsilon_{j^*s}}$ como a função densidade de ϵ_{j^*s} , vem que:

$$\begin{aligned} P_{j^*s} &= \int_{-\infty}^{+\infty} P(\epsilon_{js} < \epsilon_{j^*s} + V_{j^*} - V_j \mid \epsilon_{j^*s} = t; \forall j \in C_s, j \neq j^*) \phi_{\epsilon_{j^*s}}(t) dt = \\ &= \int_{-\infty}^{+\infty} P(\epsilon_{js} < t + V_{j^*} - V_j \mid \forall j \in C_s, j \neq j^*) \phi_{\epsilon_{j^*s}}(t) dt \end{aligned}$$

McFadden(1974) e Punj & Staelin(1978) supõem que os erros sejam identicamente distribuídos e independentes entre indivíduos e entre as alternativas avaliadas por um mesmo entrevistado. Desse modo, a expressão anterior pode ser reescrita como:

$$P_{j^*s} = \int_{-\infty}^{+\infty} \frac{\partial \Phi_s}{\partial \epsilon_{j^*s}} [t + V_{j^*} - V_1, \dots, t + V_{j^*} - V_{T_s}] dt \quad (5.3.3)$$

onde, T_s é o número de estímulos do conjunto C_s e Φ_s é a distribuição acumulada conjunta do vetor aleatório $\epsilon_s = (\epsilon_{1s}, \epsilon_{2s}, \dots, \epsilon_{n_s s})$, sendo n_s a cardinalidade do conjunto C_s .

O próximo passo é a escolha da distribuição dos erros. McFadden(1974), preocupado com aspectos computacionais do problema, sugeriu uma expressão para P_{j^*s} e provou que a expressão poderia ter sido obtida através de (5.3.3). A expressão sugerida foi:

$$P_{j^*s} = \frac{e^{V_{j^*}}}{\sum_{j \in C_s} e^{V_j}}.$$

Uma possível distribuição para ϵ_{js} que resulta na expressão anterior é a do valor extremo tipo I (Johnson & Kotz(1970))³, como é provado a seguir.

Resultado 5.3.1. A expressão anterior é válida se ϵ_{js} seguir uma distribuição do valor extremo tipo I e ϵ_{js} for independente de ϵ_{ls} , $j \neq l$, $j, l \in C_s$.

³ em geral, sua distribuição acumulada é dada por $\exp\{-\exp\{-\frac{(\epsilon_{js}-\xi)}{\theta}\}\}$; o resultado a ser provado, (5.3.1), é válido ao se utilizar $\Phi_{js}(\epsilon_{js}) = \exp\{-\exp\{-\epsilon_{js}\}\}$; $-\infty < \epsilon_{js} < +\infty$, ou seja, assumindo $\xi = 0$ e $\theta = 1$

prova -

Se ϵ_{js} não depende de ϵ_{ls} , $j \neq l$, $j, l \in C_s$, então,

$$\Phi_s(\epsilon) = \prod_{j \in C_s} e^{-e^{-\epsilon_{js}}} \quad ; \quad -\infty < \epsilon_{js} < +\infty \quad , \quad j \in C_s.$$

Assim,

$$\frac{\partial \Phi_s}{\partial \epsilon_{j^*s}} = e^{-\epsilon_{j^*s}} \prod_{j \in C_s} e^{-e^{-\epsilon_{js}}}.$$

De onde vem que:

$$\frac{\partial \Phi_s}{\partial \epsilon_{j^*s}} [t + V_{j^*} - V_1, \dots, t + V_{j^*} - V_{T_s}] = e^{-t} e^{-e^{-t} \sum_{j \in C_s} e^{V_j - V_{j^*}}}.$$

Substituindo essa expressão em (5.3.3), vem que :

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{\partial \Phi_s}{\partial \epsilon_{j^*s}} [t + V_{j^*} - V_1, \dots, t + V_{j^*} - V_{T_s}] dt &= \int_{-\infty}^{+\infty} e^{-t} e^{-e^{-t} \sum_{j \in C_s} e^{V_j - V_{j^*}}} dt = \\ &= \int_{+\infty}^0 e^u \sum_{j \in C_s} e^{V_j - V_{j^*}} du = \frac{e^{V_{j^*}}}{\sum_{j \in C_s} e^{V_j}}. \end{aligned}$$

A verossimilhança referente à escolha do indivíduo s pode ser escrita como:

$$L_s = \prod_{j \in C_s} P_{js}^{r_{js}}.$$

Assim, considerando a amostra completa, a função de verossimilhança pode ser escrita como:

$$L = \prod_s L_s = \prod_s \prod_{j \in C_s} \left[\frac{\exp\{\sum_{i=1}^p \beta_i x_{ji}\}}{\sum_{l \in C_s} \exp\{\sum_{i=1}^p \beta_i x_{li}\}} \right]^{r_{js}} \quad (5.3.4)$$

Os parâmetros estimados são obtidos através da maximização da função anterior.

Uma generalização é obtida considerando-se a amostra formada por $m = \sum_{i=1}^q m_i$ indivíduos, onde a m_i indivíduos seja apresentado o mesmo conjunto C_i de estímulos, $i = 1, 2, \dots, q$. Nesse contexto, suponha que o estímulo $j \in C_i$ tenha sido escolhido por m_{ji} sujeitos ($m_i = \sum_{j \in C_i} m_{ji}$). Desse modo, a verossimilhança para os entrevistados submetidos ao conjunto C_i será dada pela seguinte multinomial:

$$L_i = \frac{m_i!}{\prod_{j \in C_i} m_{ji}!} \prod_{j \in C_i} P_{ji}^{m_{ji}}.$$

Consequentemente, a verossimilhança para toda a amostra pode ser escrita como:

$$L = \prod_{i=1}^q L_i = \prod_{i=1}^q \frac{m_i!}{\prod_{j \in C_i} m_{ji}!} \prod_{j \in C_i} P_{ji}^{m_{ji}}. \quad (5.3.5)$$

Note que (5.3.4) é um caso particular de (5.3.5), onde $m_i = 1 \quad \forall i$.

O vetor β é estimado pelo método de máxima verossimilhança, através de algum procedimento iterativo⁴. McFadden(1974) prova que não existem máximos locais, assim, se o processo iterativo de estimação convergir, será para o máximo global. Nos resultados 5.3.2 e 5.3.3 são apresentadas, respectivamente as fórmulas da primeira e segunda derivada de $\log L$.

Resultado 5.3.2. $\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^q \sum_{j \in C_i} (m_{ji} - m_i P_{ji}) \mathbf{x}_j$, onde $\mathbf{x}_j = (x_{ji})$.

prova :Seja $K = \sum_{i=1}^q \log \frac{m_i!}{\prod_{j \in C_i} m_{ji}!}$, então

$$\begin{aligned} \log L &= \sum_{i=1}^q \log \frac{m_i!}{\prod_{j \in C_i} m_{ji}!} + \sum_{i=1}^q \sum_{j \in C_i} m_{ji} \log P_{ji} = \\ &= K - \sum_{i=1}^q \sum_{j \in C_i} m_{ji} \log \left[\sum_{l \in C_i} e^{(\mathbf{x}_l' - \mathbf{x}_j') \beta} \right] = \end{aligned}$$

⁴ ver discussão sobre procedimentos iterativos em Souza(1986)

$$\begin{aligned}
&= K - \sum_{i=1}^q \sum_{j \in C_i} m_{ji} \log[e^{-\mathbf{x}_j' \boldsymbol{\beta}} \sum_{l \in C_i} e^{\mathbf{x}_l' \boldsymbol{\beta}}] = \\
&= K + \sum_{i=1}^q \sum_{j \in C_i} m_{ji} \mathbf{x}_j' \boldsymbol{\beta} - \sum_{i=1}^q m_i \log[\sum_{l \in C_i} e^{\mathbf{x}_l' \boldsymbol{\beta}}]
\end{aligned}$$

Derivando essa expressão em $\boldsymbol{\beta}$, vem

$$\begin{aligned}
\frac{\partial \log L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^q \sum_{j \in C_i} m_{ji} \mathbf{x}_j - \sum_{i=1}^q m_i \sum_{l \in C_i} \frac{e^{\mathbf{x}_l' \boldsymbol{\beta}}}{\sum_{l \in C_i} e^{\mathbf{x}_l' \boldsymbol{\beta}}} = \\
&= \sum_{i=1}^q \sum_{j \in C_i} (m_{ji} - m_i P_{ji}) \mathbf{x}_j \quad \blacksquare
\end{aligned}$$

Resultado 5.3.3. $\frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = - \sum_{i=1}^q m_i \sum_{j \in C_i} (\mathbf{x}_j - \bar{\mathbf{x}}_i) P_{ji} (\mathbf{x}_j - \bar{\mathbf{x}}_i)'$, onde, $\bar{\mathbf{x}}_i = \sum_{l \in C_i} \mathbf{x}_l P_{li}$.

Punj & Staelin(1978) afirmam que os parâmetros podem ser obtidos de modo mais eficiente se a avaliação de cada indivíduo for realizada através da atribuição de postos. Suponha, por exemplo, a existência de 4 estímulos (E, F, G, H) e que o indivíduo s os tenha avaliado atribuindo-lhes os seguintes postos: 2, 3, 1 e 4, respectivamente. Caso fosse adotado o enfoque até agora estudado, a única informação extraída desses dados seria que o estímulo G é o preferido no conjunto $C_{1s} = \{E, F, G, H\}$. A novidade introduzida pelos autores seria acrescentar as seguintes informações:

- o estímulo E é o preferido no conjunto $C_{2s} = \{E, F, H\}$ e
- o estímulo F é o preferido no conjunto $C_{3s} = \{F, H\}$.

Assim procedendo, se o indivíduo s ordenasse n_s estímulos, originaria $n_s - 1$ novas escolhas. Esse procedimento só é válido se for possível garantir que a ordem estabelecida pelos postos corresponde realmente à ordem de escolha dos entrevistados. O que se verifica, em geral, é que a qualidade da avaliação (envolvimento do entrevistado) diminui a partir da escolha do estímulo preferido.

A verossimilhança correspondente a essa nova situação pode ser escrita como:

$$L = \prod_s \prod_{l=1}^{n_s-1} \prod_{j \in C_{ls}} \left[\frac{e^{V_j}}{\sum_{q \in C_{ls}} e^{V_q}} \right]^{r_{jls}}.$$

Em McFadden(1974), encontram-se testes para os modelos apresentados.

5.3.2. Estimação de modelos individuais

Uma maneira de tornar possível o uso do LOGIT condicional na estimação de modelos individuais seria avaliar os estímulos da maneira descrita no procedimento LIN-MAP, ou seja, apresentá-los aos pares e pedir que o entrevistado escolha o preferido em cada dupla. Cada par seria encarado como um conjunto de estímulos.

Assim, a verossimilhança para o par C_{ls} seria:

$$L_{ls} = \prod_{j \in C_{ls}} \left[\frac{\exp\{\sum_{i=1}^p \beta_{is} x_{jis}\}}{\sum_{q \in C_{ls}} \exp\{\sum_{i=1}^p \beta_{is} x_{qis}\}} \right]^{r_{jls}} \quad \text{onde:}$$

$$r_{jls} = \begin{cases} 1 & \text{se o estímulo } j \text{ for o preferido do conjunto } C_{ls} \\ 0 & \text{caso contrário} \end{cases}.$$

A verossimilhança completa do indivíduo s será:

$$L_s = \prod_{l=1}^{\frac{n_s(n_s-1)}{2}} L_{ls}$$

A suposição em que esta abordagem se baseia é a de que as comparações 2 a 2 sejam independentes entre si.

É bom ressaltar que as comparações 2 a 2 podem ser extraídas de uma avaliação via postos de modo análogo ao descrito ao final da seção 5.3.1. Por exemplo, considerando a ordenação do mais preferido para o menos preferido, se o estímulo E recebeu um posto inferior ao do estímulo F, então será considerado que o estímulo E seria o preferido do par E,F. Obviamente, neste caso, a suposição de independência entre as comparações não é válida.

5.3.3. Exemplo

A tabela 5.4 reproduz as informações necessárias à estimação de um modelo vetor pelo método LOGIT condicional, para os dados da tabela 4.11.

TABELA 5.4. Dados de entrada para a estimação do modelo vetor pelo método LOGIT condicional, baseado nos dados da tabela 4.11

estímulo	X_1	X_2	r_j
2	5	7	1
1	6	3	0
2	5	7	1
3	2	9	0
2	5	7	1
4	7	15	0
1	6	3	1
3	2	9	0
1	6	3	1
4	7	15	0
4	7	15	1
3	2	9	0

TABELA 5.5. Parâmetros estimados pelo método LOGIT condicional

parâmetro	estimativa	erro-padrão
β_1	0,515229	0,4334
β_2	-0,214453	0,1819

TABELA 5.6. Probabilidades de escolha de estímulos estimadas pelo modelo

estímulos disponíveis	Probabilidades de escolha do	
	primeiro estímulo do par	segundo estímulo do par
(1,2)	0,7979	0,2021
(1,3)	0,9660	0,0340
(1,4)	0,8868	0,1132
(2,3)	0,8781	0,1219
(2,4)	0,6649	0,3351
(3,4)	0,2159	0,7841

Estimado o modelo $y_j = \beta_1 X_1 + \beta_2 X_2$, chega-se às estimativas da tabela 5.5. Observando os parâmetros do modelo deduz-se que, quanto menor for X_2 e maior for X_1 , maior será a preferência pelo estímulo. Se as estimativas da tabela 5.5 forem comparadas com o modelo estimado na seção 4.3.1.1, nota-se uma certa equivalência em termos de seus sinais e do valor do quociente $\frac{\beta_1}{\beta_2}$, indicando uma certa similaridade entre os resultados.

A qualidade do ajuste pode ser medida através do cálculo do coeficiente de correlação de Spearman entre os postos previstos⁵ e observados. Para o modelo acima, o valor do coeficiente é de 0,80; indicando um ajuste razoável.

Na tabela 5.6, encontram-se as probabilidades previstas de escolha de cada estímulo quando apresentados dois a dois.

5.4. Comentários

Nos procedimentos introduzidos neste capítulo, o método de estimação empregado é o de máxima verossimilhança, o que garante a consistência e, sob as condições de regularidade (verificadas nos casos apresentados), a normalidade assintótica dos estimadores (Mood, Graybill & Boes(1974)). A distribuição assintótica pode ser utilizada na estimação dos erros-padrão dos estimadores. Considere θ o vetor de parâmetros a serem estimados, $L(\theta)$ a função de verossimilhança e $\hat{\theta}$ o estimador de máxima verossimilhança obtido. $\hat{\theta}$ será assintoticamente normal com média θ e matriz de covariância dada por $I^{-1}(\theta)$, onde $I(\theta) = E\left[-\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\right]$. Nos casos estudados, sugere-se que a estimativa da matriz de covariância seja⁶ $\hat{V} = \left[-\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\right]_{\theta=\hat{\theta}}^{-1}$. Na diagonal de $\hat{V}$, encontram-se estimadores para os erros-padrão de $\hat{\theta}$ (McKelvey & Zavoina(1975) e Aldrich & Nelson(1984)).

Testes da razão de verossimilhança podem ser construídos para verificação de hipóteses envolvendo vários componentes de θ .

Lembre que tanto o teste da razão de verossimilhança, quanto a normalidade de θ , com a matriz de covariância apresentada, são resultados assintóticos. Muitas vezes, em uma AP deseja-se estimar modelos individuais, a partir da avaliação de um pequeno número de estímulos, o que limita a utilização desses resultados.

⁵ do mesmo modo que no capítulo 4.3, para efeito de previsão do posto deve-se considerar o modelo $-y_j$, já que a atribuição foi realizada em ordem decrescente de preferência, e o modelo estimado supõe ordem crescente

⁶ veja uma possível justificativa em Cox & Snell(1989)

CAPÍTULO 6

Métodos Especiais

Neste capítulo serão abordados dois métodos especiais utilizados em problemas de AP, o híbrido e o bayesiano. A existência de um capítulo à parte para este estudo se justifica na medida em que eles possuem particularidades em todas as etapas da análise, ou seja, coleta de dados, regra de composição e estimação do modelo.

6.1. Método híbrido

Muitas vezes, na prática, para possibilitar a estimação de um modelo de AP a nível individual é necessário que o entrevistado avalie um grande número de estímulos. Em várias situações, mesmo a adoção de planejamentos fracionários não reduz esse número de modo satisfatório. Isso faz com que as enquetes tornem-se cansativas e passíveis de vícios, uma vez que o empenho do juiz diminui quando submetido a tarefas longas.

A estimação de modelos individuais é importante em estudos comerciais. Isso possibilita que as respostas sejam analisadas a nível de escolha de um produto (supondo que os estímulos definam produtos e que o de maior preferência seja o escolhido em uma compra), o que evita planejamentos de produtos baseados em uma "preferência média".

Na tentativa de evitar esse problema Green, Goldberg & Montemayor(1981), apresentam um procedimento simplificado de obtenção de dados, associado a um novo modelo que possibilita a estimação de preferências individuais. Este modelo foi denominado híbrido.

O modelo híbrido é formado através da união de duas classes de modelos de preferência: modelos de composição e de decomposição. Inicialmente essas classes são estudadas em separado e, em seguida, é apresentado o modelo híbrido.

Os modelos de AP, até agora estudados, pertencem à classe de decomposição. Tratam-se de modelos que têm por objetivo estimar a importância de cada atributo (utilidades dos níveis dos atributos) a partir da avaliação de seu efeito conjunto (estímulos). Resumindo, partindo-se da avaliação de combinações de níveis, quer-se conhecer a contribuição de cada atributo, em separado, na formação da preferência do entrevistado.

Os modelos de composição seguem um caminho oposto aos da classe anterior. Agora, o juiz deve avaliar os níveis dos atributos em separado. A utilidade global do perfil (preferência por um estímulo) é dada através de alguma regra de composição que envolva as avaliações em separado. Em outras palavras, partindo-se da avaliação em separado dos atributos e seus níveis, deseja-se conhecer a preferência global por um estímulo.

Para tornar mais clara a dinâmica desses modelos, é reproduzido, em seguida, o roteiro de um possível experimento indicado para estimação de um modelo de composição (Green, Goldberg & Montemayor(1981)). Este tipo de experimento recebe a denominação de auto-explicativo.

1. É mostrado a um juiz p conjuntos de níveis de atributos, um atributo por vez.
2. Para cada conjunto utiliza-se uma escala de 10 pontos. Ao nível preferido associa-se o valor 10, e os demais devem ser avaliados de acordo com essa atribuição. Empates são permitidos.
3. O juiz deve avaliar a importância relativa dos p atributos. Isto pode ser feito, por exemplo, pela distribuição de 100 pontos através dos fatores.
4. A utilidade (preferência) auto-explicada do entrevistado s para o estímulo j , será dada por:

$$U_{js} = \sum_{k=1}^p w_{ks} u_{jks} \quad \text{onde}$$

w_{ks} = importância auto-explicada do atributo k , para o indivíduo s (passo 3),

u_{jks} = avaliação do nível do atributo k , pertencente ao estímulo j , pelo indivíduo s (passo 2).

Neste modelo, w_{ks} tem a função de escalonar os atributos de modo a colocar as contribuições de seus níveis em uma mesma escala de preferência.

Exemplo numérico: Deseja-se avaliar produtos com base na marca A, B ou C, e no tempo de garantia oferecido pelo fabricante, 6 meses ou 1 ano. Inicialmente é pedido que o entrevistado avalie os níveis dos atributos em separado (passo 1):

$$\text{marca} = \begin{cases} A & \text{avaliação 1;} \\ B & \text{avaliação 10;} \\ C & \text{avaliação 7} \end{cases} \quad \text{garantia} = \begin{cases} 6 \text{ meses} & \text{avaliação 7;} \\ 1 \text{ ano} & \text{avaliação 10} \end{cases}.$$

Em seguida é pedido que 100 pontos sejam distribuídos pelos dois atributos conforme a preferência do juiz (passo 2):

$$\text{marca} \Rightarrow w_{1s} = 65 \quad \text{garantia} \Rightarrow w_{2s} = 35.$$

Deste modo a utilidade auto-explicada do estímulo (marca C; 1 ano de garantia), será dada por:

$$U = 65 \times 7 + 35 \times 10 = 455 + 350 = 805 \quad \text{onde}$$

455 = utilidade auto-explicada da marca C e

350 = utilidade auto-explicada de 1 ano de garantia.

Nesse modelo, tanto w_{ks} como u_{jks} são obtidos diretamente, o que torna a determinação da utilidade global uma tarefa aparentemente simples.

Apesar de ser facilmente aplicável, este procedimento tem alguns inconvenientes, tais como:

- os juízes podem encontrar dificuldades em determinar os valores u_{jks} e w_{ks} em uma base confiável. Shepard(1964) afirma que os entrevistados nem sempre têm uma visão acurada sobre a importância que dão aos componentes de um produto;
- a regra de combinação aditiva para U_{js} pode não ser apropriada.

6.1.1. O Modelo híbrido

A idéia básica dos modelos híbridos é unir a simplicidade da abordagem de composição e a generalidade dos modelos de decomposição em um único modelo. Sua estimação requer a passagem por duas fases de coleta de dados: fase auto-explicativa e fase conjunta.

Inicialmente essa metodologia é apresentada em sua forma mais simples (Green, Goldberg & Montemayor(1981)). Posteriormente serão tratadas variações no processo de obtenção dos dados e na forma do modelo, Green(1984).

6.1.1.1. Fase auto-explicativa

Esta fase é similar à descrita quando se tratou do modelo de composição (ver páginas 94 a 96).

6.1.1.2. Fase conjunta

Envolve a avaliação de um conjunto limitado de estímulos, em geral de 3 a 9, que são retirados de um planejamento principal, usualmente de 64 a 256 estímulos. Em geral, deseja-se que esse planejamento permita a estimação ortogonal dos efeitos principais dos atributos, β_{kj} , e de interações de primeira ordem previamente selecionadas. Definindo $\gamma_{kk'j}$ como a interação entre os atributos k e k' e V_j como a avaliação do estímulo j , o modelo de decomposição correspondente a esta fase é dado por:

$$V_j = \sum_{k=1}^p \beta_{kj} + \sum_{k < k'} \gamma_{kk'j}.$$

Trata-se de uma versão do modelo função utilidade onde se permite a existência de termos de interação.

A grande vantagem do modelo híbrido é explicitada nesta fase. Caso fosse utilizado um procedimento usual de **AP**, o entrevistado teria que avaliar uma quantidade muito maior de estímulos, o que dificultaria sua realização.

É importante ressaltar que os estímulos a serem avaliados são subconjuntos de um planejamento maior, em geral fatorial fracionário.

Os dois artigos mencionados sugerem que a atribuição dos estímulos a serem avaliados por um indivíduo se dê de modo balanceado. Por balanceamento entende-se que dentro do conjunto de estímulos a ser avaliado por um sujeito, cada par de níveis de atributos apareça com uma frequência dada pelo produto das frequências relativas marginais e o número total de perfis; por exemplo, se o nível i do atributo 1 aparece em 6 de um total de 9 perfis e o nível j do atributo 2 em 3 de um total de 9, então o par ij deve aparecer em $6/9 \times 3/9 = 2$ dos nove perfis. Green(1984) afirma que o balanceamento pode ser facilmente obtido considerando os sujeitos como um fator experimental adicional.

Ao contrário do que ocorria em modelos anteriores, diminui o sentido de avaliar os estímulos através de sua ordenação. Isto se deve ao fato de que cada juiz avaliará apenas um pequeno subconjunto de perfis. A avaliação pode ser realizada através de alguma escala de preferência.

6.1.1.3. O modelo

O modelo híbrido básico é dado por:

$$y_{js} \cong \mu + \alpha U_{js} + \sum_{k=1}^p \beta_{kj} + \sum_{k < k'} \gamma_{kk'j} \quad \text{onde:} \quad (6.1.1)$$

$y_{js} = g(r_{js})$, sendo r_{js} a avaliação dos estímulos obtida na fase conjunta;

$\cong$ aproximação por mínimos quadrados;

U_{js} utilidade auto-explicada e

μ , α , β_{kj} e $\gamma_{kk'j}$ parâmetros estimados a nível amostral.

Este modelo possui uma parte responsável por uma variação individual (U) e uma outra parte que é característica da amostra estudada. Isto permite que se continue a obter informações sobre a preferência a nível individual.

Trata-se de um modelo de análise de covariância, tendo U como covariável.

Os parâmetros estimados a nível amostral têm a função de corrigir vícios inerentes aos modelos de composição.

Os termos referentes à fase conjunta são responsáveis por explicar a variabilidade da avaliação dos estímulos que os termos da fase auto-explicativa não conseguem captar.

Uma série de modificações foram introduzidas em todas as etapas do processo. A seguir serão discutidas as variações mais importantes

6.1.2. Modificações no processo

6.1.2.1. Nível de agregação

O primeiro tipo de mudança a ser estudado se dá através de duas possíveis desagregações :

- a. o modelo pode ser modificado de modo que seja possível a estimação de parâmetros de inclinação (α) diferentes para determinados atributos;
- b. a amostra pode ser particionada em subgrupos através de uma análise de agrupamentos realizada sobre as utilidades auto-explicadas.

Shepard(1964) constatou que em experimentos auto-explicativos, os juízes tendem a superestimar a importância dos atributos menos importantes. Assim, no modelo híbrido, um modo de fazer com que o coeficiente de inclinação de U corrija este vício com maior eficiência, seria possibilitar que a cada fator fosse associado um parâmetro diferente. Então, algebricamente ter-se-ia:

$$y_{js} \cong \mu + \sum_{k=1}^p \alpha_k (w_{ks} u_{jks}) + \sum_{k=1}^p \beta_{kj} \sum_{k < k'} \gamma_{kk'j}.$$

Este modelo é denominado completamente separado.

Entre o modelo híbrido básico e o completamente separado, pode-se construir uma série de modelos com diferentes formas de desagregação. Considere, por exemplo, que os atributos sejam divididos em dois subconjuntos, A_1 e A_2 , e que por alguma razão desconfie-se que os atributos em cada subconjunto tenham preferências próximas. Desta forma um possível modelo será:

$$y_{js} \cong \mu + \sum_{k \in A_1} \alpha_1(w_{ks}u_{jks}) + \sum_{k \in A_2} \alpha_2(w_{ks}u_{jks}) + \sum_{k=1}^p \beta_{kj} + \sum_{k < k'} \gamma_{kk'j}.$$

O problema que pode surgir com o aumento do número de desagregações é a grande quantidade de parâmetros que deverão ser estimados.

O outro tipo de desagregação a ser estudado se dá em termos da amostra. Ao se trabalhar com uma amostra muito heterogênea, ou seja, quando as avaliações individuais são muito diferentes, pode acontecer dos parâmetros estimados terem uma variabilidade muito alta, o que diminui o poder preditivo do modelo. Esse problema pode ser contornado através de um procedimento simples: utilizando os dados da fase auto-explicativa, aconselha-se dividir a amostra em grupos através de uma análise de agrupamentos e, em seguida, estimar os modelos separadamente em cada grupo.

A justificativa para este procedimento encontra-se no argumento de que pessoas que demonstram preferência similar na fase auto-explicativa, também deverão demonstrá-la na fase conjunta.

Em princípio nada impede que os dois níveis de desagregação até agora estudados sejam utilizados simultaneamente.

6.1.2.2. Avaliação

Os métodos de avaliação mais comuns, tanto na fase auto-explicativa, como na conjunta, são:

- i. categorias ordenadas - exemplo: inaceitável, aceitável ou altamente preferido;

- ii. taxas - exemplo: escala de intenção de compra de 9 pontos, marcas em uma escala de preferência contínua, etc;
- iii. soma constante - este procedimento, em geral, é reservado para a avaliação da importância relativa dos atributos. Foi descrito anteriormente na descrição do processo auto-explicativo, quando se pediu para atribuir 100 pontos a um conjunto de fatores.

Os modelos híbridos poderão ser classificados de acordo com o tipo de avaliação escolhida na fase auto-explicativa. Tem-se então as seguintes classes de modelos:

1. métrica - w_{ks} e u_{jks} são expressos quantitativamente, por exemplo através de taxas ou somas constantes;
2. semi-métrica - w_{ks} expresso numericamente, mas u_{jks} expresso como variável fictícia, refletindo resposta categorizada, por exemplo categorias ordenadas;
3. não-métrica - tanto w_{ks} , quanto u_{jks} estão concatenados em um conjunto de respostas categóricas, expressas através de variáveis fictícias. Os dois valores seriam provenientes de categorias ordenadas.

Os modelos híbridos até agora vistos encaixam-se perfeitamente na categoria métrica; todavia algumas modificações devem ser introduzidas nas outras classificações.

Na classe não métrica a codificação das respostas deve ser feita através de variáveis fictícias; assim, no caso de três categorias de resposta na avaliação dos atributos (ex. não importante, indiferente e importante) e três na avaliação de seus níveis (ex. inaceitável, aceitável e altamente desejável) haveria a criação de 9 variáveis fictícias, cada qual representando uma combinação das duas avaliações.

Nos modelos semi-métricos, os níveis dos atributos também devem ser representados por variáveis fictícias. Assim, no caso de existência de três níveis, o modelo seria:

$$y_{js} \cong \mu + \sum_{k=1}^p [\alpha_k^{(1)} w_{ks} Z_{jks}^{(1)} + \alpha_k^{(2)} w_{ks} Z_{jks}^{(2)}] + V_j \quad \text{onde:}$$

$$Z_{jks}^{(1)} = \begin{cases} 1 & \text{se nível do atributo } k \text{ for avaliado na categoria 1;} \\ 0 & \text{se caso contrário} \end{cases}$$

$$Z_{jks}^{(2)} = \begin{cases} 1 & \text{se nível do atributo } k \text{ for avaliado na categoria 2;} \\ 0 & \text{se caso contrário} \end{cases}$$

6.1.3. Estimação dos parâmetros

A estimação do modelo híbrido pode ser realizada por qualquer dos métodos até agora estudados: MONANOVA, PREFMAP, LINMAP com adaptações, LOGIT, PROBIT, mínimos quadrados, etc, dependendo, é claro, do tipo de variável resposta (avaliação) utilizada na fase conjunta.

O ponto mais interessante ligado à problemática da estimação não está na escolha do método, mas sim no próprio processo de estimação. A questão levantada se resume na escolha de uma estimação em um único estágio, ou em dois estágios.

Como, em geral, o interesse de uma AP está na estimação de preferências individuais e como, em um modelo híbrido, a variável U reflete as diferenças entre os entrevistados, pode ser desejável realçar a importância dessa variável no modelo. Isto pode ser feito através de uma estimação em estágios, como por exemplo:

- a. realizar a regressão de y em U ;
 - b. realizar a regressão dos resíduos do modelo anterior nas variáveis relativas a β e γ ;
- algebricamente tem-se

$$\hat{y}_{js} = \hat{y}_{js}^{(1)} + \hat{y}_{js}^{(2)}$$

$$\hat{y}_{js}^{(1)} \cong \hat{\mu} + \hat{\beta}U_{js}$$

$$\hat{y}_{js}^{(2)} = y_{js} - \hat{y}_{js}^{(1)} \cong \sum_{k=1}^p \beta_{kj} + \sum_{k < k'} \gamma_{kk'j}$$

Procedendo desta forma as diferenças individuais serão priorizadas.

As primeiras aplicações comerciais desta metodologia, davam preferência para a estimação em um único estágio. Todavia, com o passar do tempo, a estimação em dois estágios acabou sendo mais utilizada.

Estudos empíricos (Green, Goldberg & Montemayor(1981), Akaah et al(1983) e Green(1984)) mostram evidências de que o método híbrido equivale, e em algumas situações até mesmo supera, o método de mínimos quadrados quando utilizado na estimação de um modelo tradicional de AP.

6.1.4. Exemplo

Suponha a existência de 3 atributos A, B e C, cada qual com 2 níveis. Suponha também que a amostra seja formada por 2 pessoas, S_1 e S_2 , e que os dados da fase auto-explicativa sejam os da tabela 6.1, onde os pesos dos atributos foram obtidos pelo método de soma constante e a importância de seus níveis da maneira exemplificada no início do capítulo.

Para a fase conjunta, parte-se de um planejamento principal composto por 8 estímulos, fracionado em duas partes balanceadas, cada qual analisada por um indivíduo; a avaliação desses perfis (r_j) encontra-se na tabela 6.2.

Nesse exemplo, três estimações foram efetuadas:

- a. de um modelo de regressão que não leva em conta os dados da fase auto-explicativa:

$$\hat{y}_{js} = 0.3875 + 0.1625A + 0.0875B - 0.0875C \quad (R^2 = 68,3\%).$$

- b. do modelo (6.1.1) estimado em uma fase:

$$\hat{y}_{js} = -0.5356 + 0.0011U_{js} + 0.1348A + 0.0308B + 0.0121C \quad (R^2 = 84,4\%).$$

- c. do modelo (6.1.1) estimado em duas fases:

$$\hat{y}_{js}^{(1)} = -0.7015 + 0.0013U_{js} \quad (R^2 = 54,1\%)$$

$$y_{js} - \hat{y}_{js}^{(1)} = 0.1298A + 0.0206B + 0.0301C \quad (R^2 = 64,8\%)$$

$$\hat{y}_{js} = -0.7015 + 0.0013U_{js} + 0.1298A + 0.0206B + 0.0301C \quad (R^2 = 83.8\%).$$

TABELA 6.1. Dados da fase auto-explicativa

parâmetro	indivíduo	
	S_1	S_2
w_1	20	15
w_2	20	35
w_3	60	50
u_{11}	10	10
u_{12}	8	6
u_{21}	10	10
u_{22}	5	7
u_{31}	9	4
u_{32}	10	10

TABELA 6.2. Utilidades auto-explicadas e dados da fase conjunta

sujeito	Planejamento			U	r
S_1	1	1	1	940	0,8
	1	-1	-1	900	0,5
	-1	1	-1	960	0,3
	-1	-1	1	800	0,1
S_2	1	1	-1	1000	0,7
	1	-1	1	595	0,2
	-1	1	1	640	0,1
	-1	-1	-1	835	0,4

TABELA 6.3. Correlações entre a variável resposta e os valores previstos

sujeito	U	M.Q.	1 fase	2 fases
S_1	0,627	0,899	0,974	0,975
S_2	0,961	0,717	0,928	0,938

Nota-se um incremento razoável no coeficiente de determinação com a introdução da utilidade auto-explicada, evidenciando uma melhoria no modelo.

TABELA 6.4. Valores previstos pelo modelo híbrido

estímulo	Estimação				
	mínimos quadrados S_1 e S_2	1 fase		2 fases	
		S_1	S_2	S_1	S_2
1	0,55	0,68	0,41	0,70	0,39
2	0,55	0,55	0,54	0,55	0,54
3	0,40	0,40	0,38	0,41	0,38
4	0,23	0,17	-0,12	0,16	-0,19
5	0,72	0,72	0,72	0,72	0,72
6	0,37	0,50	0,24	0,53	0,21
7	0,23	0,36	0,08	0,39	0,05
8	0,23	0,23	0,21	0,24	0,20
R^2 (entre os postos e o valor previsto)	0,686	0,621	0,640	0,587	0,626

TABELA 6.5. Importância (%) dos atributos A, B e C

sujeito	fator	modelo			
		aut.expl.	M.Q.	1 fase	2 fases
S_1	A	20	48	55	50
	B	50	26	30	28
	C	30	26	15	22
S_2	A	13	48	39	35
	B	22	26	20	18
	C	65	26	41	47

TABELA 6.6. Importância dos atributos para o indivíduo S_1

nível	import.		híbrido			
	uw	aut.expl.	0,0011uw	conjunta	soma	%
A_1	200		0,2200	0,1348	0,3548	
A_2	160		0,1760	-0,1348	0,0412	
$ w_1u_{11} - w_1u_{12} $	40	20 %	0,0440	0,2696	0,3136	50 %
$ w_2u_{21} - w_2u_{22} $	100	50 %	0,1100	0,0616	0,1712	28 %
$ w_3u_{31} - w_3u_{32} $	60	30 %	0,0660	0,0242	0,1382	22 %
TOTAL	200				0,623	

Calculado o coeficiente de correlação entre a variável dependente e as utilidades auto-explicadas (U_{js}) para cada entrevistado, chega-se no valor 0.623 para o sujeito 1 e no valor 0.961 para o sujeito 2, indicando que a introdução de U deve ter uma influência maior para S_2 do que para S_1 . Na tabela 6.3, encontram-se, para cada pessoa, informações sobre o coeficiente de correlação entre a variável dependente e o valor previsto em cada um dos modelos. A diferença existente entre a correlação do modelo estimado por mínimos quadrados e os demais deve-se somente à introdução de U . Nota-se claramente que o aumento no coeficiente é consideravelmente maior para S_2 . O coeficiente de correlação foi escolhido por estar diretamente relacionado com o coeficiente de determinação.

Por fim, na tabela 6.4 previu-se as avaliações dos 8 estímulos para os dois membros da amostra, pelos três diferentes métodos. Calculou-se também a correlação entre a variável resposta e os valores previstos.

6.1.4.1. Importância dos atributos

Da mesma forma que nos capítulos anteriores, pode-se estimar a importância dos atributos para os diferentes modelos. Na tabela 6.5 encontra-se a importância, em %, dos três fatores para os 2 elementos da amostra. Essa importância foi calculada para um modelo de composição (apenas com os dados da fase auto-explicativa), de mínimos quadrados e modelo híbrido estimado em 1 e em 2 fases.

O cálculo da importância dos fatores para o modelo de composição e para o híbrido é exemplificado na tabela 6.6, para os dados do indivíduo 1 e, em destaque, o fator A. Para o modelo de composição, inicialmente, calcula-se, para os níveis de todos os fatores, a quantidade $w_{ks}u_{jks}$; isso feito, para cada atributo em separado, calcula-se a amplitude desses valores, como pode ser visto na primeira coluna da tabela; assim, a importância de um atributo será dada pela razão entre a amplitude calculada para o fator e a soma das amplitudes de todos os atributos (ver segunda coluna da tabela).

O cálculo da importância dos atributos para os modelos híbridos se dá de maneira similar. Na tabela 6.6 está apresentado o cálculo para o modelo estimado em uma única fase. O primeiro passo consiste em calcular a contribuição de cada nível no modelo; essa contribuição está distribuída tanto na utilidade auto-explicada (através da parcela $w_{ks}u_{jks}$), quanto nos parâmetros originados na fase conjunta; esse cálculo está descrito nas colunas 3 e 4 da tabela. A contribuição total de um nível é a soma do peso relativo

à fase auto-explicativa (coluna 3) com o da fase conjunta (coluna 4). O resultado em porcentagem é também obtido através do cálculo da amplitude dos níveis de cada fator.

A importância do modelo estimado por mínimos quadrados foi obtida da maneira descrita no capítulo 1.3.

Analisando a tabela 6.5, nota-se a importância da utilidade auto-explicada para diferenciação dos indivíduos, principalmente ao se comparar a coluna dos mínimos quadrados e as colunas correspondentes aos modelos híbridos. Além disso, percebe-se que o modelo híbrido pode ser encarado como uma maneira de se corrigir o viés do modelo de composição, como já foi dito anteriormente. Outra constatação é que a introdução da covariável U trouxe, em termos da importância dos atributos, menos modificações para S_1 do que para S_2 .

6.2. Método bayesiano

Cattin, Gelfand & Danes(1983) propuseram um método de estimação bayesiano para modelos de análise conjunta. O procedimento é aplicável para dados obtidos através de um planejamento perfil completo e quando, paralelamente à avaliação conjunta dos estímulos, é realizada a tomada de dados auto-explicativos. O modelo é estimado a nível individual, por isso, afim de facilitar a notação será omitido, nas fórmulas que seguem, o índice correspondente ao entrevistado.

À primeira vista a coleta de dados deste método e do método híbrido parecem idênticas; contudo, enquanto nos modelos híbridos o número de estímulos a serem avaliados na fase conjunta é pequeno, nos modelos bayesianos esta quantidade é a mesma necessária para a estimação de um modelo usual de AP. Isto se deve ao fato de que o estimador bayesiano será uma composição entre o estimador usual de mínimos quadrados, que utilizará apenas os dados referentes à fase conjunta e um estimador referente aos dados obtidos na fase auto-explicativa.

Considere o modelo:

$$y = r = X\beta + \epsilon \quad (6.2.1)$$

ϵ é um termo de erro aleatório, que para efeito da construção do estimador

bayesiano será suposto

$$\epsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}).$$

A informação a priori a respeito de β é

$$\beta \sim N_p(\beta^*, \sigma_\beta^2 (\mathbf{X}'\mathbf{X})^{-1}),$$

onde σ_β^2 é um número real positivo. O estimador bayesiano então obtido é conhecido na literatura como estimador James-Stein. Aspectos técnicos relacionados a este estimador, bem como a sua dedução, fazem parte do apêndice 7. O estimador bayesiano é dado por:

$$\begin{aligned} \hat{\beta}_{\text{bayes}} &= \left(1 - \frac{(p-2)\hat{\sigma}^2}{(\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*)}\right) \mathbf{b} + \frac{(p-2)\hat{\sigma}^2}{(\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*)} \beta^* = \\ &\mathbf{b} + \frac{(p-2)\hat{\sigma}^2}{(\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*)} (\beta^* - \mathbf{b}) \quad , \text{ onde} \end{aligned} \quad (6.2.2)$$

$\mathbf{b}$ é o estimador de mínimos quadrados de β e $\hat{\sigma}^2$ é o estimador usual da variância residual do modelo 6.2.1.

Note que o estimador James-Stein é uma combinação linear envolvendo o estimador de regressão usual e a média da distribuição a priori de β .

6.2.1. A obtenção de β^*

β^* é obtido através de um procedimento auto-explicativo nos mesmos moldes do descrito no tópico anterior. Para facilidade de compreensão, considere que um indivíduo tenha realizado, na fase auto-explicativa, uma avaliação idêntica à do indivíduo S_1 , tabela 6.1. O vetor β^* poderia ser obtido tomando-se $\beta^* = (w_k u_{ki})$, vide tabela 6.7. No entanto, algumas modificações, visando compatibilizar as escalas de β^* e $\mathbf{b}$, serão propostas.

Na tabela 6.8, encontram-se os dados referentes à fase conjunta. A avaliação deu-se através da ordenação dos estímulos, descritos através de uma matriz de planejamento $\mathbf{X}$, em ordem decrescente de preferência. A fim de evitar a estimação de um termo constante no modelo de análise de variância, é adotada como variável dependente os postos centrados,

ou seja o posto menos a média dos postos. Nesse planejamento faz-se a restrição usual de uma análise de variância com fatores fixos, ou seja $\sum_j \beta_{kj} = 0$, onde β_{kj} é o parâmetro correspondente ao nível j do atributo k . O estimador de mínimos quadrados de $\beta' = (\beta_{11} \ \beta_{21} \ \beta_{31})$ é:

$$b' = (0.50 \quad -1.75 \quad 1.25) \quad \text{com} \quad \hat{\sigma}^2 = 0.75.$$

TABELA 6.7. Utilidades auto-explicadas dos níveis dos atributos A, B e C

níveis	wu	restrito
A_1	200	-20
A_2	160	20
B_1	200	-50
B_2	100	50
C_1	540	30
C_2	600	-30

TABELA 6.8. Dados da fase conjunta

estímulo	X			r	centrado
1	1	1	1	5	0,5
2	1	-1	-1	6	1,5
3	-1	1	-1	1	-3,5
4	-1	-1	1	8	3,5
5	1	1	-1	2	-2,5
6	1	-1	1	7	2,5
7	-1	1	1	3	-1,5
8	-1	-1	-1	4	-0,5

Do modo como foi proposto, não há uma correspondência direta entre o vetor β^* e β , uma vez que na definição do primeiro não foi imposta nenhuma restrição. Isso é facilmente corrigido impondo-se a mesma restrição do planejamento da fase conjunta aos dados da penúltima coluna da tabela 6.7, o resultado disso encontra-se na última coluna dessa tabela, resultando em:

$$b^{*'} = (-20 \quad -50 \quad 30)$$

Note que há uma visível diferença de escala entre os vetores $\mathbf{b}$ e $\mathbf{b}^*$. A fim de corrigir esse problema, sugere-se a realização da regressão de $\mathbf{y}$ em $\mathbf{Xb}^*$, ou seja, estimar, por mínimos quadrados, o fator de escala c de tal modo que:

$$\mathbf{y} = c\mathbf{Xb}^*$$

Isso feito, o vetor ajustado de parâmetros a priori será

$$\boldsymbol{\beta}^* = \hat{c}\mathbf{b}^*.$$

Para o exemplo, numericamente, obtem-se $c = 0.030263$, resultando que:

$$\boldsymbol{\beta}^{*'} = c\mathbf{b}^{*'} = (-0.6053 \quad -1.5132 \quad 0.9079).$$

Desse modo, chega-se, de (6.2.1), que:

$$\boldsymbol{\beta}_{bayes} = \mathbf{b} + 0.067217(\boldsymbol{\beta}^* - \mathbf{b}) = (0.4257 \quad -1.7341 \quad 1.2270)'.$$

6.2.2. Comentários sobre a utilização do método bayesiano

A principal crítica ao uso desta metodologia reside na suposição de que o vetor de avaliações $\mathbf{y}$ se distribua segundo uma normal, pois muitas vezes essa variável nem segue uma escala métrica. Apesar desta ressalva, em um estudo piloto desenvolvido por Cattin et al(1983), o estimador bayesiano mostrou-se superior ao de mínimos quadrados em termos da qualidade de suas previsões.

CAPÍTULO 7

Estudos Empíricos

Nesta parte da dissertação, são apresentados estudos empíricos envolvendo aspectos de uma **AP**. O capítulo é dividido em duas seções: em 7.1 alguns resultados de uma série de estudos empíricos encontrados na literatura são apresentados; na seção 7.2 são discutidos os resultados de um estudo original.

7.1. Resultados extraídos da literatura

Nos artigos levantados, três temas de interesse se destacam: comparação de aspectos ligados à fase de coleta de dados, estudo da má especificação do modelo e comparação entre diferentes métodos de estimação.

No apêndice 8, são apresentados os resumos dos artigos mencionados no que se refere aos seus objetivos e obtenção de dados.

7.1.1. Coleta de dados

a. Confiabilidade estrutural

Diz-se haver confiabilidade estrutural, quando a percepção de um atributo (ou nível de atributo) não se altera quando este é apresentado juntamente com diferentes conjuntos de fatores (níveis). Esse aspecto é importante, pois se o peso relativo de um atributo (nível) em relação a um outro muda conforme o contexto, não se pode ignorar nenhum fator quando do planejamento de uma AP, pois isso acarretaria vícios em uma eventual simulação da preferência por produtos reais.

Em McCullough & Best(1979), produtos definidos pela variação de níveis de 3 fatores foram avaliados. A seguir, um dos fatores foi substituído por um quarto, procedendo-se a uma nova avaliação. As estimativas padronizadas dos parâmetros, correspondentes aos níveis dos atributos envolvidos nas duas fases, foram comparadas. Verificou-se que, em termos médios, dois de um total de 12 parâmetros foram significativamente diferentes; no entanto, os autores consideraram a magnitude da diferença relativamente pequena se comparada com a das estimativas. A conclusão dos autores foi favorável quanto à confiabilidade estrutural.

Malhotra(1982) comparou a importância relativa de 5 fatores. Na primeira fase de seu experimento, estímulos foram construídos através da combinação de níveis dos 5 atributos e avaliados pelos entrevistados. Na segunda fase, 5 novos fatores foram adicionados aos primeiros e uma nova avaliação envolvendo níveis dos dez atributos foi realizada. Na análise dos resultados, verificou-se que, nas duas fases, um mesmo fator apresentava peso relativo bastante superior aos demais (atributo dominante). Na comparação dos pesos relativos das duas etapas notou-se que a única diferença significativa correspondia ao atributo dominante; mais ainda, que no caso de dez fatores seu valor era sensivelmente menor. Malhotra argumentou que é fato conhecido na literatura (Srinivasan(1977), Jacob et al(1978) e Newman (1977)) que na presença de um maior número de fatores a tendência dos entrevistados escolherem um atributo dominante diminui. Assim, o autor considerou-se satisfeito quanto à confiabilidade estrutural de uma AP.

b. Formas de avaliação

Carmone, Green & Jain(1978) geraram observações visando estudar o resultado

de uma **AP** sob diversas situações (apêndice 8). Um dos pontos de interesse era comparar três formas de avaliação: escala contínua, postos e escala de 6 pontos. Os resultados das análises dos dados provenientes das três variáveis resposta não diferiram, em termos da capacidade preditiva¹ quando, na presença de interação nos dados, estimou-se um modelo aditivo. Também não houve diferença quando na geração dos dados foi considerada uma parcela de erro aleatório. Por outro lado, quando inexistia interação nos dados e não havia a parcela aleatória de erro, a escala de 6 pontos mostrou-se menos eficiente. Nesse trabalho, concluiu-se que, em termos da capacidade preditiva, há uma certa equivalência ao se adotar uma variável resposta contínua ou avaliação via postos.

Cattin & Bliemel(1978) compararam duas formas de avaliação: postos e escala de 7 pontos, sendo que na última, três situações foram contempladas: juiz exagerado (tendência a atribuir valores altos ou baixos à maioria dos estímulos), juiz indeciso (tendência a atribuir valores intermediários à maioria dos estímulos) e juiz justo (tendência a atribuir valores de maneira homogênea, segundo a real preferência pelo estímulo). Os autores utilizaram a soma dos erros absolutos, SEA, na comparação e constataram que o juiz justo sempre apresentou menor SEA; o juiz indeciso, na maioria dos casos, apresentou maior SEA. As SEA's do juiz justo e postos estiveram, na maioria dos casos, bastante próximas.

Em resumo, os autores mostraram que, no que se refere ao ajuste de modelos, pode haver equivalência (ou até superioridade) entre uma escala diferencial semântica e uma avaliação via postos. Deve-se, no entanto, notar que esse resultado só foi constatado no caso de juiz justo, sendo que nas outras situações, a avaliação via postos mostrou-se superior. Ao se realizar uma pesquisa, não é possível, a priori, garantir que todos os entrevistados responderão como juízes justos. Desse modo, caso haja dúvidas quanto à adoção de postos, ou de uma escala semântica, parece razoável que a escolha recaia sobre a primeira opção.

c. Planejamento fracionário

Tanto Carmone, Green & Jain(1978) como Malhotra(1982) verificaram que a adoção de planejamentos fracionários não altera de maneira significante os resultados de

¹ a capacidade preditiva de um modelo está ligada à qualidade de suas previsões, assim, quanto mais próximas estiverem as previsões dos valores observados, maior será a capacidade preditiva do modelo. Nesse caso, essa propriedade foi medida através do coeficiente τ de Kendall

uma AP.

d. Comparação entre perfil completo (PC) e "trade-off" (TO)

Segal(1982) estudou a estabilidade dos planejamentos PC e TO, ou seja, se duas avaliações de um conjunto de perfis, separadas por um certo intervalo de tempo e realizadas por um mesmo entrevistado, serão semelhantes. Calculado o coeficiente de correlação de Spearman entre os postos das duas avaliações, constatou-se que, para PC, 72% dos casos resultou em um coeficiente superior a 0,90 e 96% superior a 0,75, sendo seu valor médio de 0,92; já para TO, 59% foi superior a 0,90 e 83% superior a 0,70, sendo o valor médio de 0,84. Comparados os valores médios, observou-se que esses diferiam significativamente. Os parâmetros estimados na primeira e segunda realizações foram comparados, constatando-se não haver diferenças significantes. Esses resultados levam a concluir pela estabilidade dos dois métodos, sendo PC ligeiramente mais estável que TO. No caso do PC a estabilidade é confirmada por McCullough & Best(1979).

Veen & Beazley(1977) detectam pequenas diferenças nas estimativas dos modelos, quando os dados são coletados por planejamento PC e TO.

A comparação entre os resultados de uma análise com dados provenientes de um planejamento PC ou TO será realizada no item relativo à comparação dos métodos de estimação.

7.1.2. Má especificação do modelo

Os dois artigos que lidaram com esse tema (Carmone, Green & Jain(1978) e Carmone & Green(1981)) geraram observações a partir de modelos com e sem termos de interação.

Carmone & Green(1981) definiram dois tipos de interação: cruzada (crossover) e não cruzada (figura 7.1.). Os resultados basearam-se em duas medidas: a correlação entre postos previstos e observados e a porcentagem de acerto do melhor estímulo de uma amostra de validação². No que se refere à correlação, percebeu-se que para os dados gerados

² nesta dissertação, entende-se por amostra de validação um conjunto de estímulos que é avaliado apenas com o intuito de testar o modelo estimado, ou seja, que não participa do processo de estimação

com interação não cruzada houve uma certa equivalência entre os resultados do modelo aditivo e do modelo com interação. Já no que se refere à previsão do melhor estímulo, o modelo aditivo mostrou-se bastante inferior na presença de qualquer tipo de interação nos dados. Essa inferioridade foi também observada por Carmone, Green & Jain(1978).

FIGURA 7.1. Representação gráfica de uma situação onde existe interação não cruzada e de outra, onde existe interação cruzada

Esses resultados mostram a importância da adoção de planejamentos que permitam a estimação de termos de interação quando se suspeita de sua existência. Green(1973) discorre sobre a importância de se considerar termos de interação em modelos de preferência.

7.1.3. Comparação dos métodos de estimação

Wittink & Cattin(1981) compararam a eficiência dos métodos MONANOVA, LINMAP, LOGIT e mínimos quadrados (MQ) ao se trabalhar com modelo atributo dominante (aquele em que há um atributo com preferência bastante superior aos demais) ou com modelo compensatório (onde a preferência de todos os atributos é próxima). Concluiu-se que MQ e MONANOVA são equivalentes tanto em termos do ajuste ("stress") quanto da capacidade preditiva (correlação entre os valores previstos e observados). No que se refere à capacidade preditiva, MQ e MONANOVA mostraram-se superiores ao LOGIT. Quanto ao ajuste, MQ e MONANOVA foram superiores ao LOGIT e LINMAP no caso de modelo compensatório e inferiores no caso de atributo dominante, onde LINMAP obteve a melhor performance.

A similaridade entre MQ e MONANOVA no que se refere à capacidade preditiva também foi observada por Carmone, Green & Jain(1978) e Jain et al(1979).

Os métodos híbrido e de mínimos quadrados foram comparados em Green, Goldberg & Montemayor(1981), Akaah & Korgaonkar(1983) e Green(1984). Akaah & Korgaonkar constataram que os pesos relativos dos atributos diferem entre os métodos. No que tange à capacidade de prever os postos de uma amostra de validação, há uma equivalência entre eles; esse fato também foi verificado por Green, Goldberg & Montemayor.

Green(1984) relata os resultados dos estudos de Green, Goldberg & Wiley(1982), onde o modelo híbrido mostrou-se superior, em termos da capacidade preditiva, ao de MQ e de Cattin, Hermet & Proche(1982) onde constatou-se uma equivalência entre os métodos.

Cattin, Gelfand & Danes(1983), comparando o método bayesiano com o de mínimos quadrados, constataram que em termos do ajuste MQ é melhor que o bayesiano. No entanto, no que se refere à capacidade preditiva, o bayesiano tem um comportamento ligeiramente melhor.

Os métodos MONANOVA, JOHNSON³, LINMAP, LOGIT e MQ foram comparados ao se adotar planejamentos PC ou TO (Jain, Acito & Mahajan(1979)). Constatou-se que a importância relativa média dos atributos para os dados coletados segundo PC ou TO, apesar de não muito distantes, diferiam significativamente. Em termos do coeficiente

³ veja apêndice 9

de correlação de Spearman médio entre os postos previstos e observados constatou-se haver equivalência entre PC e TO para cada método de estimação. Ao se adotar um plano TO, os métodos LINMAP e LOGIT mostraram-se significativamente superiores aos demais. O método de Johnson mostrou-se eficaz em planejamento TO e relativamente ruim em um plano PC (tabela 7.1).

TABELA 7.1. Coeficiente de correlação de Spearman médio (e respectivo desvio-padrão, entre parênteses) calculado entre os postos previstos e observados (amostra de validação)

Planejamento	Método de Estimação				
	MONANOVA	JOHNSON	LINMAP	LOGIT	MQ
Trade off	0,7264 (0,256)	0,7503 (0,234)	0,7789 (0,199)	0,7735 (0,209)	0,7478 (0,221)
Perfil Completo	0,7340 (0,305)	0,7149 (0,291)	0,7469 (0,282)	0,7530 (0,294)	0,7466 (0,295)

fonte: Jain, Acito, Malhotra & Mahajan (1979)

TABELA 7.2. Porcentagem de previsão correta dos postos da amostra de validação, segundo método de estimação e planejamento

postos observados	MONANOVA		JOHNSON		LINMAP		LOGIT		MQ	
	TO	PC	TO	PC	TO	PC	TO	PC	TO	PC
1	42	42	46	39	51	46	43	43	44	44
2	35	40	37	36	26	26	38	43	38	42
3	16	33	25	28	25	34	24	39	15	35
4	30	25	17	24	15	19	23	29	25	27
5	17	35	29	29	27	22	25	28	15	32
6	30	32	34	24	31	22	34	31	33	34
7	28	33	26	30	33	34	30	30	27	29
8	87	65	75	63	71	56	88	72	91	70
acerto	36	38	36	34	35	32	38	40	36	39

fonte: Jain, Acito, Malhotra & Mahajan (1979)

Em termos da predição do estímulo preferido, LINMAP apresentou as maiores porcentagens de acerto (tabela 7.2). Analisando a porcentagem de previsão correta de qualquer posto, LOGIT foi ligeiramente superior aos demais.

Resumindo, para os estudos acima, pode-se concluir que:

- a. os pesos relativos dos atributos parecem diferir quando se adota diferentes métodos de estimação;
- b. mínimos quadrados e MONANOVA equivalem em termos da capacidade preditiva;
- c. LINMAP e LOGIT obtiveram melhor desempenho em modelos atributo dominante. LINMAP destaca-se quando os dados são coletados segundo um planejamento TO;
- d. em termos da capacidade preditiva, os métodos híbrido e bayesiano mostraram-se equivalentes (em algumas situações superiores) ao de mínimos quadrados.

7.2. Estudo empírico

Realizou-se um estudo empírico visando a comparação dos métodos de estimação MONANOVA, LINMAP, LOGIT⁴, híbrido, bayesiano e MQ.

Uma amostra de 31 pessoas (15 da Fundação Getúlio Vargas (FGV) e 16 do Departamento de Estatística do IME-USP), divididas quanto ao sexo, avaliaram diferentes automóveis descritos através dos fatores marca (A, B, C ou D), consumo (baixo, moderado ou alto⁵), volume (2 ou 3), potência do motor (1.6 ou 1.8) e combustível (álcool ou gasolina). Cada indivíduo foi submetido a um questionário organizado em duas partes:

- a. auto-explicativa (AE) - onde a importância dos atributos e seus níveis foi obtida diretamente. Os atributos foram avaliados segundo o método da soma constante e seus níveis através da atribuição de nota 10 ao preferido de cada fator e, a partir daí, aos restantes⁶;
- b. conjunta - onde foram avaliados 16 estímulos (amostra principal), construídos a partir de um planejamento perfil completo fracionário. Além disso, 8 outros perfis (amostra de validação) foram avaliados com o intuito de testar os modelos estimados. Os estímulos, apresentados em cartões, foram ordenados segundo a preferência dos juízes.

Metade da amostra avaliou inicialmente a parte "a" e metade começou pela parte "b": isso foi feito para verificar do efeito da ordenação nas respostas obtidas.

⁴ Logit condicional

⁵ ver definições no apêndice 10

⁶ ver seção 6.1.2.

No apêndice 10 é apresentado o questionário utilizado na coleta de dados, o planejamento fracionário adotado, os estímulos da amostra de validação e um modelo dos cartões que descrevem os estímulos.

Ressalta-se que, do mesmo modo que em Akaah, Pradeep & Korgaonkar(1983), todos os parâmetros do modelo híbrido foram estimados a nível individual, uma vez que cada elemento da amostra avaliou a totalidade dos estímulos do planejamento fracionário. Os modelos híbridos foram estimados em dois estágios (seção 6.1.3).

O seguinte modelo utilidade⁷ foi estimado:

$$y_j = \sum_{i=1}^8 \beta_i x_{ji} + \epsilon_j \quad \text{sendo que para o estímulo } j, \text{ tem-se que:}$$

$$x_{j1} = \begin{cases} 1 & \text{se 2 volumes} \\ -1 & \text{se 3 volumes} \end{cases} \quad x_{j2} = \begin{cases} 1 & \text{se 1.6} \\ -1 & \text{se 1.8} \end{cases} \quad x_{j3} = \begin{cases} 1 & \text{se álcool} \\ -1 & \text{se gasolina} \end{cases}$$

$$x_{j4} = \begin{cases} 1 & \text{se marca A} \\ 0 & \text{se marca B ou C} \\ -1 & \text{se marca D} \end{cases} \quad x_{j5} = \begin{cases} 1 & \text{se marca B} \\ 0 & \text{se marca A ou C} \\ -1 & \text{se marca D} \end{cases}$$

$$x_{j6} = \begin{cases} 1 & \text{se marca C} \\ 0 & \text{se marca A ou B} \\ -1 & \text{se marca D} \end{cases}$$

$$x_{j7} = \begin{cases} 1 & \text{se consumo baixo} \\ 0 & \text{se consumo moderado} \\ -1 & \text{se consumo alto} \end{cases} \quad x_{j8} = \begin{cases} 1 & \text{se consumo moderado} \\ 0 & \text{se consumo baixo} \\ -1 & \text{se consumo alto} \end{cases}$$

Para o processo computacional de estimação, foi escrito um programa em Turbo Pascal (Wood(1987)) que prepara os arquivos de dados para a utilização dos seguintes pacotes: PC-MDS (Smith(1988)) para o método MONANOVA; LIMDEP (Greene(1985)) para o método LOGIT e LP88 (LP88(1987)) para o método LINMAP; além disso, construiu-se um programa no pacote SPSS/PC+ (SPSS/PC+(1989)) para o método híbrido e

⁷ estimou-se o modelo (6.1.1) no caso do modelo híbrido

no módulo CM do SOC (SOC(1989)) para os métodos bayesiano, MQ e LOGIT. Posteriormente, na avaliação dos resultados, utilizou-se o pacote SPSS/PC+.

São necessários alguns comentários a respeito das estimativas dos modelos híbrido, bayesiano, LINMAP, MONANOVA e LOGIT:

- a. Híbrido e bayesiano - não foi possível a estimação desses modelos para um juiz do sexo feminino, da FGV e que respondeu na ordem AE-conjunta, por erro no preenchimento da parte "a" do questionário;
- b. LINMAP. Nota-se, em grande parte dos modelos, a tendência de estimativas idênticas de parâmetros, o que aumenta a possibilidade de empates na previsão da utilidade total de um conjunto de estímulos. Tome como exemplo o modelo abaixo, estimado para o primeiro indivíduo da amostra: ele indica a mesma utilidade para os estímulos (2 volumes, 1.6, gasolina, marca A, baixo consumo) e (3 volumes, 1.8, álcool, marca C, consumo moderado),

$$y_j = 0,00290698x_{j1} + 0x_{j2} + 0,00290698x_{j3} + 0,00145349x_{j4} + 0,00145349x_{j5} + \\ + 0,00145349x_{j6} - 0,00387597x_{j7} - 0,00387597x_{j8};$$

- c. MONANOVA - O programa PC-MDS oferece dois critérios de parada para o processo iterativo de estimação: o valor da razão entre o "stress" anterior e atual ("default"=0,999) e a determinação de um valor aceitável para o "stress" ("default"=1%). Dos 31 modelos, em 19 houve convergência pelo critério da razão ("default"), estando o número de iterações entre zero e 99; em 4 o processo parou ao se atingir o "stress" aceitável e nos 8 restantes foi necessário um relaxamento no critério da razão para possibilitar a parada do processo iterativo em menos de 1000 iterações. Constatou-se que, mesmo para o último grupo de 8 pessoas, sempre houve convergência das estimativas dos parâmetros (pelo menos até a terceira casa decimal), sendo que no último grupo todos os coeficientes de correlação de Spearman entre os postos observados e previstos foram superiores a 0,95, o que fez com que não houvesse nenhum constrangimento em se considerar todos os modelos estimados;
- d. LOGIT - O programa LIMDEP permite a adoção de três critérios de convergência: um baseado na diferença do valor da função de verossimilhança entre iterações consecutivas,

outro baseado na diferença máxima entre as estimativas dos parâmetros de duas iterações consecutivas e um terceiro no valor da norma do gradiente. Houve convergência em todos os casos no que se refere ao valor da função e da norma do gradiente; no entanto, apenas para 2 pessoas, houve convergência nas estimativas dos parâmetros. A fim de verificar a possibilidade desses resultados terem sido causados por erro na construção do arquivo de dados de entrada do LIMDEP, construiu-se um programa no pacote SOC que exigia um outro tipo de entrada de dados; constatou-se que os resultados dos dois eram idênticos para os 31 modelos. Nas próximas seções será visto que essa falta de convergência não afeta a qualidade dos resultados da análise.

7.2.1. Análise preliminar

A amostra pode ser dividida segundo 3 fatores: local da entrevista (FGV ou IME), sexo do entrevistado e ordem do questionário (AE-conjunta ou conjunta-AE). Na tabela 7.3 encontra-se o número de pessoas em cada classificação. Antes de proceder à comparação dos métodos de estimação, verificou-se a existência dos efeitos desses fatores na importância relativa dos atributos e na correlação de Spearman entre os postos previstos e observados⁸ para modelos estimados por cada método.

TABELA 7.3. Classificação dos membros da amostra segundo local da entrevista, sexo e ordem de resposta do questionário

local	sexo	AE-Conjunta	Conjunta-AE
FGV	masculino	4	4
	feminino	4*	3
IME	masculino	4	4
	feminino	4	4

* : 3 observações no caso de estimação pelos métodos híbrido e bayesiano

a. Análise dos pesos relativos dos atributos

Na tabela 7.4. encontra-se o peso relativo médio, de cada atributo e seu correspondente erro-padrão para modelos estimados pelos 6 métodos. Em todos os métodos

⁸ tanto prevendo os postos da amostra principal como da de validação

TABELA 7.4. Pesos relativos médios (e respectivos erros-padrão) de cada atributo segundo o método de estimação

Atributo	MONANOVA	LINMAP	MQ	Híbrido	Bayesiano	LOGIT
volume	0,1048 (0,0243)	0,1061 (0,0225)	0,1208 (0,0240)	0,1220 (0,0248)	0,1158 (0,0246)	0,1027 (0,0241)
potência	0,0661 (0,0112)	0,0727 (0,0136)	0,0850 (0,0132)	0,0807 (0,0131)	0,0873 (0,0143)	0,0856 (0,0136)
combustível	0,0682 (0,0185)	0,0883 (0,0322)	0,0813 (0,0160)	0,0818 (0,0165)	0,0813 (0,0170)	0,0806 (0,0165)
marca	0,5122 (0,0524)	0,4135 (0,0434)	0,4326 (0,0359)	0,4301 (0,0370)	0,4281 (0,0367)	0,4393 (0,0421)
consumo	0,2487 (0,0338)	0,3193 (0,0484)	0,2804 (0,0308)	0,2855 (0,0314)	0,2875 (0,0311)	0,2918 (0,0344)
n.obs.	31	31	31	30	30	31

TABELA 7.5. Significância do efeito de local, segundo teste, método de estimação, para os dados da amostra principal

método	Teste	
	ANOVA	Kruskal-Wallis
MONANOVA	0,060	0,029
LINMAP	0,033	0,072
MQ	0,142	0,199
Híbrido	0,063	0,139
Bayesiano	0,108	0,235

TABELA 7.6. Significância do efeito de interação entre local de entrevista e sexo do entrevistado, para os dados da amostra de validação

método	Teste	
	ANOVA	Kruskal Wallis
MONANOVA	0,017	0,020
LINMAP	0,025	0,216
MQ	0,025	0,046
Híbrido	0,074	0,110
Bayesiano	0,064	0,214
LOGIT	0,072	0,051

marca mostrou ser, destacadamente, o fator mais importante, seguido pelo consumo e em um terceiro nível e próximos, os fatores restantes. Marca teve sua maior importância ao se adotar o método MONANOVA. Percebe-se uma semelhança nos pesos relativos médios dos atributos para todos os métodos, com exceção do MONANOVA.

Por método, aplicou-se em cada atributo uma análise de variância utilizando os fatores: local, ordem e sexo. Constatou-se a existência de interação entre ordem e sexo para combustível nos métodos: bayesiano (0,051)⁹, híbrido (0,051), MONANOVA (0,005), MQ (0,042) e LOGIT (0,020); para marca, a interação entre ordem sexo e local foi significativa para os métodos bayesiano (0,074), híbrido (0,022), MONANOVA (0,077), MQ (0,016), LINMAP (0,011) e LOGIT (0,017); para o fator consumo, a interação anterior foi significativa para os métodos híbrido (0,014), MQ (0,056), LINMAP (0,016) e LOGIT (0,022), já para os métodos bayesiano (0,011) e MONANOVA ($< 0,001$) a interação significativa foi entre ordem e sexo.

No apêndice 11, encontram-se as tabelas com as médias que permitem avaliar a influência dos efeitos significantes para combustível, marca e consumo. No que se refere a marca, nota-se, para todos os métodos, que a importância média do sexo masculino em cada local é maior ao se utilizar a sequência Conjunta-AE; o inverso ocorre com o sexo feminino. A realização de um teste de hipóteses para verificação das afirmações anteriores seria necessária, no entanto, considerou-se que o pequeno número de observações em cada casela não recomendaria essa ação.

Infelizmente, a existência dos efeitos de interação não permitem uma conclusão geral sobre o efeito da ordem nas respostas dos juízes; há apenas evidências de sua existência.

b. Análise do coeficiente de correlação de Spearman, calculado entre os postos previstos e observados

O efeito de ordem, local e sexo foi também verificado no coeficiente de correlação de Spearman entre os postos previstos e observados, para as amostras principal e de validação. Realizados uma análise de variância, e o teste de Kruskal-Wallis, constatou-se que, na amostra principal, o único efeito significativo para alguns métodos de estimação, foi o de local (tabela 7.5); já para a amostra de validação, a interação local e sexo foi significativa

⁹ os valores entre parênteses indicam o nível descritivo dos testes

para a maioria dos métodos (tabela 7.6). Nota-se que os entrevistados do IME (tabela A11.4) possuem uma correlação média maior que os da FGV; esta conclusão é significativa para os métodos MONANOVA e LINMAP da amostra principal.

Na amostra de validação encontrou-se uma observação aberrante, correspondente ao caso 8.

Na seção 7.2.3, os coeficientes de correlação são comparados 2 a 2 (amostras pareadas). Logo, torna-se necessário saber se as diferenças entre as correlações de dois métodos são, ou não, influenciadas por ordem, local e sexo. Desse estudo, conclui-se que:

- a. para a amostra principal - a interação sexo e local foi significativa para as diferenças entre MONANOVA-LOGIT (0,083), MONANOVA-EMQ (0,044) e MONANOVA-híbrido (0,034); já a interação entre sexo, local e ordem foi significativa para Bayesiano-LINMAP (0,087) e LINMAP-Híbrido (0,073);
- b. para a amostra de validação - analisando a diferença média entre os coeficientes de correlação tomados dois a dois, constatou-se que o caso 8 é também uma observação aberrante para as diferenças: MONANOVA-Bayesiano, MONANOVA-MQ, Bayesiano- LINMAP, Bayesiano-MQ, Bayesiano-Híbrido, LOGIT-Bayesiano e LOGIT-MQ; para esses casos essa observação foi removida quando da realização dos testes de hipóteses. Constatou-se a existência de interação significativa entre ordem e local para LOGIT-MQ (0,094); a interação ordem e sexo foi significativa para Bayesiano-MQ (0,077), já o efeito de ordem foi significativo para LOGIT-Híbrido (0,034) e LOGIT-Bayesiano (0,031).

Nas tabelas A11.6 a A11.10 encontram-se as correlações médias relativas aos efeitos significantes.

7.2.2. Comparação da porcentagem de acerto na previsão

Comparou-se para cada método de estimação os postos observados e previstos por cada modelo. Isso foi realizado para cada amostra, através do estudo da diferença absoluta entre os postos previstos e observados (diferença) e do nível de acerto na previsão de cada posto¹⁰.

¹⁰ no caso de ocorrer empate na previsão, por exemplo se os três menores valores previstos forem idênticos e seus postos observados forem 1, 2 ou 3, considera-se acerto na previsão

Ressalta-se que os resultados referentes à amostra principal refletem, além da capacidade preditiva do método, a qualidade do ajuste dos modelos.

TABELA 7.7. Distribuição acumulada da diferença entre os postos previstos e observados (amostra principal)

diferença	MONANOVA	LINMAP	MQ	Híbrido	Bayesiano	LOGIT
0,0	44,0	8,9	50,2	52,1	54,8	62,3
0,5	54,8	34,5	59,1	52,1	55,0	64,7
1,0	79,6	45,8	87,1	83,4	84,2	88,1
1,5	84,1	63,7	89,1	83,4	84,4	88,5
2,0	92,5	69,8	95,2	93,8	94,2	96,0
2,5	92,9	80,4	96,0	93,8	94,2	96,4
3,0	96,8	84,5	98,4	98,2	96,3	98,2
3,5	96,8	91,5	98,4	98,2	96,3	98,2
4,0	98,2	93,5	99,0	99,0	98,5	99,0
4,5	98,2	97,0	99,0	99,0	98,5	99,0
5,0	99,2	97,4	99,4	99,4	99,2	99,8
5,5	99,2	99,6	99,4	99,4	99,2	99,8
6,0	99,4	100,0	99,8	99,8	99,4	99,8
6,5	99,4	100,0	99,8	99,8	99,4	99,8
7,0	99,6	100,0	99,8	99,8	99,6	99,8
7,5	99,6	100,0	99,8	99,8	99,6	99,8
8,0	100,0	100,0	99,8	99,8	99,8	99,8
8,5	100,0	100,0	99,8	99,8	99,8	99,8
9,0	100,0	100,0	100,0	100,0	100,0	100,0

TABELA 7.8. Previsões corretas para os postos 1 e 16 (amostra principal)

posto		MONANOVA	LINMAP	MQ	Híbrido	Bayesiano	LOGIT
			limite s/limite				
1	n^0	23	12	31	25	21	22
	%	74	39	100	81	70	74
16	n^0	21	15	30	24	21	22
	%	68	48	97	77	70	74

a. Amostra principal

A tabela 7.7. descreve a distribuição acumulada da diferença entre o posto real e o previsto. A partir da tabela, observa-se que:

- a tendência de ocorrência de empates para LINMAP é facilmente constatada ao se observar as porcentagens relativas às diferenças não inteiras;
- a partir da diferença 0,5, nota-se uma certa semelhança entre MONANOVA, bayesiano, híbrido e EMQ;
- apesar da diferença máxima observada para o método LINMAP (6,0) ser menor que a de qualquer outro método, é impossível deixar de notar sua inferioridade em relação aos demais. Esta inferioridade deve-se, em grande parte, à tendência de ocorrência de empates, uma vez que em vários casos ocorreu 4 ou mais previsões de mesmo valor;
- o método LOGIT, apesar do problema de convergência das estimativas dos parâmetro, mostra-se superior aos demais.

Na tabela 7.8 encontra-se a porcentagem de previsões corretas do posto 1 (melhor estímulo) e do posto 16 (pior estímulo), para todos os métodos. Cabe ressaltar que na previsão quando houve empates nos postos de até três estímulos, por exemplo, no menor posto previsto e uma das observações empatadas correspondia ao melhor estímulo, considerou-se acerto. Caso o empate fosse entre 4, ou mais estímulos não se considerou acerto, com exceção do método LINMAP, onde são apresentados os resultados levando-se em conta o limite de 3 empates e sem considerar este limite máximo.

Analisando os resultados da tabela 7.8, percebe-se uma proximidade entre os métodos LOGIT e MQ e entre MONANOVA, híbrido e bayesiano. Já o método LINMAP mostrou-se inferior aos demais, quando se considera o limite de 3 empates. O excelente resultado do LINMAP ao se desprezar o limite deve ser encarado com cautela, uma vez que em um conjunto de 16 estímulos apontar 4 ou mais como igualmente preferidos não parece ser um bom resultado.

b. Amostra de validação

A distribuição da diferença entre os postos observados e previstos para a amostra de validação encontra-se na tabela 7.9. Novamente, a tendência de empates do método LINMAP é explicitada. Nota-se, a partir do valor 0,5, uma certa equivalência entre todos os métodos, havendo uma pequena desvantagem para o método híbrido e uma pequena vantagem para o bayesiano no que se refere à diferença máxima observada (5,0).

TABELA 7.9. Distribuição acumulada da diferença entre os postos previstos e observados (amostra de validação)

diferença	MONANOVA	LINMAP	MQ	Híbrido	Bayesiano	LOGIT
0,0	50,4	21,0	49,2	47,5	52,5	51,2
0,5	53,6	52,4	50,4	47,5	52,5	51,2
1,0	85,1	63,3	84,3	83,8	85,8	85,9
1,5	85,1	79,4	84,7	83,8	85,8	85,9
2,0	94,0	85,5	94,0	94,6	94,2	95,2
2,5	94,0	94,4	94,4	94,6	94,2	95,2
3,0	96,0	96,0	96,0	96,3	96,7	96,4
3,5	96,0	96,4	96,4	96,3	96,7	96,4
4,0	98,8	97,2	98,8	98,8	99,2	98,8
4,5	98,8	97,6	98,8	98,8	99,2	98,8
5,0	99,2	98,0	99,2	99,2	100,0	99,2
5,5	99,2	99,6	99,2	99,2	100,0	99,2
6,0	100,0	100,0	100,0	100,0	100,0	100,0

TABELA 7.10. Previsões corretas para os postos de 1 a 8 (amostra de validação)

posto		MONANOVA	LINMAP		MQ	Híbrido	Bayesiano	LOGIT
			limite	s/limite				
1	n^0	22	17	25	24	22	25	24
	%	71	55	81	77	73	83	77
2	n^0	17	17	23	17	15	19	16
	%	55	55	74	55	50	63	52
3	n^0	14	18	21	13	12	15	12
	%	45	58	68	42	40	50	39
4	n^0	11	18	23	11	11	10	11
	%	35	58	74	35	37	33	35
5	n^0	15	10	20	13	12	12	13
	%	48	32	65	42	40	40	42
6	n^0	16	14	18	14	13	14	15
	%	52	45	58	45	43	47	48
7	n^0	18	14	21	14	12	14	17
	%	58	45	68	45	40	47	55
8	n^0	20	19	21	19	17	17	19
	%	65	61	68	61	57	57	61

A porcentagem de previsão correta para os postos de 1 a 8 encontra-se na tabela 7.10. A porcentagem de acerto na previsão do melhor estímulo é maior para o método

bayesiano seguido pelo MQ e LOGIT. Na previsão do estímulo de menor preferência, MONANOVA apresenta os melhores resultados.

Comparando as tabelas 7.2 e 7.10, nota-se que o desempenho dos modelos para os dados aqui introduzidos, foi superior aos da literatura; no entanto nota-se a proximidade entre os métodos de estimação em cada tabela, com excessão do LINMAP.

Em resumo, no que se refere ao acerto na previsão, percebe-se uma inferioridade do método LINMAP seguido pelo híbrido e uma proximidade entre os restantes, sendo o LOGIT de melhor comportamento para a amostra principal, indicando uma qualidade superior no ajuste.

7.2.3. Comparação dos coeficientes de correlação de Spearman

Outro modo de medir a capacidade preditiva dos modelos estimados é o cálculo do coeficiente de correlação de Spearman entre os postos previstos e observados. Isso foi feito para os dados da amostra principal e da de validação.

a. Comparações descritivas

A fim de conhecer o comportamento das correlações através dos métodos, construiu-se as tabelas 7.11, 7.12 e 7.13. Em 7.11, apresenta-se o coeficiente médio e seu respectivo erro-padrão. Ao se comparar os coeficientes, nota-se uma semelhança entre os métodos, com excessão do LINMAP.

Em 7.12 e 7.13, dispõem-se os quartis dos coeficientes de correlação de cada método. Analisando essas tabelas, conclui-se que:

- LINMAP aparenta ser o de pior comportamento para as duas amostras;
- LOGIT apresentou os maiores valores do primeiro e terceiro quartis nas duas amostras, seguido pelos métodos bayesiano e MQ, aparentando alguma superioridade em relação aos demais.

TABELA 7.11. Médias e erros-padrão relativos ao coeficiente de correlação de Spearman, segundo amostra

Técnica	estatística	Amostra Principal	Amostra de Validação	
			completa	sem o caso 8
MONANOVA	média	0,9531	0,8379	0,8691
	erro-padrão	0,0116	0,0415	0,0283
LINMAP	média	0,8784	0,7503	0,7822
	erro-padrão	0,0147	0,0447	0,0324
MQ	média	0,9679	0,8361	0,8680
	erro-padrão	0,0085	0,0418	0,0279
Híbrido	média	0,9605	0,8333	0,8629
	erro-padrão	0,0093	0,0406	0,0287
Bayesiano	média	0,9558	0,8572	0,8736
	erro-padrão	0,0132	0,0327	0,0292
LOGIT	média	0,9711	0,8472	0,8738
	erro-padrão	0,0086	0,0377	0,0275

TABELA 7.12. Estatísticas descritivas relativas aos coeficientes de correlação de Spearman, calculadas para os dados da amostra principal

Técnica	Mínimo	Q_1	Mediana	Q_3	Máximo	N
MONANOVA	0,6882	0,9500	0,9701	0,9912	1,0000	31
LINMAP	0,7202	0,8342	0,8961	0,9393	0,9948	31
MQ	0,7971	0,9706	0,9845	0,9934	1,0000	31
Híbrido	0,7971	0,9544	0,9809	0,9912	1,0000	30
Bayesiano	0,7000	0,9581	0,9824	0,9941	1,0000	30
LOGIT	0,7971	0,9706	0,9882	0,9971	1,0000	31

TABELA 7.13. Estatísticas descritivas relativas aos coeficientes de correlação de Spearman, calculadas para os dados da amostra de validação

Técnica	Mínimo	Q_1	Mediana	Q_3	Máximo	N
MONANOVA	-0,0952	0,7619	0,9524	0,9762	0,9762	8
LINMAP	-0,2057	0,7303	0,8486	0,9157	0,9662	8
MQ	-0,1190	0,8333	0,9286	0,9701	0,9762	8
Híbrido	-0,0238	0,8036	0,9286	0,9584	0,9762	8
Bayesiano	0,3810	0,8036	0,9524	0,9762	1,0000	8
LOGIT	0,0476	0,8571	0,9286	0,9762	0,9762	8

TABELA 7.14. Níveis descritivos de testes que comparam os coeficientes de correlação de Spearman obtidos em cada técnica, para a amostra principal

Comparação	Ord	Sexo	Loc	t	Fried	Kend	Sinal	Wilc
MON - LIN				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
MON - MQ				0,002	0,002	0,001	0,002	< 0,001
		mas	FGV	0,046	0,034	0,014	0,031	0,028
		fem	FGV	0,023	0,008	0,008	0,016	0,018
		mas	IME	0,059	0,157	0,157	0,289	0,069
MON - Híb		fem	IME	0,257	1,000	1,000	1,000	0,345
		mas	FGV	0,184	0,100	0,061	0,093	0,037
		fem	FGV	0,605	1,000	1,000	1,000	0,917
		mas	IME	0,024	0,014	0,014	0,031	0,028
MON - Bay		fem	IME	0,688	1,000	1,000	1,000	0,600
		fem	IME	0,224	0,289	0,180	0,375	0,178
MON - LOG				0,382	0,361	0,317	0,424	0,119
MON - LOG				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
		mas	FGV	0,032	0,034	0,014	0,031	0,028
		fem	FGV	0,026	0,008	0,008	0,016	0,018
		mas	IME	0,049	0,034	0,034	0,070	0,050
LIN - MQ		fem	IME	0,109	0,480	0,317	0,625	0,144
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
LIN - Híb	AE-C	mas	FGV	0,105	0,046	0,046	0,125	0,068
	C-AE	mas	FGV	0,016	0,046	0,046	0,125	0,068
	AE-C	fem	FGV	0,049	0,083	0,083	0,250	0,109
	C-AE	fem	FGV	0,055	0,083	0,083	0,250	0,109
	AE-C	mas	IME	0,133	0,046	0,046	0,125	0,068
	C-AE	mas	IME	0,058	0,046	0,046	0,125	0,068
	AE-C	fem	IME	0,056	0,046	0,046	0,125	0,068
	C-AE	fem	IME	0,134	0,046	0,046	0,125	0,068
LIN - Bay				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
	AE-C	mas	FGV	0,093	0,046	0,046	0,125	0,068
	C-AE	mas	FGV	0,123	0,046	0,046	0,125	0,068
	AE-C	fem	FGV	0,662	0,564	0,564	1,000	0,593
	C-AE	fem	FGV	0,110	0,083	0,083	0,250	0,109
	AE-C	mas	IME	0,122	0,046	0,046	0,125	0,068
	C-AE	mas	IME	0,032	0,046	0,046	0,125	0,068
	AE-C	fem	IME	0,045	0,046	0,046	0,125	0,068
LIN - LOG	C-AE	fem	IME	0,129	0,046	0,046	0,125	0,068
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
MQ - Híb				0,104	0,045	0,012	0,019	0,002
MQ - Bay				0,024	0,273	0,201	0,286	0,020
MQ - LOG				0,037	0,073	0,041	0,064	0,027
Híb - Bay				0,472	0,715	0,637	0,815	0,617
Híb - LOG				0,019	0,0176	0,005	0,007	< 0,001
Bay - LOG				0,004	< 0,001	< 0,001	< 0,001	< 0,001

TABELA 7.15. Níveis descritivos de testes que comparam os coeficientes de correlação de Spearman obtidos em cada técnica, para a amostra de validação

Comparação	Ord	Sexo	Loc	t	Fried	Kend	Sinal	Wilc
MON - LIN				< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
MON - MQ				0,905	0,715	0,637	0,815	0,240
MON - Híb				0,686	0,855	0,819	1,000	0,469
MON - Bay				0,290	0,584	0,467	0,629	0,554
MON - LOG				0,400	0,590	0,491	0,647	0,644
LIN - MQ				0,001	0,001	0,001	0,001	0,001
LIN - Híb				0,003	< 0,001	< 0,001	0,001	0,003
LIN - Bay				0,002	0,002	0,002	0,003	0,001
LIN - LOG				0,001	< 0,001	< 0,001	< 0,001	< 0,001
MQ - Híb				0,673	0,855	0,655	1,000	0,893
MQ - Bay				0,441	0,458	0,206	0,344	0,386
	AE-C	mas		0,043	0,157	0,046	0,125	0,068
	C-AE	mas		0,384	0,724	0,564	1,000	0,593
	AE-C	fem		0,275	0,450	0,157	0,500	0,180
	C-AE	fem		0,182	0,450	0,157	0,500	0,180
MQ - LOG				0,368	0,584	0,405	0,581	0,328
	AE-C		FGV	0,250	0,480	0,157	0,500	0,178
	C-AE		FGV	0,199	0,257	0,178	0,375	0,178
	AE-C		IME	0,313	0,480	0,157	0,500	0,180
	C-AE		IME	0,290	0,724	0,655	1,000	0,225
Híb - Bay				0,173	0,353	0,096	0,180	0,139
Híb - LOG				0,113	0,465	0,285	0,424	0,103
	AE-C			0,531	0,796	0,564	1,000	0,593
	C-AE			0,037	0,197	0,132	0,227	0,033
Bay - LOG				0,849	0,853	0,796	1,000	0,733
	AE-C			0,207	0,211	0,096	0,180	0,139
	C-AE			0,075	0,267	0,103	0,219	0,075

b. Testes de significância

Os métodos de estimação tiveram seus coeficientes de correlação de Spearman comparados através dos testes t-pareado, Friedman, Kendall, teste do sinal e Wilcoxon pareado. A tabela 7.14 traz, para a amostra principal, o nível descritivo das diferenças entre os métodos. As principais conclusões extraídas dessa tabela são:

- LINMAP mostra-se inferior, em termos médios, aos demais;

- LOGIT mostra-se superior aos outros métodos;
- MONANOVA mostra-se inferior ao MQ;
- há equivalência entre os demais métodos.

A tabela 7.15 traz os resultados da amostra de validação. Em resumo,

- LINMAP mostrou-se inferior a todos os métodos;
- os outros métodos se equivalem.

7.3. Comentários

Para a amostra estudada, o método LOGIT apresentou, de um modo geral, o melhor desempenho e o LINMAP o pior. Os outros métodos mostraram-se bastante semelhantes em termos da capacidade preditiva.

O fraco desempenho do método LINMAP deve-se à já comentada tendência de empates nas estimativas dos parâmetros. Talvez a solução encontrada através do pacote LP88 não seja a única, existindo outras, onde essa tendência não seja tão forte. Uma alternativa para se evitar esses empates é alterando o programa linear, através da imposição de restrições que obriguem a diferença entre as estimativas de todos os parâmetros. É possível que, deste modo, o custo ótimo seja maior.

O método LOGIT, apesar de ter se mostrado, em termos médios, superior aos demais, foi, em termos computacionais, um dos mais demorados, perdendo apenas para o LINMAP. Além disso apresentou problemas de convergência nas estimativas dos parâmetros. Isto certamente diminui o seu atrativo, principalmente ao se trabalhar com uma amostra demasiadamente grande. Cabe observar, no entanto, que, com excessão do método LINMAP, quando se compara os resultados obtidos pelo método LOGIT com os dos outros procedimentos de estimação (excluindo a tabela 7.7) percebe-se que as diferenças não são tão grandes. Deste modo, a utilização de qualquer outro procedimento talvez não leve a resultados muito diferentes. Ressalta-se que essa proximidade não é apenas uma característica da amostra estudada: esse mesmo fato também ocorreu nos estudos empíricos extraídos da literatura.

Em resumo, apesar de haver um método de melhor desempenho, ele não difere em muito de outro. Isso pode ser levado em conta no momento da escolha de algum desses

procedimentos.

Obviamente, as conclusões deste estudo estariam prejudicadas pelo pequeno tamanho de amostra. No entanto, com excessão dos resultados referentes ao método LINMAP, o que foi aqui obtido encontra algum respaldo nos estudos empíricos da literatura, o que assegura certa validade nas conclusões.

7.4. Conclusões e futuros estudos

Nesta monografia, abordou-se todas as etapas envolvidas na aplicação de uma AP. Deu-se maior ênfase ao estudo de métodos de estimação de modelos de preferência, com estímulos formados segundo um planejamento perfil completo e avaliação por postos. Extensões para o planejamento "trade-off" e para escalas semânticas discretas podem ser feitas com facilidade, através da matriz de planejamento e de adaptações nas restrições de ordem a que estão sujeitos a maioria dos métodos.

Comparou-se, empiricamente, alguns métodos de estimação apresentados na dissertação. Verificou-se que, para os dados utilizados, o método LOGIT condicional mostrasse superior aos demais e o LINMAP, inferior. Apesar disso, constata-se que, na maioria dos casos, excetuando-se o procedimento LINMAP, a diferença entre os métodos, em termos da capacidade preditiva não é elevada.

Ao fim desta monografia parece que no estudo de AP, a maior ênfase deva ser dada ao seu planejamento, uma vez que os métodos de estimação, em geral, diferem pouco, quanto à capacidade preditiva. Parece, então, oportuno, avançar a pesquisa na direção de aspectos ligados ao planejamento de uma AP.

Deseja-se que este texto auxilie os interessados na utilização de AP, proporcionando, em um primeiro contacto com o tema, uma visão ampla das etapas envolvidas em sua aplicação. Acredita-se que a bibliografia levantada seja suficientemente abrangente, para aqueles que desejam um aprofundamento maior em algum tópico abordado.

APÊNDICE 1

Níveis de Mensuração

Segundo Davis(1976), em geral, as variáveis podem ser classificadas segundo 4 diferentes escalas de medida. Essas escalas são definidas de acordo com 4 propriedades, a saber:

1. classificação - se refere a possibilidade de diferenciar as realizações da variável;
2. ordenação - possibilidade de ordenar as realizações da variável;
3. distância - possibilidade de saber quantas unidades de medida separa uma realização de outra;
4. proporcionalidade - possibilidade de saber quantas vezes uma realização é maior do que outra.

A partir dessas propriedades, as variáveis podem ser classificadas nos seguintes níveis de mensuração (escalas de medida):

a. escala razão - essas variáveis possuem as propriedades 1, 2, 3 e 4. Por exemplo, suponha que um carro A tem um rendimento de 7 km/l e um carro B de 14 km/l. Sabe-se

que:

- i. A e B têm rendimentos diferentes (propriedade 1);
- ii. o rendimento de A é superior ao de B (propriedade 2);
- iii. o rendimento de B é 7 km/l a mais que o de A (propriedade 3) e
- iv. o rendimento de B é o dobro do de A (propriedade 4).

b. escala intervalar - essas variáveis devem satisfazer 1, 2 e 3. Um exemplo clássico de variável intervalar é a temperatura; se a temperatura de um corpo A é de 7 graus Celsius e a de um corpo B é de 14 graus, sabe-se que:

- i. A e B têm temperaturas diferentes;
- ii. a temperatura de B é superior a de A e
- iii. a temperatura de B é 7 graus Celsius maior que a de A.

Note que não é correto afirmar que o corpo B é duas vezes mais quente que o corpo A. Isto fica mais claro ao se converter a temperatura para graus Fahrenheit. Acontece que a temperatura possui o que se chama de zero arbitrário, ou seja, temperatura 0 representam diferentes estados físicos, de acordo com a escala de medida utilizada, assim, não faz sentido pensar em proporcionalidade entre duas temperaturas.

c. escala ordinal - possuem as propriedades 1 e 2, pense por exemplo que as possíveis avaliações de um curso são A, B e C.

- i. A, B e C são avaliações diferentes
- ii. A é melhor que B que é melhor que C.

Essas variáveis não possuem propriedades aritméticas, possuindo, no entanto a propriedade de transitividade¹.

d. escala nominal - são as que possuem apenas a propriedade 1, por exemplo, estado civil segue a escala nominal.

¹ transitividade significa que se $A > B$, $B > C$, então $A > C$

APÊNDICE 2

Planejamentos Fracionários

A utilidade básica de um planejamento fracionário é a consequente redução no número de estímulos a serem analisados. Em geral, esses planejamentos são construídos de modo a permitir a estimação dos efeitos principais de um modelo de análise de variância e de algumas interações previamente selecionadas.

A principal crítica quanto a esse tipo de planejamento é a existência de "confundimento" entre os termos a serem estimados e os que serão descartados do modelo. Para melhor entendimento dessa questão, considere os dados da tabela A.2.1. que reproduz o planejamento completo de um experimento com 3 fatores, cada qual com 2 níveis (2^3) e também um planejamento fracionário construído com base no completo.

O planejamento fracionário apresentado permite apenas a estimação dos efeitos principais do modelo. Em termos da matriz de planejamento a coluna correspondente a F_3 é idêntica à coluna correspondente à interação $F_1 F_2$, isso significa que o parâmetro estimado tanto pode se referir ao efeito do terceiro fator, quanto à interação entre os dois primeiros fatores. Diz-se então que há confundimento entre F_3 e $F_1 F_2$.

De um modo geral, entende-se por confundimento a uma colinearidade entre os

fatores do planejamento (Louviere(1988)).

O exemplo dado é recomendável quando se tem alguma garantia de que os termos descartados do modelo (as interações) realmente são não significantes. Caso isso não ocorra, as estimativas serão viciadas.

Ao se utilizar um planejamento fracionário é útil conhecer os confundimentos a ele associados para uma melhor análise dos resultados obtidos.

TABELA A2.1. Exemplo de um planejamento fracionário

efeitos principais			interações dois a dois			interação três a três
F_1	F_2	F_3	F_1F_2	F_1F_3	F_2F_3	$F_1F_2F_3$
-1	-1	-1	1	1	1	-1
-1	-1	1	1	-1	-1	1
-1	1	-1	-1	1	-1	1
-1	1	1	-1	-1	1	-1
1	-1	-1	-1	-1	1	1
1	-1	1	-1	1	-1	-1
1	1	-1	1	-1	-1	-1
1	1	1	1	1	1	1
Planejamento fracionário						
-1	-1	1	1	-1	-1	1
-1	1	-1	-1	1	1	1
1	-1	-1	-1	-1	-1	1
1	1	1	1	1	1	1

APÊNDICE 3

Regressão Isotônica

Em problemas usuais de análise de regressão simples, fixa-se uma série de valores de uma variável preditora $\tilde{x}$ e para cada valor verifica-se o comportamento de uma variável de interesse $\tilde{y}$. Nesse caso, deseja-se estimar uma função $g(x) = E[\tilde{y}/\tilde{x} = x]$.

A regressão isotônica, ou monotônica, possui os mesmos objetivos da regressão usual, com a restrição de $g(x)$ ter um comportamento monótono (crescente ou decrescente). A diferença está no fato de que na regressão usual deve-se estipular analiticamente a função g e, na monotônica, isso não é feito.

Barlow et al (1972) estudaram exaustivamente aspectos ligados à regressão isotônica partindo das seguintes definições:

Definição - Função Isotônica - Seja X um conjunto finito $x_1, \dots, x_p$ com a ordem: $x_1 \prec x_2 \prec \dots \prec x_p$. Uma função f em X é isotônica se, para $x_i, x_j \in X$ e $x_i \prec x_j \Rightarrow f(x_i) \leq f(x_j)$.

Definição - Regressão Isotônica - Considere g uma função em X e w uma função positiva em X . Uma função isotônica g^* em X é uma regressão isotônica de g com pesos w com respeito a ordem simples $x_1 \prec x_2 \prec \dots \prec x_p$, se minimizar, na classe das funções isotônicas

f em X, a soma

$$\sum_{x \in X} [g(x) - f(x)]^2 w(x).$$

Vê-se, da definição anterior, que a regressão monotônica utiliza um método de mínimos quadrados ponderados, sujeito a uma restrição de ordem.

Ao contrário do que usualmente acontece em regressão, não há necessidade de se fazer suposições sobre a forma analítica de g . Geometricamente g^* será uma poligonal.

Prova-se que a regressão isotônica existe e é única. Além disso, Barlow et al(1972) desenvolveram uma série de testes de hipóteses (para verificar qualidade do ajuste, existência de dependência positiva, etc) em situações onde se conhece a distribuição de probabilidades de $\tilde{y}$.

Os mesmos autores desenvolveram uma série de algoritmos para estimação de g^* , um dos quais será introduzido no próximo exemplo.

Algoritmo

O algoritmo de estimação de uma regressão isotônica, será introduzido através de um exemplo. Suponha que se deseja a estimar a regressão isotônica de $\tilde{y}$ em $\tilde{x}$, para os dados da tabela A3.1.

Passo 1. Ordenar $\tilde{y}$ segundo a ordem estabelecida pelos valores de $\tilde{x}$.

Passo 2. Comparar sequencialmente os valores de $\tilde{y}$ e sempre que houver uma inversão de ordem (y_{i+1} inferior a y_i) substituir esses valores por sua média ponderada por w , ou seja, substituir por:

$$\bar{y}_{(i)} = \frac{\sum_{j=0}^{N(i)} w(x_{i+j}) y_{i+j}}{\sum_{j=0}^{N(i)} w(x_{i+j})} \quad \text{onde:}$$

$y_{i+N(i)+1}$ é a primeira observação superior a y_i , na sequência: $y_i, y_{i+1}, \dots, y_n$.

Passo 3. Verificar se essa substituição acarretou uma inversão de ordem entre y_{i-k} e y_i , $k = 1, 2, \dots$. Caso isso ocorra, substituir esses valores por sua média ponderada.

EXEMPLO: Aplicação do algoritmo aos dados da tabela A.3.1. Essas observações já se encontram ordenadas segundo a ordem estabelecida por $\tilde{x}$, passo 1.

- a. A primeira inversão ocorre já nas duas primeiras observações. Na coluna (1), esses valores são substituídos por sua média, conforme indicado no passo 2.
- b. Prosseguindo, uma nova inversão ocorre para $\tilde{x} = 5$ e $\tilde{x} = 6$. O procedimento descrito no passo 2 é adotado, coluna 2.
- c. Note que após a substituição realizada em (b), há uma inversão de ordem envolvendo y_4 , y_5 e y_6 . Como descrito no passo 3, esses valores devem ser substituídos por sua média, coluna 3.
- d. Feito isso, nova inversão ocorre, coluna (3), entre y_6 e y_7 . Como $y_4 = y_5 = y_6$, deve-se substituir esses valores por sua média.
- e. Na coluna (4), encontra-se a regressão monotônica de $\tilde{y}$ em $\tilde{x}$.

TABELA A3.1. Exemplo de aplicação do algoritmo de regressão isotônica

w	x	y	(1)	(2)	(3)	(4)
1	1	2,790	2,475	2,475	2,475	2,475
1	2	2,160	2,475	2,475	2,475	2,475
1	3	3,550	3,550	3,550	3,550	3,550
1	4	6,340	6,340	6,340	5,820	5,765
1	5	7,000	7,000	5,560	5,820	5,765
1	6	4,120	4,120	5,560	5,820	5,765
1	7	5,600			5,600	5,765
1	8	7,230				7,230

APÊNDICE 4

Método Steepest Descent para MONANOVA

Kruskal(1964b), desenvolveu uma versão do método "steepest descent", a fim de minimizar o "stress" e encontrar as coordenadas de um escalonamento multidimensional.

Neste tópico o método de Kruskal será adaptado no sentido de minimizar o "stress" do método de estimação MONANOVA.

O Método "steepest descent"

"Steepest descent" é um método iterativo de minimização de funções com várias variáveis. Sua idéia básica consiste em, a partir de um ponto arbitrário do domínio da função, verificar para qual direção o ponto deve ser deslocado a fim de ocasionar uma maior diminuição no valor da função. Determinada a direção de maior decaimento, estipula-se o tamanho do deslocamento e chega-se a um novo ponto. A partir deste, repete-se o processo até que alguma regra de parada seja obedecida.

Considere $f = f(\mathbf{x})$; $\mathbf{x} = (x_1 \ x_2 \ \dots \ x_m)'$, a função a ser minimizada.

A direção de maior decaimento é dada através do cálculo do gradiente negativo

(G).

$$\mathbf{G}' = \left(-\frac{\partial f(\mathbf{x})}{\partial x_1} \quad -\frac{\partial f(\mathbf{x})}{\partial x_2} \quad \dots \quad -\frac{\partial f(\mathbf{x})}{\partial x_m} \right) = (g_1 \quad g_2 \quad \dots \quad g_m)$$

A passagem do passo (k) para o passo (k+1) é dada por:

$$x_i^{(k+1)} = x_i^{(k)} + \alpha_k g_{ik} \quad , i = 1, \dots, m \quad \text{onde,}$$

$x_i^{(k)}$ = valor da coordenada i de $\mathbf{x}$ na iteração k;

g_{ik} = coordenada i do gradiente na iteração k e

α_k = constante positiva.

O valor de α_k está relacionado com a magnitude da passagem do ponto $\mathbf{x}^{(k)}$ para o ponto $\mathbf{x}^{(k+1)}$. Deste modo, espera-se que quanto maior a distância entre o ponto da k^{a} iteração e o mínimo procurado, maior seja α_k . Um modo de saber se o ponto está próximo, ou não, do mínimo é analisar seu decaimento, quanto maior o decaimento, maior deve ser esta distância.

Um problema comum com a aplicação deste método é a possibilidade de se encontrar um mínimo local. Para se prevenir desta ocorrência, Kruskal(1964b) aconselha que o processo iterativo seja repetido várias vezes, cada uma com um ponto inicial diferente: se todos convergirem para um mesmo valor, possivelmente este será o mínimo global.

Aplicação a MONANOVA

O método MONANOVA baseia-se na minimização de uma função denominada "stress". Um de seus objetivos consiste em encontrar:

$$Str(\hat{g}, \hat{\beta}(\hat{g})) = \min_{\{y_j = g(r_j)\}} \min_{\beta} Str(g, \beta)$$

sujeito à restrição:

$$r_j < r_{j'} \Rightarrow y_j \leq y_{j'},$$

onde

$$Str(g, \beta) = \sqrt{\frac{\sum_j [y_j - y_j(\beta)]^2}{\sum_j [y_j(\beta) - \bar{y}(\beta)]^2}}.$$

No procedimento sugerido por Kruskal(1965), a aplicação do "steepest descent" é indicada quando, a partir dos valores y_j , deseja-se minimizar o "stress". Nesta etapa da análise os y_j são considerados fixos, o que faz com que:

$$Str[g, \beta] = Str[\beta].$$

Seja

$$\mathbf{G}^k = (g_{1k} \quad g_{2k} \quad \dots \quad g_{pk})'$$

o gradiente negativo na iteração k .

Assim, tem-se:

$$\beta_i^{(k)} = \beta_i^{(k-1)} + \frac{\alpha_k}{mag(\mathbf{G}^{(k)})} g_{ik}$$

onde,

$$mag(\mathbf{G}^{(k)}) = \text{magnitudo relativa do gradiente} = \frac{\sum_i g_{ik}^2}{\sqrt{\sum_i \beta_i^{(k-1)2}}}.$$

A fim de facilitar o processo iterativo, Kruskal(1964b), propõe que o quadrado médio das estimativas de β seja igualado a um: $\frac{\sum_{i=1}^p \beta_i^2}{p} = 1$. Esta padronização não influenciará o "stress", uma vez que este índice não é afetado por transformações na escala de β , como é provado no teorema A4.1.

TEOREMA A4.1. Sejam a e b constantes. Então o "stress" calculado para β será igual ao calculado para $a\beta + b$.

prova: Considere $\beta^* = a\beta + c$; tem-se que:

$$y_j(\beta^*) = \sum_{i=1}^p \beta_i^* x_{ji} = a \sum_{i=1}^p \beta_i x_{ji} + c \sum_{i=1}^p x_{ji} =$$

$$= ay_j(\beta) + c \sum_{i=1}^p x_{ji}.$$

Sendo y_j a regressão isotônica de $y_j(\beta)$, com respeito à ordem original dos dados, tem-se que a regressão isotônica de $y_j(\beta^*)$ será $y_j^* = ay_j + c \sum_{i=1}^p x_{ji}$ (propriedade da regressão isotônica).

Utilizando os fatos acima, vem

$$\begin{aligned} Str(g, \beta^*) &= \sqrt{\frac{\sum_j [y_j^* - y_j(\beta^*)]^2}{\sum_j [y_j(\beta^*) - \bar{y}(\beta^*)]^2}} = \\ &= \sqrt{\frac{\sum_j [ay_j + c \sum_{i=1}^m x_{ji} - ay_j(\beta) - c \sum_{i=1}^m x_{ji}]^2}{\sum_j [ay_j(\beta) + c \sum_{i=1}^m x_{ji} - a\bar{y}(\beta) + c \sum_{i=1}^m x_{ji}]^2}} = \\ &= \sqrt{\frac{\sum_j [y_j - y_j(\beta)]^2}{\sum_j [y_j(\beta) - \bar{y}(\beta)]^2}} = Str[f, \beta]. \quad \blacksquare \end{aligned}$$

O método para obtenção do valor de α_k , a ser exposto, é o resultado de experimentação numérica. Kruskal adverte que não há pretensão de que este valor seja ótimo em algum sentido, contudo lhe pareceu rápido e robusto em comparação a uma série de problemas encontrados em procedimentos mais antigos. Produz maiores passos nos estágios iniciais e menores nos finais, quando se espera uma maior proximidade com o mínimo.

Sugere-se, em escalonamento multidimensional, que o valor inicial para α_k esteja em torno de 0,20.

O método de obtenção consiste em:

$$\alpha_{atual} = \alpha_{anterior} \cdot (\text{fator angular}) \cdot (\text{fator de relaxação}) \cdot (\text{fator de boa sorte})$$

onde

fator angular = $4^{\cos^3 \theta}$, onde, θ é o ângulo entre o gradiente atual e o anterior. Quando se estiver próximo ao ponto de mínimo, é de se esperar que sucessivos passos resultem em valores de "stress" posicionados antes e após o mínimo. Este fato fará com que $\cos \theta$ seja negativo e, além disto, esta medida deve se aproximar de -1, nas vizinhanças do ponto procurado.

$$\text{fator de relaxação} = \frac{1,3}{1 + (\text{razão de 5 passos})^2}$$

razão de 5 passos = $\min[1 ; \frac{\text{stress atual}}{\text{stress 5 iterações atrás}}]$. Nas proximidades do mínimo, espera-se que o "stress" 5 passos atrás tenha valor semelhante ao do "stress" atual, isto fará com que o fator de relaxação assuma seu menor valor, 0,65.

Antes da quinta iteração é impossível calcular o fator de relaxação pela fórmula recém introduzida; por isso, nesta situação aconselha-se utilizar o "stress" da primeira iteração no lugar do "stress" da quinta.

$$\text{fator de boa sorte} = \min[1 ; \frac{\text{stress atual}}{\text{stress anterior}}]$$

Em resumo, próximo ao ponto de mínimo tem-se:

$$\theta \cong 180^\circ \Rightarrow \text{fator angular} \cong 0.25$$

$$\text{fator de relaxação} \cong 0.65$$

$$\text{fator de boa sorte} \cong 1.$$

$$\alpha_{\text{atual}} \cong \alpha_{\text{anterior}} \frac{1}{4} \cdot 0,65 \cdot 1 \cong 0,1625 \alpha_{\text{anterior}}$$

Gradiente do "stress" da MONANOVA

Neste tópico serão calculadas as primeiras derivadas do "stress" em relação ao parâmetro β .

Como foi visto anteriormente o "stress" é dado por:

$$Str(g, \beta) = \sqrt{\frac{\sum_j [y_j - y_j(\beta)]^2}{\sum_j [y_j(\beta) - \bar{y}(\beta)]^2}} = \sqrt{\frac{u}{v}}$$

São fornecidas as fórmulas das derivadas de u e v e, de posse desses valores, é fácil encontrar a fórmula do gradiente.

Inicialmente, considere a função v .

$$\frac{\partial v}{\partial \beta_i} = 2 \sum_{j=1}^n (y_j(\beta) - \bar{y}(\beta)) \left(\frac{\partial y_j}{\partial \beta_i} - \frac{\partial \bar{y}(\beta)}{\partial \beta_i} \right)$$

mas,

$$\frac{\partial y_j(\beta)}{\partial \beta_i} = x_{ji} \Rightarrow \frac{\partial \bar{y}(\beta)}{\partial \beta_i} = \frac{\sum_{j=1}^n x_{ji}}{n}.$$

Logo,

$$\frac{\partial v}{\partial \beta_i} = 2 \sum_{j=1}^n [y_j(\beta) - \bar{y}(\beta)] x_{ji} - 2 \sum_{j=1}^n [y_j(\beta) - \bar{y}(\beta)] \left[\frac{\sum_{j=1}^n x_{ji}}{n} \right]$$

$$\text{mas, como } \sum_{j=1}^n [y_j(\beta) - \bar{y}(\beta)] = 0,$$

vem

$$\frac{\partial v}{\partial \beta_i} = 2 \sum_{j=1}^n [y_j(\beta) - \bar{y}(\beta)] x_{ji}.$$

Considere, agora, a função u .

$$\begin{aligned} \frac{\partial u}{\partial \beta_i} &= 2 \sum_{j=1}^n (y_j - y_j(\beta)) \left[\frac{\partial y_j}{\partial \beta_i} - \frac{\partial y_j(\beta)}{\partial \beta_i} \right] = \\ &= 2 \sum_{j=1}^n (y_j - y_j(\beta)) \frac{\partial y_j}{\partial \beta_i} - 2 \sum_{j=1}^n (y_j - y_j(\beta)) x_{ji}. \end{aligned}$$

Para calcular a primeira parcela do último membro da equação acima, é necessário recordar o significado de y_j . y_j é a regressão isotônica de $y_j(\beta)$, considerando a ordem dos dados originais. Considere então os seguintes conjuntos

$$T_l = \{j \in I = \{1, 2, \dots, n\} \mid t.q. \ y_j = y_l, j \in I, l \in I\}$$

ou seja, o conjunto I está sendo particionado em subconjuntos de modo que, em cada partição os valores y_j sejam iguais. O número total de subconjuntos deste tipo será representado por L e os conjuntos serão indexados por $l=1,2,\dots,L$. Desta forma pode-se reorganizar a soma $\sum_{j=1}^n (y_j - y_j(\beta)) \frac{\partial y_j}{\partial \beta_i}$, do seguinte modo:

$$\sum_{j=1}^n (y_j - y_j(\beta)) \frac{\partial y_j}{\partial \beta_i} = \sum_{l=1}^L \sum_{j \in T_l} (y_j - y_j(\beta)) \frac{\partial y_j}{\partial \beta_i}.$$

Considere que a cardinalidade de $T_l = n_l$.

Em cada grupo T_l , $y_j = \frac{\sum_{j \in T_l} y_j(\beta)}{n_l}$, ou seja, a média. Isto leva que:

$$\frac{\partial y_j}{\partial \beta_i} = \frac{1}{n_l} \sum_{j \in T_l} \frac{\partial y_j(\beta)}{\partial \beta_i} = \frac{1}{n_l} \sum_{j \in T_l} x_{ji}.$$

Logo,

$$\sum_{j \in T_l} (y_j - y_j(\beta)) \frac{\partial y_j}{\partial \beta_i} = \sum_{j \in T_l} (y_j - y_j(\beta)) \left(\frac{1}{n_l} \sum_{j \in T_l} x_{ji} \right)$$

mas,

$$\sum_{j \in T_l} (y_j - y_j(\beta)) = 0$$

o que conduz a:

$$\frac{\partial u}{\partial \beta_i} = -2 \sum_{j=1}^n (y_j - y_j(\beta)) x_{ji}.$$

Tendo estes valores e aplicando a regra da cadeia vem que:

$$\frac{\partial Str}{\partial \beta_i} = \frac{1}{2} \frac{\frac{\partial u}{\partial \beta_i} v - u \frac{\partial v}{\partial \beta_i}}{v^2} \sqrt{\frac{v}{u}}$$

APÊNDICE 5

Teoremas relativos ao método LINMAP

TEOREMA A5.1. O programa linear definido de (4.3.6) a (4.3.10) minimiza $B(\beta)$ sujeito às restrições lineares com valor ótimo da função objetivo $F^* = B^*$.

prova:

i. Será mostrado que $F^* = B(\beta^*)$.

Seja $\beta^* = (\beta_i^*)$ e z_{jk}^* para qualquer $(j, k) \in \Omega$, a solução ótima do programa. Então, por (4.3.6), tem-se que:

$$F^* = \sum_{(j,k) \in \Omega} z_{jk}^*.$$

De (4.3.7) e (4.3.9) tira-se que:

$$z_{jk} \geq \max\{0; -[\sum_{i \in P} a_{jki} \beta_i]\} = -\delta_{jk} [a_{jki} \beta_i].$$

Logo o mínimo será atingido quando

$$z_{jk}^* = -\delta_{jk} \left[\sum_{i \in P} a_{jki} \beta_i^* \right] \quad , \text{ ou seja,}$$

$$F^* = - \sum_{(j,k) \in \Omega} \delta_{jk} \sum_{i \in P} a_{jki} \beta_i^*.$$

De (4.3.5), deduz-se que:

$$F^* = B(\beta^*).$$

ii. *Será provado que (β^*) minimiza $B(\beta)$ sobre todas as soluções (β) que satisfazem as condições do teorema.*

Suponha a existência de uma solução (β') satisfazendo as restrições do programa linear e com

$$B(\beta') < B(\beta^*) = F^* \quad (A5.1).$$

Sendo $\beta' = (\beta'_i)$, tem-se de (4.3.7) e (4.3.9) que

$$z'_{jk} = -\delta_{jk} \sum_{i \in P} a_{jki} \beta'_i \quad (j, k) \in \Omega.$$

Desse modo, o correspondente valor da função objetivo para β' será:

$$F' = - \sum_{(j,k) \in \Omega} \delta_{jk} \sum_{i \in P} a_{jki} \beta'_i = B(\beta').$$

De A5.1 tira-se que :

$$F' = B(\beta') < B(\beta^*) = F^*.$$

O que é um absurdo, uma vez que β^ é uma solução ótima.*

Conclue-se então que β^* minimiza $B(\beta)$ sobre todas as soluções factíveis nas condições do teorema. ■

TEOREMA A5.2 O programa linear (4.3.6) a (4.3.10) tem uma solução ótima finita.

prova:

O teorema estará demonstrado se for apresentada pelo menos uma solução para o sistema (4.3.7) - (4.3.10).

Considere a seguinte solução:

$$- \beta_i = 0 \quad i = 1, 2, \dots, p-1$$

$$- \beta_p = \frac{h}{|A_p|} \quad h > 0. \text{ Deste modo, as restrições (4.3.8) e (4.3.10) estão satisfeitas.}$$

$$- z_{jk} = \max\{0; -\frac{ha_{jki}}{|A_i|}\} \text{ para } (j, k) \in \Omega.$$

Assim (4.3.7) e (4.3.9) também estão satisfeitas, portanto essa é uma solução factível, o que implica que o programa linear possui uma solução ótima finita. ■

TEOREMA A5.3 Considere o programa linear (4.3.11) a (4.3.15). Defina um novo problema (denotado pelo expoente 1) igual ao anterior, exceto pela unidade de medida do primeiro atributo, que será reduzida pelo fator q , isto é, $x'_{i1} = qx_{i1}$. Então existe uma solução ótima para o novo programa, idêntica à do anterior exceto pelo fato de que $\beta'_1 = \frac{\beta_1}{q^2}$ e $v'_1 = \frac{v_1}{q}$.

prova:

As mudanças introduzidas provocarão duas modificações no programa: $a'_{jk1} = q^2 a_{jk1}$ e $b'_{jk1} = qb_{jk1}$; os outros coeficientes permanecem inalterados.

Note que:

$$a_{jk1}\beta_1 = a'_{jk1}\left(\frac{\beta_1}{q^2}\right) \quad e \quad b_{jk1}v_1 = b'_{jk1}\left(\frac{v_1}{q}\right)$$

Logo, em (4.3.8), (4.3.9) e (4.3.11), pode-se substituir β_1 por $\beta'_1 = \frac{\beta_1}{q^2}$ e v_1 por

$v'_1 = \frac{v_1}{q}$. Assim, a solução desse programa será equivalente à do anterior, exceto por esses dois valores. ■

Observação: O teorema anterior também pode ser demonstrado para o modelo vetor.

APÊNDICE 6

LINMAP - Modelos Ponto-Ideal

Nesse apêndice são apresentadas as simplificações algébricas que levam aos programas lineares relativos aos modelos ponto-ideal, descritos no capítulo 4.

A6.1. Modelo ponto-ideal ponderado

O modelo ponto ideal ponderado é dado por:

$$\begin{aligned} y_j &= \sum_{i \in P} \beta_i (x_{ji} - \pi_i)^2 = \sum_{i \in P} d_{ji}^2 = \quad j \in J \\ &= \sum_{i \in P} \beta_i x_{ji}^2 - 2 \sum_{i \in P} \beta_i \pi_i x_{ji} + \sum_{i \in P} \beta_i \pi_i^2. \end{aligned} \tag{A6.1}$$

Utilizando a expressão acima vem que:

$$y_j - y_k = \sum_{i \in P} \beta_i (x_{ji} - \pi_i)^2 - \sum_{i \in P} \beta_i (x_{ki} - \pi_i)^2 =$$

$$= \sum_{i \in P} \beta_i (x_{ji}^2 - x_{ki}^2) - 2 \sum_{i \in P} \beta_i \pi_i (x_{ji} - x_{ki}).$$

Com base nestes resultados define-se as seguintes constantes (para $(j, k) \in \Omega$ e $i \in P$),

$$a'_{jki} = x_{ji}^2 - x_{ki}^2; \quad A_i = \sum_{(j,k) \in \Omega} a_{jki};$$

$$b_{jki} = -2(x_{ji} - x_{ki}); \quad D_p = \sum_{(j,k) \in \Omega} b_{jki}.$$

Defina também:

$$V = \{v_i\} = \{\beta_i \pi_i\}, \quad \text{de onde vem que :}$$

$$y_k - y_j = \sum_{i \in P} \beta_i a'_{jki} + \sum_{i \in P} v_i b_{jki}$$

Para garantir a linearidade das restrições será considerada como solução do problema $(\beta, \mathbf{V})$ ao invés de (β, Π) .

Utilizando a notação recentemente introduzida tem-se que a restrição $G - B = h$ pode ser escrita como:

$$\sum_{i \in P} \beta_i A_i + \sum_{i \in P} v_i D_i = h \quad (\text{restrição 4.3.14}) \quad (\text{A6.2})$$

além disto pode-se escrever B como:

$$B(\beta, \mathbf{V}) = \sum_{(j,k) \in \Omega} \delta_{jk} [- (\sum_{i \in P} \beta_i a'_{jki} + \sum_{i \in P} v_i b_{jki})] = \sum_{(j,k) \in \Omega} z_{jk}. \quad (\text{A6.3})$$

Note que

$$z_{jk} \geq y_j - y_k = -\left(\sum_{i \in P} \beta_i a'_{jki} + \sum_{i \in P} v_i b_{jki}\right) \quad \text{restrição 4.3.12.}$$

Assim sendo, segue o programa linear (4.3.11) a (4.3.15).

A6.2. Modelo ponto-ideal geral

Será apresentado o desenvolvimento de um programa linear para a versão II do modelo ponto-ideal geral (ver capítulo 3.2.3.). Esse programa foi apresentado em Snivasan & Shocker(1973a).

A equação (3.2.4) diz que o modelo ponto ideal geral pode ser escrito como:

$$y_j = \sum_{i \in P} \lambda_{ii} x_{ji}^2 + \sum_{i \in P} \alpha_i x_{ji} + \sum_{i \in P} \sum_{l < i} 2\lambda_{il} x_{ji} x_{jl} + c$$

Assim,

$$y_j - y_k = \sum_{i \in P} \lambda_{ii} (x_{ji}^2 - x_{ki}^2) + \sum_{i \in P} \alpha_i (x_{ji} - x_{ki}) + \sum_{i \in P} \sum_{l < i} 2\lambda_{il} (x_{ji} x_{jl} - x_{ki} x_{kl}).$$

Considere as constantes a_{jki} , a'_{jki} e

$$c_{jkil} = 2(x_{ji} x_{jl} - x_{ki} x_{kl}) \quad , \quad E_{il} = \sum_{(j,k) \in \Omega} c_{jkil}.$$

Desse modo,

$$y_j - y_k = \sum_{i \in P} a'_{jki} \lambda_{ii} + \sum_{i \in P} a_{jki} \alpha_i + \sum_{i \in P} \sum_{l < i} c_{jkil} \lambda_{il}.$$

Utilizando um raciocínio análogo a (A6.2) e (A6.3), chega-se às restrições:

$$\sum_{i \in P} A'_i \lambda_{ii} + \sum_{i \in P} A_i \alpha_i + \sum_{i \in P} \sum_{l < i} E_{il} \lambda_{il} = 1$$

$$z_{jk} \geq -\left(\sum_{i \in P} a'_{jki} \lambda_{ii} + \sum_{i \in P} a_{jki} \alpha_i + \sum_{i \in P} \sum_{l < i} c_{jkil} \lambda_{il}\right).$$

O que resulta no programa linear (4.3.17) a (4.3.20).

A partir dos valores estimados, chega-se a Λ' e α e. Foi visto, em (4.1.8), que

$$\hat{\Pi} = -\frac{1}{2}\alpha\Lambda^{-1}$$

APÊNDICE 7

Estimador James-Stein

Neste apêndice serão apresentados aspectos técnicos ligados à regressão bayesiana, estimador James-Stein e à relação existente entre esses dois assuntos.

Estimador bayesiano para um modelo de regressão

O objetivo deste tópico é deduzir o estimador bayesiano de um vetor de parâmetros de um modelo de regressão linear. Uma análise mais detalhada deste estimador pode ser encontrada em Vinod & Ullah(1981).

Considere o modelo abaixo:

$$y = X\beta + \epsilon \quad \epsilon \sim N_n(0, \sigma^2 I).$$

A função de verossimilhança de y , considerando σ conhecido será dada por:

$$f(y/\beta) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]$$

Resultado A7.1 $(y - X\beta)'(y - X\beta) = (y - Xb)'(y - Xb) + (\beta - b)'X'X(\beta - b)$, onde b é o estimador de mínimos quadrados de β .

prova:

$$\begin{aligned} (y - X\beta)'(y - X\beta) &= (y - Xb + Xb - X\beta)'(y - Xb + Xb - X\beta) = \\ &= (y - Xb)'(y - Xb) + 2(y - Xb)'X(b - \beta) + (\beta - b)'X'X(\beta - b) \end{aligned}$$

mas,

$$\begin{aligned} (y - Xb)'X(b - \beta) &= (y - X(X'X)^{-1}X'y)'X(b - \beta) = \\ &= y'(I - X(X'X)^{-1}X')X(b - \beta) = y'(X - X)(b - \beta) = 0 \quad \blacksquare \end{aligned}$$

Utilizando este fato, pode-se escrever a verossimilhança como:

$$f(y/\beta) \propto \exp\left\{-\frac{1}{2\sigma^2}(\beta - b)'X'X(\beta - b)\right\} \quad (A7.1)$$

Teorema A7.1. Seja $y/\beta \sim N_n(X\beta; \sigma^2 I)$, σ^2 conhecido; seja também $\beta \sim N_p(\beta^*; \sigma^2 \Omega)$, Ω uma matriz não singular, a distribuição a priori de β ; então o estimador de Bayes para β , assumindo perda quadrática será dado por:

$$b^* = (X'X + \Omega^{-1})^{-1}(X'Xb + \Omega^{-1}\beta^*).$$

Além disso:

$$\beta/y \sim N(b^*, \sigma^2 \Omega^*).$$

prova: Inicialmente será encontrada a distribuição a posteriori de β : como se está assumindo perda quadrática, a média da posteriori será o estimador de Bayes para o parâmetro.

Sabe-se que a distribuição a priori de β é dada por:

$$p(\beta) \propto \exp\left\{-\frac{1}{2\sigma^2}(\beta - \beta^*)'\Omega^{-1}(\beta - \beta^*)\right\}.$$

Definidas a verossimilhança (A7.1), a priori e utilizando o resultado anteriormente provado, a distribuição a posteriori será dada por:

$$p(\beta/y) = p(\beta)f(y/\beta) \propto$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}[(\beta - b)'X'X(\beta - b) + (\beta - \beta^*)'\Omega^{-1}(\beta - \beta^*)]\right\}$$

mas,

$$(\beta - b)'X'X(\beta - b) + (\beta - \beta^*)'\Omega^{-1}(\beta - \beta^*) =$$

$$\beta'X'X\beta - 2\beta'X'Xb + b'X'Xb + \beta'\Omega^{-1}\beta - 2\beta'\Omega^{-1}\beta^* + \beta^{*'}\Omega^{-1}\beta^* = (A)$$

$$\text{Seja } c = b'X'Xb + \beta^{*'}\Omega^{-1}\beta^*.$$

$$(A) = \beta'(X'X + \Omega^{-1})\beta - 2\beta'(X'Xb + \Omega^{-1}\beta^*) + c =$$

$$= \beta'(X'X + \Omega^{-1})\beta - 2\beta'(X'X + \Omega^{-1})(X'X + \Omega^{-1})^{-1}(X'Xb - \Omega^{-1}\beta^*) + c = (B)$$

seja $\Omega^ = (X'X + \Omega^{-1})^{-1}$, então*

$$(B) = \beta' \Omega^{*-1} \beta - 2\beta' \Omega^{*-1} [(X'X + \Omega^{-1})^{-1} (X'Xb + \Omega^{-1} \beta^*)] + c = (C)$$

seja $b^* = (X'X + \Omega^{-1})^{-1} (X'Xb + \Omega^{-1} \beta^*)$, então

$$(C) = (\beta - b^*)' \Omega^{*-1} (\beta - b^*) - b^{*'} \Omega^{-1} b^* + c.$$

Portanto,

$$p(\beta/y) \propto \exp\left\{-\frac{1}{2\sigma^2} (\beta - b^*)' \Omega^{*-1} (\beta - b^*)\right\},$$

O que leva a

$$\beta/y \sim N_p(b^*; \sigma^2 \Omega^*) \quad \blacksquare$$

No caso de regressão linear a função de perda quadrática será dada por:

$$L(\hat{\beta}) = (X\beta - X\hat{\beta})'(X\beta - X\hat{\beta}) = (\beta - \hat{\beta})' X'X(\beta - \hat{\beta})$$

Estimadores James-Stein

Um resultado bastante difundido na teoria estatística é que o estimador não viado de variância mínima da média de um vetor aleatório ($y \sim N_n(\mu, \sigma^2 I)$) é a média amostral. Entretanto, existe na literatura alguns exemplos de estimadores de μ que, apesar de viciados, possuem um erro quadrático médio menor do que a média amostral. É nesta classe que se encontram os estimadores James-Stein (Draper & Nostrand(1979)).

James & Stein(1961) provaram que se $Z \sim N_p(\xi, I)$ e $p \geq 3$, então, em termos da função de perda quadrática, o estimador de ξ

$$\hat{\xi} = \left(1 - \frac{k}{Z'Z}\right)Z, \quad k \text{ uma constante,}$$

será melhor do que o usual ($\mathbf{Z}$), desde que $k \in (0, 2(p-2))$; mais do que isto, foi provado que a perda será mínima quando $k = p-2$.

$\hat{\xi}$ é o estimador James-Stein de ξ .

Uma extensão natural deste estimador para o caso $\mathbf{Z} \sim N_p(\xi, \sigma^2 \mathbf{I})$, resultou

$$\hat{\xi} = \left(1 - \frac{(p-2)\sigma^2}{\mathbf{Z}'\mathbf{Z}}\right)\mathbf{Z}.$$

Repare que $\mathbf{Z}$ pode ser um vetor de médias amostrais.

Basicamente o estimador James-Stein desloca o estimador usual de mínimos quadrados em direção ao valor zero. Pensando nisto, Lindley, discutindo o artigo Stein(1962), sugeriu a seguinte modificação no estimador:

$$\hat{\xi} = \bar{\mathbf{Z}} + \left(1 - \frac{(p-2)\sigma^2}{(\mathbf{Z} - \bar{\mathbf{Z}})'(\mathbf{Z} - \bar{\mathbf{Z}})}\right)(\mathbf{Z} - \bar{\mathbf{Z}})$$

onde $\bar{\mathbf{Z}} = \frac{1'}{p}\mathbf{Z}$

Este último estimador tem a função de deslocar $\mathbf{Z}$ em direção à média geral dos dados.

Note que:

$$\hat{\xi} = \left(1 - \frac{(p-2)\sigma^2}{(\mathbf{Z} - \bar{\mathbf{Z}})'(\mathbf{Z} - \bar{\mathbf{Z}})}\right)\mathbf{Z} + \frac{(p-2)\sigma^2}{(\mathbf{Z} - \bar{\mathbf{Z}})'(\mathbf{Z} - \bar{\mathbf{Z}})}\bar{\mathbf{Z}}$$

Aplicações à regressão

Considere o modelo de regressão apresentado no início deste apêndice. Seja $\mathbf{b}$ o estimador usual de mínimos quadrados. Sabe-se que:

$$\mathbf{b} \sim N_p(\beta; \sigma^2(\mathbf{X}'\mathbf{X})^{-1}).$$

Uma versão do estimador James-Stein obtida diretamente para este caso é:

$$\hat{\beta}_s = \left(1 - \frac{(p-2)\sigma^2}{\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}}\right)\mathbf{b}$$

Suponha, agora, que exista alguma informação disponível sobre o valor de β . Chame este valor de β^* . Levando em conta esta informação, seria razoável que o estimador usual de mínimos quadrados, $\mathbf{b}$, fosse deslocado nesta direção. Uma maneira de se conseguir este deslocamento é utilizar a seguinte adaptação do estimador (Lindley) James-Stein: teria-se:

$$\hat{\beta}_s = \left(1 - \frac{(p-2)\sigma^2}{(\mathbf{b} - \beta^*)'\mathbf{X}'\mathbf{X}(\mathbf{b} - \beta^*)}\right)\mathbf{b} + \frac{(p-2)\sigma^2}{(\mathbf{b} - \beta^*)'\mathbf{X}'\mathbf{X}(\mathbf{b} - \beta^*)}\beta^*.$$

O estimador acima, construído de maneira intuitiva, tem respaldo na teoria estatística quando se adota uma abordagem bayesiana.

Abordagem bayesiana para o estimador (Lindley) James-Stein

Resultado A7.2. Considere $\mathbf{y} \sim N_n(\mathbf{X}\beta; \sigma^2\mathbf{I})$, sendo σ^2 conhecido. Seja também $\beta \sim N_p(\beta^*; \sigma_\beta^2(\mathbf{X}'\mathbf{X})^{-1})$, com σ_β^2 um real positivo, a distribuição a priori de β . Então, sob estas condições, o estimador bayesiano de β , $\mathbf{b}^*$, é dado por:

$$\mathbf{b}^* = \left(1 - \frac{\sigma^2}{\sigma_\beta^2 + \sigma^2}\right)\mathbf{b} + \frac{\sigma^2}{\sigma_\beta^2 + \sigma^2}\beta^*.$$

prova: Utilizando o teorema A7.1 e que $\sigma_\beta = \sigma \frac{\sigma_\beta}{\sigma}$, vem

$$\begin{aligned} \mathbf{b}^* &= (\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\sigma_\beta^2}\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}\mathbf{b} + \frac{\sigma^2}{\sigma_\beta^2}\mathbf{X}'\mathbf{X}\beta^*) = \\ &= \left(1 + \frac{\sigma^2}{\sigma_\beta^2}\right)^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{b} + \frac{\sigma^2}{\sigma_\beta^2}\beta^*) = \\ &= \left(\frac{\sigma_\beta^2 + \sigma^2}{\sigma_\beta^2}\right)^{-1}\left(\frac{\sigma_\beta^2\mathbf{b} + \sigma^2\beta^*}{\sigma_\beta^2}\right) = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma^2}(\mathbf{b} + \frac{\sigma^2}{\sigma_\beta^2}\beta^*) = \end{aligned}$$

$$= (1 - \frac{\sigma^2}{\sigma_{\beta}^2 + \sigma^2})\mathbf{b} + \frac{\sigma^2}{\sigma_{\beta}^2 + \sigma^2}\beta^* \quad \blacksquare$$

O passo seguinte consiste em encontrar estimativas para σ^2 e $\sigma_{\beta}^2 + \sigma^2$.

O estimador geralmente utilizado para σ^2 é

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})}{n - p - 1}$$

O estimador para $\sigma_{\beta}^2 + \sigma^2$ será obtido a partir do resultado abaixo.

Resultado A7.3. $E[(\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*)] = p(\sigma^2 + \sigma_{\beta}^2)$

prova

$$(\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*) = \mathbf{b}' \mathbf{X}' \mathbf{X} \mathbf{b} - 2\mathbf{b}' \mathbf{X}' \mathbf{X} \beta^* + \beta^{*'} \mathbf{X}' \mathbf{X} \beta^*$$

Como,

$$\mathbf{b} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{X} \beta + \epsilon) = \beta + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \epsilon,$$

vem

$$a.) \quad \mathbf{b}' \mathbf{X}' \mathbf{X} \mathbf{b} = (\beta + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \epsilon)' \mathbf{X}' \mathbf{X} (\beta + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \epsilon) =$$

$$= \beta' \mathbf{X}' \mathbf{X} \beta + 2\beta' \mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \epsilon + \epsilon' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \epsilon =$$

$$= \beta' \mathbf{X}' \mathbf{X} \beta + 2\beta' \mathbf{X}' \epsilon + \epsilon' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \epsilon$$

$$b.) \quad 2\mathbf{b}' \mathbf{X}' \mathbf{X} \beta^* = 2\epsilon' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \beta^* = 2\beta^{*'} \mathbf{X}' \epsilon + 2\beta' \mathbf{X}' \mathbf{X} \beta^*.$$

Utilizando os resultados (a) e (b), vem que

$$\begin{aligned} (\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*) &= \beta' \mathbf{X}' \mathbf{X} \beta + 2(\beta - \beta^*)' \mathbf{X}' \epsilon + \epsilon' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \epsilon + \beta^{*'} \mathbf{X}' \mathbf{X} \beta^* = \\ &= (\beta - \beta^*)' \mathbf{X}' \mathbf{X} (\beta - \beta^*) + \epsilon' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \epsilon + 2\beta' \mathbf{X}' \epsilon - 2\beta^{*'} \mathbf{X}' \epsilon - 2\beta' \mathbf{X}' \mathbf{X} \beta^* \end{aligned}$$

Assim, utilizando resultados sobre a esperança de formas lineares e formas quadráticas (Searle(1971)), vem

$$1.) E[(\beta - \beta^*)' \mathbf{X}' \mathbf{X} (\beta - \beta^*)] = \text{tr}[\mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \sigma^2] = p \sigma^2$$

$$2.) E[\epsilon' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \epsilon] = \text{tr}[\mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \sigma^2] = p \sigma^2$$

$$3.) E[2\beta^{*'} \mathbf{X}' \epsilon] = 2\beta^{*'} \mathbf{X}' E[\epsilon] = 0$$

$$4.) E[2\beta' \mathbf{X}' \epsilon] = 2\beta' \mathbf{X}' E[\epsilon] = 0$$

O que leva a

$$E[(\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*)] = p(\sigma^2 + \sigma_{\beta}^2) \quad \blacksquare$$

Utilizando este resultado, um possível estimador não viciado para $\sigma^2 + \sigma_{\beta}^2$ é

$$\sigma^2 + \hat{\sigma}_{\beta}^2 = \frac{(\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*)}{p}$$

Assim, um possível estimador para β é

$$\mathbf{b}^* = \left(1 - \frac{\hat{\sigma}^2 p}{(\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*)}\right) \mathbf{b} + \frac{\hat{\sigma}^2 p}{(\mathbf{b} - \beta^*)' \mathbf{X}' \mathbf{X} (\mathbf{b} - \beta^*)} \beta^*$$

A única diferença entre $\mathbf{b}^*$ e $\hat{\beta}_s$ é a substituição de p por $(p-2)$. Esta troca se deve ao fato de que a perda quadrática será menor quando se adota $(p-2)$.

APÊNDICE 8

Resumo dos Estudos Empíricos Descritos na Seção 7.1.

1. Veen & Beazley (1977)

Nesse artigo, comparou-se os métodos de coleta "trade-off" e perfil completo, fazendo com que estímulos construídos a partir de 3 atributos, cada qual com 3 níveis, fossem avaliados por 600 pessoas.

Foram comparadas as utilidades médias de cada nível dos atributos, calculou-se para cada indivíduo a correlação entre os valores previstos dos modelos estimados e, por fim, utilizou-se o coeficiente de correlação de Spearman entre os postos previstos e observados de cada modelo.

2. Carmone, Green & Jain (1978)

Esse estudo tinha por objetivo avaliar os resultados de uma AP quando os dados são gerados:

- a. com e sem interação;
- b. com e sem uma parcela de erro aleatório;

- c. adotando-se planejamentos fracionários (no caso com 243, 54, 27 ou 18 observações);
- d. utilizando-se diferentes tipos de variáveis respostas (escala contínua, postos ou escala semântica de 6 pontos) e
- e. estimando-se modelos aditivos através de mínimos quadrados ou MONANOVA.

Os dados foram gerados fixando-se previamente os parâmetros de um modelo utilidade em um planejamento perfil completo com 5 atributos, cada qual com 3 níveis. Através da variação dos itens a, b, c e d acima, gerou-se 48 diferentes conjuntos de dados, o que resultou em 96 diferentes análises (item e).

O erro aleatório foi gerado através de uma distribuição normal. Para o cálculo de sua variância, estimou-se as utilidades dos 3^5 possíveis estímulos, calculando-se o seu desvio-padrão (σ); o desvio-padrão do erro foi fixado como sendo $0,2\sigma$.

Para a avaliação dos resultados, calculou-se, para cada modelo estimado, o índice τ de Kendall (Conover(1980)) entre os parâmetros reais e estimados; a seguir, considerando-se os itens a, b, c e d como fatores, construiu-se uma análise de variância para comparar os τ médios.

3. Cattin & Bliemel (1978)

O ajuste dos métodos MONANOVA e mínimos quadrados foi comparado ao se adotar como variável resposta: o posto e três tipos de escala de 7 pontos .

Foram gerados dados para um planejamento perfil completo considerando-se os seguintes casos:

- a. número de atributos: 4 (16 estímulos) ou 9 (16 estímulos), com 2 níveis cada;
- b. presença, ou não, de erro aleatório no modelo. Adotou-se o modelo: $y_j = \sum_i \beta_i x_{ji} + u_j$, $j = 1, \dots, 16$. Determinou-se que, para cada atributo, um dos parâmetros fosse zero e o outro fosse gerado através de uma normal reduzida.

No caso de presença de erro aleatório, este foi gerado a partir de uma uniforme de média zero e desvio-padrão 0,75.

c. 4 diferentes variáveis resposta. Inicialmente efetuou-se a seguinte transformação:

$$y_{aj} = 7 \frac{y_j - y_{min}}{y_{max} - y_{min}}$$

a seguir foram definidas as seguintes variáveis respostas abaixo:

- i. postos;
- ii. juiz justo: $r_j = [y_{aj}]$;
- iii. juiz exagerado: $r_j = [-\frac{2}{49}y_{aj}^3 + \frac{3}{7}y_{aj}^2]$ e
- iv. juiz indeciso: $r_j = [-\frac{4}{49}y_{aj}^3 + \frac{6}{7}y_{aj}^2 + 3y_{aj}]$ onde $[.]$ significa arredondamento para o inteiro superior.

Para cada uma das 16 combinações das opções dos itens acima gerou-se 50 observações.

A comparação deu-se através das somas dos erros absolutos.

4. Mc Cullough & Best (1979)

Desejava-se verificar empiricamente a estabilidade temporal e a confiabilidade estrutural de uma AP.

Os dados foram obtidos de uma amostra de 100 indivíduos, 55 avaliando estímulos provenientes de um planejamento perfil completo com 3 atributos, cada qual com 3 níveis e 45 de um planejamento perfil completo com 3 atributos, sendo dois com 3 níveis e um com 2 níveis. Em uma primeira fase, ordenou-se os estímulos, e em seguida, um dos atributos foi trocado por um quarto e realizou-se uma nova avaliação. A segunda fase, idêntica à primeira, realizou-se dois dias depois.

Na análise dos resultados, calculou-se a correlação entre as avaliações das duas fases, a diferença entre os parâmetros estimados nas duas etapas e a diferença entre os parâmetros comuns nas duas partes da primeira fase.

5. Jain, Acito, Malhotra & Mahajan (1979)

O objetivo desse artigo era comparar os métodos de estimação MONANOVA, JOHNSON, LINMAP, LOGIT e mínimos quadrados ao se adotar planejamentos perfil completo e "trade-off".

Estímulos, construídos a partir de 5 atributos, cada qual com 3 níveis, foram ordenados por 212 indivíduos. Desses, 106 foram submetidos a um planejamento "trade-off" e o restante a perfil completo fracionário (27 estímulos). Ao final do experimento, os entrevistados ordenaram uma amostra de 8 estímulos.

Comparou-se a importância relativa de cada atributo através de uma análise de variância e do teste de Friedman (indivíduo como bloco). A capacidade preditiva dos modelos estimados foi estabelecida através do cálculo do coeficiente de correlação de Spearman e do coeficiente de concordância de Kendall entre os postos previstos e observados.

6. Green, Goldberg & Montemayor (1981)

O trabalho apresenta o modelo híbrido e relata um estudo empírico comparando esse modelo com o de mínimos quadrados.

A partir de uma amostra de 424 pessoas, formou-se dois grupos, um com 324 entrevistados, utilizado na estimação dos modelos e outro com 100, utilizado no teste do modelo. Foram avaliados 7 atributos com 3 níveis cada e, na fase conjunta, 9 perfis extraídos de um planejamento principal de 81.

A comparação baseou-se em 4 medidas: o coeficiente de determinação dos modelos estimados, a porcentagem de acerto do estímulo preferido na amostra extra, o coeficiente τ de Kendall e a correlação entre os valores observados e previstos.

7. Wittink & Cattin(1981)

Os autores compararam os métodos de estimação LINMAP, LOGIT, MONANOVA e mínimos quadrados utilizados na estimação de modelos compensatórios e atributo dominante.

Adotado um planejamento perfil completo fracionário com 6 atributos, cada qual com 3 níveis, foram geradas observações para os dois tipos de modelo, com erros de 5, 20

e 35%, representando diferentes graus de envolvimento dos entrevistados com a pesquisa e distribuição Weibull, normal e uniforme. Utilizou-se 50 réplicas para cada combinação de modelo erro e distribuição. As avaliações geradas foram convertidas em postos.

As análises basearam-se no "stress" e nas correlações de Spearman e Pearson entre os postos observados e previstos, calculados para cada modelo estimado.

8. Carmone & Green (1981)

O efeito da má especificação do modelo, no que se refere a desprezar termos de interação e as vantagens ao se adotar um planejamento fracionário que permita a estimação de algumas interações são estudados nesse artigo.

Foram simuladas 5400 réplicas divididas em diferentes tipos de modelos (aditivo, com interação cruzada e com interação não cruzada); existência de um erro aleatório (inexistência, erro médio e erro alto) e diferentes planejamentos (o primeiro com 18 observações possibilitando apenas a estimação dos efeitos principais e o segundo com 27 casos, possibilitando a estimação de alguns termos de interação). Foi gerada também uma amostra extra de 6 perfis. Considerou-se a existência de 6 atributos, cada qual com 3 níveis.

Calculou-se a correlação de Spearman entre os valores previstos e observados e, para a amostra extra, a porcentagem de acerto do estímulo preferido.

9. Segal (1982)

Comparou-se a estabilidade temporal de dados coletados via planejamento perfil completo e "trade-off".

Os dados foram obtidos a partir de uma amostra de 50 pessoas. Os estímulos eram formados por 3 atributos, cada qual com 3 níveis. Os entrevistados ordenaram, em uma primeira fase, estímulos criados através de planejamentos perfil completo e "trade-off". Cerca de dez dias depois eles foram submetidos à mesma tarefa.

A avaliação dos resultados deu-se através da comparação do coeficiente de correlação de Spearman (calculado entre os postos das duas fases) médio e da importância média dos atributos entre as duas fases, para cada planejamento.

10. Malhotra (1982)

Verificou-se a confiabilidade estrutural de uma AP.

Estímulos construídos a partir de 5 atributos foram ordenados por 72 pessoas e, a seguir, mais 5 fatores foram acrescentados e uma nova ordenação foi realizada.

Empregou-se o método de estimação LINMAP e as conclusões basearam-se no índice de ajuste C.

11. Cattin, Gelfand & Danes (1983)

Os modelos bayesiano e de mínimos quadrados foram comparados.

Estudantes avaliaram 16 máquinas de escrever, descritas através da combinação de 6 atributos, cada qual com 2 níveis. As comparações basearam-se na soma de quadrados residual e no coeficiente de determinação dos modelos.

12. Akaah, Pradeep & Korgaonkar (1983)

Comparou-se os métodos de estimação de mínimos quadrados e híbrido com e sem interação, em termos de sua capacidade preditiva. Ressalta-se que todos os parâmetros do modelo híbrido foram estimados individualmente. O artigo compara também uma classe de modelos não estudada nessa dissertação, Huber-híbrido (Huber, Sahney & Ford(1969)).

Um grupo de 80 elementos avaliou 6 atributos, cada qual com 3 níveis. Na etapa auto-explicativa os níveis foram avaliados segundo uma escala de 9 pontos, sendo atribuído o valor 9 ao nível preferido; os atributos foram avaliados através de uma escala de 5 pontos, atribuindo-se o valor 5 ao preferido e 1 ao menos importante. Na etapa conjunta, estímulos provenientes de um planejamento perfil completo fracionário (27 estímulos) foram avaliados segundo uma escala de 9 pontos. Posteriormente, uma amostra extra de 6 estímulos foi apreciada através de uma escala de 9 pontos.

Utilizando uma análise de variância multivariada, comparou-se os pesos relativos dos diferentes modelos. Calculou-se a correlação de Spearman e os coeficientes de correlação e de concordância de Kendall entre os postos observados e previstos. Previu-se os postos da amostra extra e calculou-se a porcentagem de acertos.

13. Green (1984)

O autor relata 3 estudos que compararam o modelo híbrido com o de mínimos quadrados: Green, Goldberg & Wiley(1982); Akaah & Korgaonkar (1983) e Cattin, Hermet & Proche (1982).

APÊNDICE 9

Método de Johnson

O método de Johnson(1975), tal como o MONANOVA, se aplica à estimação de modelos utilidade com avaliação dos estímulos realizada via ordenação.

Da mesma forma que no método MONANOVA, o processo de estimação consiste na minimização de uma medida de falta de ajuste. Utilizando a mesma notação da seção 4.2, essa medida é definida por:

$$\theta = \sqrt{\frac{\sum_{i=1}^p \sum_{j<i} \delta_{ij} (y_i(\beta) - y_j(\beta))^2}{\sum_{i=1}^p \sum_{j<i} (y_i(\beta) - y_j(\beta))^2}} \quad \text{onde:}$$

$$\delta_{ij} = \begin{cases} 1 & \text{se o sinal de } (y_i - y_j) \neq \text{sinal de } (r_i - r_j) \\ 0 & \text{caso contrário} \end{cases}$$

O numerador consiste na soma dos quadrados das diferenças entre os pares (y_i, y_j) que não seguem a ordenação estabelecida pelos postos. De certo modo, o fator δ_{ij} faz o papel da restrição de ordem descrita no procedimento MONANOVA.

O denominador do índice é a soma dos quadrados de todas as diferenças. Isso faz

com que θ^2 possa ser interpretado como a proporção da variação entre os valores previstos que são inconsistentes com os postos de entrada. Se todos os valores previstos seguirem fielmente a ordem original, então $\theta^2 = 0$, caso estejam na ordem inversa, $\theta^2 = 1$. Portanto, quanto menor for o índice θ^2 , melhor será o ajuste do modelo.

Quando empates são permitidos na avaliação do entrevistado ($r_i = r_j$), duas abordagens podem ser adotadas: ou se tenta forçar que $y_i = y_j$, ou não. Isso pode ser feito através de uma redefinição de δ_{ij} , desse modo:

- a. $\delta_{ij} = 1$ se $r_i = r_j$, quando se deseja forçar o empate e
- b. $\delta_{ij} = 0$ se $r_i = r_j$, quando não se deseja forçar o empate.

A minimização de θ^2 pode ser feita através do método "steepest decent", de modo análogo à descrita no apêndice 4.

APÊNDICE 10

Estudo Empírico

O objetivo deste apêndice é apresentar alguns detalhes relativos ao estudo empírico descrito na seção 7.2.

Na tabela A10.1, encontra-se o planejamento fracionário utilizado no estudo; em seguida, na tabela A10.2, são dispostos os estímulos da amostra extra.

Os estímulos foram apresentados aos entrevistados em cartões semelhantes à figura A10.1. Esses cartões foram numerados no verso e embaralhados antes de serem oferecidos aos juízes.

Na sequência, encontra-se o questionário empregado no estudo.

TABELA A10.1. Planejamento perfil completo fracionário empregado no estudo empírico

Estímulo	Marca	Consumo	Combustível	Potência	Volumes
1	A	Baixo	Álcool	1.6	2
2	A	Moderado	Álcool	1.8	3
3	A	Moderado	Gasolina	1.6	3
4	A	Alto	Gasolina	1.8	2
5	B	Baixo	Gasolina	1.8	3
6	B	Moderado	Álcool	1.8	2
7	B	Moderado	Gasolina	1.6	2
8	B	Alto	Álcool	1.6	3
9	C	Baixo	Álcool	1.6	2
10	C	Moderado	Álcool	1.8	3
11	C	Moderado	Gasolina	1.6	3
12	C	Alto	Gasolina	1.8	2
13	D	Baixo	Gasolina	1.8	3
14	D	Moderado	Álcool	1.8	2
15	D	Moderado	Gasolina	1.6	2
16	D	Alto	Álcool	1.6	3

TABELA A10.2. Estímulos da Amostra Extra

Estímulo	Marca	Consumo	Combustível	Potência	Volumes
1	A	Alto	Álcool	1.6	2
2	A	Baixo	Álcool	1.8	3
3	B	Baixo	Gasolina	1.8	2
4	B	Moderado	Gasolina	1.6	3
5	C	Baixo	Álcool	1.6	3
6	C	Alto	Gasolina	1.8	3
7	D	Moderado	Álcool	1.8	2
8	D	Baixo	Gasolina	1.6	2

Questionário

Esta é uma pesquisa acadêmica cujos resultados farão parte de uma dissertação de mestrado.

O objetivo deste levantamento é conhecer a importância de determinadas características no momento da compra de um carro. São consideradas cinco características para o seu julgamento. São elas: **design** (2 volumes ou 3 volumes), **potência do motor** (1.6 ou 1.8), **tipo de combustível** (álcool ou gasolina) e **consumo de combustível** (baixo, moderado ou alto).

O consumo é dado em termos do gasto em cruzeiros para se percorrer 100 quilômetros em uma estrada. Assim, define-se como consumo **baixo** um gasto em torno de Cr\$ 850,00 (12 km/l de álcool ou 16 km/l de gasolina), consumo **moderado** um gasto em torno de Cr\$ 1000,00 (10,5 km/l de álcool ou 14 km/l de gasolina) e consumo **alto** um gasto em torno de Cr\$1150,00 (9 km/l de álcool ou 12 km/l de gasolina).

Responda as questões seguintes supondo que você fosse comprar um carro hoje e sempre levando em conta suas atuais necessidades.

--- Distribua 100 pontos pelas características abaixo, de modo que quanto maior for a importância de uma delas na definição de compra do carro, maior seja sua pontuação. Empates são permitidos.

DESIGN (2 ou 3 volumes).....	_____
POTÊNCIA (motor 1.6 ou 1.8).....	_____
COMBUSTÍVEL (álcool ou gasolina)...	_____
MARCA (VW, Ford, GM ou Fiat).....	_____
CONSUMO (baixo, moderado ou alto)..	_____
SOMA.....	1 0 0

--- Em cada item abaixo, escolha a opção de sua preferência atribuindo-lhe a nota 10; feito isso, atribua notas de 0 a 10 para as demais de modo que os valores atribuídos reflitam a importância que você dá a cada uma delas quando se define pela compra de um carro. Empates são permitidos.

a. Design..... 2 VOLUMES _____
3 VOLUMES

b. Potência..... motor 1.6 _____
motor 1.8

c. Combustível... álcool _____
gasolina _____

d. Marca..... VW _____
 Ford _____
 GM _____
 Fiat _____

e. Consumo..... Baixo _____
Moderado _____
Alto _____

--- Cada um desses cartões contém a descrição de um carro. Pensando no tipo de carro que você adquiriria hoje, ordene-os em ordem decrescente de preferência (do mais para o menos preferido). Para facilitar a sua tarefa, você pode primeiramente, dividir os cartões em grupos de carros com preferência similar e posteriormente, ordenar os cartões de cada grupo.

--- Ordene também esses cartões.

FIGURA A10.1. Modelo de cartão com a descrição de estímulos, utilizado na obtenção de dados do estudo empírico.

design	:	2 VOLUMES
potencia	:	1.6
combustivel	:	ALCOOL
marca	:	C
consumo	:	BAIXO

APÊNDICE 11

Tabelas referentes ao capítulo 7

TABELA A11.1. Importância relativa média do atributo "combustível" segundo método de estimação, ordem do questionário e sexo

Ordem	Sexo	Método de Estimação					
		MONANOVA	LINMAP	MQ	Híbrido	Bayesiano	LOGIT
AE-CONJ	masc	0,0849	0,0615	0,0858	0,0858	0,0839	0,1003
	fem	0,0329	0,0550	0,0574	0,0572	0,0487	0,0396
CONJ-AE	masc	0,0148	0,0528	0,0478	0,0470	0,0530	0,0551
	fem	0,1503	0,1974	0,1415	0,1416	0,1434	0,1341

TABELA A11.2. Importância relativa média do atributo "marca" segundo método de estimação, ordem do questionário, local da pesquisa e sexo

Ordem	Local	Sexo	Método de Estimação					
			MON	LIN	MQ	Híb	Bay	LOG
AE-CONJ	FGV	masc	0,2306	0,2018	0,2282	0,2282	0,2327	0,2049
		fem	0,5095	0,4485	0,4856	0,4760	0,4203	0,4785
	IME	masc	0,3606	0,3664	0,3427	0,3427	0,3381	0,3491
		fem	0,7466	0,5644	0,5240	0,5239	0,5468	0,5915
CONJ-AE	FGV	masc	0,8113	0,6330	0,5851	0,5865	0,5822	0,6328
		fem	0,0903	0,0000	0,1529	0,1531	0,1777	0,0774
	IME	masc	0,6604	0,4844	0,5049	0,5049	0,5035	0,5226
		fem	0,5826	0,5062	0,5675	0,5675	0,5590	0,5675

TABELA A11.3. Importância relativa média do atributo "consumo" segundo método de estimação, ordem do questionário, local da pesquisa e sexo

Ordem	Local	Sexo	Método de Estimação					
			MON	LIN	MQ	Híb	Bay	LOG
AE-CONJ	FGV	masc	0,4450	0,5821	0,4456	0,4456	0,4380	0,4955
		fem	0,1535	0,1565	0,1622	0,1721	0,1926	0,1523
	IME	masc	0,3450	0,3173	0,3559	0,3560	0,3658	0,3558
		fem	0,2021	0,3154	0,3294	0,3294	0,3282	0,3202
CONJ-AE	FGV	masc	0,0664	0,1635	0,1262	0,1279	0,1424	0,1178
		fem	0,4260	0,6667	0,4037	0,4033	0,3925	0,4786
	IME	masc	0,1257	0,1938	0,1719	0,1719	0,1631	0,1712
		fem	0,2704	0,2464	0,2787	0,2787	0,2802	0,2897

TABELA A11.4. Correlação de Spearman média, segundo método de estimação e local da entrevista, calculada para os dados da amostra principal

Método	Local	
	FGV	IME
MONANOVA	0,9296	0,9752
LINMAP	0,8453	0,9094
MQ	0,9541	0,9809
Híbrido	0,9412	0,9774
Bayesiano	0,9319	0,9767
LOGIT	0,9549	0,9862

TABELA A11.5. Correlação de Spearman média, segundo método de estimação, local de entrevista e sexo, calculada para os dados da amostra de validação

	Local			
	FGV		IME	
	masc	fem	masc	fem
MONANOVA	0,9626	0,7891	0,8274	0,8988
LINMAP	0,8735	0,6805	0,7420	0,8315
MQ	0,9524	0,8197	0,8093	0,8951
Híbrido	0,9286	0,8175	0,8066	0,8959
Bayesiano	0,9558	0,8135	0,8214	0,8988
LOGIT	0,9388	0,8299	0,8304	0,8988

TABELA A11.6. Correlação média de Spearman, calculada entre os postos previstos e observados, para os métodos de estimação LINMAP, híbrido e bayesiano, segundo ordem, sexo e local de entrevista, para a amostra principal

Ordem	Sexo	Local	Método de Estimação		
			LINMAP	Híbrido	Bayesiano
AE-CONJ	masc	FGV	0,8843	0,9787	0,9794
AE-CONJ	masc	IME	0,8912	0,9632	0,9581
AE-CONJ	fem	FGV	0,8764	0,9255	0,8892
AE-CONJ	fem	IME	0,9068	0,9860	0,9846
CONJ-AE	masc	FGV	0,8256	0,9111	0,9213
CONJ-AE	masc	IME	0,8945	0,9713	0,9745
CONJ-AE	fem	FGV	0,7902	0,9471	0,9255
CONJ-AE	fem	IME	0,9453	0,9890	0,9897

TABELA A11.7. Correlação média de Spearman, calculada entre os postos previstos e observados, para os métodos de estimação MONANOVA, MQ e híbrido, segundo sexo e local de entrevista, para a amostra principal

Sexo	Local	Método de Estimação			
		MONANOVA	MQ	Híbrido	LOGIT
masc	FGV	0,9530	0,9623	0,9449	0,9629
fem	FGV	0,9027	0,9446	0,9363	0,9458
masc	IME	0,9690	0,9743	0,9673	0,9783
fem	IME	0,9814	0,9874	0,9875	0,9941

TABELA A11.8. Correlação média de Spearman, calculada entre os postos previstos e observados, para os métodos de estimação MONANOVA, LINMAP, MQ e híbrido, segundo ordem de apresentação do questionário e local de entrevista, para a amostra de validação

Ordem	Local	Método de Estimação	
		MQ *	LOGIT *
AE-CONJ	FGV	0,9286	0,9137
AE-CONJ	IME	0,8406	0,8512
CONJ-AE	FGV	0,8294	0,8453
CONJ-AE	IME	0,8637	0,8780

* : sem o caso 8

TABELA A11.9. Correlação média de Spearman, calculada entre os postos previstos e observados, para os métodos de estimação MQ e híbrido, segundo ordem e sexo, para a amostra de validação

Ordem	Sexo	Método de Estimação	
		MQ *	Híbrido *
AE-CONJ	masc	0,8512	0,8325
	fem	0,9354	0,9368
CONJ-AE	masc	0,9218	0,9259
	fem	0,7891	0,7721

* : sem o caso 8

TABELA A11.10. Correlação média de Spearman, calculada entre os postos previstos e observados, para os métodos de estimação híbrido, bayesiano e LOGIT, segundo ordem de apresentação do questionário, para a amostra de validação

Ordem	Método de Estimação			
	Híbrido	Bayesiano	LOGIT	LOGIT *
AE-CONJ	0,8905	0,8706	0,8857	0,8857
CONJ-AE	0,7762	0,8773	0,8095	0,8993

*: sem o caso 8

BIBLIOGRAFIA

- Addelman, S, (1962) . Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics*, 4(1) : 21-58.
- Akaah, I.P. & Korgaonkar, P.K. (1983) . An empirical comparison of the predictive validity of self-explicated, Huber-hybrid, traditional conjoint, and hybrid conjoint models. *Journal of Marketing Research*, 20(2) : 187-97.
- Aldrich, J.H. & Nelson, F.D. (1984) .*Linear probability, logit and probit models*. Beverly-Hills, Sage Publications. 95 p. (Sage university papers series on quantitative applications in the social sciences 07-045)
- Alpert, M.I. (1971) . Identification of determinant attributes: a comparison of methods. *Journal of Marketing Research*, 8(2) : 184-91.
- Anderson, N.H. (1970) . Functional measurement and psychological judgment. *Psychological Review*, 77(3) : 153-70.
- Barlow, R.E., Bartholomew, D.J., Bremner, J.M. & Brunk, H.D. (1972) . *Statistical inference under order restrictions : the theory and application of isotonic regression*. London, John Wiley. 388p. (Wiley Series in Probability and Mathematical Statistics).

- Barron, F.H. (1977) . Axiomatic conjoint measurement. *Decision Sciences*. 8(3): 548-59.
- Box, G.E.P., Hunter, W.G. & Hunter, J. (1978) . *Statistics for experiments: an introduction to design, data analysis and model building*. New York, John Wiley. 653p. (Wiley Series in Probability and Mathematical Statistics).
- Bussab, W.O., Miazaki, E.S. & Andrade, D.F. (1990) . *Introdução à análise de agrupamentos*. São Paulo, USP. 105 p. Minicurso apresentado no 9^o Simpósio Nacional de Probabilidade e Estatística, São Paulo.
- Carmone, F.J. & Green, P.E. (1981) . Model misspecification in multiattribute parameter estimation. *Journal of Marketing Research*, 18(1) : 87-93.
- Carmone, F.J., Green, P.E. & Jain, A.K. (1978) . Robustness of conjoint analysis: some Monte Carlo results. *Journal of Marketing Research*, 15(2) : 300-3.
- Carroll, J.D. (1969) . Categorical conjoint measurement. *Meeting of Mathematical Psychology*, Annals. Arbor, MI.
- Carroll, J.D. (1972) . Individual differences and multidimensional scaling. In: Shepard, R.N., Romney, A.K. & Nerlone, S.B., eds. *Multidimensional scaling: theory and applications in the behavioral sciences*. v.1. New York, Seminar Press : 105-57.
- Cattin, P. & Bliemel, F. (1978) . Metric vs. nonmetric procedures for multiattribute modeling: some simulation results. *Decision Sciences*, 9(3) : 472-80.
- Cattin, P., Gelfand, A.E., & Danes, J. (1983) . A simple bayesian procedure for estimation in a conjoint model. *Journal of Marketing Research*, 20(1) : 29-35.
- Cattin, P., Hermet, G. & Proche, A. (1982) . Alternative hybrid models for conjoint analysis: some empirical results in *Analytical approaches to product and marketing planning: the second conference*. Cambridge, Marketing Science Institute (October) : 142-52.
- Cattin, P., Wittink, D.R. (1982) . Commercial use of conjoint analysis: a survey. *Journal of Marketing*, 46(3) : 44-53.
- Conover, W.S. (1980) . *Practical nonparametric statistics*. 2ed. New York. John Wiley. 493p.

- Cox, D.R. & Snell, E.J.(1989) . *Analysis of binary data*. London, Chapman and Hall. 236p. (Monographs on Statistics and Applied Probability).
- David, H.A. (1988) . *The method of paired comparisons*. 2ed. London, Charles Griffin. 188p. (Griffin statistical monographs, 41).
- Davis, J.A. (1976) . *Levantamento de dados em sociologia*. Rio de Janeiro, Zahar. 239p.
- Davison, M.L. (1983) . *Multidimensional scaling*. New York, John Wiley. 242p. (Wiley Series in Probability and mathematical Statistics).
- Draper, N.R. & Nostrand, R.C.V. (1979) . Ridge regression and James-Stein estimation: review and comments. *Technometrics*, 21(4) : 451-66.
- Draper, N.R. & Smith, H. (1981) . *Applied Regression Analysis*. 2ed. New York, John Wiley. 709p. (Wiley Series in Probability and Mathematical Statistics).
- Feir-Walsh, B.J. & Toothaker, L.E. (1974) . An empirical comparison of the ANOVA F-test, normal scores test and Kruskal-Wallis test under violation of assumption. *Educational and Psychological Measurement*, 34(4) : 789-99.
- Fraser, D.A.S. (1957) . *Nonparametric methods in statistics*. New York, John Wiley. 299p.
- Friedman, M. (1937) . The use of ranks to avoid the assumption of normality implicitly in the analysis of variance. *Journal of the American Statistical Association*, 32(4) : 675-701.
- Gabriel, K.R. & Lachenbruch, P.A. (1969) . Nonparametric ANOVA in small samples: a Monte Carlo study of the adequacy of the asymptotic approximation. *Biometrics*, 25(3) : 593-96.
- Green, P.E. (1973) . On the analysis of interactions in marketing research data. *Journal of Marketing Research*, 15, Nov: 410-20.
- Green, P.E. (1974) . On the design of choice experiments involving multifactor alternatives. *Journal of Consumer Research*, 1(2) : 61-8.
- Green, P.E. (1984) . Hybrid models for conjoint analysis: an expository review. *Journal of Marketing Research*, 21(2) : 155-69.

- Green, P.E. & Carmone, F.J. (1970) . *Multidimensional scaling and related techniques in marketing research*. Boston, Allyn. 203p. (Marketing Science Institute Series).
- Green, P.E., Goldberg, S.M. & Montemayor, M. (1981) . A hybrid utility estimation model for conjoint analysis. *Journal of Marketing*, 45(1) : 33-41.
- Green, P.E., Goldberg, S.M. & Wiley, J.B. (1982) . A cross validation test of hybrid conjoint models. *Proceedings of 1982 Annual Meeting of the Association for Consumer Research*. San Francisco, October.
- Green, P.E., Krieger, A.M. & Agarwal, M.K. (1991) . Adaptive conjoint analysis: some caveats and suggestions. *Journal of Marketing Research*, 28(2) : 215-22.
- Green, P.E. & Rao, V.R. (1971) . Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8(3) : 355-63.
- Green, P.E. & Srinivasan, V. (1978) . Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, 5(2) : 103-23.
- Green, P.E. & Tull, D.S. (1978) . *Research for marketing decisions*. 4ed. Englewood Cliffs, Prentice Hall.
- Green, P.E. & Wind, Y. (1975) . New way to measure consumers judgments. *Harvard Business Review*, 53(4) : 107-17.
- Greene, W.H. (1989) . *LIMDEP*. New York, New York University. 303p.
- Huber, G.P., Sahney, V. & Ford, D.L. (1969) . A study of subjective evaluation methods. *Behavioral Science*, 14(4):483-9.
- Jackson, J.E. & Shenker, K.L. (1982) . Incomplete paired comparisons for conjoint analysis. In *American Statistical Association Proceedings of Business and Economic Section*. Washington, American Statistical Association : 37-40.
- Jacoby, J., Chestnut, R.W. & Fisher, W.A. (1978) . A behavioral process approach to information acquisition in non durable purchasing. *Journal of Marketing Research*, 15(4) : 532-44.
- Jain, A.R., Acito, F., Malhotra, N.K. & Mahajan, V. (1979) . A comparison of the internal validity of alternative parameter estimation methods in decompositional multiattribute preference models. *Journal of Marketing Research*, 16(3) : 313-22.

- James, W. & Stein, C. (1961) . Estimation with quadratic loss. In Neyman, J, ed. *Proceedings of the Fourth Berkeley Symposium*. v.1. Berkeley, University of California Press. 361-79.
- Johnson, N.L. & Kotz, S. (1970) . *Distributions in Statistics* 2 ed, v.2. New York, John Wiley. 300p. (Wiley Series in Probability and Mathematical Statistics)
- Johnson, R.M. (1974) . Trade-off analysis of consumer values. *Journal of Marketing Research*, 11(2) : 121-7.
- Johnson, R.M. (1975) . A simple method for pairwise monotone regression. *Psychometrika*, 40(2) : 163-8.
- Johnson, R.M. (1987) . Adaptive conjoint analysis. In *Proceedings of the Sawtooth software conference on perceptual mapping, conjoint analysis, and computer interviewing*. Sun Valley. 253-66.
- Kamakura, W.A. (1988) . A least squares procedure for benefit segmentation with conjoint experiments. *Journal of Marketing Research*, 25(2) : 157-67.
- Kohli, R. (1988) . Assessing attribute significance in conjoint analysis: nonparametric tests and empirical validation. *Journal of Marketing Research*, 25(2) : 123-33.
- Krantz, D.M. (1964) . Conjoint measurement, the Luce-Tukey axiomatization and some extensions. *Journal of Mathematical Psychology*, 1(2) : 248-78.
- Kruskal, J.B. (1964a) . Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*. 29(1) : 1-27.
- Kruskal, J.B. (1964b) . Nonmetric multidimensional scaling: a numerical method. *Psychometrika*. 29(2) : 115-29.
- Kruskal, J.B. (1965) . Analysis of factorial experiments by estimating monotone transformation of the data. *Journal of the Royal Statistical Society, B*, 27(2) : 251-63.
- Kruskal, W.H. (1952) . A nonparametric test for the several sample problem. *Annals of Mathematical Statistics*, 23(4) : 525-40.
- Kruskal, W.H. & Wallis, W.A. (1952) . Use of ranks on one criterion variance analysis. *Journal of American Statistical Association*, 43 (Dec) : 583-621.

- Leeuw, J., Young, F.W. & Takane, Y. (1976) . Additive structure in quality data: an alternative least squares method with optimal scaling features. *Psychometrika*, 41(4) : 471-504.
- Louviere, J.J. (1988) . *Analysing decision making*. Beverly Hills, Sage Publications, 95p. (Sage university papers series on quantitative applications in the social sciences 07-067).
- LP88 (1987) . *LP88 - Linear programming for IBMPC. User's manual*. Version 3. Alexandria, Eastern Software Products, 111p.
- Luce, R.D. & Tukey, J.W. (1964) . Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology*, 1(1) : 1-27.
- Madansky, A. (1980) . On conjoint analysis and quantal choice models. *Journal of Business*, 53(2), pt 2 : S37-S50.
- Mahajan, V., Green, P.E. & Goldberg, S.M. (1982) . A conjoint model for measuring self and cross-price/demand relationship. *Journal of Marketing Research*, 19(3) : 334-42.
- Malhotra, N.K. (1982) . Structural reliability and stability of nonmetric conjoint analysis. *Journal of Marketing Research*, 19(2) : 199-207.
- McCullough, J. & Best, R. (1979) . Conjoint measurement: temporal stability and structural reliability. *Journal of Marketing Research*, 16(1) : 26-31.
- McFadden, D. (1974) . Conditional logit analysis of qualitative choice behavior. In Zarembka, P., ed. *Frontiers in Econometrics*. New York, Academic Press. 105-42. (Economic theory, econometrics and mathematical economics series).
- McKelvey, R.D. & Zavoina, W. (1975) . A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4(1) : 103-20.
- Michel, J. (1988) . Some problems in testing the double cancellation condition in conjoint measurement. *Journal of Mathematical Psychology*, 32, 466-73.
- Montgomery, D.C. & Peck, E.A. (1982) . *Introduction to linear regression analysis*. New York, John Wiley. 504p. (Wiley Series in Probability and Mathematical Statistics).

- Mood, A.M., Graybill, F.A. & Boes, D.C. (1974) . *Introduction to the theory of statistics*. 3ed. McGraw Hill. 564p.
- Moore, W.L. & Holbrook, M.B. (1990) . Conjoint analysis with environmentally correlated attributes: the questionable importance of representative designs. *Journal of Consumer Research*, 16(3) : 490-7.
- Mullet, G.M. & Karson, M.J. (1986) . Percentiles of LINMAP conjoint indices of fit for various orthogonal arrays: a simulation study. *Journal of Marketing Research*, 23(3) : 286-90.
- Newman, J.W. (1977) . Consumer external search: amount and determinants. In Woodside, A.G., Sneth, J.N. and Bennett, P.D., eds. *Consumer and industrial buying behavior*. New York, North Holland. 79-94.
- Punj, G.N. & Staelin, R. (1978) . The choice process for graduate business schools. *Journal of Marketing Research*, 15(4) : 588-98.
- Rao, V.R. & Winter, F.W. (1978) . An application of the multivariate probit model to market segmentation and product design. *Journal of Marketing Research*, 15, Aug, 361-8.
- Searle, S.R. (1971) . *Linear models*. New York, John Wiley. 531p.
- Segal, M.N. (1982) . Reliability of conjoint analysis: contrasting data collection procedures. *Journal of Marketing Research*, 19(1) : 139-43.
- Shepard, R. (1964) . On subjectively optimum selections among multiattribute alternatives. In Shelly, M.W. & Bryan, G.L., eds. *Human Judgements and optimality*. New York, John Wiley. 257-81.
- Smith, S.M. (1988) . *PC-MDS. Multidimensional statistical package*. Provo, Brigham Young University. 305p.
- SOC (1989) . *SOC. Software científico*, elaborado pelo Núcleo Tecnológico de Informática para Agropecuária da EMBRAPA (NTIA/EMBRAPA).
- Souza, D.G. (1986) . *Algumas considerações sobre regressão não linear*. São Paulo. 185p. Dissertação (Mestrado). Instituto de Matemática e Estatística da USP.
- SPSSPC+(1989) . *Statistical Package for Social Sciences*. 2v. Chicago, SPSS Inc.

- Srinivasan, V., Jain, A.K. & Malhotra, N.K. (1983) . Improving predictive power of conjoint analysis by constrained parameter estimation. *Journal of Marketing Research*: 20(4) : 433-8.
- Srinivasan, V. & Shocker, A.D. (1973a) . Linear programming techniques for multi-dimensional analysis of preferences. *Psychometrika*, 38(3) : 337-69.
- Srinivasan, V. & Shocker, A.D. (1973b) . Estimating the weights for multiattributes in a composite criterion using pairwise judgments. *Psychometrika*, 38(4) : 473-93.
- Stein, C. (1962) . Confidence sets for means of a multivariate normal distribution. *Journal of the Royal Statistical Society, B*, 24, 265-96.
- Tversky, A. (1967) . A general theory of polynomial conjoint measurement. *Journal of Mathematical Psychology*, 4(1) : 1-20.
- Veen, W.M.O. & Beazley, D. (1977) . An investigation of alternative methods of applying the trade-off model. *Journal of the Marketing Research Society*, 19(1) : 2-11.
- Vinod, H.D. & Ullah, A. (1981) . *Recent Advances in Regression Methods*. New York, Marcel Dekker. 231p.
- Wittink, D.R. & Cattin, P. (1981) . Alternative estimation methods for conjoint analysis: a Monté Carlo study. *Journal of Marketing Research*, 18(1) : 101-6.
- Wittink, D.R. & Cattin, P. (1989) . Commercial use of conjoint analysis: an update. *Journal of Marketing*, 53(3) : 91-6.
- Wood, S. (1987) . *Turbo Pascal: guia do usuário*. São Paulo, McGraw Hill. 205p.
- Young, F.W. (1969) . Polynomial conjoint analysis of similarities: definitions for a special algorithm. *Research paper n. 76. Psychometric Laboratory, University of North Carolina*.
- Yoshida, L.K. (1987) . *Programação linear*. São Paulo, Atual. 168p. (Métodos Quantitativos).