

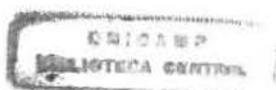
UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO  
CIENTÍFICA  
DEPARTAMENTO DE ESTATÍSTICA

*Uma Estratégia de Análise Psicométrica para Instrumentos  
Baseados em Escala de Likert via Análise de Correspondência  
Múltipla*

Fernando Antonio Basile Colugnati

Campinas - SP

Agosto, 2001



UNICAMP  
BIBLIOTECA CENTRAL  
SEÇÃO CIRCULANTE

UNIDADE Bc  
N.º CHAMADA:  
T/ UNICAMP  
C723e  
V. \_\_\_\_\_ Ex. \_\_\_\_\_  
TOMBO BR/ 47027  
PROC. 16-392/07  
C  D   
PREÇ. R\$ 11,00  
DATA 20/11/07  
N.º CPD \_\_\_\_\_

CM00161522-8

**FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DO IMECC DA UNICAMP**

Colugnati, Fernando Antonio Basile

C723e Uma estratégia de análise psicométrica para instrumentos baseados em escala de Likert via análise de correspondência múltipla / Fernando Antonio Basile Colugnati. Campinas, [S.P. :s.n.], 2001.

Orientador : Regina Célia Carvalho Pinto Moran

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Análise multivariada. 2. Psicometria. 3. Análise de correspondência. I. Moran, Regina Célia Carvalho Pinto. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

**Uma Estratégia de Análise Psicométrica para Instrumentos Baseados em Escala de Likert via Análise de Correspondência Múltipla**

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Fernando Antonio Basile Colugnati e aprovada pela comissão julgadora

Campinas, 7 de agosto de 2001



---

Profa. Dra. Regina C.C.P. Moran  
Orientadora

**Banca Examinadora**

1. Profa. Dra. Anita Liberalesso Neri (FE/ UNICAMP)
2. Prof. Dr. Armando Infante (IMECC/ UNICAMP)
3. Prof. Dr. Emanuel Pimentel Barbosa (IMECC/ UNICAMP) - Suplente

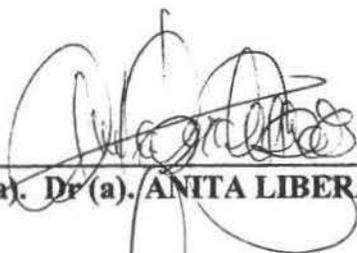
Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do título de Mestre em Estatística

Dissertação de Mestrado defendida em 07 de agosto de 2001 e aprovada

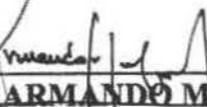
Pela Banca Examinadora composta pelos Profs. Drs.



Prof (a). Dr (a). REGINA CÉLIA DE CARVALHO PINTO MORAN



Prof (a). Dr (a). ANITA LIBERALESSO NERI



Prof (a). Dr (a). ARMANDO MARIO INFANTE

## Agradecimentos

- A Deus por tudo e por todos;
- Aos meus pais, Walter e Elizabeth pelo amor, carinho e incentivo. Aos meus irmãos e familiares próximos, por serem minha família, e se comportarem como uma;
- Aos meus amigos, irmãos de luta, Carlos “Borjão” e Odair “Dadálho” pela compreensão e o apoio nas transformações nada lineares da minha vida. A suas “respectivas” Lu e Claudia pelos mesmos motivos e por ainda darem o alvará de sexta;
- Ao pessoal da Pós-Graduação em Nutrição EPM/UNIFESP, Dra. Dirce e Taddei principalmente, pelo apoio, confiança e incentivo. Vocês não têm idéia de como são importantes na conclusão deste trabalho;
- A PrOfa. Dra. Regina pela orientação, confiança e amizade;
- A Diana Tosello Laloni pelo apoio e auxílio, pelos dados, pelo problema que originou este trabalho;
- A Luis Laloni pela lição de vida;
- Aos amigos Paulo “Maciêra”, Zédu, Zero, Endrigo “Minêro” (que Deus o tenha...), Mariana, Anderson, Marcelo, Nanci, Edmilson, Juli, Bar do Zé, etc... e todo o povo dos botecos e festas (ninguém é de ferro);
- Ao pessoal do TACO/ NEPA, Profª. Galeazzi, Profª. Déila, Dag, Renata, Lúlu, Arlinda, etc, etc, etc...Muita gente!
- Ao Zappa, Hendrix, Lou Reed, Ivan Vilela, Paulo Freire, Mingus, Raulzito, Stevie Ray Vaughan, Albert King, etc... pelos momentos de inspiração;
- Ao CNPq por conceder minha bolsa de estudos;
- Ao atual presidente da república Dr. Fernando Henrique Cardoso e sua magnânima equipe, por me deixar claro que Caráter, Respeito e Competência são virtudes pessoais que independem de títulos acadêmicos.

*“Nem a mais sedutora verosimilhança pode nos proteger contra o erro, ainda que todos os elementos de um problema pareçam ordenar-se como as peças de um quebra-cabeças, teremos de recordar que o provável não é necessariamente certo, nem a verdade sempre provável”*

Sigmund Freud

(Moisés e o monoteísmo)

---

## Resumo

Este trabalho explora a utilização da técnica de Análise de Correspondência Múltipla no campo da psicometria como uma ferramenta de avaliação e validação de instrumentos baseados em Escala de Likert, e procura disponibilizar uma estratégia para este tipo de estudo.

É apresentada uma generalização das técnicas baseadas em Decomposição de Valores Singulares e a particularização para a Análise de Correspondência Múltipla. Além disso é sugerida a utilização de um índice de interpretação análoga à carga fatorial, além da Escala Ótima, importante no estudo de ordenação das categorias de resposta dos itens.

Um estudo de simulação é feito aplicando-se a nova estratégia e a forma tradicional, através de Análise Fatorial solucionada por Componentes Principais e os resultados são comparados.

Observou-se uma vantagem na abordagem sugerida quanto à identificação de construtos e na discriminação de grupos de indivíduos com diferentes perfis, principalmente quando a relação das variáveis e o fator não é linear.

Também é feita uma aplicação em dados reais de um estudo realizado em Campinas para a validação do instrumento SCL-90-R.

---

# ÍNDICE ANALÍTICO

<b>1. INTRODUÇÃO .....</b>	<b>13</b>
1.1- ESCALAS PSICOMÉTRICAS .....	15
1.2- A ESCALA DE LIKERT .....	18
<b>2. PROPOSIÇÃO DO PROBLEMA E GENERALIZAÇÃO DA SOLUÇÃO .....</b>	<b>20</b>
2.1- O AJUSTE DA NUVEM DE PONTOS EM UM SUB-ESPAÇO ÓTIMO .....	21
2.2- A INÉRCIA – UMA MEDIDA DE VARIABILIDADE .....	23
2.3- O AJUSTE DO SUB-ESPAÇO ÓTIMO VIA DVS .....	25
2.3.1- <i>As relações entre os dois espaços</i> .....	27
2.4- COMENTÁRIO FINAL.....	29
<b>3. ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA.....</b>	<b>30</b>
3.1- A MATRIZ INDICADORA E SUAS PROPRIEDADES.....	32
3.1.1 – <i>As Nuvens de Pontos</i> .....	33
3.1.2- <i>O Referencial da Nuvem de Pontos</i> .....	36
3.1.3- <i>As distâncias em G</i> .....	36
3.2 – A MATRIZ TRANSFORMADA A SER APROXIMADA .....	39
3.3- A INTERPRETAÇÃO.....	41
3.3.1- <i>O Princípio Baricêntrico</i> .....	41
3.3.2- <i>A Escala Ótima</i> .....	42
3.3.3- <i>A Inércia na ACM</i> .....	42
3.3.4- <i>A contribuição de uma variável à inércia de um eixo</i> .....	44
3.3.5- <i>O coeficiente de discriminação</i> .....	45
3.3.6- <i>Uma sugestão de carga fatorial para variáveis qualitativas</i> .....	46
3.3.7 - <i>O Efeito Ferradura (EF) na Análise de Correspondência Múltipla e sua Importância na Psicometria</i> .....	50
3.3.7- <i>A escolha do número de eixos</i> .....	55
3.3.8- <i>A utilização dos pontos ilustrativos</i> .....	57
<b>4. ESTUDO DE SIMULAÇÃO .....</b>	<b>59</b>
4.1- ESTRUTURA DE DADOS UTILIZADA .....	61
4.2- GERAÇÃO DOS DADOS .....	62
4.3- MATRIZES E VETORES UTILIZADOS.....	64
4.4- SIMULAÇÕES.....	66
4.4.1- <i>Caso da escala unidimensional</i> .....	66

---

4.4.2- Casos de instrumentos com duas escalas .....	71
<i>Simulação 2</i> .....	71
<i>Simulação 3</i> .....	78
<i>Simulação 4</i> .....	83
<i>Simulação 5</i> .....	89
4.5- ESCALAS ÓTIMAS NAS SIMULAÇÕES .....	95
4.6- <i>Comentários e Conclusões</i> .....	98
<b>5. APLICAÇÃO</b> .....	<b>100</b>
5.1- O QUESTIONÁRIO SCL-90-R .....	101
5.1.1- <i>Utilizações e Amostras apropriadas</i> .....	103
5.1.2- <i>Índices e escores utilizados</i> .....	104
5.1.3- <i>A Amostra</i> .....	104
5.2- ANÁLISE DO INSTRUMENTO VIA ACM .....	110
5.3- CONCLUSÃO SOBRE A VALIDAÇÃO DA ESCALA GSI DO SCL 90 R4 .....	122
<b>6. CONSIDERAÇÕES FINAIS</b> .....	<b>123</b>
6.1- A ESTRATÉGIA SUGERIDA .....	125
<b>APÊNDICE A</b> .....	<b>127</b>
A.1- BREVE INTRODUÇÃO A PSICOMETRIA .....	127
A.1.1 <i>Medidas Psicológicas</i> .....	127
A.1.2- <i>Breve Histórico</i> .....	128
A.2- SOBRE VALIDAÇÃO .....	130
A.2.1- <i>A validação na psicometria</i> .....	131
A.3- ANÁLISE FATORIAL .....	135
A.3.1 <i>Breve Históricos</i> .....	135
A.3.2 <i>O Modelo de Análise Fatorial Ortogonal</i> .....	137
A.3.3- <i>A solução por Componentes Principais</i> .....	140
<b>APÊNDICE B</b> .....	<b>141</b>
B.1- DECOMPOSIÇÃO EM VALORES SINGULARES (DVS) .....	141
<b>APÊNDICE C</b> .....	<b>146</b>
C.1- PROGRAMAS EM <i>Ox 2.1</i> UTILIZADOS NAS SIMULAÇÕES .....	146
<b>REFERÊNCIAS</b> .....	<b>153</b>
<b>ANEXOS</b> .....	<b>157</b>

---

DISTRIBUIÇÕES MARGINAIS DOS ÍTENS .....	157
INSTRUMENTO APLICADO.....	158

---

## ÍNDICE DE FIGURAS E GRÁFICOS

fig. 2.1- Ajuste de uma reta no $R^2$	22
fig. 3.1- Tabela original e Matriz de Petrie sobre a mesma (fonte: Greenacre (1984), p.228-229 )	53
graf. 3.2- Efeito Ferradura na Análise de Correspondência	54
fig. 3.3- Diagramas de auto-valores. Em a) provenientes de uma análise sem qualquer estrutura latente. Em b) há evidência de estrutura nos primeiro e segundo eixos.	56
graf. 4.1- Mapa da ACM para Simulação 1	69
graf. 4.2- Dispersão dos indivíduos na ACM (a) e AF (b) para Simulação 1	70
graf. 4.3- Mapa da ACM para Simulação 2	76
graf. 4.4- Dispersão dos indivíduos na ACM (a) e AF (b) para Simulação 2	77
graf. 4.5- Mapa da ACM para Simulação 3	80
graf. 4.6- Dispersão dos indivíduos na ACM (a) e AF (b) para Simulação 3	82
graf. 4.7- Mapa da ACM para Simulação 4	85
graf. 4.8- Dispersão dos indivíduos na ACM (a) e AF (b) para Simulação 4	88
graf. 4.9- Mapa da ACM para Simulação 5	92
graf. 4.10- Dispersão dos indivíduos na ACM (a) e AF (b) para Simulação 5	94
fig. 4.11- Gráfico de linha entre $\varphi_1$ x Escala Original para Simulação 1	95
fig. 4.12- Gráficos de linha entre $\varphi_1$ (1º eixo) ou $\varphi_2$ (2º eixo) x Escala Original para Simulação 2	96
graf. 4.13- Gráfico de linha entre $\varphi_1$ (1º eixo) ou $\varphi_2$ (2º eixo) x Escala Original para Simulação 4	97
graf. 4.14- Gráfico de linha entre $\varphi_1$ (1º eixo) ou $\varphi_2$ (2º eixo) x Escala Original para Simulação 5	98
graf. 5.1 – Mapa da ACM para variáveis demográficas da amostra	109
graf. 5.2 – Mapa de ACM das Escalas Ótimas nos eixos 1 e 2	112
graf. 5.3 – Mapa de ACM das Escalas Ótimas nos eixos 1 e 3.	112
graf. 5.4 – Mapa de ACM das Escalas Ótimas nos eixos 1 e 4.	113
graf. 5.5- Mapa de ACM para os Escores do Indivíduos no primeiro plano	118
fig. 5.6 – Escores Ótimos do primeiro eixo da ACM contra escala de resposta do instrumento (por dimensão de sintoma).	119
fig. 5.6 (continuação) – Escores Ótimos do primeiro eixo da ACM contra escala de resposta do instrumento (por dimensão de sintoma)	120

---

 ÍNDICE DE TABELAS E QUADROS

tab. 4.1 – Estrutura de dados utilizada nas simulações	62
tab. 4.2- Estrutura de dados da Simulação 1	67
tab. 4.3- (a) Dez primeiros autovalores da ACM para Simulação 1. (b) Dez primeiros autovalores da AF para Simulação 1.	67
tab. 4.4- Cargas fatoriais obtidas para Simulação 1	68
tab. 4.5- Estrutura de geração de indivíduos na Simulação 2	72
tab. 4.6- (a) Dez primeiros autovalores da ACM para Simulação 2. (b) Dez primeiros autovalores da AF para Simulação 2.	73
tab. 4.7- Cargas fatoriais obtidas para Simulação 2	75
tab. 4.8- Estrutura de geração de indivíduos na Simulação 3	78
tab. 4.9- (a) Dez primeiros autovalores da ACM para Simulação 2. (b) Dez primeiros autovalores da AF para Simulação 2.	79
tab. 4.10- Cargas fatoriais obtidas para Simulação 3.	81
tab. 4.11- Estrutura de geração de indivíduos na Simulação 4	83
tab. 4.12- (a) Dez primeiros autovalores da ACM para Simulação 2. (b) Dez primeiros autovalores da AF para Simulação 4.	84
tab. 4.13- Cargas fatoriais obtidas para Simulação 4	87
tab. 4.14- Estrutura de geração de indivíduos na Simulação 5	89
tab. 4.9- (a) Dez primeiros autovalores da ACM para Simulação 2. (b) Dez primeiros autovalores da AF para Simulação 5.	90
tab. 4.10- Cargas Fatoriais obtidas para Simulação 5	93
tab. 5.1- Itens que compõem as dimensões de sintoma do SCL-90-R	103
tab. 5.2- Distribuição de Sexo	105
tab. 5.3- Distribuição de Escolaridade	106
tab. 5.4- Distribuição de Estado Civil	107
tab. 5.5- Distribuição de Sit. de Trabalho	107
tab 5.6- Autovalores da ACM para variáveis demográficas	108
tab. 5.7- 10 primeiros autovalores para análise dos 90 itens	110
tab. 5.8- Coeficientes de carga fatorial para o primeiro plano da ACM.	115
tab. 5.9- Estatísticas resumo dos Escores Brutos de cada dimensão de sintoma para respostas completas	121
quadro c.1 – Programa da função <i>GeraConstruto</i> (continuação)	148
quadro c.2 – Programa da função <i>GeraMultInd</i>	149
quadro c.3- Programa para Simulação 5	150
quadro c.3- Programa para Simulação 5	152
quadro c.3- Programa para Simulação 5 (continuação)	152

---

## 1. Introdução

O emprego de questionários como instrumento de coleta de informações nas ciências humanas e da saúde é uma prática freqüente. A intenção geralmente é traçar perfis de populações sobre atitudes e opiniões em relação a determinado assunto (ciências políticas e sociais), ou levantar dados que permitam prognósticos e/ou diagnósticos (no caso da saúde). Estes instrumentos variam quanto às formas de abordagem e das respostas às questões, entretanto sempre buscando uma forma simples e objetiva que minimize erros (do entrevistador, dos itens, etc...) e permita uma análise estatística interpretável e fidedigna.

Um tipo de instrumento muito utilizado são os chamados baseados em “Escala de Likert” (Likert, 1932). Esta denominação não é correta na definição formal de Likert, uma vez que esta engloba não só as formas das respostas, mas também o tratamento posterior da mesma, muito empregado na psicometria. Geralmente o instrumento é classificado como tal apenas pela escala de resposta qualitativa ordinal, geralmente de cinco categorias, onde o respondente escolhe as respostas quanto à concordância com o estímulo apresentado (*eg.* Extremamente Desfavorável, Desfavorável,....). A estas categorias são atribuídos números inteiros tais como 0, 1, 2, 3, 4, obedecendo o gradiente de interesse.

Quando falamos em psicologia clínica, estes instrumentos ou os chamados Testes Mentais têm papel fundamental, e são de grande utilização em triagens de pacientes, acompanhamento de evolução de tratamentos, etc... . Cabe à psicometria então estudar, avaliar e validar este tipo de instrumento para sua utilização.

---

A forma clássica de estudo e avaliação dos instrumentos baseados em escala de Likert trás alguns questionamentos, uma vez que os algarismos empregados na escala de resposta são tratados como valores, ou números, de uma forma contínua, adotando como “verdadeira” a métrica imposta por esta escala. Ou seja, toma-se como mensuráveis e verdadeiras as distâncias entre as categorias de resposta. Assim, um indivíduo que responde com a categoria 2 responde ao estímulo com uma intensidade duas vezes maior que o indivíduo que responde com a categoria 1.

Grande parte dos métodos empregados na Psicometria são os chamados de Análise Multivariada, principalmente a Análise Fatorial, que teve seu desenvolvimento nesta área da ciência a partir de Spearman (1904) e Thurstone (1967).

Tendo como premissa a natureza categórica ordinal da escala de resposta, sem uma métrica definida, este trabalho tem como objetivo a exploração de uma forma alternativa de análise psicométrica deste tipo de instrumentos através da Análise de Correspondência Múltipla, técnica adequada à natureza destes dados. Esta metodologia, além da adequação aos dados, traz resultados que permitem um estudo mais aprofundado das propriedades do instrumento, além da análise de construto que é usualmente feita.

Os próximos tópicos deste capítulo tratam das definições de escalas e dimensões na psicometria (1.1) e sobre a utilização da Escala de Likert.

O capítulo 2 trás a generalização da análise baseada em redução de dimensionalidade através da Decomposição em Valores Singulares.

O capítulo 3 traz a apresentação da Análise de Correspondência Múltipla (ACM), com uma breve introdução histórica. Em 3.1 é definida a matriz indicadora e suas propriedades e em seguida a formalização da ACM, também tratada como um caso particular da generalização no capítulo 2. É apresentado um item sobre interpretação da ACM no caso da psicometria (3.3), sendo feita a construção de cargas fatoriais derivadas a partir das contribuições relativas das variáveis.

O capítulo 4 traz um estudo de simulação e a comparação entre a ACM e a solução da Análise Fatorial por Componentes Principais (CP) apenas na questão das cargas fatoriais e dos escores dos indivíduos, destacando-se a detecção de construtos e a discriminação de grupos. As simulações mostram melhores resultados na ACM para a validação de construtos que a CP, principalmente em casos onde as relações entre a escala de resposta e

---

---

o fator encontrado não são lineares. Na discriminação de grupos também se obtêm melhores resultados com a ACM.

No capítulo 5 é apresentada uma aplicação da metodologia proposta sobre a versão brasileira do instrumento SCL 90 R (Laloni, 2001) utilizado em triagem de pacientes. As análises indicam que nesta primeira aplicação o instrumento demonstrou-se válido ao que se propõe enquanto escala unidimensional. No entanto a discriminação de indivíduos necessitaria de um novo estudo com populações mais heterogêneas.

O apêndice A traz uma introdução e histórico do desenvolvimento da psicometria e dos critérios de validação classicamente adotados, a Análise Fatorial e sua solução via Componentes Principais, bem como referências para um estudo mais aprofundado.

No Apêndice B estão os teoremas e resultados básicos sobre a DVS, e no Apêndice C os programas em *Ox 2.1* (Durnick, 1998) utilizados nas simulações.

Em Anexo estão as distribuições marginais de todos os itens do instrumento da aplicação.

## 1.1- Escalas Psicométricas

Podemos definir como processo de medição *“qualquer processo que assinale números a determinado objeto de estudo, de acordo com alguma regra”* (Kruskall, 1978). Esta regra de forma geral especifica aspectos quantitativos sobre a observação de um objeto, definindo a escala. Podemos dizer então que podemos defini uma *escala* sempre que *“há uma relação biunívoca entre as propriedades de um conjunto de algarismos e um conjunto de operações (o processo de medição) que podem ser realizadas ou observadas no objeto de estudo”*. Segundo Pasquali (1999) *“A escala psicométrica visa escalonar estímulos que expressam um construto psicológico”*.

No entanto mesmo na psicometria a escala pode estar sendo utilizada de diferentes formas e para diferentes propósitos. Para se entender quais são estas formas e propósitos, e objetivar-se qual será enfocado neste trabalho precisamos nos aprofundar um pouco mais no problema psicométrico.

---

---

Ao se elaborar ou utilizar um instrumento de escalonamento psicométrico estamos intrinsecamente abordando três dimensões: os estímulos que compõem o instrumento (os itens no caso de um questionário), os indivíduos, ou a população em que o mesmo será aplicado, e a resposta aos estímulos. Podemos realizar o escalonamento em qualquer uma das três dimensões do problema. Por exemplo, pode-se escalonar os estímulos quanto a grau de dificuldade, ou alguma outra classificação. Este enfoque é mais utilizado em problemas de psicofísica, como no caso das Escalas de Thurstone (Pasqualli, 1999; Thurstone, 1927), baseada quase sempre na idéia de contínuos psicológicos, e de pontos limiares de detecção à diferenciação de uma resposta a um estímulo. Um exemplo deste emprego são testes audiométricos.

Podemos também escalonar os indivíduos quanto às respostas aos estímulos. Este é o propósito das Escalas de Likert (Likert, 1932) (EL), e este será o enfoque deste trabalho. As EL tem como preocupação escalonar o indivíduo de acordo com uma série de afirmações que expressem atitude favorável ou desfavorável em relação ao construto psicológico estudado, em termos de respostas pontuadas em uma escala de resposta que pode variar de 3 a mais de 10 pontos, sendo as de 5 e 7 as mais utilizadas, onde a ordem nos pontos expressa o grau de concordância do indivíduo com o atributo do objeto apresentado. Classicamente os indivíduos são escalonados através da soma ou média dos pontos das respostas, o que a faz ser conhecida também como Método dos Pontos Somados. Note que mesmo aqui temos dois empregos do termo escala.

Guttman (1944, 1945, 1947, 1950) ainda apresenta um enfoque onde o escalonamento é feito sobre o estímulo e o indivíduo. Os estímulos são expressos em diferentes magnitudes seguindo uma série monotônica crescente, de forma que o conjunto destes estímulos representam o contínuo psicológico estudado, sendo que a aceitação de um estímulo de maior hierarquia implica na aceitação dos de menor hierarquia. A partir das respostas a estes itens, escalona-se o indivíduo.

Neste trabalho, o termo Escala será utilizado como sendo o contínuo final onde se escalonam os indivíduos através das somas (ou médias) dos pontos das escalas de resposta, conforme proposto por Likert.

Os tipos de escala acima descritas são consideradas unidimensionais, ou seja, definem um traço ou estado latente (construto) apenas, escalonando o indivíduos em uma

---

---

única dimensão, por exemplo se fosse desejada a mensuração da Ansiedade. Mesmo em um instrumento que deseja captar mais de um tipo de característica psicológica, o enfoque dado é de várias escalas unidimensionais, como no instrumento analisado no capítulo 5. Ou seja, teoricamente o instrumento apresentaria várias sub-escalas unidimensionais.

A literatura em escalas psicológicas é vasta, e uma exploração maior do assunto fugiria do escopo deste trabalho. Referências para este assunto, além das já citadas, são: Guilford (1954), Torgerson (1958), Zeller & Carmines (1980).

A utilização da Análise Multivariada no contexto apresentado é a de escalonamento, ora do indivíduos, ora das categorias de resposta, ou simultaneamente. Desta forma, este trabalho visa a generalização destes métodos através do que Gifi (1990) chama de Análise de Homogeneidade, e que a escola francesa (Benzécri (1973), Lebart (1995)) denomina de Métodos Fatoriais (*Méthodes Factorielles*).

Segundo Gifi, a idéia de homogeneidade está “*historicamente muito relacionada com a idéia de que as variáveis podem medir ‘a mesma coisa’* “. Esta é a idéia central quando tentamos reduzir um conjunto de variáveis a um número menor de variáveis latentes, ou fatores, como buscam as ciências sociais quantitativas e a psicometria desde o final do século XIX na mensuração de seus construtos teóricos. Neste período surgem os trabalhos sobre os coeficientes de correlação de Pearson e de Spearman, e a idéia de uma correlação média entre um conjunto maior de variáveis. Nesta época o enfoque era a redução a apenas uma dimensão, principalmente nos trabalhos de Spearman.

Pensando nas variáveis em questão, geralmente qualitativas, a grande preocupação sempre foi com relação a forma ótima de representação das variáveis, ou a métrica adequada das categorias de resposta no espaço dos indivíduos, o que na literatura é chamado de “pesos” (*weights*), que na verdade seria o peso atribuído a cada item para soma final da escala. Vários autores trataram deste assunto (Edgeworth, 1888; Spearman, 1913, Guilford, 1954). A conclusão é de que pouco se ganha na busca do peso ótimo, ou seja, este esforço não é compensado na medida que se temos uma estrutura de correlação uniforme, com um coeficiente médio alto (algo maior que 0.80) esta não seria afetada por uma combinação linear das respostas que tomasse os pesos adequados para cada variável. Este é o enfoque dado por Likert por exemplo, que soma os valores respondidos nos itens do instrumento uniformemente (sem ponderações) criando o contínuo final da escala.

---

---

No entanto quando esta estrutura não é homogênea podemos ter problemas se não adequarmos a representação à métrica imposta pelas heterogeneidades (ou pela falta de homogeneidade). No caso de variáveis qualitativas ordinais o problema pode ser maior, pois a métrica da escala dos itens depende da forma da distribuição marginal de resposta dos itens, que define a métrica do espaço onde deverão ser representados os indivíduos. Desprezar estas informações pode levar a uma perda de poder de discriminação, como será visto no capítulo 4.

Neste trabalho parte-se do pressuposto do benefício da dúvida, aplica-se as técnicas que consideram diferentes métricas e é feita uma comparação da eficiência.

## 1.2- A Escala de Likert

A Escala de Likert (Likert, 1932), como citado no tópico anterior, é formada a partir da soma (ou média) dos pontos das respostas do indivíduo. Então, utilizada como variável contínua os métodos estatísticos aplicados nos estudos analíticos, de validação, ou descritivos são aqueles para variáveis desta natureza, como ANOVAS, correlações lineares e Análise Fatorial.

As escalas de resposta são formadas por pontos que representam a discordância ou concordância do indivíduo com o estímulo, sendo portanto uma representação numérica de uma variável qualitativa. Por exemplo, suponha uma pergunta sobre algum sintoma de uma doença que deve ser respondida em uma escala de resposta de 3 pontos. O indivíduo poderia responder **Nenhum** sintoma (0), **Pouco** (1), **Muito**(2). A variável de resposta é tipicamente qualitativa ordinal. Temos o gradiente partindo de nenhum sintoma até um extremo oposto (esta oposição semântica uma das maiores vantagens das EL segundo diversos autores (Pasquali, 1999)). No entanto quando empregamos os algarismos em parênteses e trabalhamos com os mesmos como números, realizando somas por exemplo, podemos estar assumindo uma métrica não existente. Podemos assumir que Muito é duas

---

vezes mais forte que Pouco. Ou seja um indivíduo que respondera Muito está duas vezes mais doente que algum que tenha respondido Pouco.

Segundo Pereira (1998) *“a criação de uma escala envolve o estabelecimento de premissas de relação entre atributos de um objeto e uma representação simbólica deste.”* Extrapolar a representação simbólica pode ser perigoso neste caso.

Quando temos um número grande de itens no instrumento, este problema pode ser maior. Indivíduos com respostas totalmente diferentes em todos os itens, constituindo um quadro ou perfil diferente também, podem ser escalonados em uma mesma posição, sendo classificados da mesma maneira. No capítulo 4 deste trabalho as simulações mostram claramente este tipo de comportamento mesmo nos Escores Fatoriais obtidos através de Componentes Principais.

---

---

## 2. Proposição do Problema e Generalização da Solução

Considere que sobre  $I$  indivíduos sejam observadas  $J$  variáveis e estes resultados sejam representados por uma matriz de dados  $\mathbf{H}$ ,  $I \times J$  de posto  $p$ .

Originalmente podemos pensar na representação das  $J$  colunas no espaço  $I$ -dimensional, neste caso o  $\mathcal{R}^I$  por isomorfismo, ou de forma simétrica na representação dos  $I$  indivíduos no espaço isomorfo  $\mathcal{R}^I$ . As relações entre os indivíduos ou entre variáveis podem ser analisadas através do formato das nuvens de pontos nos respectivos espaços.

Mas este estudo torna-se problema se estas dimensões são muito grandes, ou simplesmente maiores que 3, surgindo a necessidade de redução da dimensionalidade. O objetivo então é reduzir a dimensionalidade desta matriz com o mínimo possível de perda de informação, de forma que as inter-relações nas respectivas nuvens de ponto possam ser interpretadas. Por redução de dimensionalidade entende-se a projeção dos pontos em um espaço de menor dimensão, ou a aproximação de  $\mathbf{H}$  por uma matriz  $\mathbf{H}^*$  de posto  $p^* < p$ .

Neste trabalho, o espaço denotado  $\mathcal{R}^I$  será denominado de *Espaço das Variáveis*, uma vez que está se representando as variáveis tendo os indivíduos como eixos. Analogamente o espaço denotado por  $\mathcal{R}^J$  será denominado o *Espaço dos Indivíduos*.

O desenvolvimento será feito sobre a redução de dimensionalidade do Espaço das Variáveis através da projeção dos pontos representantes das colunas de  $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_J]$ , ou os pontos representantes das variáveis. Então será buscada uma projeção com a menor deformação possível destes pontos em um espaço de menor dimensão. Esta projeção é feita

---

sobre um sistema de coordenadas ortonormais cujos eixos tenham as direções de maior variabilidade.

O formato da nuvem de pontos é definida pelas distâncias mútuas de todos os  $J$  pontos, logo se queremos a projeção de menor deformação procura-se um sub-espço  $\mathcal{S}$ ,  $p^*$ - dimensional tal que a soma do quadrado das distâncias 2 a 2 de todos os pontos projetados sobre  $\mathcal{S}$  seja máxima, ou seja

$$\text{Max}_S \left\{ \sum_{h_i} \sum_{h_{i^*}} d^2(\mathbf{h}_i, \mathbf{h}_{i^*}) \right\} = \text{Max}_S \left\{ \sum_{h_i} \sum_{h_{i^*}} \langle \mathbf{h}_i - \mathbf{h}_{i^*} \rangle^T \langle \mathbf{h}_i - \mathbf{h}_{i^*} \rangle \right\} \text{ para todo } i \neq i^*.$$

Generalizando, suponha que estes pontos são afetados por uma massa  $\omega_i$  e  $\omega_{i^*}$  respectivamente, então buscaremos  $\mathcal{S}$  tal que:

$$\text{Max}_S \left\{ \sum_{h_i} \sum_{h_{i^*}} \omega_i \omega_{i^*} d^2(\mathbf{h}_i, \mathbf{h}_{i^*}) \right\}.$$

## 2.1- O Ajuste da Nuvem de Pontos em um Sub-Espaço Ótimo

Considere a matriz  $\mathbf{H}$  definida anteriormente onde as variáveis são afetadas por massas determinadas pela matriz diagonal  $\Phi$ ,  $J \times J$ , e onde  $\Omega$ ,  $I \times I$ , contém as massas de cada indivíduo. Desta forma  $\Omega^{-1}$ , ou o inverso das massas dos indivíduos, define a métrica em  $\mathcal{R}^I$ . O desenvolvimento partirá do ajuste de um sub-espço vetorial unidimensional, sendo este caso particularmente interessante na psicometria como será visto adiante.

Seja  $\mathbf{m}$  o vetor de ordem  $I \times 1$  que define a direção do sub-espço a ser encontrada, tal que  $\mathbf{m}^T \Omega^{-1} \mathbf{m} = 1$ . Seja  $\mathbf{h}_j$  o vetor representante da  $j$ -ésima variável, e  $h_j^*$  sua projeção na reta de ajuste (fig. 2.1). O comprimento da projeção  $h_j^*$  em relação à origem  $O$ , denotada por  $l(O, h_j^*)$ , é o produto escalar de  $\mathbf{h}_j$  por  $\Omega^{-1} \mathbf{m}^T$ , ou seja

$$l(O, h_j^*) = (\mathbf{h}_j^T \Omega^{-1})^T \mathbf{m} = \sum_{i=1}^I h_{ij} m_i \omega_i^{-1}. \quad (2.1.1)$$

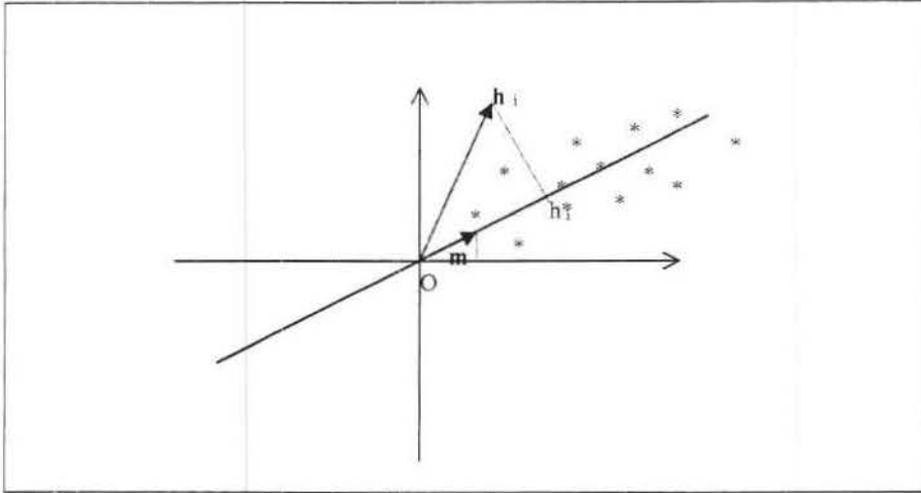


fig.. 2.1- Ajuste de uma reta no  $\mathbb{R}^2$

Qualquer uma das  $J$  colunas de  $\mathbf{H}$  é um vetor em  $\mathcal{R}^l$ , logo o produto escalar  $\mathbf{H}^T \Omega^{-1} \mathbf{m}^T$  definirá o vetor com as coordenadas das projeções de todas as variáveis sobre o sub-espço.

O ajuste do subespaço parte do princípio de Mínimos Quadrados, ou seja, se buscará a direção da maior variabilidade da nuvem de pontos, minimizando a soma ponderada dos quadrados das distâncias (ou desvios)  $d(\mathbf{h}_j, h_j^*)$ , ou seja,  $\text{Min}(\sum_{j=1}^J \phi_j d^2(\mathbf{h}_j, h_j^*))$ .

Aplicando o teorema de Pitágoras em cada um dos  $J$  triângulos retângulos contidos na nuvem temos a relação

$$\sum_j \phi_j d^2(\mathbf{h}_j, h_j^*) = \sum_j \phi_j d^2(O, \mathbf{h}_j) - \sum_j \phi_j d^2(h_j^*, O), \quad (2.1.2)$$

e uma vez que  $\sum_j \phi_j d^2(\mathbf{h}_j, O)$  independe de  $\mathbf{m}$  que queremos encontrar a minimização

citada é análoga à maximização apenas de  $\sum_j \phi_j d^2(h_j^*, O)$ , ou

$$\text{Max}(\mathbf{m})[\mathbf{m}^T \Omega^{-1} \mathbf{H} \Phi \mathbf{H}^T \Omega^{-1} \mathbf{m}^T] = \text{Max}(\mathbf{m}) [\mathbf{H} \Phi \mathbf{H}^T \Omega^{-1}] \quad (2.1.3)$$

sob a restrição e  $\mathbf{m}^T \Omega^{-1} \mathbf{m} = 1$ .

Para o ajuste de um subespaço de mais de uma dimensão realiza-se os mesmos procedimentos analíticos, sendo que há mais uma restrição, de que a base encontrada seja ortogonal. No caso de  $p^* = 2$  temos:

$$\begin{aligned} 1^\circ \text{ passo: } & \text{Max}_{(m_1)} [\mathbf{m}_1^T \Phi \mathbf{H} \Omega^{-1} \mathbf{H}^T \mathbf{m}_1] \\ & \mathbf{m}_1^T \Omega^{-1} \mathbf{m}_1 = 1 \end{aligned}$$

$$\begin{aligned} 2^\circ \text{ passo: } & \text{Max}_{(m_2)} [\mathbf{m}_2^T \Phi \mathbf{H} \Omega^{-1} \mathbf{H}^T \mathbf{m}_2] \quad (2.1.4) \\ & \mathbf{m}_2^T \Omega^{-1} \mathbf{m}_2 = 1 \\ & \mathbf{m}_1^T \mathbf{m}_2 = 0 \end{aligned}$$

e assim sucessivamente até se encontrar  $\mathbf{u}_{p^*}$  para o sub-espço  $p^*$ -dimensional.

É possível encontrarmos simultaneamente  $p^*$  soluções. Seja  $\mathbf{M}$ ,  $I \times P^*$ , matriz. Maximizar (2.1.3) é equivalente também a maximizar

$$\begin{aligned} \text{tr}[\mathbf{M}^T \Omega^{-1} \mathbf{H}^T \Phi \mathbf{H}^T \Omega^{-1} \mathbf{M}] \quad (2.1.5) \\ \mathbf{M}^T \Omega^{-1} \mathbf{M} = \mathbf{I} \end{aligned}$$

A Decomposição em Valores Singulares (DVS) fornece esta solução, como será visto adiante. O próximo item trata de uma medida de dispersão da nuvem e sua relação com o ajuste do subespaço.

## 2.2- A Inércia – Uma Medida de Variabilidade

Seja  $O$  a origem do espaço das variáveis. Podemos pensar em uma medida da dispersão da nuvem de pontos que seria a média ponderada pelos respectivos pesos de todos os desvios quadráticos dos pontos em relação à origem do espaço.

Se centralizarmos a matriz original  $\mathbf{H}$  pelo centro de massa da nuvem de pontos, ou seja

$$\ell \mathbf{H} = \mathbf{H} - \bar{\mathbf{H}}$$

onde as colunas de  $\bar{\mathbf{H}}$  são as médias ponderadas das variáveis teremos que  $O$  coincidirá com o centro de massa.

Então a expressão (2.1.5) define uma medida de variabilidade que será chamada de Inércia. Esta denominação é bastante utilizada pelos autores no caso da AC. No entanto a

escola francesa trata desta medida como sendo geral, inclusive em Componentes Principais (Lebart *et al.*, 1995). O conceito é análogo ao utilizado em física, onde a inércia de um ponto afetado por uma massa em relação a qualquer outro é o produto da distância ao quadrado entre estes pontos pela massa dos mesmos.

Então se  $O$  a origem do espaço, a inércia de um ponto coluna  $\mathbf{h}_j$  qualquer em relação ao centro será definida por

$$In_O(\mathbf{h}_j) = \phi_j d^2(\mathbf{h}_j, O) \quad (2.2.6).$$

Logo a inércia total de uma nuvem de pontos  $\mathbf{H}$  em relação ao centro de massa é dada por

$$In_O(\mathbf{H}) = \sum_{j=1}^J \phi_j d^2(\mathbf{h}_j, O) \quad (2.2.7).$$

A inércia tem uma interpretação análoga à de variância, outra medida de dispersão. Então faz sentido pensarmos em decompor a inércia total em componentes, de forma que sirvam de coeficientes comparáveis de variabilidade.

De forma matricial a inércia pode ser escrita como

$$In_O(\mathbf{H}) = tr[\Omega^{-1} \mathbf{H} \Phi \mathbf{H}^T] \quad (2.2.8).$$

Logo a maximização proposta em 2.1 tem uma interpretação estatística, ou seja, estamos encontrando o sub-espaço que concentra a maior parte da variabilidade total da nuvem de modo que as relações existentes sejam evidenciadas.

### 2.3- O Ajuste do Sub-Espaço Ótimo via DVS

Se está se buscando um sistema de coordenadas ortogonais com direções de máxima variabilidade, maximiza-se a inércia total da nuvem projetada no sub-espço  $\mathcal{S}$ ,  $p^*$ -dimensional. Então critério de maximização se torna

$$\text{Max}_{\mathcal{S}} [In_{\mathcal{O}}(\mathbf{H})] = \text{Max}_{\mathcal{S}} [\text{tr}\{(\Omega^{-1} \mathbf{H} \Phi \mathbf{H}^T)\}] = \text{Max}_{\mathcal{S}} [\text{tr}\{(\Omega^{-1/2} \mathbf{H} \Phi^{1/2})^T (\Omega^{-1/2} \mathbf{H} \Phi^{1/2})\}]$$

Fazendo

$$\mathbf{X} = (\Omega^{-1/2} \mathbf{H} \Phi^{1/2}) \quad \text{e} \quad \mathbf{R} = \mathbf{X} \mathbf{X}^T$$

temos

$$\text{Max}_{\mathcal{S}} [In_{\mathcal{O}}(\mathbf{H})] = \text{Max}_{\mathcal{S}} \{\text{tr}[\mathbf{X} \mathbf{X}^T]\} = \text{Max}_{\mathcal{S}} \{\text{tr}[\mathbf{R}]\}.$$

Encontrar o sub-espço ótimo então é equivalente a encontrarmos a matriz  $\mathbf{R}^*$  de posto  $p^* < p$ , onde  $p$  é o posto de  $\mathbf{R}$ . Então devemos minimizar a função  $\sigma(\mathbf{R}^*)$  onde

$$\sigma(\mathbf{R}^*) = \text{tr}\{\mathbf{R} - \mathbf{R}^*\} \quad (2.3.9)$$

Este problema foi tratado pela primeira vez por Eckart e Young (1936) onde os autores provam o teorema da “Aproximação de matrizes por uma de posto menor”. O enunciado e prova deste teorema está no Apêndice B. Gifi trata o problema de uma forma semelhante chamando a expressão em (2.3.9) de “função perda”, no sentido já apresentado de perda de informação no ajuste da nuvem original de dados em um sub-espço de menor dimensão.

Neste ponto utilizaremos o conceito de DVS generalizada, onde as matrizes resultantes obedecem à restrição de ortonormalização em relação às métricas das linhas e colunas das matrizes transformadas, e sua extensão para o teorema de Eckart e Young.

Queremos então representar esta matriz  $\mathbf{H}$  com um número menor de dimensões, ou representa-la por outra de posto menor de forma ótima.

**Teorema 2.1** (Decomposição em Valores Singulares Generalizada):

Seja  $\mathbf{H}_{I \times J} = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_J]$  a matriz das  $J$  variáveis transformadas representadas pelos vetores  $\mathbf{h}_j$ , com posto  $p$ . Sejam  $\Omega_{I \times I}$  e  $\Phi_{J \times J}$  matrizes simétricas, positivo-definidas. Então a matriz  $\mathbf{H}$  pode ser expressa como:

$$\mathbf{H} = \mathbf{N} \mathbf{D}_{\sqrt{\lambda}} \mathbf{M}^T = \sum_s \sqrt{\lambda_s} \mathbf{n}_s \mathbf{m}_s^T \quad s = 1, \dots, p \quad (2.3.10),$$

onde as colunas de  $\mathbf{N}_{I \times I}$  e  $\mathbf{M}_{J \times J}$  são ortonormalizadas com respeito a  $\Omega^{-1}$  e  $\Phi$  respectivamente, ou

$$\mathbf{M}^T \Omega^{-1} \mathbf{M} = \mathbf{N}^T \Phi \mathbf{N} = \mathbf{I} \quad (2.3.11)$$

, e  $\mathbf{D}_{\sqrt{\lambda}}$  é a matriz diagonal com os valores singulares, raízes quadradas do autovalores de  $\mathbf{H}\mathbf{H}^T$ , ordenados de forma decrescente .

**prova** : A prova é direta aplicando a DVS ordinária (Apêndice B) fazendo  $\mathbf{H} \equiv \Omega^{1/2} \mathbf{H} \Phi^{1/2}$ , utilizando-se o fato de que se  $\Omega$  tem DVS  $\Omega = \mathbf{W} \mathbf{D} \mathbf{W}^T$ , então  $\Omega^{1/2} = \mathbf{W} \mathbf{D}^{1/2} \mathbf{W}^T$ . Então:

$$\begin{aligned} \Omega^{-1/2} \mathbf{H} \Phi^{1/2} &= \mathbf{U} \mathbf{D}_{\lambda} \mathbf{V}^T, \text{ onde } \mathbf{U}\mathbf{U} = \mathbf{V}\mathbf{V} = \mathbf{I}. \\ \Rightarrow (\Omega^{1/2})^T \Omega^{-1/2} \mathbf{H} \Phi^{1/2} (\Phi^{-1/2})^T &= (\Omega^{1/2}) \mathbf{U} \mathbf{D}_{\lambda} (\Phi^{-1/2})^T \mathbf{V}^T \end{aligned}$$

Fazendo

$$\mathbf{N} \equiv \Omega^{1/2} \mathbf{U} \quad \text{e} \quad \mathbf{M} \equiv \Phi^{-1/2} \mathbf{V}$$

temos a decomposição definida em (2.3.10) e (2.3.11).

**Teorema 2.2** (Eckart e Young generalizado):

Se os  $p - p^*$  termos de (1.1) são desprezados, então

$$\mathbf{R}^* = \mathbf{X}^* \mathbf{X}^{*T} = \mathbf{N}^* \mathbf{D}_{\lambda}^* \mathbf{M}^{*T} = \sum_s^{p^*} \lambda_s \mathbf{n}_s \mathbf{m}_s^T \quad (2.3.11)$$

é a “aproximação por mínimos quadrados generalizados de ordem  $p^*$ ” que minimiza a função perda dada por

$$\sigma(\mathbf{H}^*) \equiv \text{tr}\{ (\Omega^{-1} (\mathbf{H} - \mathbf{X})^T \Phi (\mathbf{H} - \mathbf{X})) \} \quad (2.3.12)$$

entre todas as matrizes  $\mathbf{R}^*$  de posto  $p^*$ .

**prova:** A prova é consequência imediata do teorema de Eckart e Young (Apêndice B) à decomposição da matriz  $\mathbf{X} = \mathbf{\Omega}^{-1/2} \mathbf{H} \mathbf{\Phi}^{1/2}$ , como  $\mathbf{H}$ .

Se fixarmos uma dimensão  $s$ , com  $s = 1, \dots, p^*$ , temos a relação

$$\mathbf{X}\mathbf{X}^T \mathbf{m}_s = \lambda_s \mathbf{m}_s, \quad (2.4.13)$$

portanto o eixo fatorial<sup>1</sup> que define a  $s$ -ésima dimensão no espaço das variáveis é dado por

$$\boldsymbol{\psi}_s = \mathbf{m}_s \sqrt{\lambda_s} \quad (2.4.14)$$

e consequentemente as coordenadas das  $J$  variáveis no  $s$ -ésimo eixo fatorial são dadas por

$$\varphi_s = \sqrt{\lambda_s} \mathbf{X}^T \mathbf{m}_s = \mathbf{X}^T \boldsymbol{\psi}_s. \quad (2.4.15)$$

### 2.3.1- As relações entre os dois espaços

O desenvolvimento até o presente tópico foi feito sobre o ajuste de um sub-espaço do original  $\mathfrak{R}^J$ , ou seja, a representação da nuvem de pontos das  $J$  variáveis no espaço dos indivíduos. No entanto o desenvolvimento pode ser feito de forma inversa, encontrando um sub-espaço de  $\mathfrak{R}^J$  para se representar a nuvem de pontos dos indivíduos. Serão mostradas a seguir as relações entre estes dois subespaços.

De forma análoga a (2.3.13) temos para a  $s$  - ésima dimensão do sub-espaço encontrado a partir das colunas de  $\mathbf{H}$  temos:

$$\mathbf{X}^T \mathbf{X} \mathbf{n}_s = \mu_s \mathbf{n}_s \quad (2.4.16)$$

onde  $\mu_s$  é o  $s$ -ésimo auto vetor de  $\mathbf{X}^T \mathbf{X}$ . Consequentemente o eixo é definido por

$$\boldsymbol{\varphi}_s = \mathbf{n}_s \sqrt{\mu_s}.$$

Multiplicando-se (2.4.14) por  $\mathbf{X}$  nos dois termos temos

$$(\mathbf{X}\mathbf{X}^T) \mathbf{X} \mathbf{n}_s = \mu_s \mathbf{X} \mathbf{n}_s \quad (2.4.17).$$

---

<sup>1</sup> Os eixos do espaço de representação, neste caso o Espaço das Variáveis, serão chamados de Eixos Fatoriais ou simplesmente Fatores, tomando a nomenclatura utilizada pela escola francesa (Lebart, 1995).

---

A relação (2.4.15) diz que para todo auto-vetor  $\mathbf{n}_s$  de  $\mathbf{X}^T \mathbf{X}$  existe um auto-vetor  $\mathbf{Xn}_s$  de  $\mathbf{XX}^T$  correspondente ao mesmo auto-valor  $\mu_s$ . Fazendo  $\lambda_1$  o maior auto-valor não trivial de  $\mathbf{XX}^T$  temos que necessariamente  $\mu_1 \leq \lambda_1$ .

Se procedermos da mesma forma com (2.3.13) chegaremos em  $\lambda_1 \leq \mu_1$ , logo são iguais. Daí as fórmulas de transição

$$\begin{aligned} \mathbf{m}_s &= \frac{1}{\sqrt{\lambda_s}} \mathbf{Xn}_s \\ \mathbf{n}_s &= \frac{1}{\sqrt{\lambda_s}} \mathbf{X}^T \mathbf{m}_s \end{aligned} \quad (2.4.18)$$

e obtêm-se o  $s$ -ésimo eixo fatorial

$$\boldsymbol{\varphi}_s = \mathbf{n}_s \sqrt{\lambda_s} \quad (2.4.19)$$

e as coordenadas dos indivíduos neste eixo fatorial por

$$\psi_s = \mathbf{X n}_s \sqrt{\lambda_s} = \mathbf{X} \boldsymbol{\varphi}_s \quad (2.4.20).$$

Desta forma temos definida uma generalização das técnicas multivariadas baseadas em DVS para a redução de dimensionalidade. As técnicas variam de acordo com as matrizes adotadas para os pesos e as transformações na matriz original.

## 2.4- Comentário Final

Podemos então definir estes métodos em três passos (Greenacre, 1984):

1. Transformar a matriz original de dados. Estas transformações são em geral centralizações e/ou normalizações das observações das variáveis envolvidas na busca de uma representação adequada dos dados. Esta é a matriz  $\mathbf{X}$ ;
2. Calcular a DVS para  $\mathbf{X}$  escolhendo-se apropriadamente  $\Omega$  e  $\Phi$ , selecionando-se as  $p^*$  dimensões de interesse.
3. Escolher a melhor representação em termos gráficos das linhas e/ou colunas (indivíduos e/ou variáveis) através de

$$\Psi \equiv \mathbf{N}^* \mathbf{D}_{\lambda^*} \quad \text{e/ou} \quad \vartheta \equiv \mathbf{M}^* \mathbf{D}_{\lambda^*}$$

para o hiperplano  $p^*$ - dimensional.

São inúmeras as técnicas que se encaixam no esquema proposto. Entre elas Componentes Principais (nas mais diversas formas), Escalonamento Multidimensional, Médias Recíprocas, Escalonamento Dual, Análise de Correspondência, etc... .

O próximo capítulo trata da Análise de Correspondência Múltipla (ACM) através deste enfoque. No Apêndice A é introduzida a Análise de Componentes Principais como solução da Análise Fatorial, muito utilizada nos problemas de psicometria.

---

### 3. Análise de Correspondência Múltipla

As técnicas de Análise de Correspondência, por exemplo Simples, Múltipla, em Grupos, (Greenacre 1984, Benzécri 1973, Pamplona 1998) são conhecidas por seu poder em análise exploratória através de gráficos que representam projeções ótimas, de espaços multidimensionais de ordens elevadas em um número reduzido de dimensões, definidas pelos “eixos principais” na concepção de Pearson (1901). Esta representação permite que sejam analisadas inter-relações entre linhas e colunas de uma matriz de dados, onde as interpretações sobre as posições relativas entre os pontos das linhas e/ou colunas dependem justamente da estrutura desta tabela. No entanto esta técnica permite enfoques alternativos inclusive de aspecto inferencial (De Leeuw 1984).

Tendo a psicometria como escopo deste trabalho, a técnica de Análise de Correspondência Múltipla (ACM) além das inter-relações citadas, explora a estrutura fatorial, propriedades de escala e conseqüente classificação de indivíduos através do instrumento estudado. O problema pode ser pensado como a busca de uma quantificação ou escala ótima para as categorias das variáveis qualitativas, e através destas se encontrar escores dos indivíduos induzidos por esta escala (Gifi, 1990). Desta forma estudam-se evidências do comportamento do instrumento sobre a população estudada, desde que a origem e representatividade destes dados permitam estas conclusões.

Os primeiros trabalhos de DVS em matrizes indicadoras, ainda sem a denominação de Análise de Correspondência Múltipla podem ser atribuídos a Guttman(1941), Burt

---

---

(1950), Hayashi (1950, 1956), e seu posterior desenvolvimento a Carroll (1968) e Benzécri (1973).

A escola francesa, tendo Benzécri como expoente, é notadamente a que mais se refere e aplica esta técnica, incluída em uma classe de técnicas chamadas de Métodos Fatoriais (*Méthodes Factorielles*)<sup>2</sup>, dando um enfoque quase estritamente geométrico. Deve-se a este grupo também o nome Análise de Correspondência Múltipla (*Analyse de Correspondences Multiples*).

A escola holandesa, tendo Gifi como maior representante, e outros como De Leeuw e Van Rijckevorsel, apresentam a técnica como pertencente a classe de Análise de Homogeneidade (AH), onde se inclui também técnicas generalizadas de Componentes Principais. A AH aplicada a matriz indicadora equivale à Análise de Correspondência Múltipla (ACM), no entanto o enfoque é de escalonamento. Este mesmo enfoque é dado por Weller e Rommey (1990), onde também recebe o nome de ACM a aplicação da Análise de Correspondência Simples (ACS) em uma justaposição de tabelas, sejam linhas e/ou colunas justapostas. No entanto os resultados desta análise e sua interpretação diferem muito da aplicação sobre a matriz indicadora, como os próprios autores indicam.

As referências citadas trazem um histórico mais detalhado desta técnica e sua relação com as demais técnicas multivariadas. Nas seções seguintes serão desenvolvidas a teoria, álgebra e resultados da ACM, bem como a forma de sua interpretação na psicometria.

---

<sup>2</sup> Não se deve confundir com a Análise Fatorial conforme descrita no Apêndice B, e proposta por Thurstone (1938) por exemplo.

---

### 3.1- A matriz indicadora e suas propriedades

Seja a matriz  $\mathbf{H}$  como definida no capítulo 2, contendo apenas variáveis qualitativas onde são observadas nos  $I$  indivíduos as respostas às  $J$  variáveis codificadas numericamente. Definiremos uma matriz de códigos binários sobre a qual será aplicada a ACM. A utilização desta matriz indicadora, também chamada de disjuntiva completa (Crivisqui, 1996), traz uma série de vantagens para o tipo de dados e análise discutidos neste trabalho. Esta codificação não utiliza a codificação numérica das categorias em  $\mathbf{H}$  como uma variável contínua de forma que podemos estudar e definir as ponderações de forma conveniente.

**Definição 3.1** Seja  $\mathbf{H} = [\mathbf{h}_1 \dots \mathbf{h}_J]$  a matriz das respostas observadas das  $J$  variáveis categóricas. Então definiremos  $\mathbf{G}_j, I \times k_j$  como a matriz indicadora da variável categórica  $\mathbf{h}_j$  que possui  $k_j$  categorias de resposta, a  $j$ -ésima coluna de  $\mathbf{H}$ . Logo  $\mathbf{G}_j = \{ g_{(j)ir} \}$  onde:

$$g_{(j)ir} = \begin{cases} 1 & \text{se o } i\text{-ésimo indivíduo apresentou a } r\text{-ésima categoria de resposta da } j\text{-ésima variável} \\ 0 & \text{caso contrário} \end{cases}$$

Tomando as  $J$  variáveis de  $\mathbf{H}$  podemos definir uma matriz  $\mathbf{G} = [\mathbf{G}_1 \dots \mathbf{G}_J]$  de dimensão  $I \times k = \sum_{j=1}^J k_j$  que representa  $\mathbf{H}$  em sua totalidade.

A matriz  $\mathbf{G}_j$  é dita completa se todas as células são completas, ou se apresentam 0 ou 1, sem dados faltantes. Desta forma as linhas de  $\mathbf{G}_j$  somam 1, ou  $\mathbf{G}_j \mathbf{1}_{k_j} = \mathbf{1}_I$ , onde  $\mathbf{1}_n$  é um vetor coluna com  $n$  elementos iguais a 1. Dado que todas as  $\mathbf{G}_j$  são completas dizemos então que  $\mathbf{G}$  é completa, e então  $\mathbf{G} \mathbf{1}_k = J \mathbf{1}_I$ , ou as linhas somam  $J$ <sup>3</sup>.

---

<sup>3</sup> A falta de dados não será discutida neste trabalho por tratar-se de um assunto extenso, dado o grande número de enfoques para a estimação das caselas faltantes. Referências para este assunto estão em Gifi (1990), Greenacre (1984), Lebart (1995). Então nas definições e discussões subentende-se que as matrizes são todas completas. Quando presentes serão tratados como categorias ilustrativas, como visto na seção 3.3.8 e aplicado no capítulo 4.

### 3.1.1 – As Nuvens de Pontos

Definiremos agora as nuvens de pontos em seus respectivos espaços de representação. A partir destas definições será possível chegar à matriz que se pretende decompor. Os passos e a álgebra utilizada no caso da ACM na matriz indicadora são análogos aos realizados na ACS, como se a matriz indicadora fosse uma tabela de dupla entrada. Logo algumas interpretações possíveis na ACS tornam-se sem sentido no caso da ACM devido às características da matriz estudada. Estas diferenças serão apontadas e apresentam-se também interpretações plausíveis.

**Definição 3.2** Seja  $\mathbf{F} = \frac{1}{IJ} \mathbf{G} = \left\{ \frac{g_{(j)ir}}{IJ} \right\}$  para todo  $j=1, \dots, J$ ;  $r=1, \dots, k_j$  e  $i=1, \dots, I$ . Esta é chamada a Matriz de Correspondência.

Na Análise de Correspondência Simples (ACS) de uma tabela de contingência esta matriz representaria a distribuição conjunta das variáveis de linha e coluna da tabela, ou as frequências relativas ao total  $IJ$  em cada casela. No caso da matriz indicadora perde-se esta interpretabilidade uma vez que não estamos trabalhando com frequências, e sim com um código binário. Desta forma, a matriz  $\mathbf{F}$  torna-se necessária apenas para efeito de ajuste das nuvens de pontos, e como uma simplificação matemática, como será visto adiante.

Foi visto anteriormente que as linhas de  $\mathbf{G}$  somam  $J$  para qualquer um dos  $I$  indivíduos. Desta forma, em  $\mathbf{F}$  estas linhas somarão  $I^{-1}$ , pois se  $\mathbf{F} = (IJ)^{-1} \mathbf{G}$  então  $\mathbf{F} \mathbf{1}_k = (IJ)^{-1} \mathbf{G} \mathbf{1}_k = (IJ)^{-1} (J \mathbf{1}_I) = I^{-1} \mathbf{1}_I$ .

**Definição 3.3** A métrica do espaço  $I$  – dimensional dos indivíduos é uniforme, definida pela matriz diagonal  $\mathbf{D}_I = 1/I \mathbf{I}_I$ , que definirá a representação das  $k = \sum_{j=1}^J k_j$  categorias das  $J$  variáveis neste espaço.

Portanto a métrica do espaço dos indivíduos é uniforme, sendo que esta apenas “encolhe” ou padroniza as distâncias entre os pontos das colunas em relação ao total de indivíduos.

Tendo definido a métrica do espaço dos indivíduos cabe verificar as propriedades do espaço  $k$  – dimensional das categorias.

**Definição 3.4** Seja  $\mathbf{g}_j = \{g_{(j)r}\} = \left\{ \sum_{i=1}^I g_{(j)ir} \right\}$  o vetor  $k_j \times 1$  dos totais das colunas de  $\mathbf{G}_j$ , ou seja, a marginal de frequência total da variável  $\mathbf{h}_j$  tal que  $\mathbf{1}_I^T \mathbf{g}_j = I$  para todo  $j$ . Então  $\mathbf{D}_j = \text{diag}(\mathbf{g}_j)$  é a matriz diagonal com esta marginal total de coluna, para  $j = 1, \dots, J$ .

**Definição 3.5** Seja  $\mathbf{B} = \mathbf{G}^T \mathbf{G}$ ,  $k \times k$ , a matriz que contém todos os cruzamentos possíveis  $2 \times 2$  entre as  $J$  variáveis envolvidas. Esta matriz também é chamada de “Tabela de Burt”. Então  $\mathbf{D}_j = \mathbf{B}_{jj}$  é a matriz diagonal que contém as marginais de coluna de  $j$  e

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_J \end{bmatrix}. \quad (3.1.1)$$

é a matriz que contém todas as marginais totais das colunas.

**Definição 3.5** Então a matriz  $\mathbf{D}_k = (\mathbf{I}J)^{-1} \mathbf{D}$  define a métrica do espaço  $k$  – dimensional que definirá a representação dos  $I$  pontos da nuvem dos indivíduos. A matriz  $\mathbf{D}_k$  define as massas que afetam os pontos das  $k$  categorias (colunas) das  $J$  variáveis.

**Definição 3.6** Os perfis de linha da matriz indicadora  $\mathbf{G}$  são dados por

$$\mathbf{R} = \mathbf{D}_I^{-1} \mathbf{F} = 1/J \{g_{(j)ir}\} \quad (3.1.2)$$

enquanto os perfis de coluna são dados por

$$\mathbf{C} = \mathbf{D}_k^{-1} \mathbf{F}^T = \mathbf{D}_k^{-1} [\mathbf{F}_1 \dots \mathbf{F}_k] = [\mathbf{c}_{(1)l} \dots \mathbf{c}_{(J)k}] \quad \text{onde } \mathbf{c}_{(j)r} = \{g_{(j)ir} / g_{(j),r}\} \quad (3.1.3)$$

para todo  $j = 1, \dots, J$ ;  $r = 1, \dots, k_j$  e  $i = 1, \dots, I$ .

Os perfis definidos em 3.5 são as representações das nuvens de pontos das  $I$  linhas no  $\mathfrak{R}^k$  e das  $k$  colunas (categorias) no espaço  $\mathfrak{R}^I$  respectivamente. Novamente fazendo um paralelo com a ACS, estes perfis seriam interpretados como as distribuições da tabela condicionadas nas variáveis de linha e nas variáveis de colunas respectivamente, ou as frequências relativas das linhas e colunas respectivamente. No caso da ACM também se

perde esta interpretação direta. Os perfis linhas são uma “recodificação” da matriz  $\mathbf{G}$  original, onde todos elementos que não são 0 são iguais a  $J^{-1}$ , não há uma interpretação útil nestes perfis, servindo apenas como artifício para o cálculo dos eixos principais. Os perfis colunas também não possuem a mesma interpretação, mas já podem ser considerados como uma primeira quantificação das categorias das variáveis qualitativas, uma vez que os valores são os pesos relativos de cada categoria dentro da respectiva variável.

Desta forma a definição da matriz a ser decomposta, e o ajuste do subespaço ótimo será desenvolvido a partir dos perfis colunas da matriz original, que pode ser interpretada como a busca pela escala ótima, ou a melhor quantificação para as variáveis qualitativas. O problema é simétrico, e se chegaria aos mesmos resultados através dos perfis linhas, mas optou-se por este caminho pelo melhor apelo didático à proposta deste trabalho.

Fazendo  $\mathbf{C}_j^T = [\mathbf{c}_{(j)1}, \dots, \mathbf{c}_{(j)k_j}]$ , a matriz dos perfis coluna da variável  $j$  temos que

$$\mathbf{1}_{k_j}^T \mathbf{C}_j \mathbf{1}_{k_j} = \sum_{r=1}^{k_j} \sum_{i=1}^I \frac{g_{(j)ir}}{g_{(j),r}} = \frac{1}{J},$$

ou seja, cada variável  $j$  é afetada pela massa  $1/J$ , distribuída em suas categorias. Dada a dependência causada pela soma fixa, as categorias de  $j$  ocupam um sub-espaço de no máximo  $k_j - 1$  dimensões. Logo a representação conjunta das  $J$  variáveis ocupará um espaço de no máximo  $\sum_{j=1}^J (k_j - 1) = K - J$  dimensões.

### 3.1.2- O Referencial da Nuvem de Pontos

Definiremos neste ponto um referencial, com papel de ponto médio, ou o centro de massa, que desempenhará papel importante no ajuste do subespaço através de critérios de distância, como será visto no próximo tópico.

Cada elemento do centro de massa é dada pela média ponderada pelas massas das colunas das respostas da respectiva categoria da variável  $j$ .

Seja  $\mathbf{d}_j^T = (\mathbf{C}_j \mathbf{1}_{k_j})^T = (IJ)^{-1} [g_{0j1} \dots, g_{0jk_j}]$ , ou seja, o vetor  $k_j \times 1$  contendo as massas das categorias de cada categoria da variável  $j$ , contidas na diagonal de  $\mathbf{D}_k$  definida em 3.5. Logo  $\mathbf{d}_j^T \mathbf{1}_{k_j} = J^j$ . Então em linguagem matricial, fixando-se  $j$  teremos:

$$\bar{\mathbf{C}} = \frac{\mathbf{G}_j \mathbf{d}_j}{\mathbf{d}_j^T \mathbf{1}_{k_j}} = \frac{(IJ)^{-1} \mathbf{1}_{k_j}}{J^{-1}} = \frac{1}{I} \mathbf{1}_{k_j} \quad (3.1.3)$$

para qualquer  $j$ . Então o centróide das colunas será denotado por  $\bar{\mathbf{C}} = [\bar{\mathbf{C}}_1 \dots \bar{\mathbf{C}}_j]$ , e é o mesmo para qualquer uma das categorias das variáveis qualitativas uma vez que o resultado independe de  $j$  e  $k_j$ .

Tendo definido este centro de massa, podemos então centralizar os perfis coluna de  $\mathbf{G}$ , ditos  $\mathbf{C}$  de forma que facilite a interpretação e o estudo da matriz conforme visto no capítulo 2. Então estaremos trabalhando com os perfis centralizados  $\ell\mathbf{C} = \mathbf{C} - \bar{\mathbf{C}}$ .

### 3.1.3- As distâncias em $\mathbf{G}$

Conforme visto na Análise Generalizada (cap. 2) a matriz a ser decomposta com o objetivo de redução dimensional é definida a partir das distâncias entre os pontos das nuvens projetadas. No entanto a idéia de distância entre pontos de uma matriz indicadora parece pouco intuitiva, uma vez que não temos frequências, mas codificações.

Uma forma mais plausível de se apresentar este problema é através da idéia de dissimilaridade. Os pontos existentes em  $\mathbf{G}$ , sejam de linhas ou colunas, são na verdade

padrões de respostas binárias, logo é muito mais intuitivo pensar em padrões dissimilares do que em padrões distantes, no sentido geométrico.

Mas matematicamente estes dois conceitos podem ser equivalentes, uma vez que distância é reconhecidamente uma medida de dissimilaridade, dentre muitas outras, e o desenvolvimento será feito através da distância Euclidiana ponderada. Será evitada aqui a denominação de “Distância Qui-quadrado” como no caso da ACS, uma vez que este termo remete à análise de frequências em tabelas de contingência, porém os termos distância e dissimilaridade serão utilizados como sinônimos neste caso.

A distância entre duas categorias no espaço, ou a dissimilaridade entre dois padrões de resposta, ou ainda entre dois perfis coluna centralizados é dada por

$$d^2(\mathbf{c}_{(j)r}, \mathbf{c}_{(j)r'}) = \sum_{i=1}^I I \left( \frac{g_{(j)ir}}{g_{.r}} - \frac{g_{(j)ir'}}{g_{.r'}} \right)^2 = I \langle \ell \mathbf{C}_j - \ell \mathbf{C}_{j'} \rangle^T \langle \ell \mathbf{C}_j - \ell \mathbf{C}_{j'} \rangle \quad (3.1.4)$$

para  $r \neq r'$  e para quaisquer  $j$  e  $j'$ , incluindo  $j=j'$ .

Desenvolvendo o termo quadrático em 3.1.4 temos

$$d^2(\mathbf{c}_{(j)r}, \mathbf{c}_{(j)r'}) = \sum_{i=1}^I I \left( \frac{g_{(j)ir}}{g_{.r}} - \frac{g_{(j)ir'}}{g_{.r'}} \right)^2 = \sum_{i=1}^I I \left( \frac{g_{(j)ir}^2}{g_{.r}^2} + \frac{g_{(j)ir'}^2}{g_{.r'}^2} - 2 \frac{g_{(j)ir}}{g_{.r}} \frac{g_{(j)ir'}}{g_{.r'}} \right) \quad (3.1.5)$$

e os seguintes resultados:

Se  $g_{rr'}$  é o número de indivíduos que apresentam as categorias  $r$  e  $r'$ , para  $j \neq j'$ , o valor do termo entre parênteses em (3.1.5) pode ser obtido pela soma das seguintes parcelas (Crivisqui, 1997):

- 1)  $\frac{1}{g_{(j),r}^2}$  para os  $(g_{(j),r} - g_{rr'})$  indivíduos que apresentam somente a categoria  $(j)r$ ;
- 2)  $\frac{1}{g_{(j),r'}^2}$  para os  $(g_{(j),r'} - g_{rr'})$  indivíduos que apresentam somente a categoria  $(j')r'$ ;
- 3)  $\frac{1}{g_{(j),r}^2} + \frac{1}{g_{(j),r'}^2} + \frac{2}{g_{(j),r} g_{(j),r'}}$  para os  $g_{rr'}$  indivíduos que apresentam ambas categorias simultaneamente;
- 4) 0, para os indivíduos que não apresentam nenhuma das categorias envolvidas.

Substituindo a somatória em (3.1.5) pelos valores 1) a 4) temos uma expressão fechada para a distância entre duas categorias quaisquer dada por:

$$d^2(\mathbf{c}_{(j)r}, \mathbf{c}_{(j)r'}) = \sum_{i=1}^I I \left( \frac{g_{(j)ir}}{g_{.r}} - \frac{g_{(j)ir'}}{g_{.r'}} \right)^2 = I \left[ \left( \frac{g_{(j)ir} - g_{rr'}}{g_{(j).r} g_{(j).r'}} \right) + \left( \frac{g_{(j)ir'} - g_{rr'}}{g_{(j).r} g_{(j).r'}} \right) \right] \quad (3.1.6)$$

A interpretação de (3.1.6) como distância, em termos geométricos, torna-se realmente pouco intuitiva, mas pode se dizer que a dissimilaridade entre as categorias representadas pelos perfis  $\mathbf{c}_{(j)r}$  e  $\mathbf{c}_{(j)r'}$  aumenta na medida em que os indivíduos apresentam apenas uma das duas, e é inversamente proporcional à importância relativa de cada categoria no instrumento inteiro, ou seja, se ambas possuem a mesma importância relativa (leia-se frequências relativas de magnitudes próximas) então têm padrões de resposta similares.

Se  $j=j'$  então a dissimilaridade será entre categorias de uma mesma variável, e estaremos identificando a métrica verdadeira da variável qualitativa ordinal. Mas se  $j \neq j'$  então estaremos realmente analisando a dissimilaridade entre os padrões, principalmente se as duas categorias das duas variáveis representam intensidades de resposta semelhantes. Espera-se que caso haja uma estrutura latente, as categorias de mesma ordem de intensidade serão similares quanto ao perfil de resposta dos indivíduos.

### 3.2 – A Matriz Transformada a ser Aproximada

Apesar de possuírem o mesmo centro de massa, os perfis coluna, ou seja a representação das categorias em  $\mathbf{G}$ , são afetadas por massas diferentes que se não consideradas podem levar a uma deformação na projeção da nuvem de pontos.

Uma vez determinada a distância, como visto no capítulo 2 o que se quer é o sub-espço que maximize as distâncias entre os pontos e que passa pela origem  $O$ , no caso representada pelo centro de gravidade  $\bar{\mathbf{C}}$  definido anteriormente, ou seja,

$$\text{Max}_{\mathbf{V}_{p^*}} \left( \sum_{j=1}^J \sum_{r=1}^{k_j} \frac{g_{(j)r}}{IJ} d^2(\mathbf{c}_{(j)r}, \bar{\mathbf{c}}) \right) = \text{Max}_{\mathbf{V}_{p^*}} (\mathbf{D}_k \ell \mathbf{C} \mathbf{D}_I^{-1} \ell \mathbf{C}^T) \quad (3.2.7)$$

onde  $\mathbf{V}_{p^*}$  é a matriz dos vetores com direção de máxima variabilidade, ou seja, é a matriz dos autovetores associados aos  $p^*$  maiores autovalores  $\lambda$ , não triviais (diferente de 1),  $\mathbf{D}_I$  e  $\mathbf{D}_k$  de definidas em 3.3 e 3.5 respectivamente.

Então a matriz que deve ser reduzida, conforme visto na sessão 2.3 é dada por

$$\begin{aligned} \mathbf{R} &= \mathbf{V}_{p^*} \mathbf{D}_I^{-1} \ell \mathbf{C} \mathbf{D}_k \ell \mathbf{C}^T \mathbf{D}_I^{-1} \mathbf{V}_{p^*}^T = \mathbf{V}_{p^*} \mathbf{D}_I^{-1} \ell (\mathbf{F} \mathbf{D}_k^{-1}) \mathbf{D}_k \ell (\mathbf{F} \mathbf{D}_k^{-1})^T \mathbf{D}_I^{-1} \mathbf{V}_{p^*}^T = \\ &= \mathbf{V}_{p^*} \mathbf{D}_I^{-1} \ell \mathbf{F} \mathbf{D}_k^{-1} \ell (\mathbf{F} \mathbf{D}_k^{-1})^T \mathbf{D}_I^{-1} \mathbf{V}_{p^*}^T = \ell \mathbf{F} \mathbf{D}_k^{-1} \ell (\mathbf{F} \mathbf{D}_k^{-1})^T \mathbf{D}_I^{-1} = \frac{1}{J} \ell \mathbf{G}^T \ell \mathbf{G} \mathbf{D}^{-1} \end{aligned} \quad (3.2.8)$$

de termo geral

$$R_{rr^*} = \frac{1}{J g_{(j)r^*}} \sum_{i=1}^I g_{(j)ir} g_{(j)ir^*} \quad (3.2.9)$$

Então, seguindo a notação da generalização (cap. 2) temos que a função perda a ser minimizada é dada por

$$\sigma(\mathbf{R}^*) = \text{tr}\{(\mathbf{R} - \mathbf{R}^*)(\mathbf{R} - \mathbf{R}^*)^T\} \quad (3.2.10)$$

Logo  $\mathbf{R}^* = \mathbf{V}^* \mathbf{D}_{\lambda^*} \mathbf{V}^{*T}$  é a melhor aproximação de posto  $p^*$  de  $\mathbf{R}$ . Em  $\mathfrak{R}^k$  temos que a equação que define o  $s$ -ésimo autovetor  $\mathbf{v}_s$  é

$$\frac{1}{J} \mathbf{G}^T \mathbf{G} \mathbf{D}^{-1} \mathbf{v}_s = \lambda_s \mathbf{v}_s. \quad (3.2.11)$$

As coordenadas dos perfis coluna sobre  $\mathbf{v}_s$ , definidas por  $\varphi_s = \mathbf{D}^{-1} \mathbf{v}_s$  fornecem então a equação

$$J^{-1} \mathbf{D}^{-1} \mathbf{G}^T \mathbf{G} \varphi_s = \lambda_s \varphi_s \quad (3.2.12)$$

De forma análoga, a equação em  $\mathfrak{R}^l$  das coordenadas  $\psi_s$  no  $s$ -ésimo eixo dos pontos de perfis coluna é

$$J^{-1} \mathbf{G} \mathbf{D}^{-1} \mathbf{G}^T \psi_s = \lambda_s \psi_s \quad (3.2.13)$$

fornecendo as relações de transição

$$\begin{cases} \varphi_s = \frac{1}{\sqrt{\lambda_s}} \mathbf{D}^{-1} \mathbf{G}^T \psi_s \\ \psi_s = \frac{1}{J \sqrt{\lambda_s}} \mathbf{G} \varphi_s \end{cases} \quad (3.2.14)$$

Desta forma as coordenadas do  $i$ -ésimo indivíduo no  $s$ -ésimo eixo são

$$\psi_{si} = \frac{1}{\sqrt{\lambda_s}} \sum_{j=1}^J \sum_{r=1}^{k_j} \frac{g_{(j)ir}}{g_{(j)i.}} \varphi_{s(j)r} \quad (3.2.15)$$

e as coordenadas da  $r$ -ésima categoria da variável  $j$ , ou seja, o valor desta categoria na escala otimizada são

$$\varphi_{s(j)r} = \frac{1}{J \sqrt{\lambda_s}} \sum_{i=1}^l g_{(j)ir} \psi_{si} \quad (3.2.16)$$

Neste ponto temos definida a técnica até as expressões que definem a representação simultânea das linhas e colunas de  $\mathbf{G}$ . A seguir serão apresentados resultados que auxiliam a interpretação.

### 3.3- A Interpretação

A partir da decomposição apresentada até este ponto parte-se para a análise e interpretação dos resultados. Os conceitos envolvidos na Decomposição em Valores Singulares, como o Princípio Baricêntrico e as coordenadas das categorias; medidas resumo como as contribuições e coeficientes de discriminação e resultados gráficos na forma de mapas de representação simultânea, por exemplo, são as ferramentas que a Análise de Correspondência Múltipla fornece para este propósito.

Além destes é apresentada uma relação entre a ACM e a Análise Fatorial solucionada através dos Componentes Principais que permite uma interpretação e estudo em nível de construto, análogo à forma clássica, que permite uma comparação direta com o enfoque explorado neste trabalho.

#### 3.3.1- O Princípio Baricêntrico

Nas expressões (3.2.15) e (3.2.16), bem como as relações de transição em (3.2.14) pode-se notar um fato interessante. A menos do coeficiente de dilatação  $\frac{1}{\sqrt{\lambda_s}}$  a coordenada do  $i$ -ésimo indivíduo no  $s$ -ésimo eixo é exatamente o ponto médio da nuvem de pontos das categorias que escolheu, ou seja, é a média ponderada dos valores da escala ótima das categorias que escolheu. Simetricamente, o valor de  $r$ -ésima categoria qualquer é a média aritmética das coordenadas dos indivíduos que a escolheu. Apesar da simetria a primeira relação tem maior apelo no tipo de problema estudado.

A ponderação em (3.2.15) é dada pelos totais de cada categoria, portanto a massa das categorias que o indivíduo escolheu no instrumento como um todo. Espera-se então que o agrupamento e posterior classificação dos indivíduos via ACM seja mais eficiente que aquela ocorrida no caso da ACP, uma vez que este desequilíbrio não é respeitado nesta última técnica, onde todas as variáveis têm mesmo peso.

Alguns autores como Gifi (1990) apresentam a técnica a partir desta relação, ou seja, ele parte das relações baricênicas para definir a sua função perda e utiliza as

propriedades de otimização da DVS para encontrar o valor mínimo desta função através de um algoritmo de Mínimos Quadrados Alternados (*Alternating Least Squares*).

### 3.3.2- A Escala Ótima

Nas aplicações de estudo das propriedades psicométricas de instrumentos de pontos somados, como no caso da EL, é extremamente importante conhecermos a quantificação das categorias no primeiro eixo.

As coordenadas das categorias no primeiro eixo podem ser interpretadas como uma quantificação ótima da escala ordinal de resposta. Desta forma o vetor  $\varphi_1^T = [\varphi_{(1)1}, \dots, \varphi_{(j)1}]$ , como definido em (3.2.14), pode ser interpretado como a nova escala de resposta ordinal das  $J$  variáveis otimizada.

O gráfico de um  $\varphi_{(j)1}$  contra a escala de resposta original é uma boa ferramenta de análise do comportamento do item  $j$ , como será visto nas aplicações.

### 3.3.3- A Inércia na ACM

A distância entre uma categoria  $\mathbf{c}_{(j)r}$  qualquer e o centro de gravidade  $\bar{\mathbf{c}}$  é dada por

$$d^2(\mathbf{c}_{(j)r}, \bar{\mathbf{c}}) = \sum_{i=1}^I I \left( \frac{g_{(j)ir}}{g_r} - \frac{1}{I} \right)^2 = \frac{I}{g_{(j)r}} - 1 \quad (3.3.17)$$

Note que (3.3.17) é o inverso do efetivo na respectiva categoria sobre o total de indivíduos subtraindo 1. Portanto esta distância será maior quanto menor for o efetivo nesta categoria, tendo seu máximo, no ponto de vista prático, quando  $g_{(j)r}$  for igual a 1 e não definido quando o efetivo é zero. Logo uma categoria rara pode trazer conseqüências na interpretação dos resultados.

### A inércia de uma categoria

Conforme a definição de inércia em (2.2.6) a inércia de uma categoria é dada por

$$In_{\bar{c}}(\mathbf{c}_{(j)r}) = \frac{g_{(j)r}}{IJ} d^2(\mathbf{c}_{(j)r}, \bar{\mathbf{c}}) = \frac{g_{(j)r}}{IJ} \left( \frac{I}{g_{(j)r}} - 1 \right) = \frac{1}{J} \left( 1 - \frac{g_{(j)r}}{I} \right) \quad (3.3.18)$$

logo, proporcionalmente à sua distância ao centro de massa em (3.3.17), a inércia de uma categoria será grande quando a frequência for pequena. O máximo  $1/J$  será atendido quando  $g_{(j)r}$  for 0. Como consequência as categorias raras podem ter grande importância no ajuste do subespaço de uma maneira não conveniente. No tratamento destas categorias agrupa-se esta com outra categoria ou torna-a ilustrativa, conceito que será visto na sessão 3.3.8, de modo que as categorias de frequência muito baixa não interfiram na direção dos eixos que se busca.

### A inércia de uma questão

A inércia de uma questão é dada pela soma das inércias de suas categorias como apresentado em (3.3.18), ou seja:

$$In_{\bar{c}}(\mathbf{c}_{(j)}) = \sum_{r=1}^{k_j} \frac{g_{(j)r}}{IJ} d^2(\mathbf{c}_{(j)r}, \bar{\mathbf{c}}) = \sum_{r=1}^{k_j} \frac{g_{(j)r}}{IJ} \left( \frac{I}{g_{(j)r}} - 1 \right) = \sum_{r=1}^{k_j} \frac{1}{J} \left( 1 - \frac{g_{(j)r}}{I} \right) = \frac{1}{J} (k_j - 1) \quad (3.3.19)$$

a inércia de uma variável depende exclusivamente do número de categorias que apresenta.

Este resultado trás um benefício prático principalmente se todas as variáveis têm o mesmo número de categorias, como no casos de instrumentos que utilizam a Escala de Likert. Desta forma, uma vez que todas as questões teriam a mesma inércia total, torna-se imediata a comparação das inércias nos eixos, indicando quais variáveis estariam contribuindo mais na direção deste eixo. Outro benefício direto neste caso é a comparação direta entre as contribuições das categorias na inércia total da variável. A comparação destas magnitudes é direta.

Caso as variáveis não tenham o mesmo número de categorias podemos compará-las através da inércia relativa  $IR((j)r) = \frac{In_{\bar{c}}(\mathbf{c}_{(j)r})}{In_{\bar{c}}(\mathbf{c}_{(j)})}$ .

### A inércia total

Conseqüentemente a inércia total da nuvem, dada pela soma das inércias de todas as categorias, o que resulta em

$$In(\mathbf{G}) = \frac{1}{J} \sum_{j=1}^J (k_j - 1) = \frac{k}{J} - 1 \quad (3.3.20)$$

também é fixa e depende do número  $k$  total de categorias e do número  $J$  de variáveis.

Este resultado revela a artificialidade das magnitudes totais da inércia, uma vez que independe dos indivíduos, não trazem informação sobre o formato da nuvem, ao contrário da ACS.

Uma contribuição dos resultados (3.3.19) e (3.3.20) é a comparação entre instrumentos do mesmo tipo, com o mesmo número de itens, ou no caso da adaptação de alguns itens. O fato dos instrumentos terem a inércia fixa à priori leva à comparação direta da inércia dos eixos de interesse, sendo que se a soma das inércias destes eixos é maior em um instrumento, há mais evidência de discriminação e latência que no outro comparado.

Através das contribuições das variáveis à inércia contida em cada eixo fatorial, podemos interpretar as inter-relações das variáveis sobre o contexto de variabilidade.

#### **3.3.4- A contribuição de uma variável à inércia de um eixo**

Conforme visto na generalização da análise (cap. 2) o autovalor  $\lambda_s$  é a norma do  $s$ -ésimo eixo e pode ser entendido como a parte da inércia que coube a este eixo. Logo, podemos interpretar os comportamentos das variáveis nestes eixos, e entender sua importância na nuvem  $\mathbf{G}^*$  de modo geral.

A contribuição da  $j$ -ésima variável no eixo  $s$  é dada por

$$C_s(\mathbf{h}_j) = \frac{1}{\lambda_s} \sum_{r=1}^{k_j} \frac{g_{(j)r}}{IJ} \varphi_{s(j)r}^2 \quad (3.3.21)$$

Note que o termo na somatória é a soma ponderada das coordenadas ao quadrado de cada categoria. Logo temos a relação

$$C_s(\mathbf{H}) = \sum_{j=1}^J C_s(\mathbf{h}_j) = \sum_{j=1}^J \frac{1}{\lambda_s} \sum_{r=1}^{k_j} \frac{g_{(j)r}}{IJ} \varphi_{s(j)r}^2 = \frac{1}{\lambda_s} \sum_{j=1}^J \sum_{r=1}^{k_j} \frac{g_{(j)r}}{IJ} \varphi_{s(j)r}^2 = \frac{1}{\lambda_s} \boldsymbol{\varphi}_S^T \mathbf{D}_k \boldsymbol{\varphi}_S = 1 \quad (3.3.22)$$

Então a interpretação é feita em forma de contribuição relativa em cada eixo. É comum expressar (3.3.21) em termos de porcentagem. A importância da variável será maior quanto maior for sua contribuição.

### 3.3.5- O coeficiente de discriminação

Um outro coeficiente de grande importância é apresentado por Gifi (1990, p.113) e é chamado de Medida de Discriminação (*Discrimination Measures*). A medida de discriminação da variável  $j$  no  $s$ -ésimo eixo é dada por

$$\eta_{(j)s}^2 = \frac{1}{I} \boldsymbol{\varphi}_{(j)s}^T \mathbf{D}_j \boldsymbol{\varphi}_{(j)s} = \frac{1}{I} \sum_{r=1}^{k_j} g_{(j)r} \varphi_{(j)rs}^2 \quad (3.3.23)$$

onde  $\varphi_{(j)rs}$  estão definidos em (3.2.16).

Note que a soma de todas as medidas de discriminação é

$$\eta_s^2 = \frac{1}{I} \sum_{j=1}^J \sum_{r=1}^{k_j} g_{(j)r} \varphi_{(j)rs}^2 = \frac{1}{I} IJ\lambda_s = J\lambda_s \quad (3.3.24).$$

Este índice pode ser interpretado da mesma forma que a contribuição em 3.3.3. No entanto suas propriedades mais interessantes não são relativas a esta identidade.

Seja  $\mathbf{q}_{(j)s} = \mathbf{G}_j \boldsymbol{\varphi}_{(j)s}$  o vetor recodificado das respostas originais apresentadas em  $\mathbf{h}_j$ , como definida no capítulo 2, dos  $I$  indivíduos de acordo com as escalas ótimas (coordenadas das categorias) obtidas no  $s$ -ésimo eixo, ou seja, o vetor representante da variável  $j$  no eixo  $s$  neste novo subespaço.

Então a medida apresentada em (3.3.23) é o quadrado da correlação entre variável quantificada  $\mathbf{q}_{(j)s}$  com o  $s$ -ésimo eixo fatorial  $\psi_s$ , definido em (3.2.14) em uma interpretação semelhante às cargas das variáveis na ACP.

**prova:** A correlação ao quadrado  $r^2(\mathbf{q}_{(j)s}, \psi_s)$  é o cosseno ao quadrado do ângulo formado entre estes dois vetores. O índice  $s$  será omitido uma vez que mantém-se fixo. Logo, dado  $s$

$$r^2(\mathbf{q}_{(j)}, \psi) = \frac{(\psi^T \mathbf{q}_{(j)})^2}{(\psi^T \psi)(\mathbf{q}_{(j)}^T \mathbf{q}_{(j)})} = \frac{(\psi^T \mathbf{G}_j \boldsymbol{\varphi}_{(j)})^2}{(\psi^T \psi)(\boldsymbol{\varphi}_{(j)}^T \mathbf{G}_j^T \mathbf{G}_j \boldsymbol{\varphi}_{(j)})} \stackrel{1}{=} \frac{(\sqrt{\lambda} \boldsymbol{\varphi}_{(j)}^T \mathbf{D}_j \boldsymbol{\varphi}_{(j)})^2}{I \lambda (\boldsymbol{\varphi}_{(j)}^T \mathbf{G}_j^T \mathbf{G}_j \boldsymbol{\varphi}_{(j)})} =$$

$$\frac{\lambda (\boldsymbol{\varphi}_{(j)}^T \mathbf{D}_j \boldsymbol{\varphi}_{(j)})^2}{I \lambda (\boldsymbol{\varphi}_{(j)}^T \mathbf{D}_j \boldsymbol{\varphi}_{(j)})} = \frac{(\boldsymbol{\varphi}_{(j)}^T \mathbf{D}_j \boldsymbol{\varphi}_{(j)})}{I} = \eta_{(j)}^2$$

Em 1 utilizou-se o fato de  $\psi_s^T \mathbf{G} = \sqrt{\lambda} \mathbf{D} \boldsymbol{\varphi}_s$  como visto em (3.2.14).

Desta forma o próximo item sugere uma forma de interpretação de estrutura fatorial, de forma análoga a feita na Análise Fatorial tradicional.

### 3.3.6- Uma sugestão de carga fatorial para variáveis qualitativas

Uma vez que identificamos que as medidas  $\eta_s^2$  são na verdade o quadrado das correlações entre as variáveis e os eixos, podemos interpretar a raiz destas medidas como a correlação entre estes vetores, o que nos dá uma interpretação de carga fatorial, semelhante a utilizada na AF, que pela solução por Componentes Principais são as chamadas cargas (*loads*).

No entanto há o problema dos sinais destas correlações. Gifi prova que há uma relação muito interessante entre as soluções da ACM e da ACP para uma análise de fatores. Apresentaremos o resultado, sua prova pode ser encontrada em Gifi (1990, p.120-123).

**Relação entre a ACM e ACP na Primeira Dimensão da Análise Fatorial**

Seja  $\mathbf{Q}_1 = [\mathbf{G}_1 \varphi_{(1)1} \mid \dots \mid \mathbf{G}_J \varphi_{(j)1}]$ , onde  $\varphi_{(j)1}$  é o vetor com as coordenadas das categorias da  $j$ -ésima variável no primeiro eixo. A decomposição em valores singulares de  $\mathbf{Q}_1$  pode ser escrita  $\mathbf{Q}_1 = \mathbf{K} \Lambda_{\mathbf{Q}}^{-1/2} \mathbf{L}^T$  e sua decomposição espectral como  $\mathbf{R}_1 = \mathbf{Q}_1^T \mathbf{Q}_1 = \mathbf{L} \Lambda_{\mathbf{Q}} \mathbf{L}^T$  onde  $\Lambda_{\mathbf{Q}} = [\lambda_{1\mathbf{Q}}, \dots, \lambda_{p\mathbf{Q}}]$  (ver Apêndice B). Então as medidas de discriminação  $\eta^2$  obtidas para o primeiro eixo de  $\mathbf{G}$  são iguais ao quadrado das cargas (loads) do primeiro eixo da solução para Componentes Principais de  $\mathbf{Q}_1$ , logo iguais ao quadrado das cargas fatoriais da AF, no primeiro fator, solucionada por Componentes Principais.

Ou seja, a raiz das medidas de discriminação do primeiro eixo de  $\mathbf{G}$  são as cargas fatoriais do primeiro eixo da decomposição em Componentes Principais de  $\mathbf{Q}$ . Aplicando-se então ACP em  $\mathbf{Q}$  obtêm-se as cargas fatoriais para as variáveis qualitativas em  $\mathbf{H}$ . Equivale a uma análise de Componentes Principais sobre a matriz de escalas ótimas.

Seja  $\vartheta_1$   $k \times J$  a matriz contendo as escalas ótimas da primeira dimensão da decomposição, conforme definidos em (3.2.14), tal que a 1ª coluna apresenta as primeiras  $k_1$  linhas diferentes de zero, contendo  $\varphi_{(1)1}$ , a segunda coluna apresenta as linhas  $k_1+1$  a  $k_1+k_2$  contendo os elementos  $\varphi_{(2)1}$  e os demais diferentes de zero, e assim por diante. Então  $\mathbf{Q}_1 = \mathbf{G}\vartheta_1$  e  $\vartheta_1 \mathbf{1}_J = \varphi_1$ . A diagonal da matriz  $(\mathbf{Q}_1^T \mathbf{Q}_1) / I = (\mathbf{G}\vartheta_1)^T (\mathbf{G}\vartheta_1) / I$  são as medidas de discriminação conforme definidas em (3.3.23). Seja  $\eta$  o vetor  $J \times 1$  contendo os elementos  $\eta_j$ , raízes das medidas de discriminação, e  $\mathbf{D}_\eta$  matriz diagonal  $J \times J$  com os elementos de  $\eta$ .

Então podemos escrever  $\mathbf{R} = \mathbf{D}_\eta^{-1} \mathbf{Q}_1^T \mathbf{G}^T \mathbf{G} \mathbf{Q}_1 \mathbf{D}_\eta^{-1}$  como a matriz das correlações entre as variáveis transformadas (com as respostas codificadas pela escala ótima) de  $\mathbf{Q}$ . Então  $\eta$  é um autovetor de  $\mathbf{R}$  associado com o autovalor  $\lambda_1$  (primeiro autovalor de  $\mathbf{G}$ ) pois

$$\begin{aligned} \mathbf{D}_\eta^{-1} \mathbf{Q}_1^T \mathbf{G}^T \mathbf{G} \mathbf{Q}_1 \mathbf{D}_\eta^{-1} \eta &= \mathbf{D}_\eta^{-1} \mathbf{Q}_1^T \mathbf{G}^T \mathbf{G} \mathbf{Q}_1 \mathbf{1}_J = \mathbf{D}_\eta^{-1} \mathbf{Q}_1^T \mathbf{G}^T \mathbf{G} \varphi_1 = \\ &= \lambda_1 \mathbf{D}_\eta^{-1} \mathbf{Q}_1^T \mathbf{D} \varphi_1 = \lambda_1 \eta \end{aligned}$$

onde  $\mathbf{D}$  está definida em (3.3.1) e utilizou-se (3.3.12) no segundo passo.

Então  $\lambda_{1Q} = \lambda_1$  e uma vez que  $\eta^T \eta = \eta_1^2 = J\lambda_1$  como visto em (3.3.24), segue que  $\eta$  é o autovetor de  $\mathbf{R}$  normalizado para que sua norma ao quadrado seja igual ao seu autovalor associado, tal que  $\eta = \lambda_{1Q} \mathbf{I}$ , onde  $\mathbf{I}$  é a primeira coluna de  $\mathbf{L}$  definida acima.

Este resultado é válido apenas na primeira dimensão. Para as demais dimensões existe a prova de que as medidas de discriminação obtidas no  $s$ -ésimo eixo coincidem com o quadrado das cargas fatoriais de alguma dimensão da AF, não necessariamente a  $s$ -ésima (Gifi, 1990, p.124-126).

Então se queremos analisar as cargas fatoriais das variáveis qualitativas em  $p^*$  dimensões, devemos realizar  $p^*$  ACPs nas matrizes  $\mathbf{Q}_s$  onde  $s = 1, \dots, p^*$ . Este procedimento pode ser bastante custoso do ponto de vista computacional, principalmente se forem utilizados pacotes estatísticos pouco flexíveis.

Neste trabalho sugerimos uma forma de encontrarmos os sinais de uma forma simples que foi comprovada ao menos empiricamente, como será visto no cap. 4, nas simulações. A sugestão é a de se atribuir o sinal da carga fatorial da  $j$ -ésima variável no  $s$ -ésimo eixo,  $\eta_{(j)s}$ , igual ao sinal da soma das coordenadas das categorias da variável  $j$  no  $s$ -ésimo eixo, ou seja, a soma da escala de resposta ótima obtida na ACM. A soma deve ser direta, uma vez que a soma ponderada é igual a zero, pois nestas coordenadas estes pesos são parte de sua construção.

O argumento se dá pelo que se apresenta em (3.2.16) e (3.3.21). Se a soma das coordenadas é positiva, é imediato que a contribuição à inércia da variável foi maior no lado positivo do eixo, e o mesmo ocorre se a soma for negativa, ou seja, as categorias da variável têm maior importância no lado negativo do eixo. Logo a carga fatorial segue a mesma direção.

Tanto na AF, quanto na ACP, as diferenças de sinais das cargas são interpretadas como oposições entre as variáveis. Em variáveis qualitativas estas oposições são pouco intuitivas, a menos que se assuma a escala de resposta utilizada como real. Desta forma se temos uma escala ótima obtida pela ACM, que pela centralização apresentará necessariamente oposições entre as categorias, é razoável utilizarmos a sugestão

apresentada. Se  $j$  contribui mais no lado positivo do eixo e  $j^*$  no negativo é plausível pensar em oposição entre  $j$  e  $j^*$ .

Esta oposição tem papel importante na separação dos indivíduos e posterior classificação dos mesmos.

Ainda resta o problema da soma ser zero. No entanto, pelos mesmo argumentos, se uma soma é zero indica que esta variável não contribui no eixo, e mesmo seu valor em módulo será pouco significativo na análise. Sugere-se no entanto, para fins de apresentação de resultado, adotar-se o sinal positivo.

---

### 3.3.7 - O Efeito Ferradura (EF) na Análise de Correspondência Múltipla e sua Importância na Psicometria

É muito comum nos mapas de ACM, principalmente no primeiro plano (1º e 2º eixos), que os pontos projetados (linhas e/ou colunas) se dispersem em forma de uma parábola, ou algo próximo a um efeito quadrático. Este efeito também ocorre com frequência em outras técnicas que utilizam a DVS como Componentes Principais, Escalonamento Múltiplo e na própria Análise de Correspondência Simples. O enfoque será dado sobre a ACM, mas analogias podem ser feitas facilmente, desde que tratem de dados qualitativos.

Este efeito, que recebe o nome de Efeito Ferradura (EF) (Kendall, 1971) utilizado neste trabalho, é encontrado com maior frequência nos trabalhos da escola holandesa e de Greenacre. Também é chamado de Efeito de Guttman (Bénzecri, 1973), na escola francesa, e Efeito Arco (e.g. Gauch, 1982) no jargão da arqueologia, ecologia, e ciências da terra.

Segundo Van Rijckevorsel (1985, p.377-388), uma vez que se tem apenas variáveis qualitativas a presença deste efeito ferradura de natureza quadrática, ou outro de grau menor como a ordenação correta das categorias no caso de uma variável ordinal, reflete uma permutação única das categorias da matriz de dados. Esta permutação única que reflete a ordenação das categorias é a chamada *Matriz de Petrie* (tab. 3.1) que a partir da matriz indicadora **G** através de permutações de linhas e colunas chega a uma matriz indicadora quase diagonal que evidencia este gradiente. A figura 1 (Greenacre 1984, p. 228-229) mostra um exemplo onde **G** tem dimensão **10x15** sendo as linhas Tipos de Locais (quanto a clima) e as colunas espécies de plantas. O objetivo deste estudo, detalhado na referência, é escalonar as espécies quanto aos locais classificados pelo tipo de clima. Logo uma casela com valor 1 indica a presença da planta no respectivo local e 0 (caselas vazias na tabela 3.1) caso contrário. O mapa da ACM resultante deste exemplo está apresentado no gráfico 3.1.

A ordem e o espaçamento entre os elementos são boas fontes de informação sobre o potencial de ordem e espaçamento destes em uma estrutura latente intrinsecamente unidimensional (Van Rijckevorsel, 1985). Para entendermos este efeito do ponto de vista matemático, com um pouco mais de formalismo devemos introduzir o conceito de “ordem-

dependência" (*order dependence*). Esta propriedade está definida para tabelas de frequência de dupla entrada, mas podemos pensar em  $G$  como uma tabela de dupla entrada de pseudo-frequências e estender esta definição, sem perda de compreensão, uma vez que as definições baseiam-se apenas em somas parciais.

**Definição 3.3.1:** Dada uma matriz de indicadora  $G$   $I \times J$  (conforme já definido anteriormente), dizemos que  $G$  é **ordem-dependente de ordem 1 (OD1)** se as linhas e colunas podem ser permutadas de tal forma que os novos índices para as linhas  $\Delta i = 1, \dots, I$  e os novos índices para as colunas  $\Delta j = 1, \dots, J$  sejam tais que:

Dadas duas linhas distintas  $\Delta i$  e  $\Delta i'$  com  $\Delta i < \Delta i'$  as somas parciais

$$S^k(\Delta i) = \frac{1}{g_{\Delta i}} \sum_{\Delta j=1}^k g_{\Delta i \Delta j}, \text{ para } k=1, \dots, J \text{ obedecem } S^k(\Delta i) < S^k(\Delta i');$$

e

ii) Dadas duas colunas distintas  $\Delta j$  e  $\Delta j'$  com  $\Delta j < \Delta j'$  as somas parciais

$$S^k(\Delta j) = \frac{1}{g_{\Delta j}} \sum_{\Delta i=1}^k g_{\Delta i \Delta j}, \text{ para } k=1, \dots, I \text{ obedecem } S^k(\Delta j) < S^k(\Delta j').$$

É bastante claro que a permutação que resulta na matriz de Petrie satisfaz estas condições. Em seus trabalhos Schriever (1983, 1985) prova ainda que esta indexação é única se as linhas (colunas) não são proporcionais 2 a 2. Prova ainda que se  $G$  é OD1, então ocorrerá a ordenação correta. Mas esta definição não explica todo o efeito ainda. A ferradura só aparece se  $G$  é também OD2. No entanto a definição de ordem-dependência de ordem maior que 1 fugiria do escopo deste trabalho, e não traz grande benefício para a compreensão e interpretação deste efeito. Referências para esta teoria são os já citados trabalhos de Schriever. O importante neste ponto é entender a importância deste fenômeno se o mesmo ocorre no primeiro plano.

Schriever prova que se  $G$  é OD2 então é OD1. Na verdade a presença da ferradura indica uma forte ordenação das categorias das variáveis envolvidas na matriz indicadora. Não há nenhuma garantia, segundo vanRijckevorsel, de que o segundo eixo seja aquele que expressa a OD2, mas se este fato ocorre, então é muito plausível que esta dependência exista, implicando OD1 e evidenciando a consistência da ordenação no primeiro eixo.

Lebart (1995) sustenta que a presença do efeito ferradura indica que quase toda explicação encontra-se no primeiro eixo, tornando o problema trivial, e que o mesmo se torna mais interessante se a parábola não é perfeita, onde os “pontos de inflexão” são interessantes de ser analisados.

Outro argumento (Greenacre, 1984, p.231) é dado em termos de distância (similaridade). Na tabela de exemplo, fig.3.1, a similaridade entre os Locais 2 e 4 é idêntica à similaridade entre 3 e 7, uma vez que cada par tem cinco espécies presentes em cada Local, sem espécies em comum. Assim o mapa tenta recuperar e representar simultaneamente os pares de Locais distantes como equidistantes, e os pares vizinhos como progressivamente dissimilares. Esta “tentativa” é marcante no segundo eixo, resultando em uma curvatura, uma vez que o primeiro eixo opõe as categorias de maior distância, ou menor similaridade.

Então na psicometria podemos entender que a presença deste fenômeno no primeiro plano evidencia uma estrutura latente unidimensional, ou um construto, e dado que a ordenação das categorias da escala de resposta está correta, também indica a validade da utilização do instrumento enquanto escala unidimensional. Torna-se útil também para a identificação de itens que não apresentem a mesma ordem, ou apresentem comportamento atípico.

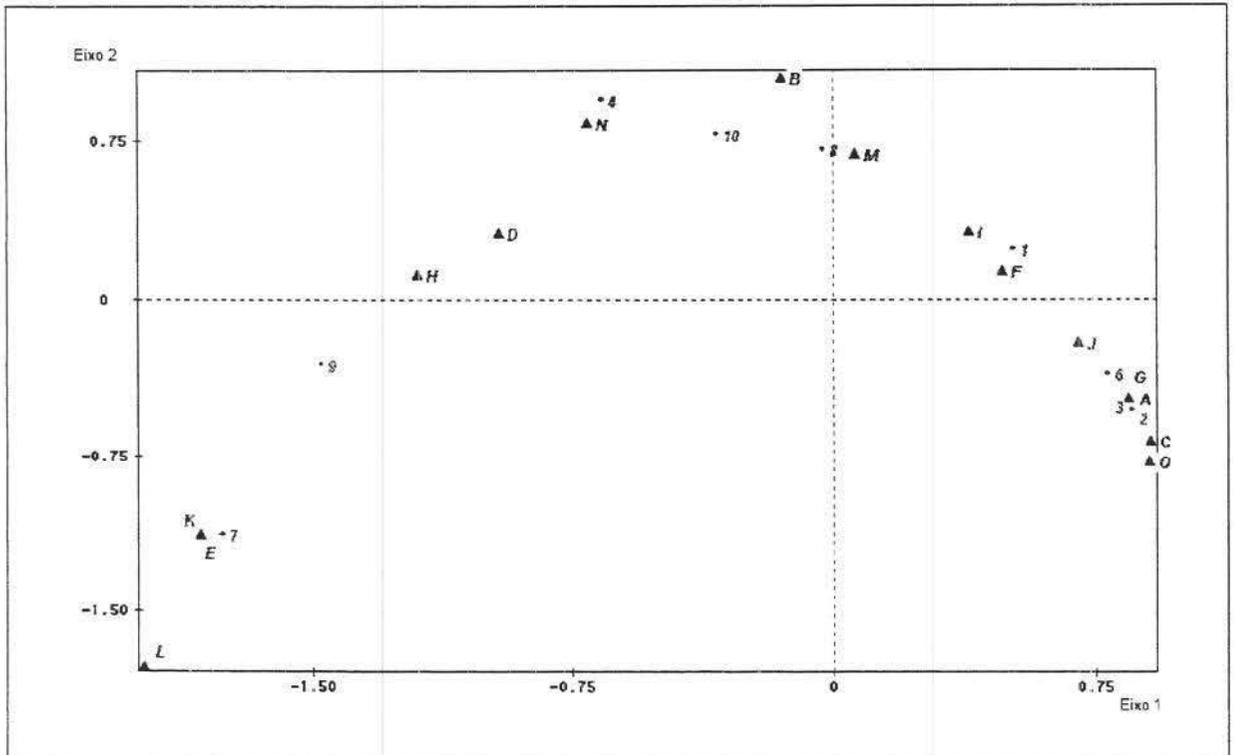
---

<i>Espécies</i>															
Locais	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>	<i>m</i>	<i>n</i>	<i>o</i>
1	1	1				1	1		1	1			1		
2	1		1			1	1			1					
3	1		1				1			1			1		1
4		1		1				1					1	1	
5	1		1			1	1		1	1					
6	1		1			1	1		1	1					
7				1	1			1			1	1			
8		1		1		1			1	1			1	1	
9				1	1			1			1			1	
10		1		1		1		1	1				1	1	

	<i>o</i>	<i>c</i>	<i>a</i>	<i>g</i>	<i>j</i>	<i>f</i>	<i>i</i>	<i>m</i>	<i>b</i>	<i>n</i>	<i>d</i>	<i>h</i>	<i>k</i>	<i>e</i>	<i>l</i>
3	1	1	1	1	1										
2		1	1	1	1	1									
6		1	1	1	1	1	1								
5		1	1	1	1	1	1	1							
1			1	1	1	1	1	1	1						
8					1	1	1	1	1	1	1				
10						1	1	1	1	1	1	1			
4								1	1	1	1	1			
9										1	1	1	1	1	
7												1	1	1	1

fig. 3.1- Tabela original e Matriz de Petrie sobre a mesma (fonte: Greenacre (1984), p.228-229 )



vgraf. 3.1- Efeito Ferradura na Análise de Correspondência

### 3.3.7- A escolha do número de eixos

Nas técnicas multivariadas descritivas baseadas em DVS este é sempre um problema bastante discutido. Quantos eixos, ou dimensões, devem ser analisadas? Qual a importância relativa entre os eixos e qual a influência na interpretação de resultados?

Inúmeras referências tratam deste assunto (Greenacre, Lebart, Benzécri, Anderson, etc...), alguns com métodos inferenciais, encontrando distribuições para os autovalores, e conseqüentemente testes de significância. Outros propõem apenas métodos empíricos.

Uma discussão mais aprofundada não seria adequada aos propósitos deste trabalho. O enfoque aqui é sobre esta escolha na ACM aplicada à psicometria. Partindo-se do problema em questão, o de escalas hipoteticamente unidimensionais, é imediato que a análise principal deva ser sobre o primeiro eixo. Como visto até aqui, este eixo, além da estrutura fatorial que evidenciaria a latência da escala (a medição de um construto) também é o que certifica a ordenação ótima e a métrica das categorias qualitativas ordinais, pelo próprio fato de que este é o eixo de direção de maior variabilidade.

No entanto há o problema das importâncias relativas entre os eixos. É claro que se o segundo auto-valor é relativamente próximo ao primeiro a importância deste segundo na interpretação é clara. As simulações no cap. 4 trazem contribuições importantes nesse aspecto.

O tópico anterior, sobre o EF, também sugere que o número máximo de eixos a serem analisados, com algum tipo de informação relevante, não deve ultrapassar o número máximo de categorias das variáveis. No caso de um instrumento em EL, onde todas tem o mesmo número de pontos de resposta, este número seria o número máximo de dimensões que pudessem trazer alguma informação.

Cattell (1966) sugere um método empírico, de utilização não só na psicometria, baseada no formato do diagrama decrescente dos auto-valores. Uma descontinuidade, ou um salto (fig. 3.2 b), nos valores ordenados dos auto-valores em forma decrescente seria uma evidência de que os eixos representados pelos auto-valores anteriores ao salto são aqueles de maior importância no hiper-elipsóide formado pela nuvem de pontos. Ou seja, são os eixos mais alongados da nuvem (eixos principais).

Podemos interpretar este fato pela contribuição de cada eixo na inércia total. A inércia de um respectivo eixo é dado pelo autovetor associado ao mesmo. Se a inércia total é fixa e depende apenas do número de colunas de  $\mathbf{G}$  e variáveis em estudo, como visto em (3.3.20), então a inércia relativa do  $s$  – ésimo eixo é dada por  $\lambda_s / \text{In}(\mathbf{G})$ . Logo se os  $p^*$  primeiros eixos  $\psi_s$  absorvem inércia de forma mais marcante, o que ocorre quando estas direções da nuvem são mais alongadas em relação às  $p - p^*$  restantes, então ocorrerá uma descontinuidade na monotonicidade decrescente dos autovalores. Uma forma de se interpretar as importâncias dos saltos também é analisar as razões  $\lambda_s / \lambda_{s+1}$ .

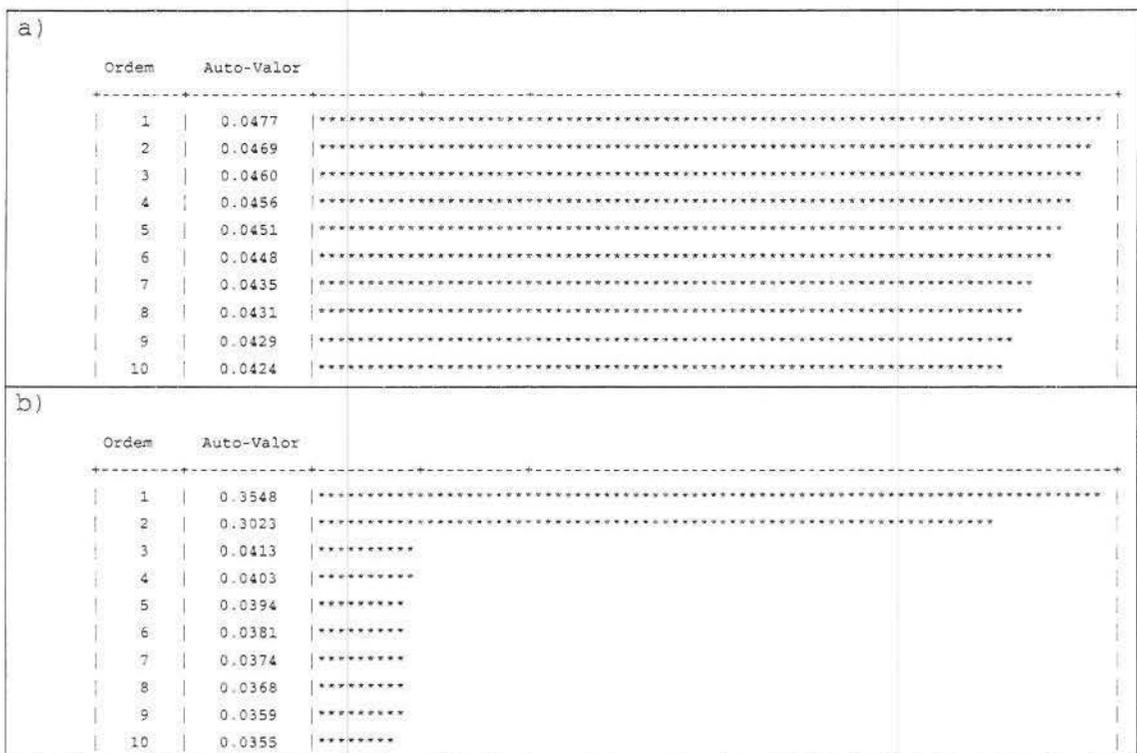


fig. 3.3- Diagramas de auto-valores. Em a) provenientes de uma análise sem qualquer estrutura latente. Em b) há evidência de estrutura nos primeiro e segundo eixos.

### 3.3.8- *A utilização dos pontos ilustrativos*

Podemos representar linhas e/ou colunas ilustrativamente, representando-as sobre os novos eixos, ou seja suas projeções no novo sub-espço. No caso da ACM estamos falando de indivíduos (linhas) e categorias de respostas (colunas) ilustrativas. A seguir serão discutidos os dois casos no escopo da psicometria.

#### *A representação de indivíduos ilustrativos*

A representação de indivíduos ilustrativos é um artifício interessante para um estudo de repetibilidade do instrumento, bem como de reteste. Suponha que o instrumento foi aplicado, validado e calibrado de acordo com uma população específica. A representação dos indivíduos de uma nova aplicação como ilustrativos sobre os eixos encontrados no primeiro estudo nos mostrara a repetibilidade do instrumento.

Suponha que os novos indivíduos sejam da mesma população que a primeira, ou bastante semelhante (quanto à característica estudada, ou perfil sócio-econômico, etc...). Espera-se que estes se apresentem inseridos na nuvem de pontos dos indivíduos anteriores, o que seria uma repetição da medida.

No caso de populações diferentes deve-se observar o comportamento. O caso de uma discrepância ou comportamento atípico destes novos indivíduos sugeriria uma validação e calibração específica para esta população.

No caso do reteste, onde os indivíduos são os mesmos analisados após um período de tempo, espera-se uma concordância na disposição dos pontos em relação à primeira aplicação.

#### **Representação Algébrica**

Seja  $\mathbf{G}_+$ ,  $n \times k$ , a matriz contendo as  $n$  linhas com os dados dos indivíduos que se deseja representar como ilustrativos, definida como em 3.1. Então as coordenadas destes indivíduos no  $s$ -ésimo, segundo notação até aqui utilizada, eixo é dado por

$$\psi_{s+} = \frac{1}{J\sqrt{\lambda_s}} \mathbf{G}_+ \varphi_s \quad (3.3.25)$$

onde o  $\psi_s$  é o vetor  $n \times 1$  contendo as novas coordenadas e os demais elementos são os já definidos na sessão 3.2 .

### A representação de categorias e/ou variáveis ilustrativas

A ACM permite o recurso de se representar apenas uma categoria de uma variável como ilustrativa. Como visto em (3.1.6) categorias com pouca frequência, ou seja raras, tendem a contribuir de uma forma enganosa na inércia da nuvem de pontos da variável a que pertence. Representa-la como ilustrativa é uma solução para não retirá-la por completo da análise e entender sua posição relativa às demais.

Pode-se representar por completo uma variável como ilustrativa se tomarmos  $G_j$ . No caso de itens atípicos no instrumento esta é uma solução. Ou no caso da inclusão ou substituição de um item também. Da mesma forma que os indivíduos podemos aqui estudar a adaptação de um item ao instrumento, em um sentido de reproducibilidade do construto estudado.

### Representação Algébrica

Seja  $G^+$ ,  $I \times m$ , a matriz contendo as  $m$  colunas que definem os pontos a serem representados como ilustrativos. Então as coordenadas destes pontos no novo sistema são dadas por:

$$\varphi_s^+ = \frac{1}{\sqrt{\lambda_s}} \mathbf{D}^{+^{-1}} \mathbf{G}^{+T} \psi_s \quad (3.3.26)$$

onde  $\varphi_s^+$  é o vetor  $m \times 1$  contendo as novas coordenadas e  $\mathbf{D}^+$  é a matriz diagonal dos totais das colunas, ou a diagonal de  $(\mathbf{G}^{+T} \mathbf{G}^+)$ .

Outra aplicação interessante de variáveis ilustrativas diz respeito a variáveis demográficas, sociais, ou de controle em estudo que podem ser representadas para auxiliar entender o comportamento do instrumento quando aplicado em uma população heterogênea. As simulações (cap. 4) e a aplicação real (cap. 5) trazem exemplos de variáveis ilustrativas.

---

## 4. Estudo de Simulação

O avanço computacional dos últimos anos disponibilizou máquinas e linguagens de programação bem como pacotes estatísticos mais eficientes, e conseqüentemente o desenvolvimento de algoritmos mais complexos. Desta forma um estudo de simulação torna-se viável e vantajoso no estudo comparativo do desempenho de técnicas e métodos estatísticos, pois parte-se de um conhecimento a priori da estrutura dos dados (ex: distribuições, estruturas de grupo), valores de parâmetros, e estuda-se a adequação e eficiência das aplicações das técnicas comparadas.

Tendo este trabalho como motivação a análise de propriedades e validação de instrumentos de escalas psicométricas, a idéia neste ponto é a de representar através de simulações uma estrutura semelhante a um instrumento deste tipo. No desenvolvimento de um instrumento real estas estruturas são definidas a partir da teoria da característica psicológica estudada, sendo estes itens estímulos que em conjunto a caracterizariam. Em termos de simulação podemos definir esta estrutura através da modelagem probabilística de uma estrutura de dependência entre estes itens, como será visto mais adiante.

Há também a preocupação da discriminação de grupos de indivíduos quanto aos perfis de resposta dos itens destes instrumentos, ou seja, a capacidade de classificação de indivíduos através deste instrumento quanto às diferenças de padrão das respostas aos estímulos. Logo a simulação de diferentes padrões de resposta permite a criação de estruturas de grupos reais.

---

Existem poucos estudos de simulação quando se trata de análise multivariada de dados discretos, e quando encontrados sempre focalizam o comportamento dos coeficientes da Análise Fatorial em diferentes métodos de estimação. Exemplos destes trabalhos são Fuller & Hemmerle (1983), Boomsma (1983), Muthén & Kaplan (1985, 1992).

Estes autores trataram mais especificamente dos efeitos de estimadores de Máxima-Verossimilhança e de variações dos métodos de Mínimos Quadrados em situações de não normalidade em variáveis de resposta tipo Likert. Os dados foram gerados sem qualquer modelagem da dependência entre os itens, com observações geradas a partir de distribuições Normal categorizadas posteriormente seguindo sempre um mesmo critério, ou seja, as marginais de resposta são sempre as mesmas. Não houve em nenhum momento preocupação com o desempenho da AF e seus diferentes estimadores na detecção de variáveis latentes, são comparados apenas os estimadores e suas propriedades.

---

#### 4.1- Estrutura de dados utilizada

Com base na introdução acima podemos imaginar uma estrutura genérica de dados que representariam a aplicação de um instrumento de escala psicométrica em diferentes grupos de indivíduos.

Suponha um instrumento contendo  $J$  itens com resposta tipo *Likert* de  $p$  pontos relativos à intensidade da resposta ao estímulo, sendo que estes itens teoricamente formam  $C$  escalas, sendo que cada item pertence exclusivamente a uma escala. Então o conjunto de itens da  $c$ -ésima escala é formado por  $J$  itens ( $c = 1, \dots, C$ ) e  $K = JC$  é o número total de estímulos no instrumento. Suponha agora que este instrumento foi aplicado em  $I$  indivíduos, e que estes indivíduos pertencem, também de forma mutuamente exclusiva, a  $G$  grupos, sendo  $n$  o número de indivíduos pertencentes ao  $g$ -ésimo grupo, e  $I = nG$ .<sup>4</sup>

Esta é a estrutura que será utilizada nas simulações descritas a seguir.

---

<sup>4</sup> Neste trabalho serão assumidos que o número de itens que compõe cada construto teórico do instrumento são iguais, bem como o número de indivíduos em cada grupo. Em aplicações de instrumentos reais estas suposições podem ou não serem satisfeitas. No entanto uma discussão quanto à influência do número de itens e da influência do número de indivíduos na verificação do construto fugiria do escopo deste trabalho.

---

## 4.2- Geração dos dados

A mensuração de uma característica psicológica (traços ou estados) em um determinado indivíduo através de um instrumento baseado em escala de Likert se dá pelo padrão das intensidades de resposta aos estímulos. Espera-se que o indivíduo responda aos estímulos de uma mesma característica de forma semelhante, ou seja, que haja uma dependência entre as respostas a estes estímulos. Suponha a caracterização de um distúrbio psicológico onde os estímulos são perguntas sobre a ocorrência de sintomas. Espera-se que um indivíduo que sofra deste distúrbio responda positivamente aos sintomas, ou a grande parte deles, com intensidades (dada pela escala de resposta) semelhantes. Assim as observações segundo a estrutura descrita na tabela 4.1 pode ser representada conforme a matriz  $\mathbf{H}$  (como no capítulo 2), onde  $h_{ij}^{(g,c)}$  é a resposta do  $i$ -ésimo indivíduo, pertencente ao  $g$ -ésimo grupo, no item  $j$  da característica  $c$ , sendo que  $h_{ij}^{(g,c)} \in \{1, 2, \dots, p\}$ .

tab. 4.1 – Estrutura de dados utilizada nas simulações

	Escala 1			...	Escala C		
	item 1	...	item J	...	item (CJ-J+1)	...	item CJ
<b>Grupo 1</b>	$h_{11}^{(1,1)}$	...	$h_{1J}^{(1,1)}$	...	$h_{1(CJ-J+1)}^{(1,C)}$	...	$h_{1(KJ)}^{(1,C)}$
	$\vdots$		$\vdots$		$\vdots$		$\vdots$
	$h_{n1}^{(1,1)}$	...	$h_{nJ}^{(1,1)}$	...	$h_{n(CJ-J+1)}^{(1,C)}$	...	$h_{n(KJ)}^{(1,C)}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$
<b>Grupo G</b>	$h_{11}^{(G,1)}$	...	$h_{1J}^{(G,1)}$	...	$h_{1(CJ-J+1)}^{(G,C)}$	...	$h_{1(KJ)}^{(G,C)}$
	$\vdots$		$\vdots$		$\vdots$		$\vdots$
	$h_{n1}^{(G,1)}$	...	$h_{nJ}^{(G,1)}$	...	$h_{n(CJ-J+1)}^{(G,C)}$	...	$h_{n(KJ)}^{(G,C)}$

No caso de um estudo de validação deste instrumento o construto só será evidenciado se existirem indivíduos na amostra tomada que realmente apresentem a característica medida pela escala. Sabe-se que a identificação de um construto através de fatores depende da amostra onde foi aplicado.

É necessário então que na simulação esta estrutura de dependência entre as respostas seja modelada de maneira que permita julgamentos e conclusões sobre o desempenho das técnicas. Então os diferentes perfis (grupos) de indivíduos a serem simulados podem ser caracterizados por um modelo pré-definido para sua geração.

Podemos pensar na escala de resposta como estados de um processo estocástico, de forma que a dependência entre as respostas pode ser modelada a partir de Matriz de Probabilidades de Transição (MT) em uma Cadeia de Markov (CM) (Ross 1990) com espaço de estados discreto e finito definido por  $S = \{1, 2, \dots, p\}$ .

Logo é possível se definir uma MT tal que se um indivíduo manifesta a presença de um atributo da característica psicológica a ser medida através da resposta de um item, a probabilidade da resposta positiva e com intensidade próxima em outro atributo da mesma característica seja maior. Assim cada indivíduo é representado por um número de passos de uma CM, definido pelo número de estímulos (itens) que se pretende simular.

Seja  $p_{(j, j+1)}(s) = P(h_{j+1}^{(g,c)} = s \mid h_j^{(g,c)} = s^*)$  a probabilidade da resposta ao  $(j+1)$ -ésimo ( $j=1, \dots, J-1$ ) item do construto  $c$ , do indivíduo  $i$  do grupo  $g$  ser igual a  $s \in S$ , dado que a resposta ao  $j$ -ésimo é  $s^* \in S$ . Então as respostas aos  $j+1$  itens da característica  $c$  de um indivíduo do grupo  $g$  podem ser geradas como  $J$  passos de uma CM com matriz de transição  $p \times p$   $\mathbf{P}^{(g,c)} = \{ p_{(j, j+1)}(s) \}$ . Replicando este procedimento  $n$  vezes temos as respostas do grupo  $g$  ao conjunto de itens da característica  $c$ . Espera-se que as respostas simuladas de grupos de indivíduos a partir deste modelo evidenciem construtos nas técnicas comparadas neste trabalho.

Segundo os objetivos deste estudo de simulação, já citados, é interessante que também sejam caracterizados indivíduos que não apresentem a característica psicológica, ou seja, que não contribuem na detecção do construto (como dito no início deste tópico).

Para a geração deste perfil de indivíduos optou-se em não modelar a dependência entre as respostas, sendo que estas foram geradas segundo distribuições de probabilidade Multinomiais com diferentes vetores de parâmetros (probabilidades de ocorrência das

categorias) tal que estes vetores representassem marginais de resposta de itens usualmente encontradas em aplicações reais (Anexo 1).

Seja  $p_{ij}(s) = P(h_{ij}^{(g,c)} = s)$  a probabilidade da resposta ao  $j$ -ésimo item do construto  $c$ , do indivíduo  $i$  do grupo  $g$  ser igual a  $s \in S$ . Então  $h_{ij}^{(g,c)} \sim \text{Multinomial}(1, \mathbf{p})$ , onde  $\mathbf{p}$  é o vetor  $l \times p$  de probabilidades de ocorrência das respostas e  $p_{ij}(s)$  independe de  $h_{ij-1}^{(g,c)}$ .

### 4.3- Matrizes e vetores utilizados

Os instrumentos simulados são de escala de resposta de 5 pontos ( $p=5$ ) por ser este o número de pontos mais comum em instrumentos psicométricos. Logo as matrizes e vetores aqui descritos serão de ordem  $5 \times 5$  e  $l \times 5$  respectivamente.

Foram utilizadas duas matrizes de transição, que serão denominadas **A** (4.1) e **B** (4.2), e três vetores de probabilidade multinomiais, denominados **i**, **u** e **a** (4.3).

$$\mathbf{A} = \begin{pmatrix} 0.10 & 0.20 & 0.20 & 0.25 & 0.25 \\ 0 & 0.20 & 0.20 & 0.30 & 0.30 \\ 0 & 0.20 & 0.20 & 0.30 & 0.30 \\ 0 & 0.10 & 0.20 & 0.35 & 0.35 \\ 0 & 0.10 & 0.10 & 0.40 & 0.40 \end{pmatrix} \quad 4.1$$

$$\mathbf{B} = \begin{pmatrix} 0.05 & 0.10 & 0.15 & 0.30 & 0.40 \\ 0 & 0.15 & 0.15 & 0.30 & 0.40 \\ 0 & 0.05 & 0.20 & 0.35 & 0.40 \\ 0 & 0 & 0.25 & 0.40 & 0.35 \\ 0 & 0 & 0 & 0.40 & 0.60 \end{pmatrix} \quad 4.2$$

A matriz **A** fornece uma relação de dependência que favorece de forma um pouco mais uniforme os estados 3, 4 e 5, o que caracterizaria um traço menos marcante quando presente, ou que apresentasse dificuldade no discernimento de intensidade do atributo. Já em **B** os estados 4 e 5 são fortemente favorecidos, caracterizando um traço muito marcante, de fácil discernimento de intensidade. Também representam diferentes perfis de respostas de indivíduo, sendo **B** de indivíduos com sintomas mais fortes.

$$\begin{aligned} \mathbf{i} &= (0.20 \quad 0.20 \quad 0.20 \quad 0.20 \quad 0.20) \\ \mathbf{u} &= (0.35 \quad 0.15 \quad 0.05 \quad 0.15 \quad 0.30) \quad 4.3 \\ \mathbf{a} &= (0.70 \quad 0.20 \quad 0.05 \quad 0.05 \quad 0.00) \end{aligned}$$

O vetor **i** representa uma uniforme discreta, sendo que a presença deste perfil de resposta confunde-se com os demais. O vetor **u** apresenta uma distribuição bimodal nos extremos, bastante encontrado em aplicações reais. Enquanto **a** caracteriza indivíduos não sintomáticos, ou seja, com moda no ponto 1 da escala utilizada, e seriam indivíduos controle em um experimento por exemplo.

A partir da combinação destas matrizes e vetores na geração dos dados serão caracterizados grupos distintos de indivíduos quanto à resposta de um instrumento virtual. O tópico a seguir descreverá as situações estudadas.

## 4.4- Simulações

Serão estudados dois casos gerais em especial: o caso de uma escala unidimensional (um construto) e casos com duas escalas em um mesmo instrumento (medindo dois construtos).

As amostras foram geradas em linguagem *Ox 2.1* (Doornik, 1998) e os algoritmos e programas estão no Apêndice C, seguindo os modelos de geração de indivíduos descritos.

Em seguida foram aplicadas e comparadas as técnicas de Análise Fatorial (AF) e Análise de Correspondência Múltipla (ACM).

Para a Solução por Componentes Principais e Análise de Correspondência foi utilizado o *software* SPAD 3.5 (CISIA-CERESTA, 1998). Para os gráficos de indivíduos utilizou-se o *software* STATISTICA 5.0 (StatSoft, 1996), enquanto para os gráficos de linhas utilizou-se STATA 5.0 (Stata, 1995), assim como os cálculos para os coeficientes de discriminação.

### 4.4.1- Caso da escala unidimensional

Neste ponto será estudado um instrumento que contemple apenas uma característica psicológica, ou um instrumento de escala unidimensional. Existe uma grande variedade destes em uso na psicologia clínica, como as escalas de atitude, o que torna este caso interessante para usuários deste tipo de instrumento.

Suponha um questionário constituído de 20 itens ( $J = 20$ ,  $C = 1$ ) com escala de resposta de 5 pontos indicando a intensidade de resposta ao atributo contemplado no item. Suponha uma aplicação em 4000 indivíduos distribuídos uniformemente em 4 grupos ( $G=4$ ,  $I = 4000$ ) distintos quanto ao perfil de resposta.

O Grupo I (GI) é formado por indivíduos gerados a partir da CM definida pela MT  $\mathbf{A}$  ( $\mathbf{P}^{(1,1)} \equiv \mathbf{A}$ ), enquanto o Grupo II (GII) pela CM definida por  $\mathbf{B}$  ( $\mathbf{P}^{(2,1)} \equiv \mathbf{B}$ ). O Grupo III (GIII) é formado por realizações de uma distribuição Multinomial com o vetor de probabilidades dado por  $\mathbf{p}=\mathbf{i}$  (em 4.3), e GIV pela mesma Lei de Probabilidade mas com o vetor  $\mathbf{p} = \mathbf{a}$ . A estrutura final utilizando-se a notação dada em 4.2 está na tabela 4.2.

tab. 4.2- Estrutura de dados da Simulação 1

$C=1 \quad G=4$	$i = 1, \dots, 1000 \quad j = 1, \dots, 20$
GI	$p_{(i, j+1)}(s)$ segundo <b>A</b>
GII	$p_{(i, j+1)}(s)$ segundo <b>B</b>
GIII	$h_{ij-1}^{(3,1)} \sim \text{Multinomial}(1, \mathbf{i})$
GIV	$h_{ij-1}^{(4,1)} \sim \text{Multinomial}(1, \mathbf{a})$

A tabela 4.3(a) apresenta os autovalores resultantes da ACM. Observa-se que o primeiro, 0.60, é muito maior que os demais, sendo que o segundo sugere alguma informação, apesar de pequena, segundo o formato do diagrama. O mesmo não acontece com os auto-valores resultantes da AF (tab. 4.3 (b)), sendo o 1º autovalor 11.24, e o segundo 0.56. Ou seja, o primeiro é cerca de 20 vezes maior que o seguinte, sugerindo apenas um fator.

Observa-se a forte estrutura fatorial dos itens na primeira dimensão, tanto nas cargas da AF quanto nas da ACM (tab. 4.4). As cargas apresentam-se altas em ambos casos, sendo que nos coeficientes  $\eta_1$  estes valores tendem a ser um pouco maiores, indicando uma estrutura fatorial mais consistente.

tab. 4.3- (a) Dez primeiros autovalores da ACM para Simulação 1. (b) Dez primeiros autovalores da AF para Simulação 1.

(a)				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.6005	15.01	15.01	*****
2	0.1007	2.52	17.53	*****
3	0.0681	1.70	19.23	*****
4	0.0678	1.70	20.93	*****
5	0.0661	1.65	22.58	*****
6	0.0647	1.62	24.20	*****
7	0.0640	1.60	25.80	*****
8	0.0627	1.57	27.36	*****
9	0.0624	1.56	28.92	*****
10	0.0618	1.54	30.47	*****

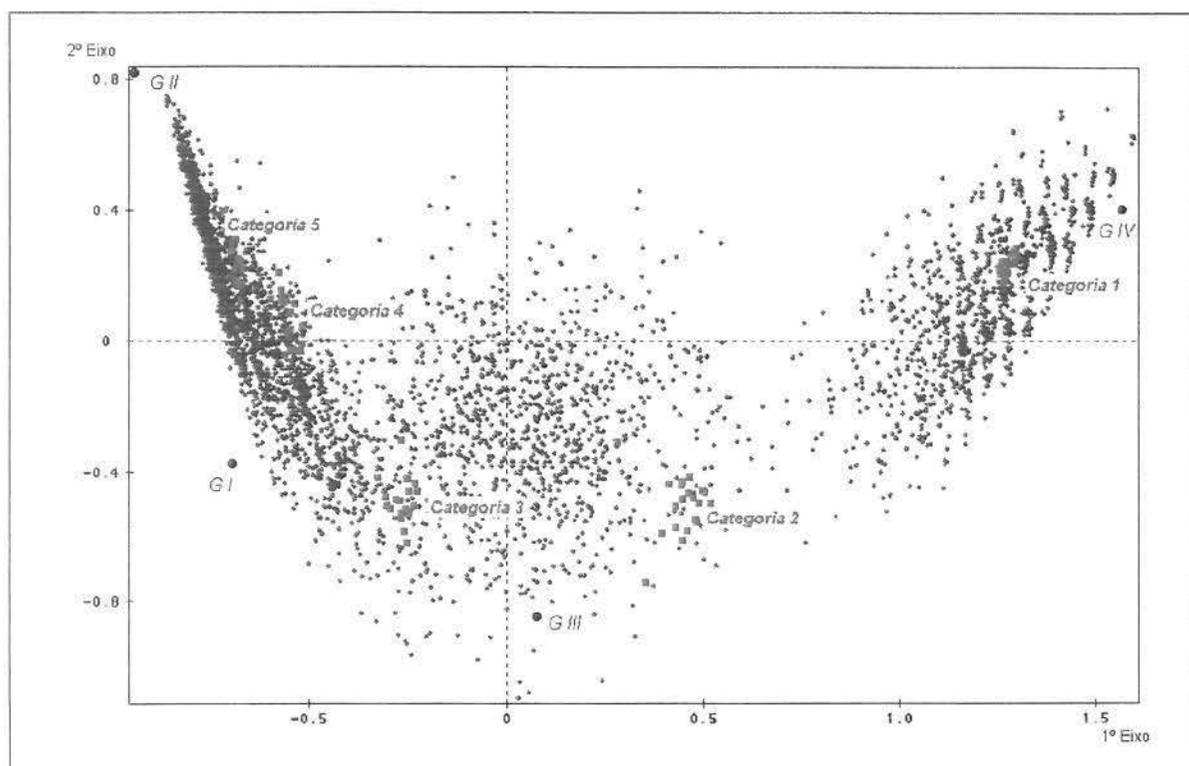
(b)				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	11.2359	56.18	56.18	*****
2	0.5684	2.84	59.02	*****
3	0.5543	2.77	61.79	*****
4	0.5510	2.76	64.55	*****
5	0.5341	2.67	67.22	*****
6	0.5183	2.59	69.81	*****
7	0.4994	2.50	72.31	*****
8	0.4841	2.42	74.73	*****
9	0.4798	2.40	77.13	*****
10	0.4640	2.32	79.45	*****

O gráfico (graf. 4.1) mostra que há um efeito Ferradura bem definido nas nuvens de pontos das categorias de resposta (pontos em vermelho). Tal fato indica uma ótima ordenação no primeiro eixo, sendo o segundo apenas uma representação auxiliar das categorias intermediárias. Nos pontos dos indivíduos ocorre uma maior abertura e espalhamento dos pontos no lado direito do mapa, refletindo a maior variabilidade entre os indivíduos, de acordo com as estruturas de grupo simuladas. Há também uma boa discriminação dos indivíduos, principalmente nos perfis de menor intensidade de sintoma, enquanto os perfis sintomáticos (GI e GII) não a apresentam.

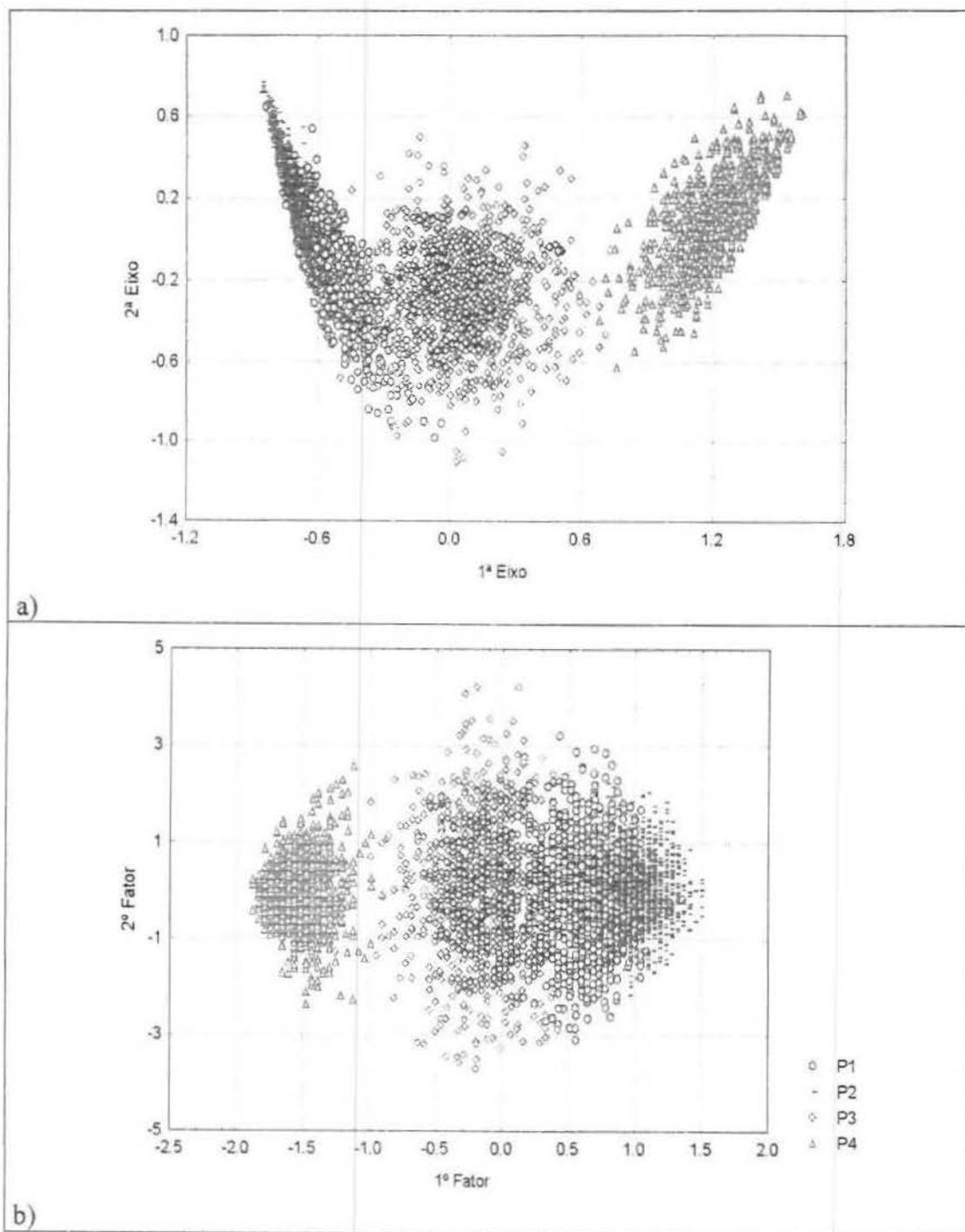
No entanto nota-se que na ACM (graf. 4.2a) a separação dos perfis torna-se efetivamente eficiente apenas com a presença do segundo eixo. Na AF (graf. 4.2b) a presença do segundo fator praticamente não acrescenta nada na separação dos perfis, onde observa-se inclusive um confundimento maior entre GI e GII em relação à ACM.

tab. 4.4- Cargas fatoriais obtidas para Simulação 1

	<b>Fator 1</b>	$\eta_1$
<b>1</b>	0.7148	0.7510
<b>2</b>	0.7559	0.7823
<b>3</b>	0.7533	0.7746
<b>4</b>	0.7509	0.7823
<b>5</b>	0.7559	0.7746
<b>6</b>	0.7498	0.7746
<b>7</b>	0.7572	0.7746
<b>8</b>	0.7544	0.7823
<b>9</b>	0.7603	0.7823
<b>10</b>	0.7671	0.7899
<b>11</b>	0.7459	0.7746
<b>12</b>	0.7528	0.7746
<b>13</b>	0.7360	0.7589
<b>14</b>	0.7557	0.7823
<b>15</b>	0.7501	0.7746
<b>16</b>	0.7473	0.7746
<b>17</b>	0.7421	0.7668
<b>18</b>	0.7498	0.7746
<b>19</b>	0.7387	0.7668
<b>20</b>	0.7501	0.7823



graf. 4.1- Mapa da AFCM para Simulação I



#### 4.4.2- Casos de instrumentos com duas escalas

Nos próximos casos serão estudados instrumentos virtuais que contemplam duas características psicológicas distintas, ou duas escalas. Nestes casos de simulação não houve qualquer preocupação de uma modelagem probabilística entre as respostas de construtos diferentes, ou seja, não se considera nem se exclui uma estrutura de associação entre as escalas.

Suponha então um instrumento com 40 itens que procurasse identificar dois traços (ou estados) psicológicos distintos, sendo que os itens que contemplam exclusivamente cada uma destas características estão divididos em igual número ( $j = 20, C = 2$ ).

Suponha novamente a aplicação em 4000 indivíduos distribuídos uniformemente em 4 grupos ( $G = 4, I = 4000$ ) distintos quanto ao perfil de resposta. Foram realizadas 3 “aplicações virtuais” seguindo esta estrutura, e mais uma que difere das demais por apresentar apenas 3 grupos de indivíduos ( $G = 3, I = 3000$ ). Procurou-se partir de um caso mais simples até o de maior complexidade em relação às estruturas geradas, que estão descritas a seguir.

#### *Simulação 2*

Esta é a situação mais simples, onde teremos uma definição clara entre os grupos devido aos modelos impostos. Em GI temos caracterizados indivíduos que apresentam a Característica 1 (C1), com respostas aos itens modeladas através da MT **A** ( $\mathbf{P}^{(1,1)} \equiv \mathbf{A}$ ), e não apresentam a Característica 2 (C2) onde estas respostas obedecem uma distribuição Multinomial (1, **a**). Em GII temos esta mesma distribuição para as respostas à C1, enquanto as repostas de C2 obedecem à MT **B**. GIII seriam indivíduos que apresentam C1 e C2, sendo as respostas geradas segundo **A** e **B** respectivamente. GIV é gerado a partir da distribuição Multinomial (1, **i**) em todos os itens.

A simplicidade esta na estrutura de Presença/Ausência das características muito bem definidas, o que propiciaria uma discriminação de grupos mais fácil e a identificação de dois construtos de forma muito clara.

tab. 4.5- Estrutura de geração de indivíduos na Simulação 2

$C=1$ $G=4$	$i = 1, \dots, 1000$ $j = 1, \dots, 20$	
	<i>Escala 1</i>	<i>Escala 2</i>
GI	$p_{(j,j+1)}(s)$ segundo <b>A</b>	$h_{ij}^{(1,2)} \sim \text{Multinomial}(1, \mathbf{a})$
GII	$h_{ij-1}^{(2,2)} \sim \text{Multinomial}(1, \mathbf{a})$	$p_{(j,j+1)}(s)$ segundo <b>B</b>
GIII	$p_{(j,j+1)}(s)$ segundo <b>A</b>	$p_{(j,j+1)}(s)$ segundo <b>B</b>
GIV	$h_{ij-1}^{(4,1)} \sim \text{Multinomial}(1, \mathbf{a})$	$h_{ij-1}^{(4,2)} \sim \text{Multinomial}(1, \mathbf{a})$

A tabela 4.6 apresenta os autovalores resultantes da ACM e AF. Observa-se claramente a presença de duas dimensões fatoriais apenas, sendo um indício da detecção dos construtos simulados. Observa-se que as razões entre o primeiro e segundo autovalor são bastante semelhantes em ambos casos

As cargas fatoriais (tab. 4.7) também confirmam claramente a presença de dois fatores conforme a simulação sendo que cada Fator define apenas um dos construtos. Dado um fator, há praticamente a ausência de cargas fatoriais nos itens que não pertencem ao construto predominante neste fator. Observa-se que em ambas cargas estimadas, da AF e da ACM, o primeiro Fator marca a presença do construto relativo à Característica **B** simulada, enquanto o segundo marca a presença da característica **A**. No entanto os valores obtidos pelas cargas  $\eta_1$  são muito mais contrastantes, definindo de modo muito mais claro a presença de construtos, tendo inclusive cargas zero nos itens pertencentes a características diferentes. Tal fato pode ser explicado devido ao coeficiente  $\eta_1$  ser derivado das contribuições de cada item no respectivo eixo fatorial (ver 3.3.4).

tab. 4.6- (a) Dez primeiros autovalores da ACM para Simulação 2. (b) Dez primeiros autovalores da AF para Simulação 2.

(a)				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.3548	8.87	8.87	*****
2	0.3023	7.56	16.43	*****
3	0.0413	1.03	17.46	*****
4	0.0403	1.01	18.47	*****
5	0.0394	0.98	19.45	*****
6	0.0381	0.95	20.41	*****
7	0.0374	0.94	21.34	*****
8	0.0368	0.92	22.26	*****
9	0.0359	0.90	23.16	*****
10	0.0355	0.89	24.04	*****

(b)				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	4.8191	12.05	12.05	*****
2	3.4018	8.50	20.55	*****
3	1.0791	2.70	23.25	*****
4	1.0501	2.63	25.88	*****
5	1.0404	2.60	28.48	*****
6	1.0013	2.50	30.98	*****
7	0.9840	2.46	33.44	*****
8	0.9765	2.44	35.88	*****
9	0.9513	2.38	38.26	*****
10	0.9414	2.35	40.61	*****

O primeiro eixo pode ser interpretado como sendo o construto **B**, sendo que as categorias de resposta deste apresentam contribuição apenas neste eixo, sem qualquer contribuição no segundo. Ocorre algo semelhante no segundo eixo, com a caracterização de **A**, sendo que este também apresenta contribuição somente neste eixo. Logo as cargas fatoriais  $\eta_s$  serão nulas onde as contribuições também forem.

Um fato interessante, observado no mapa da ACM (graf. 4.2), é a forma como as categorias se agruparam nos dois eixos. Apesar de em ambos eixos estas terem se agrupado em três nuvens, em **B**, eixo 1, as categorias 1 e 2 formaram uma nuvem, seguida pela nuvem formada pela categoria 3 e por último a nuvem formada por 4 e 5. Em **A** a categoria 1 forma uma nuvem, enquanto a nuvem seguinte é formada por pontos de 2 e 3, seguida pela nuvem formada por 4 e 5. Pode-se explicar esta diferença na formação das nuvens pela estrutura de dependência imposta. Como dito em 4.2, a matriz **A** é menos rigorosa quanto às transições de uma categoria (estado) maior para uma menor, sendo que desta forma a distância entre 2 e 3 é menor do que a que ocorre em **B**, mais rigorosa, forçando a categoria 2 estar mais próxima de 1 que de 3.

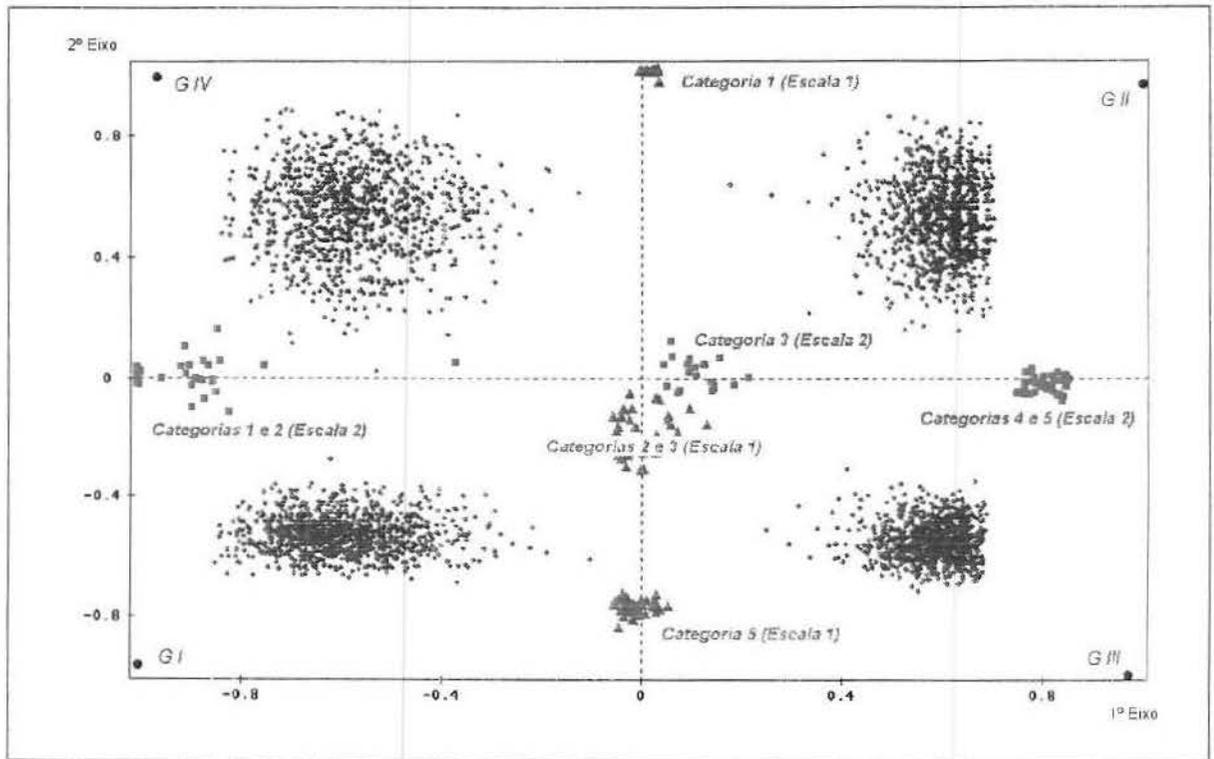
Quanto aos indivíduos a discriminação neste caso é total, como apontado pelas variáveis ilustrativas dos perfis. O interessante neste caso é notar a disposição em quadrantes e o formato das nuvens. Os quadrantes podem ser definidos de acordo com as categorias de resposta presentes nas extremidades dos eixos. Logo, seguindo a convenção (sentido horário), o quadrante 1 é caracterizado pela presença de categorias maiores de **B** e de categorias menores de **A**, coerente com os indivíduos de GII descritos anteriormente. O mesmo ocorre nos quadrantes seguintes. Na AF esta discriminação apesar de clara é menos contrastante que a ACM (graf. 4.4(b)).

As diferenças e a importância da métrica adotada neste tipo de problema está bem ilustrada neste caso. A resolução com a métrica obtida pela Escala Ótima permite observar melhor as diferenças entre os grupos. Podemos explicar esta melhora pelo fato desta métrica sempre levar a informação das marginais de resposta dos indivíduos no cálculo dos escores fatoriais (coordenadas), que de fato é o que os diferencia neste estudo. Esta informação é perdida quando calculamos as cargas e posteriormente os escores fatoriais dos indivíduos na AF.

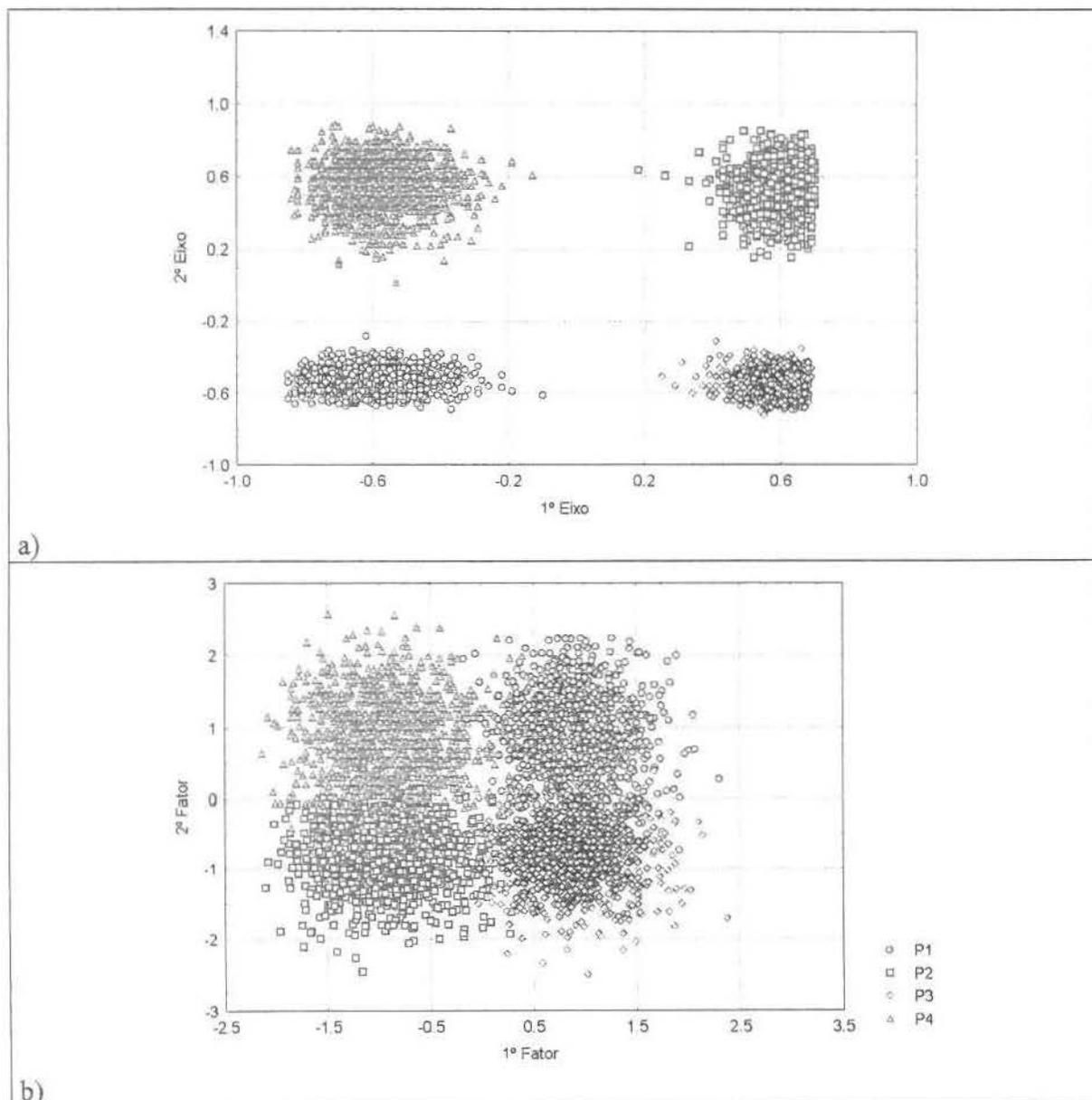
---

tab. 4.7- Cargas fatoriais obtidas para Simulação 2

	Fator 1	Fator 2	$\eta_1$	$\eta_2$
1	-0.0416	-0.3671	0.0000	0.7539
2	-0.0091	-0.4208	0.0000	0.7853
3	0.0025	-0.3961	0.0000	0.7853
4	-0.0222	-0.4081	0.0000	0.7776
5	0.0048	-0.4267	0.0000	0.7697
6	0.0061	-0.4078	0.0000	0.7776
7	-0.0051	-0.4154	0.0000	0.7776
8	0.0001	-0.4347	0.0000	0.7776
9	-0.0129	-0.3962	0.0000	0.7619
10	0.0011	-0.4144	0.0000	0.7776
11	0.0247	-0.3997	0.0000	0.7697
12	-0.0099	-0.4170	0.0000	0.7853
13	-0.0172	-0.4096	0.0000	0.7776
14	0.0024	-0.4173	0.0000	0.7853
15	0.0015	-0.3882	0.0000	0.7776
16	-0.0024	-0.4060	0.0000	0.7776
17	-0.0001	-0.4205	0.0000	0.7930
18	-0.0036	-0.4267	0.0000	0.7776
19	-0.0170	-0.4073	0.0000	0.7697
20	-0.0072	-0.4044	0.0000	0.7697
21	0.4645	0.0345	0.7902	0.0000
22	0.5541	-0.0001	0.8424	0.0000
23	0.5655	-0.0221	0.8508	0.0000
24	0.5503	0.0038	0.8508	0.0000
25	0.5387	0.0120	0.8424	0.0000
26	0.5423	-0.0152	0.8339	0.0000
27	0.5464	0.0159	0.8339	0.0000
28	0.5672	-0.0177	0.8508	0.0000
29	0.5568	-0.0358	0.8508	0.0000
30	0.5511	-0.0244	0.8424	0.0000
31	0.5552	0.0088	0.8508	0.0000
32	0.5581	0.0076	0.8591	0.0000
33	0.5236	0.0180	0.8254	0.0000
34	0.5454	-0.0117	0.8508	0.0000
35	0.5680	-0.0014	0.8508	0.0000
36	0.5602	-0.0028	0.8339	0.0000
37	0.5447	0.0071	0.8424	0.0000
38	0.5637	-0.0293	0.8424	0.0000
39	0.5780	-0.0266	0.8508	0.0000
40	0.5524	0.0146	0.8424	0.0000



graf. 4.3- Mapa da AFCM para Simulação 2



graf. 4.4- Dispersão dos indivíduos na AFM (a) e AF (b) para Simulação 2

### Simulação 3

A única diferença da estrutura desta simulação em relação a anterior é a mudança do vetor de parâmetros da distribuição Multinomial utilizada. O fato de se utilizar um vetor de probabilidades uniformes nas respostas que não obedecem a uma MT deve deixar a discriminação de grupos menos clara que o caso anterior, uma vez que os indivíduos podem responder com categorias de intensidade mais alta com igual probabilidade nestes itens.

tab. 4.8- Estrutura de geração de indivíduos na Simulação 3

$C=2 \quad G=4$	$i = 1, \dots, 1000 \quad j = 1, \dots, 20$	
	<i>Escala 1</i>	<i>Escala 2</i>
GI	$p_{(j,j+1)}(s)$ segundo <b>A</b>	$h_{ij}^{(1,2)} \sim \text{Multinomial}(1, \mathbf{i})$
GII	$h_{ij-1}^{(2,2)} \sim \text{Multinomial}(1, \mathbf{i})$	$p_{(j,j+1)}(s)$ segundo <b>B</b>
GIII	$p_{(j,j+1)}(s)$ segundo <b>A</b>	$p_{(j,j+1)}(s)$ segundo <b>B</b>
GIV	$h_{ij-1}^{(4,1)} \sim \text{Multinomial}(1, \mathbf{i})$	$h_{ij-1}^{(4,2)} \sim \text{Multinomial}(1, \mathbf{i})$

Na aplicação das técnicas, observa-se uma maior diferença entre os 2 primeiros autovalores resultantes da ACM, apesar de ainda indicarem a análise de apenas duas dimensões (tab. 4.9). Na situação anterior (**Simulação 2**) os dois eixos apresentaram importância equivalente em termos dos autovalores, o que não ocorre neste caso, onde o primeiro autovalor (0.16) é quase duas vezes maior que o segundo (0.09). Nos autovalores da AF a razão também se mantém semelhante (6.05/ 3.36).

tab. 4.9- (a) Dez primeiros autovalores da ACM para Simulação 2. (b) Dez primeiros autovalores da AF para Simulação 2.

(a)			
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE
1	0.1620	4.05	4.05
2	0.0964	2.41	6.46
3	0.0378	0.95	7.41
4	0.0374	0.93	8.34
5	0.0365	0.91	9.25
6	0.0359	0.90	10.15
7	0.0357	0.89	11.04
8	0.0353	0.88	11.92
9	0.0346	0.86	12.79
10	0.0342	0.86	13.64

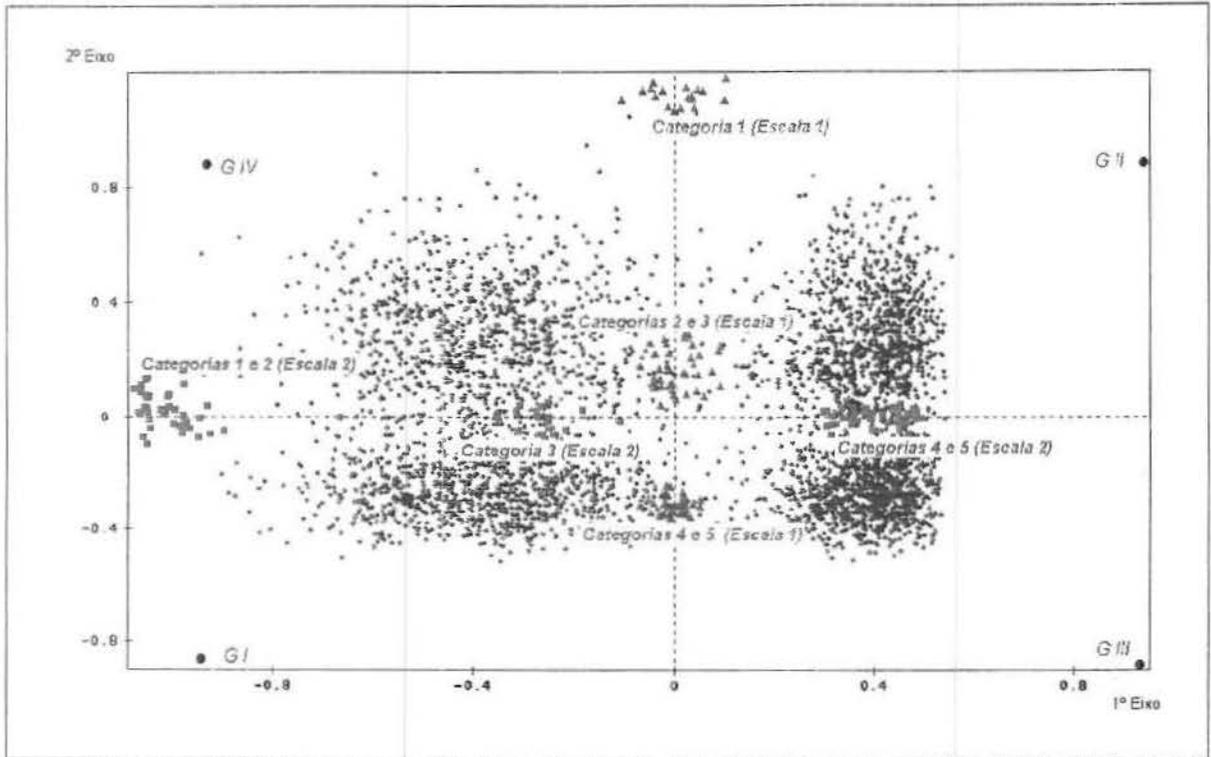
  

(b)			
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE
1	6.0532	15.13	15.13
2	3.3611	8.40	23.54
3	1.0238	2.56	26.09
4	1.0162	2.54	28.64
5	0.9995	2.50	31.13
6	0.9812	2.45	33.59
7	0.9762	2.44	36.03
8	0.9660	2.42	38.44
9	0.9447	2.36	40.80
10	0.9307	2.33	43.13

As cargas fatoriais no entanto se apresentaram iguais à situação anterior (tab. 4.10), ou seja, os itens de **B** contribuindo apenas no eixo 1 e os de **A** apenas no segundo eixo. Novamente temos itens que na ACM resultaram contribuição nula para a inércia do eixo, apresentando cargas fatoriais iguais a zero, e valores mais significativos para as cargas  $\eta$  em relação às cargas da AF.

O que se observa no gráfico (graf. 4.3) é uma maior proximidade das categorias de resposta, principalmente entre as categorias de maior intensidade, ou seja, as categorias 4 e 5 se aproximaram mais de 2 e 3 nos dois construtos identificados. A presença de respostas de intensidade alta nos indivíduos nos itens sem estrutura de dependência modificou as distâncias entre as intensidades de sintoma devido à mudança das marginais de resposta. Observa-se assim uma maior dispersão dos pontos nas nuvens de categorias de maior intensidade.

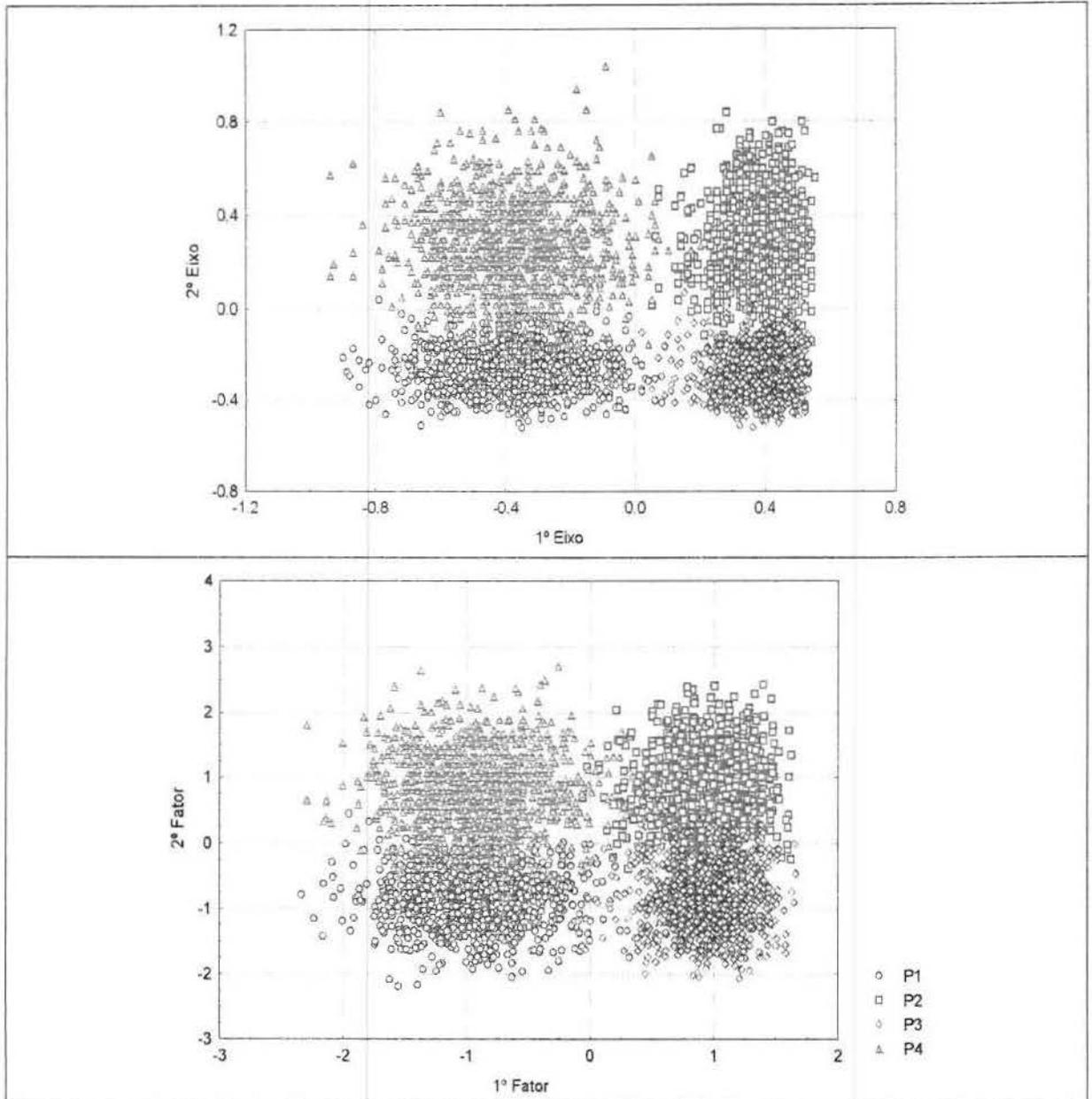
Quanto aos indivíduos a discriminação ainda é boa (graf. 4.6(a)), apesar de já não ser mais total como no caso anterior. Os perfis de indivíduos se confundiram principalmente em relação ao segundo eixo, de forma coerente com a relação entre os autovalores já citada. Na AF ocorre o mesmo (graf. 4.6(b)), atentando-se apenas ao fato da inversão de dois perfis quanto a disposição no gráfico anterior (graf. 4.4(b)).



graf. 4.5- Mapa da AFCM para Simulação 3

tab. 4.10- Cargas fatoriais obtidas para Simulação 3.

	<b>Fator 1</b>	<b>Fator 2</b>	$\eta_1$	$\eta_2$
1	-0.5002	-0.0268	0	0.7617
2	-0.5246	0.0195	0	0.8742
3	-0.5159	0.0231	0	0.8156
4	-0.4977	0.0048	0	0.8496
5	-0.4926	0.0014	0	0.8496
6	-0.4670	-0.0109	0	0.8496
7	-0.4574	-0.0087	0	0.8412
8	-0.4941	0.0020	0	0.8496
9	-0.4841	0.0352	0	0.8578
10	-0.4994	0.0224	0	0.8496
11	-0.5190	0.0053	0	0.8327
12	-0.4741	-0.0128	0	0.8242
13	-0.4834	0.0173	0	0.8242
14	-0.5157	-0.0108	0	0.8578
15	-0.5025	-0.0094	0	0.7891
16	-0.4895	-0.0167	0	0.8327
17	-0.4821	-0.0021	0	0.8578
18	-0.4544	0.0298	0	0.8578
19	-0.4777	-0.0025	0	0.8496
20	-0.4716	0.0199	0	0.8156
21	0.0076	0.4174	0.7138	0
22	0.0129	0.4264	0.8412	0
23	-0.0257	0.4076	0.8578	0
24	-0.0121	0.4291	0.8412	0
25	-0.0068	0.4109	0.8242	0
26	0.0054	0.4188	0.8327	0
27	0.0402	0.3806	0.8327	0
28	0.0065	0.4301	0.8742	0
29	0.0070	0.3710	0.8412	0
30	0.0161	0.3983	0.8412	0
31	0.0319	0.4145	0.8496	0
32	0.0076	0.4296	0.8412	0
33	0.0160	0.4186	0.8156	0
34	-0.0125	0.4090	0.8327	0
35	-0.0111	0.4204	0.8661	0
36	-0.0170	0.4175	0.8578	0
37	-0.0098	0.3815	0.8242	0
38	0.0017	0.4194	0.8578	0
39	0.0087	0.4186	0.8822	0
40	0.0304	0.4147	0.8412	0



graf. 4.6- Dispersão dos indivíduos na AFM (a) e AF (b) para Simulação 3

### Simulação 4

Neste caso será introduzida uma marginal de resposta bastante encontrada em aplicações reais e que apresentam maior dificuldade de interpretação. Trata-se de uma distribuição bimodal nos extremos (em forma de U), fornecidas pelas distribuições Multinomiais com vetor de parâmetros  $\mathbf{u}$  (4.3). A tabela 4.11 apresenta a estrutura que neste caso é de 3 grupos e apenas um perfil com estrutura de dependência modelada, por facilitar a interpretação dos efeitos da distribuição introduzida.

tab. 4.11- Estrutura de geração de indivíduos na Simulação 4

$C=2 \quad G=3$	$i = 1, \dots, 1000 \quad j = 1, \dots, 20$	
	<i>Escala 1</i>	<i>Escala 2</i>
GI	$h_{ij-1}^{(1,1)} \sim \text{Multinomial}(1, \mathbf{u})$	$h_{ij-1}^{(1,2)} \sim \text{Multinomial}(1, \mathbf{u})$
GII	$h_{ij-1}^{(2,1)} \sim \text{Multinomial}(1, \mathbf{i})$	$p_{(j,j+1)}(s)$ segundo <b>B</b>
GIII	$h_{ij-1}^{(3,1)} \sim \text{Multinomial}(1, \mathbf{i})$	$h_{ij-1}^{(3,2)} \sim \text{Multinomial}(1, \mathbf{i})$

A tabela 4.3 apresenta os 10 primeiros autovalores resultantes. Ainda é marcante a presença de duas dimensões. A relação entre os dois primeiros autovalores na ACM é semelhante ao ocorrido na simulação 3. No entanto na AF o histograma indica apenas um fator, sendo que a razão entre os dois primeiros autovalores (aproximadamente 4) é maior que no caso da ACM.

Na ACM os dois eixos não definem dois construtos como nos casos anteriores, ambos compartilham do mesmo plano sem formar dimensões próprias. Não há uma interpretação clara e objetiva. A presença da bimodalidade reorienta os extremos de intensidade de cada construto gerando uma "circularidade", ou seja, a "ferradura" se fecha. O construto **B** domina o plano, imprimindo a ferradura (como no caso de um construto, **Simulação 1**), e **A** fecha esta ferradura exibindo a quebra de gradiente presente na bimodalidade.

tab. 4.12- (a) Dez primeiros autovalores da ACM para Simulação 2. (b) Dez primeiros autovalores da AF para Simulação 4.

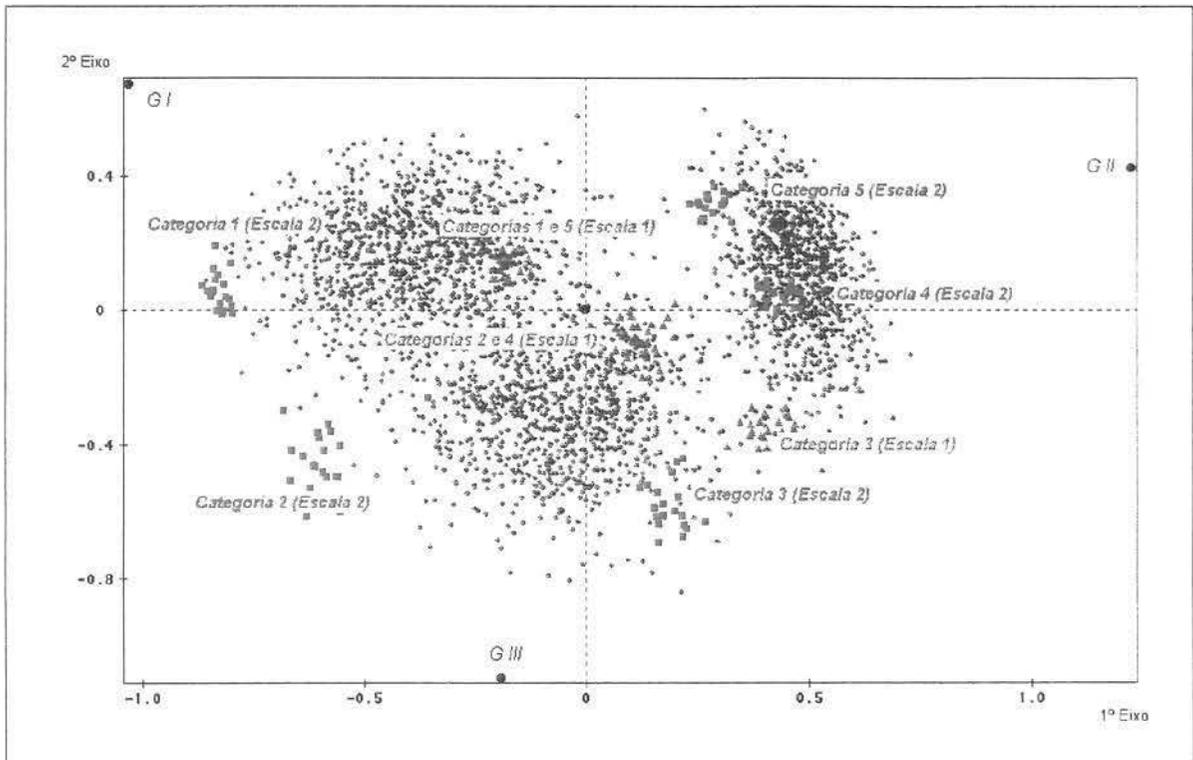
(a)				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.1479	3.70	3.70	.....
2	0.0700	1.75	5.45	.....
3	0.0378	0.95	6.39	.....
4	0.0367	0.92	7.31	.....
5	0.0360	0.90	8.21	.....
6	0.0356	0.89	9.10	.....
7	0.0355	0.89	9.99	.....
8	0.0351	0.88	10.87	.....
9	0.0349	0.87	11.74	.....
10	0.0347	0.87	12.61	.....

(b)				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	4.6498	11.62	11.62	.....
2	1.1590	2.90	14.52	.....
3	1.1376	2.84	17.37	.....
4	1.1281	2.82	20.19	.....
5	1.0925	2.73	22.92	.....
6	1.0800	2.70	25.62	.....
7	1.0659	2.66	28.28	.....
8	1.0577	2.64	30.93	.....
9	1.0561	2.64	33.57	.....
10	1.0479	2.62	36.19	.....

Mais detalhadamente, observa-se que o eixo 1 indica a oposição entre a Presença e a Ausência de sintoma nos pontos pertencentes ao construto **B**. O segundo eixo seria a ordenação das intensidades de resposta deste construto. Os pontos de **A** apresentam comportamento bastante diferente de **B**, uma vez que não foi imposta qualquer estrutura de dependência. Observa-se que os pontos da categoria 5 e 1 de **A** formaram uma nuvem única, efeito da distribuição bimodal imposta (quebra do gradiente). Nota-se também que a relação dos pontos destes itens apresenta uma relação linear com os dois eixos, formando uma diagonal de quase 45°. As nuvens de **B** também se separaram em 5 diferente dos casos anteriores, uma para cada categoria, apesar das distâncias entre as mesmas se mostrarem coerentes ainda com as anteriores, ou seja, as que estavam em uma mesma nuvem continuam bastante próximas.

As cargas fatoriais encontradas também não possuem mais a estrutura simples de anteriormente. A identificação dos construtos se dá mais de forma visual, onde há clara discriminação entre as nuvens de pontos das duas estruturas impostas. Tal fato é bastante comum em problemas práticos na psicologia.



graf. 4.7- Mapa da AFCM para Simulação 4

---

No entanto no caso dos coeficientes da ACM observa-se um contraste no primeiro eixo, enquanto na AF não há exatamente este contraste, mas apenas a falta de carga dos 20 primeiros itens.. O segundo eixo é de difícil interpretação em ambos casos, pois em ambos casos as dimensões não definem construtos de forma exclusiva. Mas nas cargas da ACM os itens de 21 a 40 apresentam valores maiores em relação aos itens anteriores.

Quanto aos indivíduos, há clara discriminação daqueles que apresentam a dominância do construto **B** dos demais (GII, graf. 4.8(a)). Observa-se ainda assim duas nuvens nos quadrantes 3 e 4 do gráfico, que no entanto não correspondem literalmente aos dois grupos restantes neste caso de simulação, pois estes se confundem.

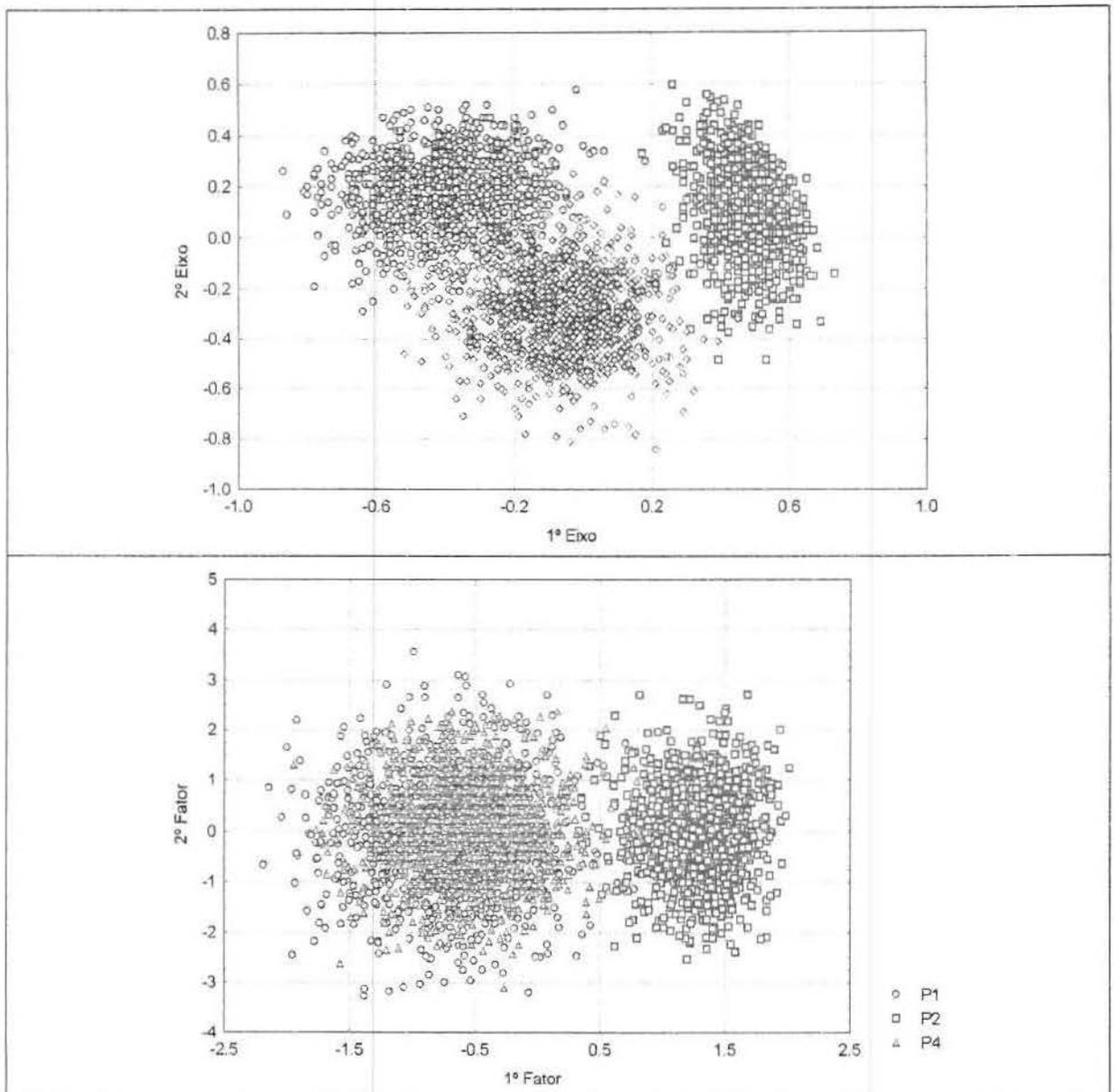
Na AF a separação dos grupos é bem menos clara, sendo que os perfis 1 e 2 se confundem totalmente (graf. 4.8(b)). Podemos atribuir este fato à perda de informação causada pela métrica utilizada na AF, ou seja, as marginais de resposta uniformes e bimodais trazem a mesma informação para a estrutura de correlação do modelo descrito em 2.1.

Devido ao formato das marginais, principalmente no caso bimodal, é praticamente impossível que se extraia fatores baseados em correlações lineares, como pode-se observar nas cargas fatoriais do **Fator 1** da AF dos 20 primeiros itens (tab. 4.13), onde não há nenhuma estrutura de dependência de resposta imposta sendo gerados apenas pelas marginais em questão. Logo estes itens não contribuem na discriminação dos grupos no primeiro eixo. O mesmo não ocorre na ACM, onde a diferença de forma destas marginais é importante na métrica Qui-quadrado e trazem contribuições à inércia dos eixos. A separação dos grupos ocorre portanto quanto às intensidades das categorias de respostas, ou seja, indivíduos de perfis de resposta semelhantes tendem a formar grupos.

---

tab. 4.13- Cargas fatoriais obtidas para Simulação 4

	Fator 1	Fator 2	$\eta_1$	$\eta_2$
1	-0.0024	0.3505	0.2773	0.2117
2	-0.0241	0.3067	0.2432	0.1058
3	0.0459	0.2555	0.3077	0.2245
4	0.0293	-0.1151	0.2664	0.1673
5	0.0127	-0.0001	0.2878	0.2245
6	0.0115	-0.2545	0.2307	0.1673
7	0.0097	-0.0005	0.2878	0.1980
8	0.0144	-0.0027	0.2551	0.2698
9	0.0242	0.2146	0.3263	0.2117
10	-0.0091	-0.1138	0.3077	0.1587
11	0.0526	0.0379	0.2773	0.3960
12	0.0331	0.0885	0.2551	0.2898
13	0.0138	-0.1509	0.2878	0.2245
14	0.0399	0.2529	0.2878	0.1058
15	0.0290	0.0717	0.3263	0.1296
16	0.0207	0.2953	0.2551	0.1058
17	0.0062	0.1117	0.3171	0.3219
18	0.0125	-0.4191	0.2773	0.2049
19	0.0273	0.2519	0.2664	0.0917
20	-0.0085	0.2095	0.2432	0.2117
21	0.4295	-0.0742	-0.4418	0.2482
22	0.4901	0.1057	-0.4485	0.3305
23	0.4939	0.1351	-0.4418	0.4643
24	0.4952	-0.0328	-0.4925	0.3510
25	0.4766	0.0663	-0.4865	0.2538
26	0.4861	-0.0188	-0.5273	0.2482
27	0.4621	-0.0140	-0.4679	0.4733
28	0.4731	-0.0313	-0.4351	0.4490
29	0.4926	0.0557	-0.4679	0.1755
30	0.4893	0.0054	-0.4615	0.1497
31	0.4618	-0.0497	-0.4615	0.2592
32	0.4889	-0.0302	-0.4485	0.4363
33	0.4842	0.0673	-0.4679	0.4427
34	0.4725	0.0543	-0.4741	0.2425
35	0.4819	-0.1153	-0.4865	0.1980
36	0.5061	-0.0747	-0.5160	0.1673
37	0.4736	-0.0359	-0.4351	0.2117
38	0.4879	-0.1191	-0.4550	0.2182
39	0.5204	0.1079	-0.4741	0.2182
40	0.4562	-0.1055	-0.3997	0.2117



graf. 4.8- Dispersão dos indivíduos na AFCM (a) e AF (b) para Simulação 4

### Simulação 5

Este é o caso mais complexo estudado. Todas as estruturas de dependência e de marginais de resposta estão presentes, o que se espera que cause alguma dificuldade na interpretação dos resultados, principalmente quanto à discriminação dos grupos. A tabela 4.14 apresenta a estrutura dos dados simulados. Os grupos GI e GIV entram como confundidores na estrutura simples que formariam GII e GIII pelas Ausência/Presença marcadas da característica representada pelo construto, como no caso das **Simulação 2**.

tab. 4.14- Estrutura de geração de indivíduos na Simulação 5

$C=2$ $G=4$	$i = 1, \dots, 1000$ $j = 1, \dots, 20$	
	<i>Escala 1</i>	<i>Escala 2</i>
GI	$h_{ij-1}^{(1,1)} \sim \text{Multinomial}(1, \mathbf{u})$	$h_{ij-1}^{(1,2)} \sim \text{Multinomial}(1, \mathbf{u})$
GII	$p_{(j,j+1)}(s)$ segundo <b>A</b>	$h_{ij}^{(2,2)} \sim \text{Multinomial}(1, \mathbf{a})$
GIII	$h_{ij-1}^{(3,2)} \sim \text{Multinomial}(1, \mathbf{a})$	$p_{(j,j+1)}(s)$ segundo <b>B</b>
GIV	$h_{ij-1}^{(4,1)} \sim \text{Multinomial}(1, \mathbf{i})$	$h_{ij-1}^{(4,2)} \sim \text{Multinomial}(1, \mathbf{i})$

Os autovalores já mostram um comportamento bem diferente (tab. 4.15). O primeiro é discrepantemente maior que os demais, apesar do formato do diagrama sugerir a presença de mais 2 dimensões, apesar de seus valores pouco significativos em comparação ao primeiro. Na AF apenas 1 dimensão a mais é indicada.

tab. 4.9- (a) Dez primeiros autovalores da ACM para Simulação 2. (b) Dez primeiros autovalores da AF para Simulação 5.

(a)				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.3095	7.74	7.74	*****
2	0.0720	1.80	9.54	*****
3	0.0546	1.37	10.90	*****
4	0.0356	0.89	11.79	*****
5	0.0347	0.87	12.66	*****
6	0.0344	0.86	13.52	*****
7	0.0339	0.85	14.37	*****
8	0.0338	0.84	15.21	*****
9	0.0330	0.82	16.04	*****
10	0.0327	0.82	16.85	*****

(b)				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	11.6180	29.05	29.05	*****
2	1.8023	4.51	33.55	*****
3	1.0260	2.57	36.12	*****
4	0.9652	2.41	38.53	*****
5	0.9572	2.39	40.92	*****
6	0.9258	2.32	43.24	*****
7	0.9047	2.26	45.50	*****
8	0.8766	2.19	47.69	*****
9	0.8582	2.15	49.84	*****
10	0.8420	2.11	51.94	*****

Os autovalores poderiam sugerir um único fator. As cargas fatoriais em ambos casos apresentaram comportamento semelhante, sendo que na segunda dimensão a ACM apresentou cargas um pouco maiores nos itens de 21 a 40 em comparação a AF. A AF também apresentou a oposição no primeiro eixo como a ACM. Pode ser entendido como o formato cruzado entre Ausência/Presença nos grupos GII e GIII das estruturas de dependência.

Ao observarmos o gráfico entendemos melhor o que está acontecendo. Os construtos teóricos neste caso apresentaram comportamento totalmente oposto em relação ao primeiro eixo, ou seja, as direções das intensidades de sintoma estão invertidas, de forma coerente com a estrutura imposta na simulação. O segundo eixo ainda neste caso apresenta a distinção entre os níveis de intensidade de sintomas, neste caso na mesma direção para ambos construtos. No entanto a presença do perfil de resposta bimodal faz com que no segundo eixo as categorias 1 e 5 estejam próximas, revelando novamente a quebra de gradiente presente nas marginais bimodais. Observa-se também que a ferradura formada pelos itens do construto **B** (21 a 40) é mais aberta, mostrando um distanciamento maior principalmente entre as categorias de maior intensidade de resposta. Em **A** estas categorias estão mais confundidas, estando muito próximas principalmente na primeira dimensão. O

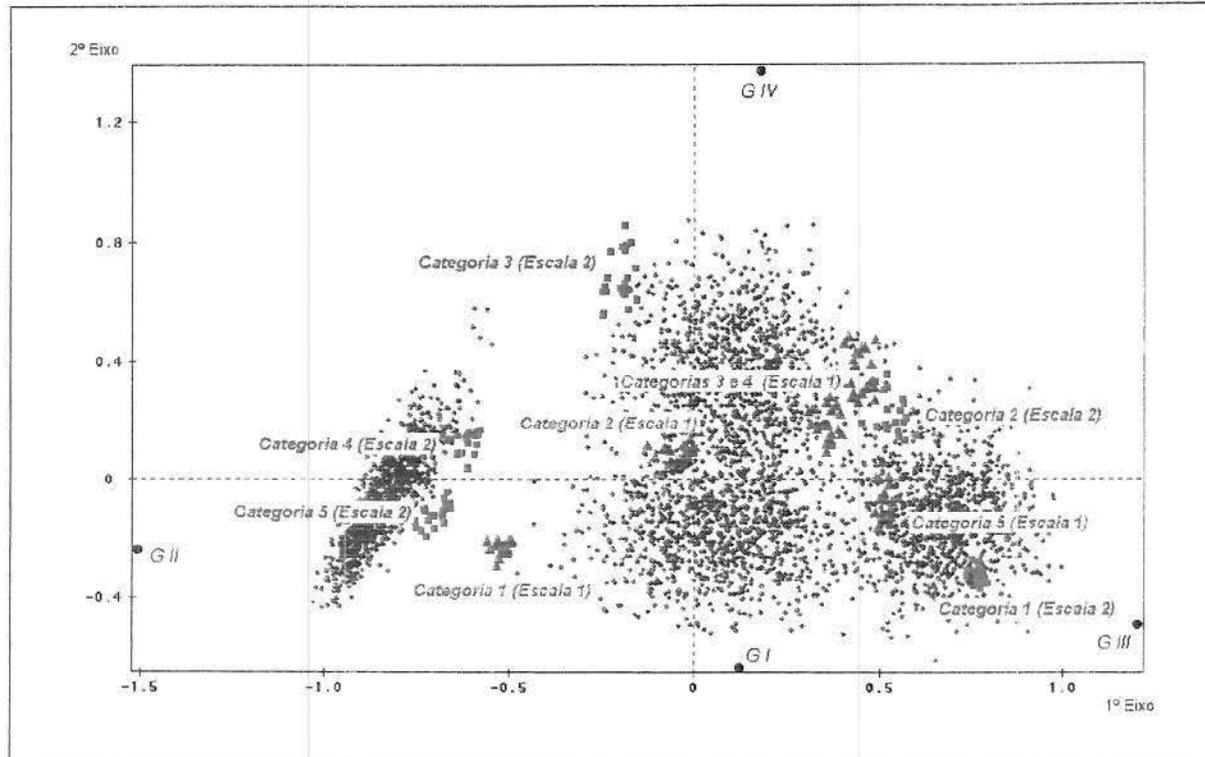
---

segundo eixo pode ser interpretado como a distinção entre as categorias de maior intensidade, e a revelação da bimodalidade marginal das respostas.

O terceiro eixo não revelou importância quanto às caracterizações dos construtos, mas terá papel importante na discriminação dos indivíduos na ACM.

Quanto aos indivíduos, aqueles com o construto **B** predominante (GII) se apresentaram bem distantes dos demais de forma coerente com a localização das categorias de maior intensidade de **B**. Aqueles com predominância de **A** (GIII) apresentam uma separação um pouco mais fraca, sendo confundido com os grupos uniforme e bimodal. Na AF (graf. 4.10(b)) a separação existe entre os perfis GII e GIII, sendo que GI e GIV se confundem como no caso anterior (**Simulação 5**).

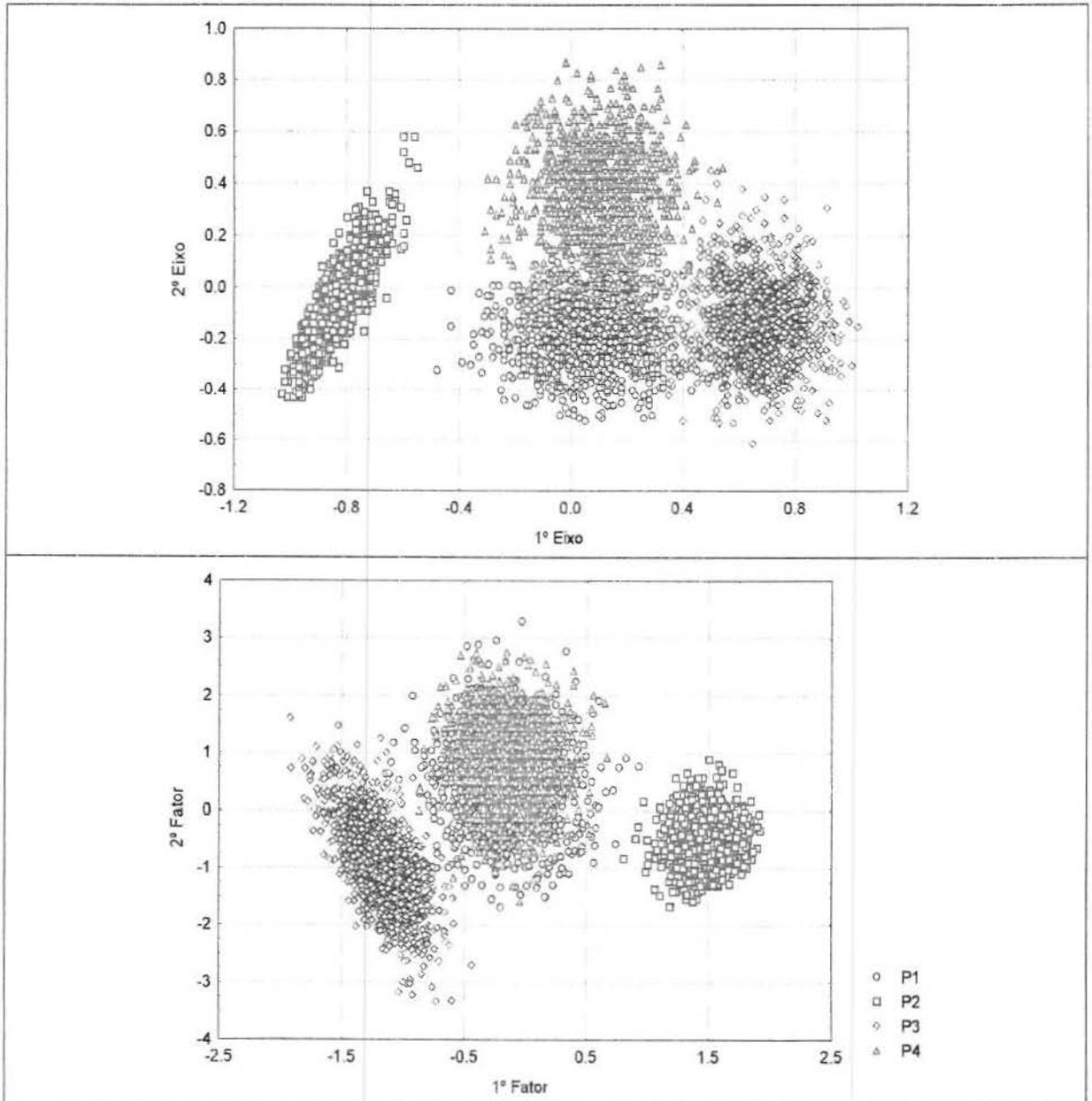
---



graf. 4.8- Mapa da AFCM para Simulação 5

tab. 4.10- Cargas Fatoriais obtidas para Simulação 5

	Fator 1	Fator 2	$\eta_1$	$\eta_2$
1	-0.4197	0.2553	0.4451	0.2147
2	-0.4265	0.2216	0.4588	0.2147
3	-0.3851	0.2347	0.4163	0.2008
4	-0.3994	0.2766	0.4309	0.1780
5	-0.4105	0.2851	0.4451	0.2339
6	-0.4371	0.2501	0.4588	0.2008
7	-0.4055	0.2830	0.4451	0.2400
8	-0.4046	0.2970	0.4309	0.2277
9	-0.4086	0.2602	0.4309	0.1859
10	-0.4019	0.2537	0.4309	0.2517
11	-0.4089	0.2378	0.4309	0.1859
12	-0.4268	0.2533	0.4451	0.2147
13	-0.4208	0.2268	0.4451	0.1935
14	-0.4123	0.2036	0.4309	0.2277
15	-0.4174	0.2301	0.4451	0.2400
16	-0.4045	0.2437	0.4309	0.2277
17	-0.4121	0.2737	0.4309	0.1935
18	-0.4233	0.2915	0.4451	0.2277
19	-0.4113	0.2318	0.4309	0.2277
20	-0.4105	0.2051	0.4451	0.2459
21	0.5958	0.1577	-0.6094	0.2939
22	0.6385	0.1441	-0.6488	0.2890
23	0.6484	0.1822	-0.6583	0.3640
24	0.6538	0.1160	-0.6583	0.3478
25	0.6468	0.1743	-0.6676	0.3308
26	0.6307	0.1555	-0.6488	0.2890
27	0.6370	0.1677	-0.6583	0.2683
28	0.6409	0.1495	-0.6583	0.2840
29	0.6384	0.1983	-0.6488	0.3220
30	0.6387	0.1598	-0.6488	0.3264
31	0.6404	0.2063	-0.6583	0.3175
32	0.6376	0.1390	-0.6488	0.2840
33	0.6372	0.1751	-0.6488	0.2736
34	0.6399	0.1599	-0.6488	0.3351
35	0.6466	0.1339	-0.6583	0.3264
36	0.6509	0.1585	-0.6583	0.3436
37	0.6408	0.1556	-0.6488	0.3308
38	0.6482	0.1349	-0.6583	0.2988
39	0.6599	0.1849	-0.6676	0.3036
40	0.6432	0.1734	-0.6583	0.2736



graf. 4.10- Dispersão dos indivíduos na AFM (a) e AF (b) para Simulação 5

## 4.5- Escalas Ótimas nas Simulações

Esta item trata do estudo de comparação da Escala Ótima (EO) obtida, ou seja, o vetor  $\phi_j$  (definido em (3.2.14)) com a Escala Original (EO) adotada, (1, 2, 3, 4, 5). A representação de ambas em um gráfico de linhas é uma boa forma de estudarmos o comportamento. O tratamento da escala adotada como variável contínua é perfeitamente justificado quando o gráfico de linhas exibir uma reta de 45°, considerando que a ordenada e a abscissa do gráfico estejam na mesma escala.

Na **Simulação 1**, com 20 itens, são mostrados todos eles (graf. 4.11). Observa-se que são praticamente iguais em seu comportamento. Observamos uma boa ordenação nas categorias de 0 a 3, mas sem distinção entre 3 e 4. Ou seja, esta escala, nesta população não discrimina entre as categorias 3 e 4. Note também que de 2 para 3 há uma descontinuidade na métrica, ou seja, a distância entre as categorias 2 e 3 é menor que entre 1 e 2, e 1 e 0.

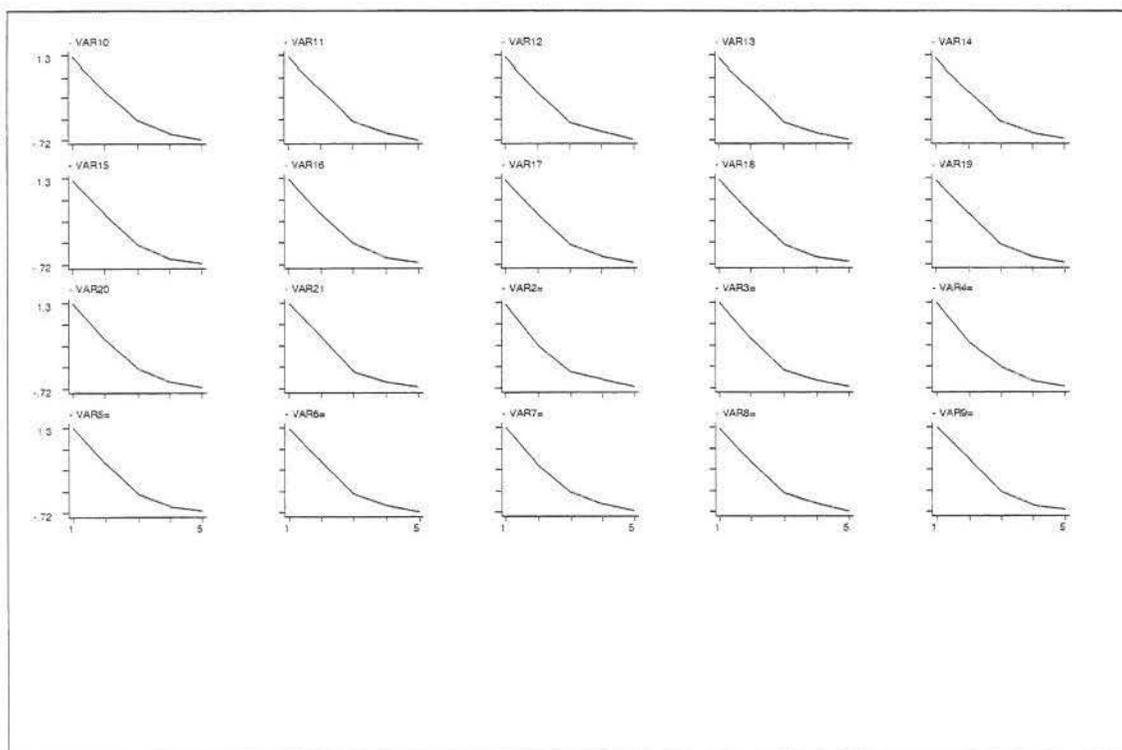


fig. 4.11- Gráfico de linha entre  $\phi_1$  x Escala Original para Simulação 1

Para as demais simulações são apresentados os gráficos de um item apenas sendo que os demais apresentam-se praticamente idênticos, como na Simulação 1. A Simulação 3 também apresentou comportamento muito semelhante a 2 e não é apresentada.

Na Simulação 2 (graf. 4.12) observa-se nitidamente a separação das duas escalas, uma em cada eixo, fato explicado pelas discriminações nulas apresentadas (tab. 4.7). Observe que a Escala A (EA) no eixo 1 e a B (EB) no eixo 2 são praticamente constantes. Do ponto de vista prático pode-se dizer que as escalas A e B são unidimensionais e podem ser aplicadas de forma independente se for o caso. Note ainda que as direções das escalas se invertem, sendo a relação de EB com EO positiva e de EA com EO negativa. Ainda podemos dizer que na EB os níveis de resposta 3 e 4 não se diferenciam, bem como em EA. Em EA também observa-se que as categorias intermediárias (1, 2 e 3) são mais próximas que a distância entre 0 e 1.

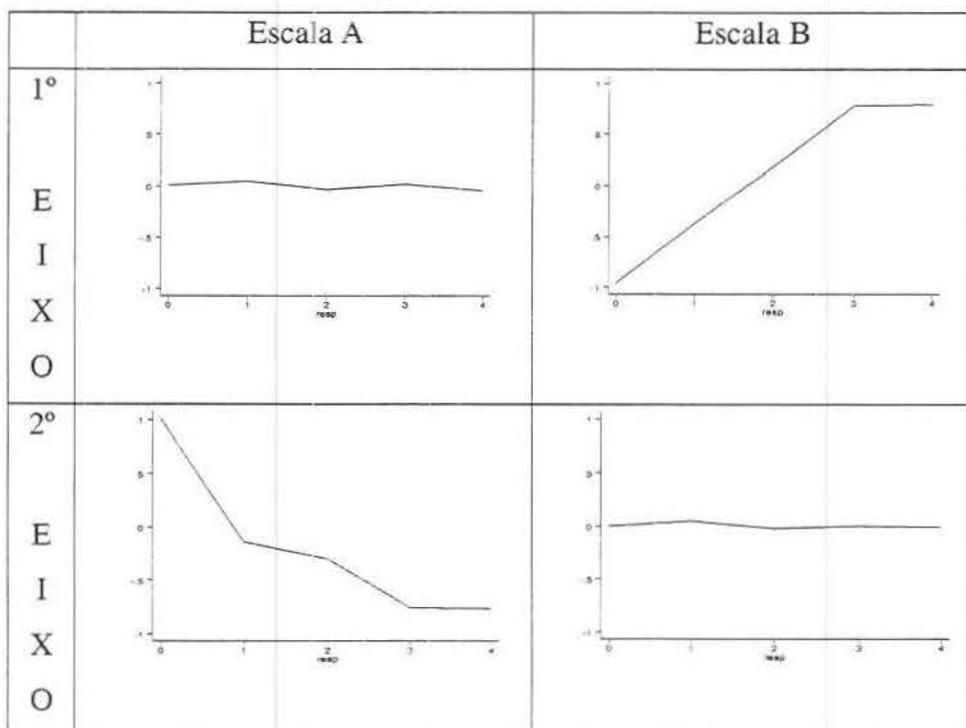


fig. 4.12- Gráficos de linha entre  $\phi_1$  (1° eixo) ou  $\phi_2$  (2° eixo) x Escala Original para Simulação 2

Na Simulação 4 (graf. 4.13) ocorre um caso interessante. A Escala A se apresenta uma relação não-linear entre as escalas ótima e original, tanto no primeiro como no segundo eixo. Esta escala, conforme já visto anteriormente, seria aquela cujos respondentes

não tiveram qualquer modelagem de dependência, caracterizando uma escala não válida para a população aplicada. A EB, no primeiro eixo apresenta uma relação linear fraca, havendo discriminabilidade apenas entre as categorias 0, 1 e 2. Em 2, 3 e 4 não ocorre uma discriminação em termos práticos. O segundo eixo também mostra comportamento pouco útil. Mas de acordo com o que foi visto na análise do mapa de ACM (graf. 4.7) esta escala pode ser considerada bidimensional. Note no gráfico citado que a ordem ocorre no plano (pontos vermelhos), e não em uma reta.

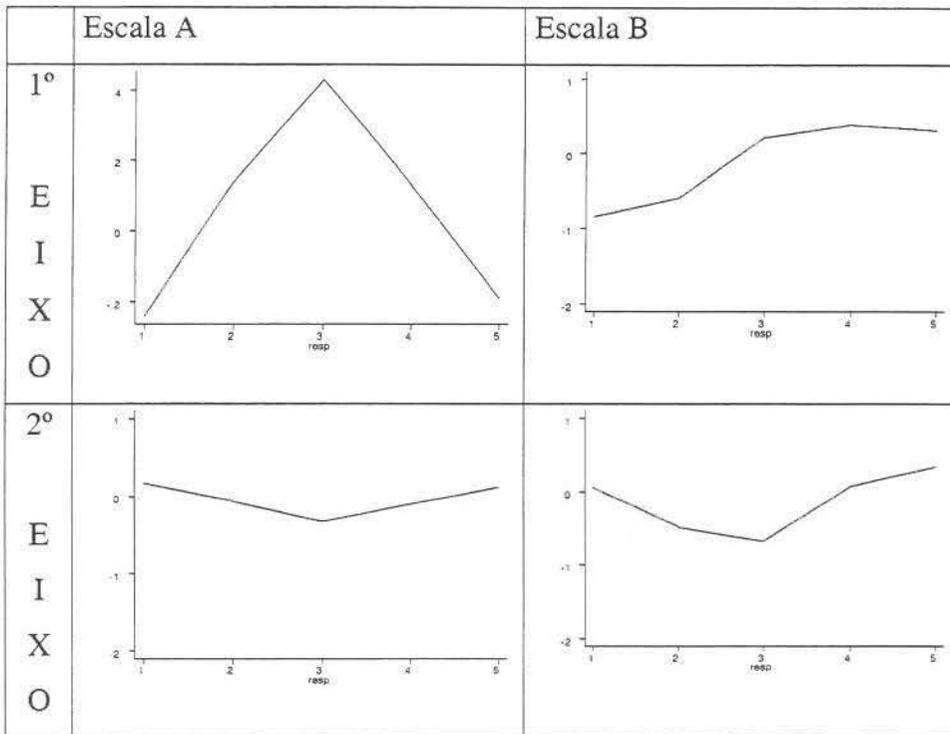


fig. 4.13- Gráfico de linha entre  $\varphi_1$  (1º eixo) ou  $\varphi_2$  (2º eixo) x Escala Original para Simulação 4

A última simulação, Simulação 5 (graf. 4.14) mostra que o segundo eixo traz pouca informação. No entanto no primeiro a ordenação se apresenta bastante razoável, no entanto com ambas escalas no mesmo eixo, mas no sentido contrário, sendo que EA tem uma relação linear positiva com a escala original e EB negativa.

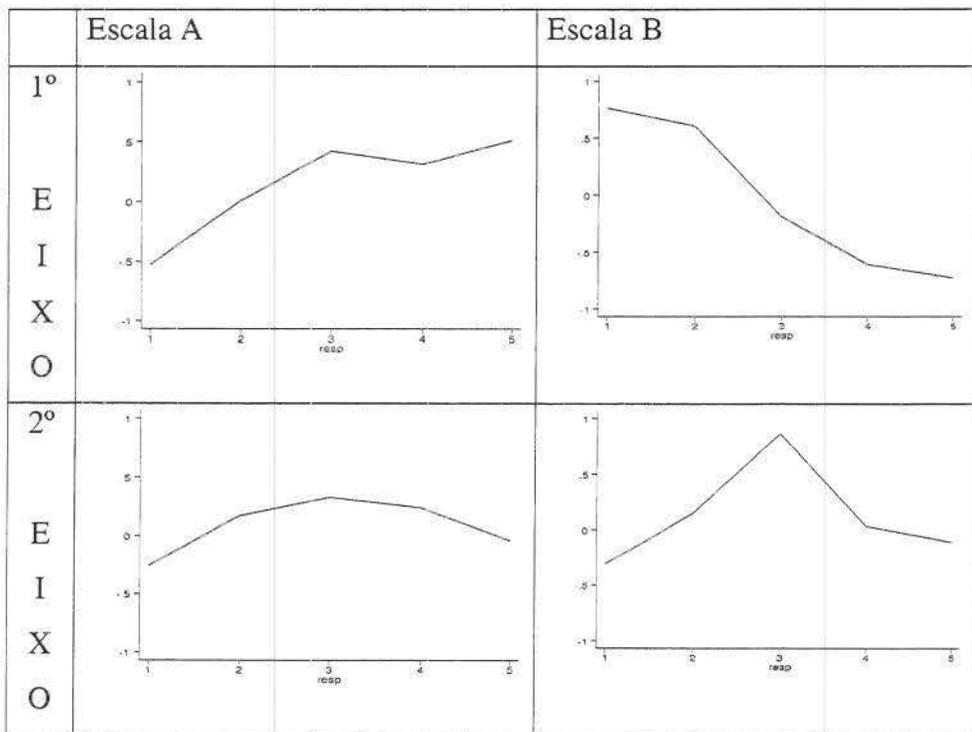


fig. 4.14- Gráfico de linha entre  $\phi_1$  (1º eixo) ou  $\phi_2$  (2º eixo) x Escala Original para Simulação 5

#### 4.6- Comentários e Conclusões

A aplicação das técnicas sobre a estrutura simulada permite observar um melhor desempenho da Análise de Correspondência Múltipla no tratamento de respostas em escala ordinal. O exame das marginais multinomiais e a ausência e presença de uma estrutura de dependência entre as respostas dos itens refletem a adequação dos instrumentos sempre levando em conta as sobre-estruturas dos indivíduos gerados.

É claro também o poder de interpretação ganho com a representação, seja ela simultânea aos indivíduos ou não, da escala de resposta dos itens. Esta representação permite identificar o comportamento desta escala métricamente assim como exibir sua configuração simultânea aos indivíduos dando então interpretabilidade às separações dos grupos observadas. Quando a estrutura de grupos quanto às marginais de resposta não é conhecida a priori, como em qualquer aplicação real, esta representação é de grande auxílio.

---

O mesmo na análise da métrica da Escala Ótima contra a original, que como visto também certifica ou não o uso de somas e médias dos pontos originais.

No caso do instrumento com apenas uma escala os desempenhos de ambas técnicas são equivalentes tanto na identificação do fator quanto na discriminação dos diferentes grupos.

Construtos avaliados por itens altamente associados sobre grupos que se diferenciam quanto à presença ou ausência dos atributos destes construtos, de forma fatorial completa, ilustram o caso mais simples, como nas **Simulações 2 e 3**, com discriminação total entre estes grupos. Nestes casos a AF apresentou cargas fatoriais mais baixas que a ACM. Quanto a discriminação fica clara a superioridade da ACM sobre a AF, onde esta última mesmo no caso mais simples não apresenta discriminação total.

Construtos que não ocorrem de forma fatorial completa, sendo que a presença de um não implica a ausência completa de outro nos diferentes grupos, tem menos poder discriminatório além de maior complexidade de interpretação e avaliação, como nos casos **4 e 5**. Nestes casos a ACM também apresentou resultados melhores na detecção dos construtos pelas cargas fatoriais, principalmente no caso **4**, enquanto a discriminação dos grupos ainda é claramente superior na ACM. A AF confunde totalmente os grupos de perfis uniforme e bimodal, como se fossem um único grupo. Vale a pena lembrar que este fato ocorre devido à métrica utilizada pelo modelo adotado na AF, que não releva o formato das marginais na estrutura de correlação.

---

## 5. Aplicação

Este capítulo trata da aplicação da técnica de ACM em dados reais. Os dados são provenientes de um estudo, coordenado pela Profa. Dra. Diana Toselo Laloni, da PUCC, com o objetivo da validação do instrumento psicométrico SCL-90-R no Brasil.

O objetivo da aplicação é estudar as propriedades psicométricas do instrumento como uma primeira etapa de validação do mesmo. Identificação de construtos, discriminação de indivíduos e dimensionalidade das características psicológicas contempladas pelo instrumento são os objetivos principais.

Não há pretensão neste trabalho de se produzir uma norma nacional nos moldes da americana, apesar de ser tratado o aspecto da métrica da escala em termos da ACM. Tal intuito abrange outros métodos (Teoria de Resposta ao Item (Andrade (2000), por exemplo) e iria além do escopo deste trabalho.

## 5.1- O questionário SCL-90-R

O *Symptom Check-list* (SCL) é um instrumento americano desenvolvido na década de 70 e de larga aplicação, utilizado principalmente com três propósitos: indicador de gravidade de distúrbios mentais, um dispositivo de triagem e uma medida de psicopatologias específicas.

É constituído de 90 itens na forma de perguntas sobre sintomas, onde o respondente classifica a intensidade deste sintoma em uma escala tipo Likert de 0 a 4. É perguntado ao indivíduo quanto a intensidade do sintoma apresentado, que responde na escala:

0. Nenhum Sintoma
1. Pouco
2. Moderado
3. Bastante
4. Muito

Nas versões europeias e americana o instrumento é auto-respondido. No caso deste estudo o mesmo é aplicado por um profissional em psicologia devido ao perfil sócio-cultural da população estudada que inviabiliza a aplicação na forma original proposta..

O SCL pode ser interpretado em 9 dimensões primárias de sintomas (distúrbios psicológicos), conforme descrito abaixo:

1. **Somatização:** reflete uma angústia surgida de percepções de disfunções corporais. Dores e desconfortos da musculatura, distúrbios gastrointestinais, respiratórios e cardiovasculares e outros somáticos equivalentes de ansiedade, são contemplados nesta dimensão.
2. **Obsessão-Compulsão:** inclui sintomas que são freqüentemente padrões da síndrome clínica de mesmo nome. Focaliza a medida em pensamentos, impulsos e ações que são entendidas pelo respondente como irresistíveis e incontroláveis, de uma natureza estranha à do indivíduo.
3. **Sensibilidade Intrapessoal:** focaliza sentimentos de inadequação e inferioridade, particularmente em comparações com outras pessoas. Indivíduos com grande intensidade destes sintomas apresentam autoconsciência aguda e

---

expectativas negativas no que diz respeito ao comportamento intra e interpessoal.

4. **Depressão:** reflete uma representação da amplitude das manifestações da depressão clínica. Afeto e ânimo disfóricos são representados como sinais de um distanciamento do interesse em viver, falta de motivação e pela perda de energia vital. São contemplados ainda sentimentos de desesperança, pensamentos suicidas, e outros correlatos somato-cognitivos com depressão.
5. **Ansiedade:** sinais gerais de ansiedade tais como nervosismo, tensão e tremores, bem como ataques de pânico, sentimentos de terror e apreensão. Alguns correlatos somáticos de ansiedade também estão incluídos.
6. **Hostilidade:** reflete pensamentos, sentimentos ou ações característicos de estados de afetividade negativa e raiva.
7. **Ansiedade Fóbica:** ansiedade fóbica é definida como um medo persistente a respostas de estímulos– a uma pessoa, lugar situação ou objeto específico – que é irracional e desproporcional ao estímulo, levando à rejeição ou fuga.
8. **Idéias Paranóides:** representa o comportamento paranóico como pensamentos desordenados. Pensamentos projectivos, hostilidade, suspeitas, grandiosidade, centralização, medo de perda de autonomia e desilusões, são tidas como reflexos primários deste distúrbio.
9. **Psicoticismo:** esta dimensão foi delineada para representar um *continuum* da experiência humana. Itens indicativos de afastamento, isolamento, estilo de vida esquizóide foram incluídos, bem como sintomas de primeira ordem de esquizofrenia tais como alucinações e sentimento de controle mental externo.

Os itens correspondentes a cada dimensão estão alocados pelo questionário, conforme a tabela 5.1. Não há informação na referência quanto a escolha da disposição dos itens. Além dos itens que contemplam dimensões, num total de 83, estão presentes ainda 7 itens adicionais, que poderiam muito bem estar em mais de uma das dimensões apresentadas. Estes itens contribuem para o índice global de severidade (discutido adiante), e auxiliam na interpretação das dimensões primárias.

---

tab. 5.1- Itens que compõem as dimensões de sintoma do SCL-90-R

Dimensão	Item
<b>Somatização</b>	1, 4, 12, 27, 40, 42, 48, 49, 52, 53, 56, 58
<b>Obsessão-Compulsão</b>	3, 9, 10, 28, 38, 45, 46, 51, 55, 65
<b>Sensibilidade Intra-Pessoal</b>	6, 21, 34, 36, 37, 41, 61, 69, 73
<b>Depressão</b>	5, 14, 15, 20, 22, 26, 29, 30, 31, 32, 54, 71, 79
<b>Ansiedade</b>	2, 17, 23, 33, 39, 57, 72, 78, 80, 86
<b>Hostilidade</b>	11, 24, 63, 67, 74, 81
<b>Ansiedade Fóbica</b>	13, 25, 47, 50, 70, 75, 82
<b>Idéias Paranóides</b>	8, 18, 43, 68, 76, 83
<b>Psicoticismo</b>	7, 16, 35, 62, 77, 84, 85, 87, 88, 90
<b>Adicionais</b>	19, 44, 59, 60, 64, 66, 89

### *5.1.1- Utilizações e Amostras apropriadas*

O SCL é designado a refletir os estados momentâneos dos sintomas psicológicos de pacientes psiquiátricos, pacientes de outras áreas e indivíduos na comunidade que não sejam pacientes correntes. Pode ser aplicado em adolescentes maiores de 13 anos, uma vez que haja um profissional capacitado para interpretar ao respondente os itens de maior dificuldade.

Pacientes com sintomas clínicos extremos, vítimas de delírios por exemplo, são considerados como fracos candidatos.

O SCL-90-R pode ser utilizado em uma única medição, ou em repetições de forma a acompanhar alguma tendência no indivíduo.

---

### 5.1.2- Índices e escores utilizados

O manual do SCL indica o uso de três tipos de escores, além do escore bruto, calculado dentro de cada uma das dimensões, que nada mais é do que a média aritmética das respostas. O Índice Global de Severidade (GSI, *Global Severity Index*), o Total de Sintomas Positivos (PST, *Positive Symptom Total*) e Índice de Sintoma Positivo de Angústia (PSDI, *Positive Symptom Distress Index*).

O GSI é a média das respostas do questionário. Ou seja, a soma dos valores dividido pelo número de respostas, sem contar as respostas em branco, e é a forma mais comum de utilização do SCL. O PST é o número de itens que apresentaram resposta não-nula. O PSDI é a razão entre a soma de todas as respostas e o PST.

Estes são os escores brutos, que após calculados são padronizados segundo 8 normas diferentes (na versão americana, Derogatis, 1994) para tipos diferentes de pacientes (Psiquiátricos, Psiquiátricos Internados, Não Psiquiátricos e Adolescentes) diferenciadas para cada sexo.

### 5.1.3- A Amostra

Para a proposta de validação foi projetado um estudo comparativo entre duas populações, sendo uma que apresenta sintomas de distúrbio mental, coletadas na triagem do ambulatório de Psiquiatria do Hospital Universitário Celso Piero de Campinas, e outra constituída de indivíduos coletados de diversos ambulatórios, que não o já citado, do mesmo hospital. A idéia deste último grupo é de controle, sem no entanto atender as definições e suposições formais de um estudo Caso-Controlle. Pretende-se estudar as propriedades psicométricas em duas populações teoricamente diferentes. Por facilidade e convenção os grupos serão referidos ao longo do trabalho como CASOS e CONTROLES.

A amostra, neste caso sistemática pois os pacientes são incluídos no estudo por ordem de chegada uma vez que atendam ao perfil, foi planejada para atender a 200 Casos e

---

400 controles, não pareados, e procurando manter uma proporção igual entre sexos, no período de Set./1999 e Julho/2000.

As tabelas a seguir descrevem algumas características sócio-econômicas da amostra. A tabela 5.1 mostra que a divisão por sexo dentro dos grupos foi equilibrada no grupo Controle, 54% mulheres e 46% de homens aproximadamente, sendo que no grupo dos Casos há um certo desequilíbrio, contando com 68.5% de mulheres e 31.5% de homens.

tab. 5.2- Distribuição de Sexo

freq. % coluna			Total
	Controles	Casos	%
<b>Feminino</b>	216	137	353
	54.14	68.50	58,93
<b>Masculino</b>	183	63	246
	45.86	31,50	41.07
<b>Total</b>	399	200	599
<b>%</b>	66.6	33.39	100.0

Com relação a escolaridade a amostra é predominantemente de indivíduos com 1º grau incompleto, sendo 72% do total da amostra. Esta proporção se mantém dentro dos grupos, conforme apontado na tabela 2. Em seguida temos os analfabetos, contando com 8.85% da amostra. Os demais níveis possuem pouca representatividade, e estão equilibrados entre os grupos de estudo. Houve um índice significativo de não respostas, com 22 não respostas nos Casos e 47 nos Controles.

tab. 5.3- Distribuição de Escolaridade

<b>freq.</b>	<b>Controles</b>	<b>Casos</b>	<b>Total</b>
<b>% coluna</b>			
<b>Analfabeto</b>	41	12	53
	10.28	6.00	8.85
<b>1º grau incompleto</b>	287	148	435
	71.93	74.00	72.62
<b>1º grau completo</b>	5	4	9
	1.25	2.00	150
<b>2º grau incompleto</b>	9	5	14
	2.26	2.50	2.54
<b>2º grau completo</b>	5	3	8
	1.25	2.50	2.54
<b>Sup. Incompleto</b>	1	1	2
	0.25	0.50	0.53
<b>Sup. Completo</b>	4	5	9
	1.00	2.50	1.50
<b>Não Resposta</b>	47	22	69
	11.78	23.50	11.52
<b>Total</b>	399	200	530
	100.0	100.0	100.0

Quanto ao estado civil, temos a predominância dos Casados, 62.1% da amostra, seguido pelos Solteiros, 21.7%. Estes percentuais se mantêm bem equilibrados entre os grupos. Observamos no entanto 12% de Separados no grupo dos Casos contra 6.8% nos Controles, ao passo que temos 9.5% de viúvos nos Controles e apenas 3% nos Casos.

A pergunta sobre trabalho apresentou um grande número de não respostas no grupo Casos, sendo que apenas 45 responderam. Tal fato inviabiliza qualquer análise sobre esta variável no grupo citado. Já nos Controles foi obtido um total de 385 respostas, sendo que 61% trabalham.

tab. 5.4- Distribuição de Estado Civil

<b>freq.</b> <b>% coluna</b>	<b>Controles</b>	<b>Casos</b>	<b>Total</b>
<b>Casado/ Amigado</b>	244 61.15	128 64.00	372 62.10
<b>Separado</b>	27 6.77	24 12.00	51 8.51
<b>Solteiro</b>	89 22.31	41 20.50	130 21.70
<b>Viúvo</b>	38 9.52	6 3.00	44 7.35
<b>Não Resposta</b>	1 0.25	1 0.50	2 0.33
<b>Total</b>	399 100.0	200 100.0	597 100.0

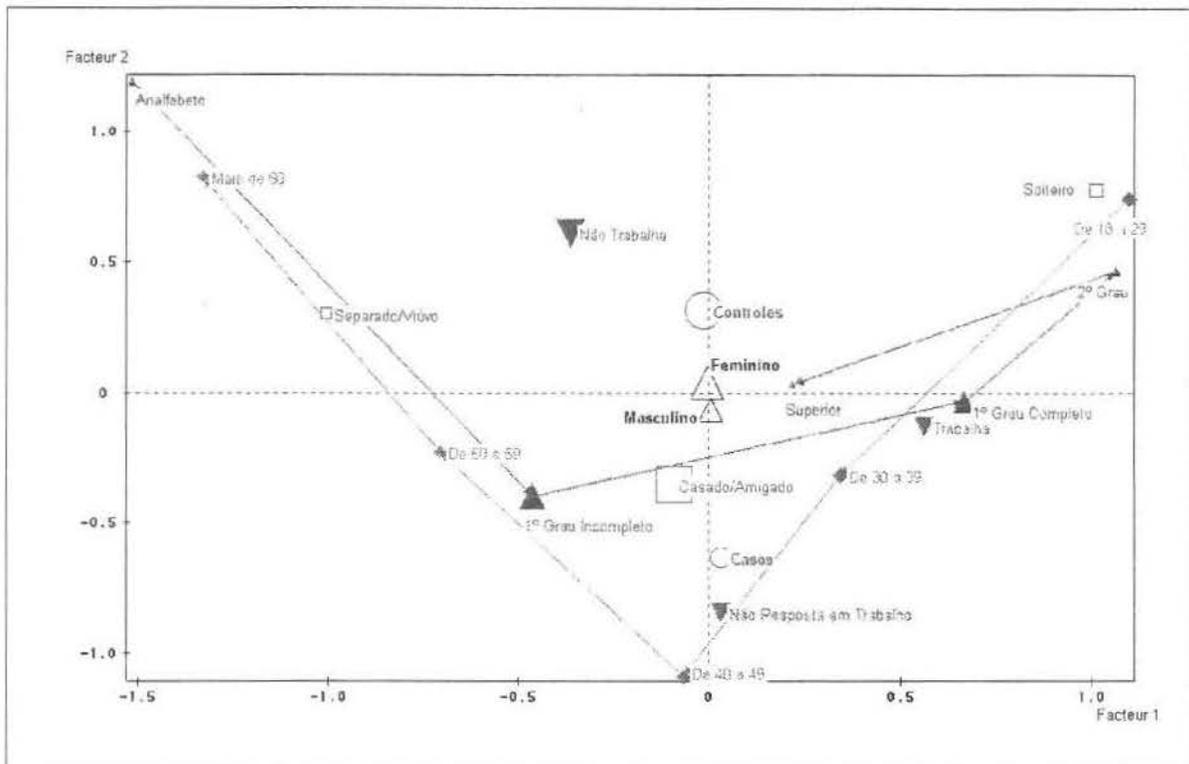
tab. 5.5- Distribuição de Sit. de Trabalho

<b>freq.</b> <b>% coluna</b>	<b>Controles</b>	<b>Casos</b>	<b>Total</b>
<b>Não Trabalha</b>	240 60.15	27 13.50	267 44.57
<b>Trabalha</b>	145 36.34	18 9.00	163 27.21
<b>Não Resposta</b>	14	155	169
<b>%</b>	3.51	77.50	28.21
<b>Total</b>	399	200	499
<b>%</b>	89.53	10.47	100.0

tab 5.6- Autovalores da ACM para variáveis demográficas

NUMERO	VALEUR PROPRE	POURCENT	POURCENT CUMULE
1	0.4510	15.03	15.03
2	0.3338	11.13	26.16
3	0.2828	9.43	35.59
4	0.2697	8.99	44.58
5	0.2603	8.68	53.25
6	0.2450	8.17	61.42
7	0.2368	7.96	69.37
8	0.2277	7.59	76.97
9	0.2140	7.13	84.10
10	0.1911	6.37	90.47
11	0.1726	5.75	96.22
12	0.1133	3.78	100.00

Os autovalores da aplicação da ACM indicam a presença de 2 dimensões. O mapa da ACM para estas variáveis (graf. 5.1) nas duas primeiras dimensões nos dá uma idéia das inter-relações. O tamanho dos símbolos é proporcional à massa de cada categoria. Os gradientes interpretáveis de forma direta, por se tratarem de variáveis ordinais, são os de Escolaridade e Faixa Etária, que seguem na mesma direção, mostrando uma relação plausível entre escolaridade e idade. Acompanhando estes dois estão as categorias de Estado Civil, também de forma coerente. Podemos distinguir os perfis de jovens (18 a 29 anos) cursando o 2º grau, ou com este completo, solteiros e que não trabalha. Este perfil é de maior predominância no grupo controle. Nota-se que os poucos analfabetos estão associados a pessoas com mais de 60 anos. Outro perfil distinto é o de pessoas na faixa dos 30 aos 49 anos, casadas/amigadas, com escolaridade distribuída entre 1º grau Completo e Incompleto sendo que estas categorias estão fortemente associadas com Trabalhar e Não Trabalhar, respectivamente no primeiro eixo. É interessante a forte presença de Não Resposta ao Trabalho neste perfil. Interessante notar também que este perfil, teoricamente mais socialmente estressado é predominante no grupo Caso.



graf. 5.1 – Mapa da ACM para variáveis demográficas da amostra

## 5.2- Análise do instrumento via ACM

Foi aplicada a ACM no instrumento todo, contemplando desta forma a escala do GSI conforme Derrogatis (1990), que se trata na verdade de um índice geral de desconforto (*distress index*), na amostra toda. A variável Grupo, que determina os grupos de estudo, foi utilizada como variável suplementar ou ilustrativa. A idéia neste posto é verificar a dimensionalidade do instrumento, bem como a evidência do construto medido pelo GSI. Foram utilizados os mesmos *softwares* referenciados no capítulo 4 (ver 4.4).

tab. 5.7- 10 primeiros autovalores para análise dos 90 itens

NUMERO	VALEUR PROPRE	POURCENT	POURCENT CUMULE	
1	0.3204	8.01	8.01	.....
2	0.1753	4.38	12.39	.....
3	0.1287	3.22	15.61	.....
4	0.0942	2.36	17.97	.....
5	0.0429	1.07	19.04	.....
6	0.0383	0.96	20.00	.....
7	0.0362	0.91	20.90	.....
8	0.0344	0.86	21.76	.....
9	0.0328	0.82	22.58	.....
10	0.0320	0.80	23.39	.....

As Não Respostas foram tratadas como categorias ilustrativas, como definido no capítulo 3 (ver 3.3.8) e são representadas como triângulos nos gráficos. A tabela 5.7 apresenta o diagrama de auto-valores da decomposição da matriz indicadora 599 x 450. O resultado sugere que os quatro primeiros eixos devem ser analisados. O primeiro plano (graf. 5.2) é sempre o de maior interpretabilidade por se tratar das direções de maior variabilidade. Nota-se que o primeiro eixo traz a escala ótima, ordenada como a original, mas trazendo as distâncias plausíveis entre as mesmas. A separação entre as categorias intermediárias de intensidade de sintoma neste eixo é bastante pobre, sendo que 1 e 2 (Pouco e Razoável) se interseccionam em uma grande área, próxima à origem do plano. No entanto a presença do segundo eixo traz uma contribuição importante, evidenciando que o eixo 1 absorve quase totalmente as contribuições das categorias Nenhum e Muito (0 e 4), sendo que as categorias 1 e 2 se apresentam na origem do gráfico, sem poder de discriminação neste plano, destacando a oposição entre as categorias 3 e 4 no segundo eixo. Logo a interpretação direta do primeiro plano é que o primeiro eixo esta separando as

---

categorias de maior distância entre si, ou Nenhum e Muito, enquanto o segundo eixo separa a segunda maior distância entre as categorias, que ocorre entre Bastante e Muito.

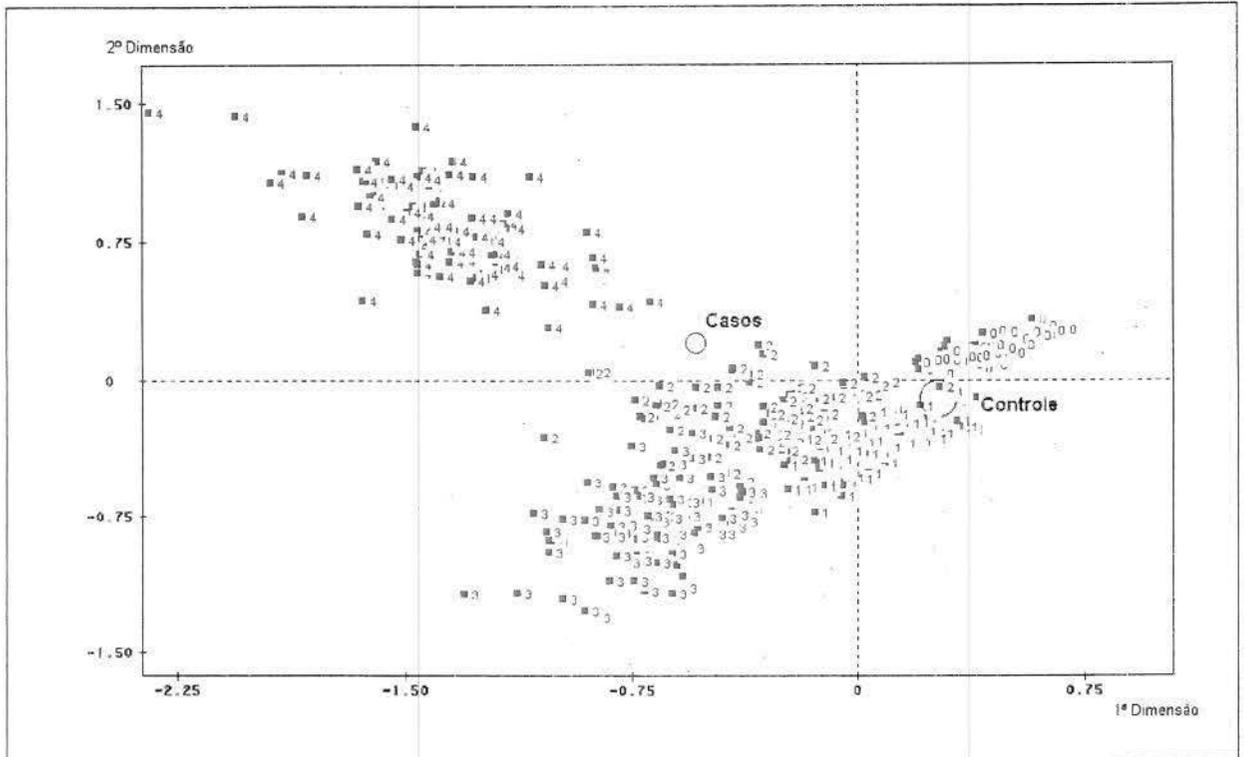
A análise dos eixos 3 e 4 como abscissas do eixo 1 (graf. 5.3 e 5.4 respectivamente) revelam que a variabilidade representadas nestes é relativa às separações entre as categorias de resposta 1 (Pouco) e as demais no eixo 3, e categoria 2 (Razoável) e demais no eixo 4, sendo esta mais marcante.

Logo podemos concluir que cada eixo opõe duas categorias, apresentando uma certa ordem de grandeza em termos de distância entre as mesmas, ou seja, o primeiro eixo opõe 0 e 4, ou a presença forte e ausência de sintoma, o segundo as categorias 3 e 4 (Bastante e Muito), terceiro as categorias 1 e 3 e o quarto opõe 2 (Razoável) em relação às demais. Estes 4 eixos permitem então discriminar indivíduos quanto à predominância de categorias em suas respostas, sendo possível teoricamente então classificarmos os indivíduos em pelo menos três grupos diferentes. Cabe ao pesquisador do instrumento definir a importância e viabilidade de contrastes tão definidos na prática clínica para a escolha dos eixos a serem utilizados em um método de classificação afim de se chegar a uma escala final útil para uma triagem clínica, objetivo principal do instrumento.

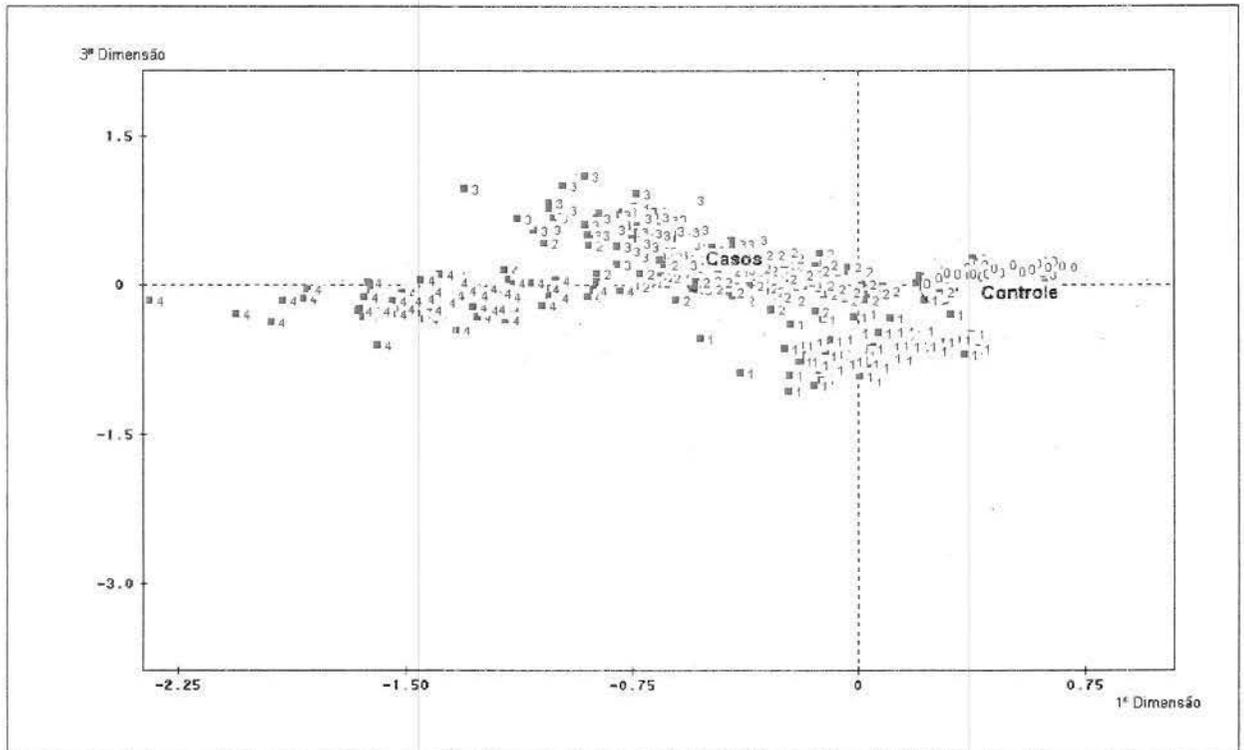
Nota-se a presença de apenas um grande construto realmente, uma vez que não há distinção entre os padrões de representação dos itens, como ocorria nas simulações. A variável ilustrativa Grupo posicionou-se de forma coerente com o esperado, sendo que os Controles foram atraídos para a categoria Muito no primeiro eixo, enquanto Casos posicionou-se sobre a categoria Nenhum e Pouco. Nota-se uma evidência de forte poder classificação de não ocorrência de sintoma do instrumento pelo formato das nuvens de pontos da categoria Nenhum, formando um bloco compacto, atestando também a predominância, em termos de frequência, desta categoria.

As coordenadas (ou escalas ótimas) para os 90 estão no Anexo 1, bem como as medidas de contribuição (cap. 3) utilizadas para o cálculo dos coeficientes  $\eta$ . Decidiu-se neste trabalho analisar-se apenas o primeiro plano (primeiro e segundo eixos). Os coeficientes  $\eta_1$  e  $\eta_2$ , ou os coeficientes de carga fatorial para o primeiro e segundo eixos, equivalentes ao primeiro e segundo fatores em uma AF na matriz de escala ótima de resposta (3.3.6) estão apresentados na tabela 5.7, sendo que os valores maiores ou iguais a 0.6 estão em negrito.

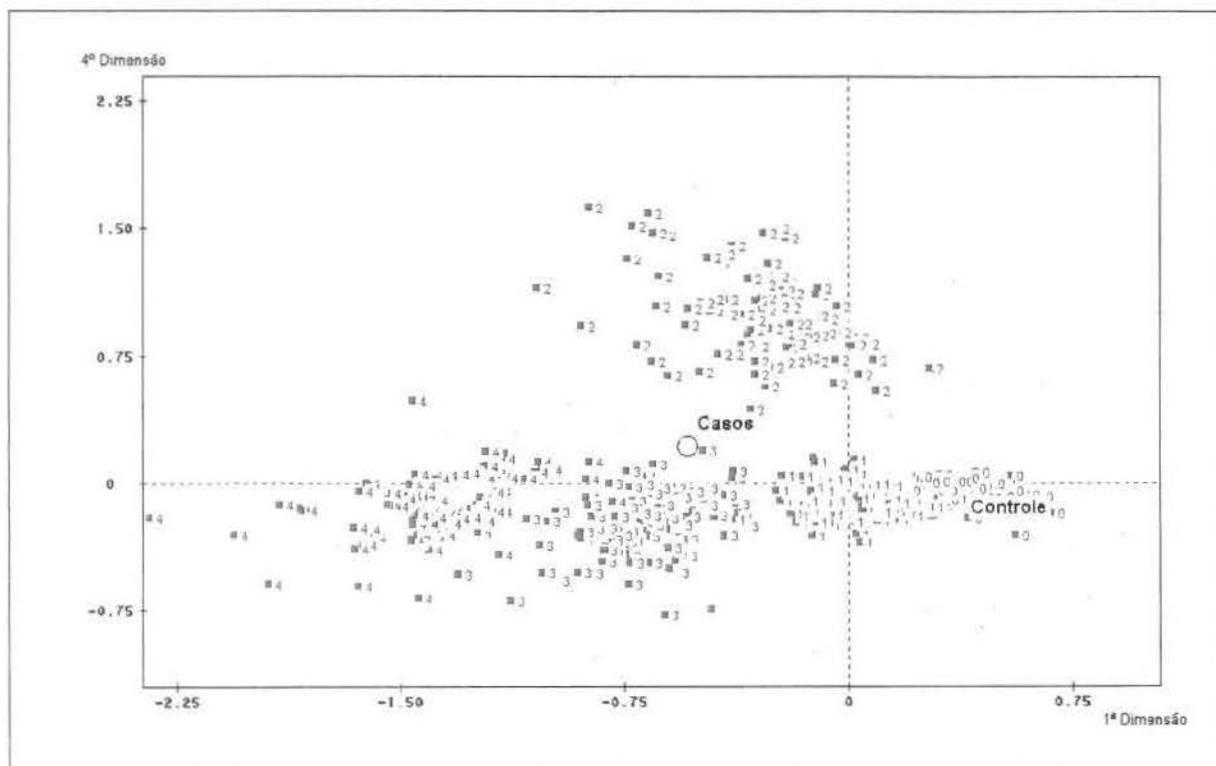
---



graf. 5.2 – Mapa de ACM das Escalas Ótimas nos eixos 1e 2



graf. 5.3 – Mapa de ACM das Escalas Ótimas nos eixos 1e 3



graf. 5.4 – Mapa de ACM das Escalas Otimas nos eixos 1e 4

---

O primeiro fator é bastante consistente, sendo que alguns itens apresentam valores bem baixos por serem de baixa discriminação. Chama atenção o item 31, da pertencente à Depressão, pela inversão do sinal. Este item é o único a possuir Muito como moda em sua distribuição marginal (ver Anexo 1), ao contrário de todas as demais que sempre apresentam Nenhum como moda. Desta forma este item teria uma correlação com o fator em uma ordem inversa. Observe que no mapa de ACM (graf. 5.1 (a)) o gradiente de ordem das categorias esta no sentido negativo – positivo, sugerindo uma correlação negativa com o fator, como todos os itens com exceção do 31. O segundo fator opõe alguns itens, e a oposição se dá justamente quanto ao formato da distribuição marginal novamente (as marginais dos itens estão apresentadas no Anexo 1), sendo que sinal negativo indica itens com maior resposta em 3, em comparação com 4, e os de sinal positivo o inverso.

---

tab. 5.8- Coeficientes de carga fatorial para o primeiro plano da ACM.

Item	$\eta_1$	$\eta_2$	Item	$\eta_1$	$\eta_2$	Item	$\eta_1$	$\eta_2$
1	0.45	0.18	41	<b>0.64</b>	0.42	81	0.56	0.44
2	<b>0.61</b>	-0.36	42	0.54	0.36	82	0.54	0.42
3	<b>0.64</b>	0.38	43	0.59	0.50	83	0.59	0.47
4	0.54	0.25	44	0.56	0.44	84	0.48	0.45
5	0.51	0.31	45	0.54	0.42	85	0.48	0.47
6	0.48	0.31	46	0.59	0.44	86	<b>0.66</b>	0.52
7	0.48	0.33	47	0.42	0.33	87	0.56	0.49
8	0.51	0.28	48	0.42	0.28	88	0.64	0.44
9	0.59	-0.36	49	0.51	-0.40	89	0.59	0.44
10	0.51	-0.33	50	0.56	0.40	90	<b>0.72</b>	0.56
11	<b>0.66</b>	0.44	51	0.70	0.50			
12	0.51	0.33	52	0.51	-0.40			
13	0.54	0.36	53	0.59	0.44			
14	0.59	-0.38	54	0.61	0.45			
15	0.54	0.42	55	<b>0.70</b>	0.52			
16	0.51	0.38	56	0.59	0.42			
17	0.59	0.33	57	<b>0.66</b>	0.47			
18	0.51	0.47	58	0.51	0.40			
19	0.29	0.36	59	<b>0.61</b>	0.45			
20	0.48	0.40	60	0.29	0.28			
21	0.45	0.38	61	<b>0.61</b>	0.49			
22	0.59	0.42	62	<b>0.64</b>	0.47			
23	<b>0.66</b>	0.42	63	0.48	0.33			
24	<b>0.66</b>	0.50	64	0.51	0.33			
25	0.48	0.28	65	0.51	0.38			
26	0.56	0.42	66	0.68	0.50			
27	0.48	0.38	67	0.54	0.40			
28	<b>0.66</b>	0.44	68	0.54	0.44			
29	<b>0.61</b>	0.44	69	0.54	0.40			
30	<b>0.64</b>	0.53	70	0.56	0.47			
31	-0.45	-0.36	71	0.51	-0.38			
32	0.59	0.52	72	<b>0.68</b>	0.49			
33	0.59	0.45	73	0.48	0.44			
34	<b>0.68</b>	0.52	74	0.56	0.49			
35	0.45	0.33	75	0.59	0.47			
36	0.56	-0.38	76	0.59	0.52			
37	0.56	0.42	77	<b>0.66</b>	0.49			
38	0.42	-0.38	78	<b>0.64</b>	0.50			
39	0.56	0.38	79	<b>0.64</b>	0.47			
40	0.51	-0.40	80	0.72	0.49			

Assim podemos interpretar o plano principal da ACM da seguinte forma: o primeiro e segundo quadrantes (sentido horário a partir do quadrante positivo) apresentarão indivíduos com pouco ou nenhum sintoma, sendo que quanto maior o escore no primeiro eixo, menor a intensidade ou presença de sintomas no indivíduo. Nestes quadrantes o segundo eixo pouco contribui. No terceiro quadrante apresentará indivíduos com intensidades de sintomas Razoável e Bastante, sendo que quanto menor o escore do indivíduo no segundo eixo, maior será a presença destas categorias. Logo o quarto quadrante apresenta os indivíduos com predominância em Muito. Note que o gradiente dos escores está em ordem inversa à intensidade de sintomas. Este problema poderia ser resolvido multiplicando-se este eixo por  $-1$ , e o segundo também conseqüentemente.

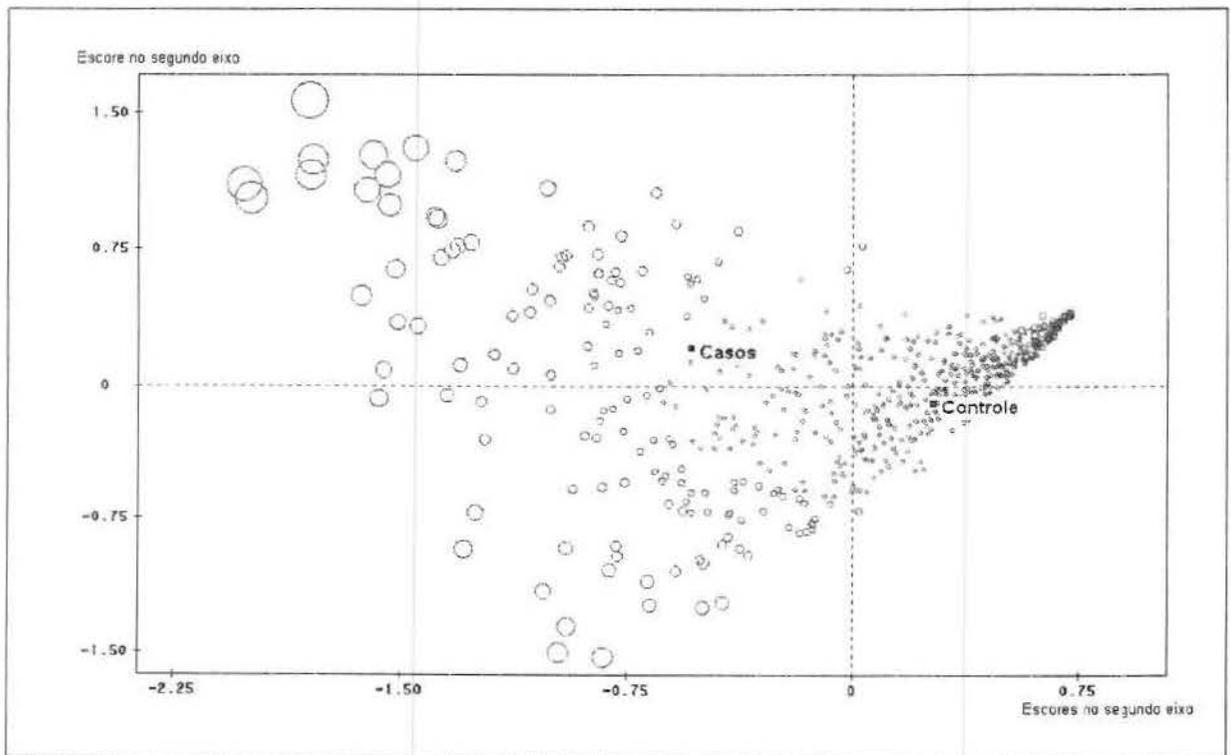
Então quanto maior a distância à origem na direção do quarto quadrante, mais intenso o sintoma geral do indivíduo. Quanto maior a distância à origem na direção dos primeiro e segundo quadrantes, menor a intensidade de sintomas gerais, e conseqüentemente, no terceiro quadrante temos a intensidade também aumentando a medida que o indivíduo se afasta da origem em seus escores. O gráfico 5.2 apresenta os escores dos indivíduos do estudo no primeiro plano. O tamanho dos círculos é proporcional à contribuição do indivíduo na inércia dos eixos 1 e 2, e pode ser interpretado também como a intensidade do sintoma: quanto maior o círculo, maior o desconforto geral do indivíduo. Note a posição coerente das categorias da variável Grupo.

Uma vez que não há uma clara diferença de padrão de escala entre os itens, evidenciando um único construto, espera-se que o mesmo se repita para as sub-escalas, ou dimensões de sintoma, que compõem o instrumento.

O gráfico 5.3 apresenta as coordenadas das categorias de cada variável (ordenadas) contra a escala de resposta empregada (0, 1, 2, 3, 4) nas abscissas. Quanto mais próximo de um segmento de reta a  $45^\circ$  a linha do gráfico se apresentar, maior a evidência de validade da métrica da escala de resposta empregada, para a utilização dos escores propostos (médias das categorias de resposta). Percebemos que a grande maioria dos itens apresenta métrica coerente com a escala de resposta original. As retas são descendentes pelo fato do primeiro fator se apresentar em ordem inversa ao gradiente da escala original, como já citado acima.

Logo a utilização do critério de somas e médias para obtenção dos escores e classificação dos indivíduos sobre esta nova escala final pode ser considerada válida neste instrumento.

---



graf. 5.5- Mapa de ACM para os Escores do Individuos no primeiro plano

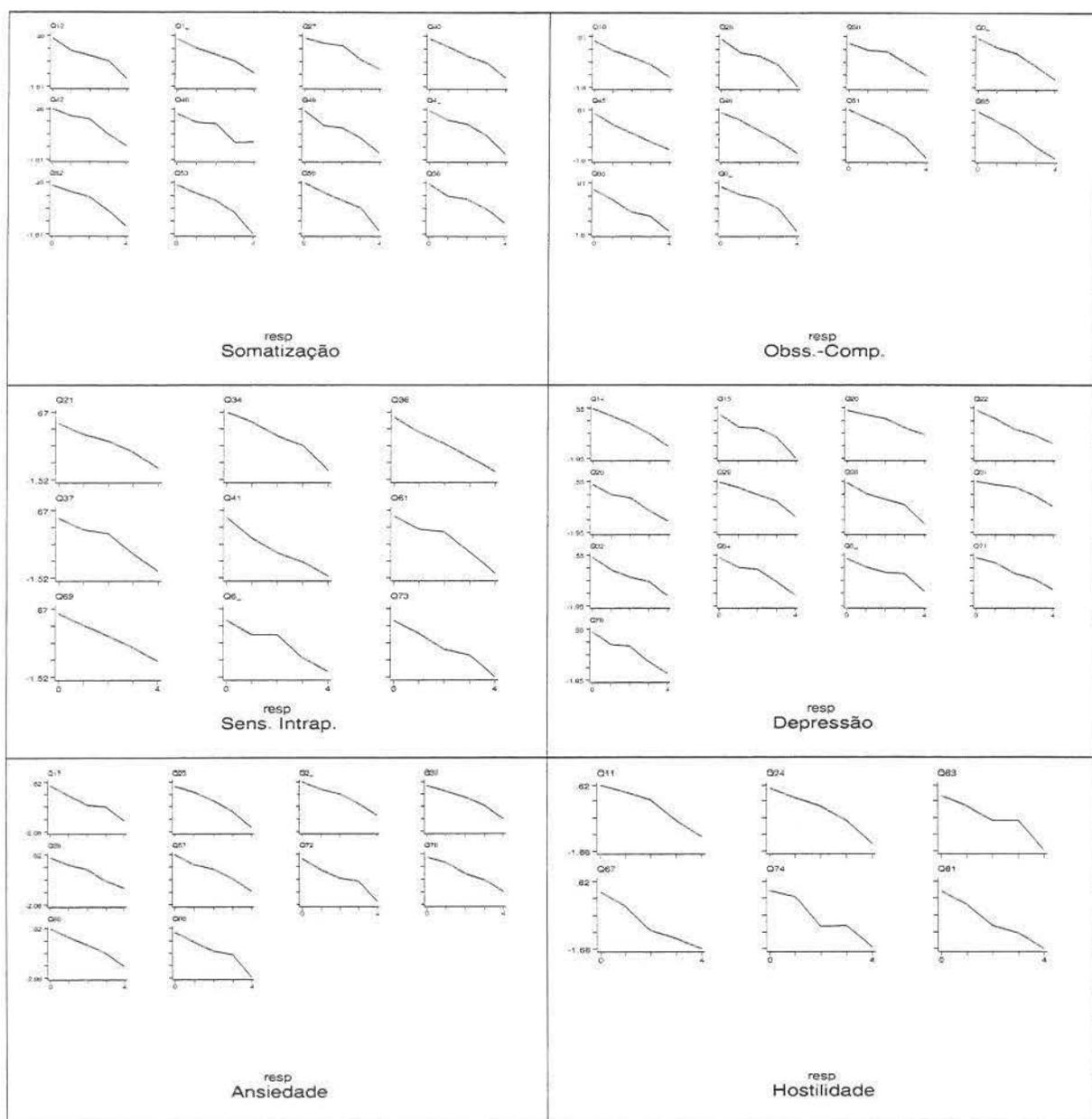


fig. 5.6 – Escores Ótimos do primeiro eixo da ACM contra escala de resposta do instrumento (por dimensão de sintoma).

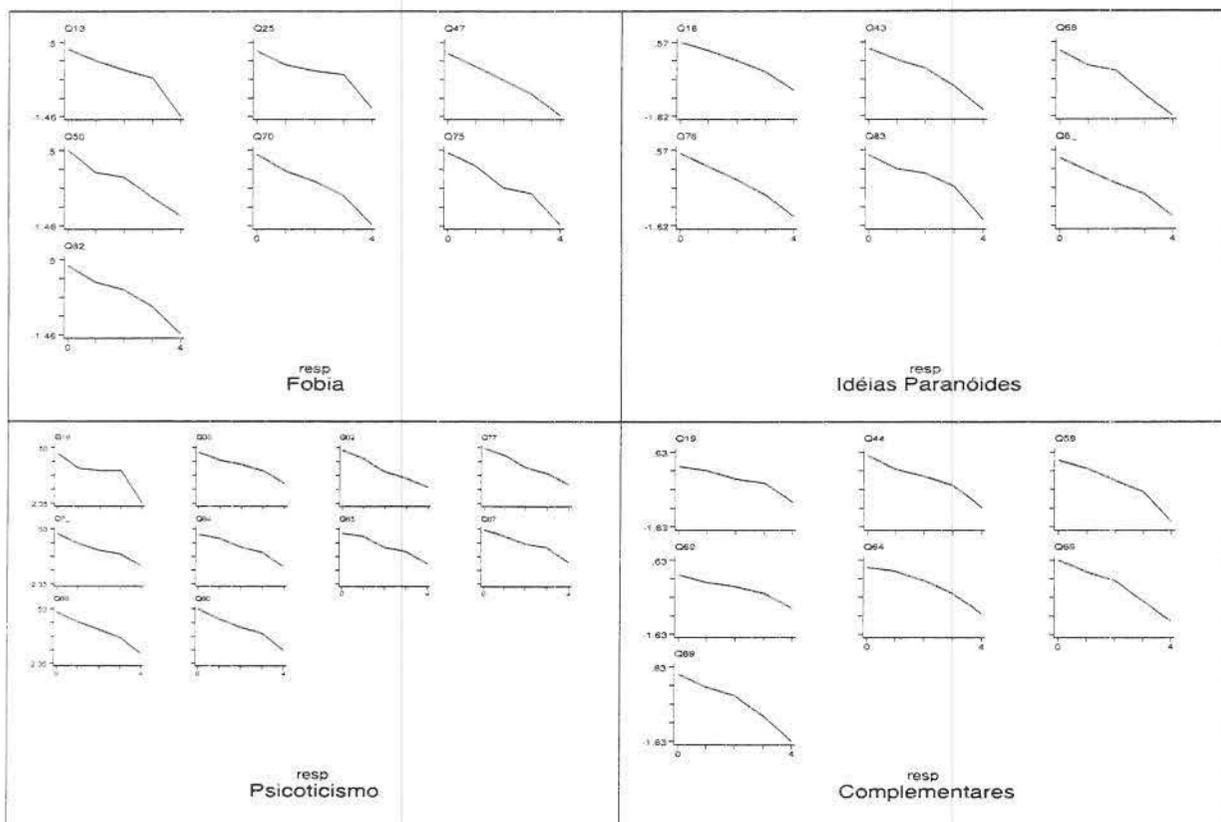


fig. 5.6 (continuação) – Escores Ótimos do primeiro eixo da ACM contra escala de resposta do instrumento (por dimensão de sintoma)

Desta forma podemos analisar e interpretar os escores propostos pelo instrumento, como descrito em 5.1.2. A tabela 5.2 apresenta as estatísticas descritivas para os escores das nove dimensões de sintoma, divididos por sexo e grupo de estudo, compondo 4 grupos comparáveis, de um total de 8 normas americanas propostas (Derogatis, 1994). No Anexo 1 encontram-se os histogramas com as distribuições dentro de cada grupo formado por Grupo de Estudo e Sexo.

tab. 5.9- Estatísticas resumo dos Escores Brutos de cada dimensão de sintoma para respostas completas

<b>Controle / Fem.</b>						<b>Controle/ Masc.</b>					
	Obs	Mean	DP	Min	Max		Obs	Média	DP	Min	Max
somato	213	.99	.75	0	3.33	somato	181	.84	.75	0	3.33
interp	213	.94	.76	0	3.22	interp	181	.76	.68	0	2.77
obss	211	.99	.75	0	3.40	obss	183	.87	.74	0	3.6
depre	210	.93	.72	0	3.61	depre	183	.87	.66	0	2.92
ansied	213	.92	.83	0	3.70	ansied	182	.75	.74	0	3.6
host	216	.78	.81	0	4.00	host	183	.64	.74	0	3.16
fobia	214	.81	.80	0	3.57	fobia	182	.66	.73	0	3.42
paranoia	213	.98	.91	0	4.00	paranoia	182	.97	.84	0	4.00
psico	214	.67	.76	0	4.00	psico	178	.74	.75	0	3.50
adicio	215	1.07	.80	0	3.28	adicio	183	.86	.72	0	3.42
<b>Caso/ Fem.</b>						<b>Caso/ Masc.</b>					
Variable	Obs	Média	DP	Min	Max	Variable	Obs	Média	DP	Min	Max
somato	134	1.62	.99	0	4.00	somato	62	1.34	.91	0	3.50
interp	137	1.51	.98	0	3.55	interp	63	1.39	.90	0	3.66
obss	134	1.58	.96	0	3.90	obss	61	1.51	.96	0	3.80
depre	136	1.71	.91	0	3.84	depre	62	1.43	.91	0	3.84
ansied	133	1.53	1.00	0	3.90	ansied	62	1.41	.99	0	4.00
host	136	1.25	1.02	0	4.00	host	63	1.11	.98	0	4.00
fobia	136	1.27	1.06	0	3.85	fobia	62	1.13	.89	0	4.00
paranoia	136	1.41	.99	0	3.83	paranoia	62	1.35	.90	0	3.00
psico	136	1.15	.83	0	3.40	psico	62	1.20	.82	0	3.40
adicio	137	1.64	.90	0	3.42	adicio	63	1.44	.89	0	3.14

### 5.3- Conclusão sobre a validação da escala GSI do SCL 90 R4

O instrumento propõe como resultado uma escala final unidimensional na forma de um Índice Global de Gravidade (*Global Severity Index*). Pressupõe-se então que este instrumento apresente como estrutura fatorial apenas um fator bem definido, além de uma correta classificação dos indivíduos em seu respectivo grupo. A interpretação dos resultados da validação via ACM podem ser sintetizados da seguinte forma:

1. A estrutura fatorial acusa um fator consistente no primeiro eixo da ACM, fato comprovado pelos coeficientes fatoriais calculados (tab. 5.8). Os demais eixos, 2 a 4, seriam contrastes entre as intensidades de sintoma representados pela escala de resposta, uma espécie de refinamento da escala. A utilização destes eixos com intuito de classificação de indivíduos se torna mais eficiente se deseja-se a classificação de uma intensidade intermediária de sintomas do distúrbio;
2. A validação de critério, via comparação de indivíduos em dois grupos, apesar de apresentar resultados coerentes conforme já citado, não mostrou grande poder de discriminação. Talvez o fato do grupo denominado Controle ser formado de pacientes de outros ambulatorios, que não o psiquiátrico, tenha causado esta perda de contraste, uma vez que um hospital enquanto ambiente hostil ao cotidiano pode gerar o desconforto (*distress*) medido pelo instrumento independente do ambulatorio;
3. A aplicação do GSI, a média das respostas dos indivíduos como instrumento de triagem clínica é valido conforme observado na comparação entre a escala ótima e a de resposta (graf. 5.3). No entanto deve-se criar critérios e pontos de corte para a população estudada especificamente. A aplicação de Análise de Agrupamento e Discriminante seriam de grande importância no aspecto exploratório, cabendo a Teoria de Resposta ao Item (Andrade, D. *et al.* (2000)) uma estimação mais consistente destes pontos de corte em uma escala unidimensional.

## 6. Considerações Finais

Este trabalho procurou explorar a aplicação de uma técnica multivariada, a Análise de Correspondência Múltipla (ACM), no campo da psicometria aproveitando basicamente as abordagens de duas escolas tradicionais no estudo de Análise Multivariada, a francesa e a holandesa. A primeira, como dito no capítulo 1 aborda a interpretação geométrica dos gráficos, respaldada por resultados numéricos voltadas à qualidade de representação. A segunda utiliza um enfoque de escalonamento, de uma *“redução puramente numérica, ou uma compressão de dados”* (Lebart (1995), p. 17) na chamada Análise de Homogeneidade. Busca-se uma quantificação ótima que têm sua utilização apropriada como valores de categorias ordinais qualitativas.

Na exploração utilizou-se a solução da Análise Fatorial, método clássico na psicometria, através de Componentes Principais como um critério de comparação. Observou-se no capítulo 4 que em algumas situações a técnica de ACM fornece uma melhor resolução no estudo da estrutura fatorial dos instrumentos, como no caso da simulação 5 onde ocorre a situação mais complexa. Na separação de grupos de indivíduos a ACM mostrou-se sempre superior, indicando que esta técnica em um estudo de definição de normas para um instrumento, por exemplo, tem maior potencial como *input* para técnicas classificatórias.

Além do desempenho da técnica em relação à forma clássica de análise são apresentadas outras vantagens, uma vez que a ACM permite um estudo em nível de escala de resposta, ou seja, é possível analisar as propriedades de ordenação de resposta ao

---

---

estímulo através da escala ótima obtida no eixo de interesse (aquele que representou o construto, como visto nas simulações).

A representação simultânea de indivíduos e categorias (linhas e colunas da matriz de dados) é outra vantagem, uma vez que permite uma interpretação direta das posições dos indivíduos em relação às categorias. A representação de indivíduos e variáveis ilustrativas também é de grande auxílio, principalmente no estudo psicométrico em grupos de diferentes perfis culturais e sócio-econômicos.

Uma aplicação semelhante de Análise de Correspondência Múltipla em instrumentos com escala de Likert é feita por Gifi (Gifi (1990), p. 438-440). Nesta aplicação são apontadas as vantagens desta técnica na identificação de itens com ordem diferentes através da Escala Ótima. É feita uma comparação também com a forma tradicional de análise através da correlação média inter-item, sendo que a matriz recodificada (a matriz  $Q$  apresentada em 3.3.6) apresentou um aumento neste índice. Gifi salienta que na forma tradicional os indivíduos são escalonados de acordo com a soma dos pontos considerando a mesma importância para todos os itens, enquanto nos escores induzidos pela escala ótima, itens com maior poder de discriminação contribuem mais. A Análise de Correspondência Múltipla (ACM) é tratada também como um teste para a suposição de que a escala de resposta empregada pode realmente ser considerada um contínuo. Termina concluindo que a abordagem via ACM apresenta vantagens em relação a forma tradicional de tratamento e análise.

Alguma discussão a respeito é feita também em Weller e Romney (1990) mas em uma comparação da Análise de Correspondência Simples (ACS) e Componentes Principais. Salientam que a ACS apresenta as relações de interação, na linguagem de modelos log-lineares (Bishop, 1975), e não os efeitos principais, ou as diferenças das marginais totais. Se pensarmos na ACM como uma ACS aplicada a uma matriz indicadora onde as variáveis de linha são os indivíduos e as colunas representam as categorias de resposta dos itens, esta é uma vantagem, uma vez que não se explora esta questão de efeitos principais, e sim as interrelações entre as colunas e/ou linhas de modo a se obter um escalonamento ótimo.

---

## 6.1- A Estratégia sugerida

Em vista da exploração tratada até aqui se pode traçar uma estratégia para avaliação de instrumentos psicométricos com o intuito de validação dos mesmos.

Parte-se do pressuposto de que o instrumento foi aplicado sobre uma amostra representativa da população alvo, suficientemente grande, com os controles adequados, nos moldes dos Ensaio Clínicos. A aplicação em grupos de indivíduos com perfis diferentes é de grande interesse no estudo do comportamento. No caso de instrumentos que visam diagnosticar determinado distúrbio é aconselhável a aplicação em um estudo tipo Caso-Controle afim de se verificar o poder de classificação da escala final.

Tendo os dados em mãos, parte-se à análise estatística, que de forma geral pode seguir os seguintes passos:

1. É importante a descrição demográfica da amostra estudada de forma a se ter bem definida a população sobre qual foi aplicado o instrumento, bem como o controle das fontes observáveis de variação. Em um primeiro momento através de tabelas de contingência marginais, e através da Análise de Correspondência Múltipla;
2. Aplica-se a Análise de Correspondência Múltipla aos dados do instrumento na amostra como um todo. No caso de um estudo com interesse de grupos determinados a priori a representação das categorias de grupo deve ser feita como variável ilustrativa. Uma vez verificada alguma latência nos dados, indicado pelo **Diagrama de Autovalores**, identifica-se os eixos de interesse e parte-se para a interpretação dos dados;
3. Analisa-se o formato da nuvem de pontos dos itens. Este formato já fornece evidências sobre a ordenação das escalas de resposta e da dimensionalidade da escala do instrumento (ou das escalas quando o instrumento apresenta múltiplas escalas). Os resultados numéricos de contribuições dos itens à inércia dos eixos de interesse bem como o cálculo das cargas fatoriais (ver 3.3.6);
4. Analisa-se as Escalas Ótimas obtidas nos eixos fatoriais de interesse contra a escala de resposta original. Este é um estudo de ordenação mais detalhado, item a item, onde

---

identifica-se os itens problemáticos quanto a ordenação e discriminação através da forma do gráfico de linha.

5. Parte-se para a análise no nível dos indivíduos. Estuda-se a nuvem de pontos dos indivíduos na procura de agrupamentos destes que possa trazer alguma informação sobre o instrumento. No caso de um estudo do tipo Caso-Controle por exemplo, espera-se encontrar uma discriminação destes indivíduos. A aplicação de outras técnicas multivariadas como Análise de Agrupamento e Análise Discriminante (Johnson e Wichern, 1998) sobre os escores do indivíduos pode diagnosticar a presença destes grupos.
6. Representa-se simultaneamente os indivíduos (linhas) e as categorias (colunas). Esta representação ajuda no entendimento e justificativas das separações dos indivíduos.
7. Caso certificada a linearidade da escala de resposta dos itens que compõem o instrumento pode se partir para as análise convencionais de comparação de escores entre grupos através de teste estatísticos usuais.

Até o passo 4 temos elementos suficientes sobre a estrutura fatorial do instrumento e da validade de uma utilização através de somas de pontos. No caso de se encontrar itens inadequados a este uso, exclui-se estes itens e retorna-se ao passo 2, início da análise. A exclusão de itens que não possuam a relação linear com a escala de resposta original tende a deixar mais significativa a estrutura fatorial. Gifi sugere a aplicação da análise fatorial na matriz de dados recodificada, ou seja, com os valores de resposta originais dos indivíduos substituídos pela Escala Ótima (Gifi, 1990). Nos passos 5 e 6 procura-se entender as relações entre os indivíduos e o instrumento, bem como as questões relativas à discriminação e classificação.

Estes elementos são suficientes para um bom entendimento do instrumento; uma exploração inicial que serve de base para a justificativa de uso de técnicas inferenciais em uma etapa posterior, baseada na escala final obtida através de soma ou média dos pontos.

---

## Apêndice A

### A.1- Breve introdução a Psicometria

Segundo Guilford (1954) "*Os métodos psicométricos são procedimentos para medidas psicológicas*". Esta simples definição já nos dá idéia do complexo problema que envolve a psicometria.

Partindo do princípio que estes são métodos de medição de objetos abstratos como sentimentos, julgamentos, traços de personalidade, dentre outros, este tópico buscará introduzir os conceitos principais da psicometria através de um histórico, não aprofundado, traçando os devidos paralelos com os conceitos habituais da teoria estatística.

#### A.1.1 Medidas Psicológicas

Uma medição de forma geral é definida como "*um procedimento para atribuírmos números ao objetos de estudo, de modo que representem quantidades dos atributos deste objeto*" (Numnaly, 1967), ou de modo mais simples "*a descrição da informação através de números de modo a usufruir das características e propriedades que os números fornecem*" (Guilford, 1954).

O problema de definição de medidas, questões filosóficas sobre sua representação dos fenômenos naturais e outras questões possuem larga bibliografia que remete ao começo do século XX final do XIX. Podemos pensar como este sendo o principal problema das ciências naturais. Uma vez aceita, validada, uma medida<sup>5</sup> torna-se padrão para discussões, comparações e interpretações de diferentes problemas.

Em fenômenos físicos a associação do objeto de estudo e sua representação em termos de medida é quase imediato. Ao se estudar altura, temperatura, velocidade, tamanho

---

<sup>5</sup> O termo "medida" será utilizado como o resultado de uma medição, e não deve ser confundido, no âmbito deste trabalho, com o termo utilizado na Teoria da Medida em probabilidade.

---

---

de populações, o próprio objeto de estudo se confunde com seus atributos a serem medidos, e por sua vez a medida se confunde com o atributo em termos racionais. Os atributos são quantitativos e não é necessário grande esforço para entendê-lo e representá-lo em uma dimensão numérica.

No entanto quando se estudam problemas cujos atributos são qualitativos a discussão e compreensão de “o quê medir e como medir” para representar o objeto de interesse passa a ser mais complexo.

Para melhor entendermos as complexidades envolvidas, podemos partir da visão de que o processo de medição envolve, seja qual for a área, três passos principais (Thorndike 1961):

1. *Identificar e definir o(s) atributo(s) a ser(em) medido(s)*: Nunca se mede o objeto de estudo. A medição sempre se dá sobre atributos deste objeto. Como dito acima, esta identificação nem sempre é clara em medições de variáveis subjetivas, como é o caso das medidas psicológicas.
2. *Determinar um processo mediante o qual o atributo possa ser medido (sistema de medição)*: Define-se a maneira e os métodos e instrumentos a serem empregados na medição. No caso da psicologia estes métodos devem manifestar os atributos de forma a serem perceptíveis. Os chamados estímulos, podem ser físicos, julgamentos ou sentimentos descritos pelo indivíduo geralmente de forma padronizada, através de respostas fechadas.
3. *Estabelecer procedimentos e técnicas que traduzam as observações de indicações qualitativas em termos de quantidades*.

### ***A.1.2- Breve Histórico***

Até final do século XIX, início do XX a psicologia era entendida como parte da filosofia, buscando explicações para os eventos psicológicos na relação entre a “mente e alma” (Thorndike *et al.*, 1961), quase totalmente não experimental. A idéia de que fosse possível medir atributos psicológicos era vista com certa relutância e até hostilidade pelos pesquisadores da área.

---

---

Com o avanço das ciências físicas e biológicas, baseadas principalmente em experimentação, a psicologia encontrou o apoio e embasamento que precisava para desenvolver seus métodos específicos.

Herbart (1786-1841) , Weber (1795-1878) e Fechner (1801-1887), escrevem trabalhos no final do século XIX, tentando atribuir medidas a eventos psicológicos. A preocupação nesta época era estudar e quantificar as relações entre eventos físicos e psicológicos, tais como reações a estímulos visuais e auditivos. Desenvolvia-se a *psicofísica*: busca-se um modelo, matemático ou probabilístico, que “explique” a relação entre um estímulo (físico) e uma resposta (atributo psicológico). As perguntas básicas a serem respondidas são “como a resposta depende do estímulo?”, “quanto a resposta varia quando variamos o estímulo, em termos de unidades de medida?” e “qual é a unidade de medida?”. Algumas leis são formuladas, como a “*lei de Weber*”, ou a “*lei dos julgamentos*” (Nunnally, 1967).

O primeiro laboratório experimental para psicologia foi fundado em 1879, em Leipzig, Alemanha, por Wilhelm Wundt (1832-1920) e tinha como interesse principal estudos psicofísicos.

Enquanto isso, uma outra linha de estudo começava a se desenvolver baseada nas diferenças individuais, levantadas nos aspectos físicos por Darwin em seus trabalhos sobre a origem das espécies. Quem estuda estes aspectos do posto de vista mental, incentivado pelo próprio Darwin é Francis Galton (1822-1911). Seus estudos também incluíam o desenvolvimento de um ferramental estatístico para as comparações individuais, contando com a ajuda de Karl Pearson e Charles Spearman.

Contemporâneo a Galton, Francis Binet desenvolve a primeira escala de inteligência conhecida como “*Escala de Binet-Simon*” (Binet, 1905), parte de um projeto do governo francês que visava detectar crianças com dificuldade de aprendizado.

Esta linha de trabalho ganhou o nome de “*Teoria de Testes Mentais*”. O termo foi utilizado pela primeira vez por James McKeen Cattell (1860-1944) (Cattell, 1890), discípulo “rebelde” de Wundt, que apesar da oposição do mestre levou adiante seus estudos sobre diferenças individuais, desenvolvendo metodologias sobre testes mentais.

Desde então os testes mentais fazem parte das mais diversas áreas, em seleção de pessoal, encaminhamento especializado, dentre outras. O grande “boom” dos testes mentais

---

---

ocorreu entre 1915 e 1930 (Freeman, 1967) com o advento da primeira grande guerra, sendo o exército americano um dos seus maiores percussores. Após esta fase o uso destes testes passou a ser indiscriminado, nas mais diversas áreas para os mais diversos propósitos, o que levou a muitas más utilizações, sem qualquer critério, fornecendo argumentos para críticas severas das linhas de trabalho opostas aos testes, que caíram em certo descrédito, tendo o período entre 1930 e 1945 como de reavaliação e reestruturação dos testes. Novamente com a Segunda Grande Guerra houve o segundo “boom”. Neste período a preocupação foi principalmente educacional e de medir aptidões. Em 1935 é criada a revista *Psychometrika*, por L. L. Thurstone, especializada no assunto de aplicações matemáticas e estatísticas na área de psicologia. Thurstone desenvolve trabalhos tanto na psicofísica (“*lei dos julgamentos comparativos*”), como na área de testes mentais, escrevendo muitos trabalhos em *Análise Fatorial Múltipla*.

A utilização dos testes mentais fora do objetivo de classificação do indivíduo em termos de aptidões, como um instrumento na psicologia clínica é mais recente. Mesmo os testes de inteligência não podem ser considerados puramente clínicos, pois sempre têm o objetivo de alocar indivíduos em funções ou grupos específicos.

## **A.2- Sobre Validação**

A grosso modo, validar uma escala de medida consiste em verificar se esta está realmente medindo aquilo que se propõe a medir. Apesar da aparência óbvia desta “definição” diversos aspectos, nem sempre tão óbvios, estão envolvidos na validação de uma escala de medida.

Se pensarmos em um caso simples, como por exemplo a validação da escala utilizada em uma régua, o problema da validação consiste em uma comparação das medidas realizadas com este instrumento, com padrões físicos já devidamente estabelecidos, uma vez controlada fontes de variação sistemática (operadores, método de medição, etc...). O problema neste caso está bem delimitado, a medida em questão é de uma dimensão objetiva, e um julgamento final se torna de fácil percepção e compreensão, tendo inclusive parâmetros de precisão e exatidão bem estimados.

---

No entanto o problema passa a se tornar mais complexo na medida em que o grau de subjetividade da dimensão a ser medida aumenta. Então uma simples comparação com padrões conhecidos nem sempre é suficiente, ou o que temos como critério de comparação não se pode chamar de padrão.

No caso de escalas psicométricas, foco deste trabalho, a validação não é algo facilmente definida e acessível. As dimensões a serem medidas não são físicas, são em grande parte teóricas, até comprovadas clinicamente, mas em um senso estrito não existem. Para medi-las são utilizados recursos indiretos de estímulo-resposta, questões sobre sintomas aptidões, sentimentos, julgamentos dentre outros, de forma que um conjunto destes estímulos, inter-relacionados de alguma forma já conhecida, consiga representar esta dimensão subjetiva de interesse.

Pelos aspectos apresentados acima fica bastante claro que um estudo de validação se baseia em investigações empíricas. Não há como validar algo, nestes aspectos, meramente pela teoria ou imposição de idéias. Também podemos dizer que na verdade o que se valida não é a escala em si, e sim a utilização desta escala para determinado problema, em determinado universo. Ou seja, a régua pode ser válida para se medir comprimento em superfícies retas e regulares, mas não para superfícies irregulares ou curvilíneas.

#### ***A.2.1- A validação na psicomетria***

A preocupação com a validação de escalas psicométricas surge junto com o desenvolvimento das escalas e a teoria de Testes Mentais.

As medidas psicológicas devem servir a três importantes funções (Nunnaly, 1967): 1) estabelecimento de relações estatísticas com uma determinada variável, 2) representar o conteúdo de um domínio específico, e 3) medir traços psicológicos. Correspondendo a estas três funções temos três tipos principais de validação: 1) validação de relacionada ao critério, 2) validação de conteúdo e 3) validação de construto. Os nomes destes três tipos de validação variam de autor para autor. Uma discussão maior sobre estes problemas pode ser encontrada na literatura relacionada (Nunnaly, 1967; Freeman 1967; Guilford, 1954). A seguir a definição e discussão sobre cada um dos tipos.

---

### 1) Validação relacionada aocritério- (ou Preditiva)

Este tipo de validação diz respeito à capacidade do instrumento utilizado de estimar algum aspecto ou comportamento importante comparado a um critério externo ao próprio instrumento. Seria o mais próximo ao caso simples da régua. Ou seja, trata-se da associação empírica da escala do instrumento com algum critério já conhecido e utilizado.

No caso da psicologia este critério pode ser alguma outra escala já utilizada, e a validade é acessada quase exclusivamente através de coeficientes de correlação entre as variáveis estudadas nos dois instrumentos.

É argumentado por alguns autores (DeVellis, 1991) que este tipo de validação é independente de qualquer teoria envolvida, ou seja se um instrumento tem grande correlação com o critério não importa a incoerência com a teoria. No entanto o problema de critério não deve ser encarado desta forma, ou estaremos recorrendo à espuricidade dos fenômenos naturais ou não que tanto nos cercam (Nunnaly, 1967). Mais uma vez fica claro que a validação se dá quanto ao modo de uso da escala.

### 2) Validação de Conteúdo

Ao se estudar comportamento, sentimentos, traços de personalidade o pesquisador deve ter definido o domínio sobre o qual está trabalhando. Sobre domínio podemos dizer que seria o universo de todos estímulos que dizem respeito à variável que está sendo medida.

Validar o conteúdo de um instrumento seria avaliar o quanto adequado são os estímulos presentes na representação do domínio.

Claramente este não é um problema fácil. A delimitação de um domínio nem sempre é clara. Um exemplo objetivo é um teste fonético, onde o domínio seriam todas as palavras do vocabulário de determinado idioma. A validação seria feita sobre as palavras envolvidas no teste, e o quanto elas estariam representando os fonemas daquele idioma. No entanto para instrumentos que procuram medir traços de personalidade este domínio é praticamente impossível de ser delimitado. Qual seria o

---

---

universo dos estímulos para se verificar depressão através de um instrumento, por exemplo?

Alguns autores também colocam nesta classe de validação os problemas semânticos e estruturais do instrumento, argumentando que o conteúdo na verdade é mais do que representar. O domínio pode estar bem representado, mas o teste apresenta dificuldades nas respostas dos estímulos, ou o teste é de difícil aplicação.

Esta validação é normalmente feita na construção do instrumento, e não na aplicação.

### 3) Validação de Construto

Esta é tida como a validação mais importante de um instrumento na opinião da maioria dos autores, principalmente após a década de 50 (Nunnally, 1967). Com os avanços dos métodos de Análise Fatorial esta passou a ser a principal preocupação dos pesquisadores da área.

Basicamente busca saber se os estímulos envolvidos no instrumento se inter-relacionam de forma adequada com a teoria, e se podemos interpretar estas relações como a formação de um construto. É bastante semelhante com a definição grosseira de validação apresentada no começo deste tópico: “O instrumento mede o que se propõe?”.

Um construto representa uma hipótese de que uma variedade de comportamentos irão se relacionar de uma forma esperada de acordo com a teoria em estudos de diferenças entre indivíduos (Nunnally, 1967). Então podemos entender a validação de construto como um estudo da estrutura de variação de resposta aos estímulos. Se a associação entre os estímulos é relativamente forte, então podemos dizer algo sobre a representação de um construto.

Podemos delinear este tipo de validação por três objetivos principais:

- 1) Especificar o domínio das variáveis observáveis (resposta aos estímulos);
  - 2) Analisar através de estudos empíricos e metodologia estatística quanto estas variáveis observáveis medem a mesma coisa ou não, em relação às medidas de um mesmo construto e de construtos diferentes;
-

- 3) Realizar estudos de diferenças individuais e/ou experimentos controlados de forma a acessar o quanto as medidas utilizadas no construto são previsíveis, no que diz respeito à teoria envolvida.

Ainda podemos resumir estes três objetivos em dois atributos que espera-se que o construto em questão tenha: *Estrutura Interna* e *Estrutura Cruzada*. A primeira, também chamada *consistência*, diz respeito aos objetivos 1 e 2 citados anteriormente. Ou seja, se as variáveis observáveis estão medindo as mesmas coisas, o que certamente é mais plausível de acontecer uma vez que domínio esteja bem representado pelas mesmas. Este atributo e os termos utilizados possuem estreita relação com a fidedignidade do instrumento quanto ao construto, como será discutido adiante. No entanto o fato de um conjunto de itens apresentar consistência interna é uma característica **necessária** no entanto não **suficiente**. Pode-se estar medindo a mesma coisa com os itens, mas nada indica neste caso que o que está sendo medido seja realmente o intencionado. É de se esperar que estas medidas se relacionem de forma coerente com a teoria de base com relação a outros construtos e/ou critérios. Neste ponto estamos analisando o cruzamento da estrutura interna com outras fontes, através de comparações com outros instrumentos ou de intervenções experimentais, conforme salientado no objetivo 3.

De certa forma a validação de construto engloba as demais validações citadas anteriormente, uma vez que precisamos de uma definição de domínio e de critérios de comparação. Como classificar um indivíduo quanto à determinada aptidão (validade preditiva), por exemplo, se não existe um conteúdo válido e/ou um construto medido de forma coerente? Novamente temos a questão da validação do uso do instrumento.

Vale a pena salientar ainda que uma validação, seja qual for o tipo, é um processo contínuo de aprimoramento de um instrumento de medida, e nunca um estudo único.

A literatura cita ainda outros tipos de validações, variações ou mesmo apenas nomes diferentes das aqui apresentadas. Nomes como *Validade Facial*, *Validade Fatorial*, dentre outras. Referências para um estudo mais aprofundado são Nunnally (1967), Guilford (1954), Freeman (1962).

---

## A.3- Análise Fatorial

### A.3.1 Breve Históricos

A Análise Fatorial (AF) tem como berço o estudo de problemas psicológicos, mais especificamente a psicometria. A busca pela explicação de aptidões e habilidades é de grande interesse da ciência, e este método veio como uma tentativa de se especificar modelos estatísticos para explicar as teorias psicológicas das habilidades e comportamentos humanos. Por este motivo talvez esta técnica é muitas vezes considerada como uma sub-área da psicologia. Não devemos no entanto esquecer o forte apelo matemático e estatístico que a justifica como área de estudo das ciências estatísticas.

Para entender-se o desenvolvimento e as idéias deste método é preciso voltar ao final do século XIX, quando o cientista britânico Sir Francis Galton buscava respostas e explicações sobre as diferenças individuais em traços humanos como estatura, força e inteligência. A pergunta principal de Galton era: "*Quanto estas diferenças em traços humanos podem ser geneticamente determinadas?*". Recorreu então ao estudo entre pais e filhos. A proximidade com o matemático Karl Pearson fez com que este também se interessasse pelo problema, em termos de quantificações destas diferenças em traços.

Motivado com os problemas de quantificação da biologia, antropologia e psicologia do colega Galton, Pearson, voltou-se ao estudo de um índice que indicasse a natureza e o grau de relação entre duas variáveis métricas. Tratando as variáveis como vetores centrados em suas médias, sugere um índice de relação entre dois vetores muito utilizada já em álgebra linear, o cosseno do ângulo entre estes vetores. Surge o coeficiente de correlação de Pearson.

Com esta idéia lançada e aceita pelo meio, desde que baseada em uma estrutura teórica conhecida (evitando espuricidade), em 1897 G.U. Yule publica o *paper* clássico sobre as Correlações Parciais e Múltiplas (Yule, 1897). A idéia era, dada uma variável fixa em uma matriz de variáveis, procurá-se a combinação linear das demais variáveis que tenha a máxima correlação com a primeira. Esta seria a base para a AF.

---

---

Em 1901 Pearson publica o artigo sobre o “Método dos Eixos Principais” (Pearson, 1901), a base para o que conhecemos hoje por Componentes Principais, tanto a técnica quanto a forma de estimação das cargas fatoriais dos modelos atuais de AF.

Nesta época os psicólogos ingleses, inclusive Galton e Herbert Spencer, teorizavam sobre um fator geral que fosse comum a todas as habilidades de um indivíduo, em qualquer área de estudo. Charles Spearman, partidário desta teoria, escreve em 1904 o primeiro trabalho sobre AF: “*General Intelligence, Objectively Determined and Measured*” . O modelo proposto se baseava na equação

$$Z_j = a_j G + U_j$$

onde  $Z_j$  é a  $j$ -ésima resposta a um estímulo (score de um teste, por exemplo),  $a_j$  é a carga fatorial indicando o grau com que esta se associa com a habilidade-geral latente  $G$ , e  $U_j$  não correlacionada com  $G$  mas apenas com  $Z_j$  representando a parte específica desta. Este modelo ficou conhecido como o modelo de dois fatores, um geral e um específico.

Neste ponto a discussão passa a ser quanto a modelagem de fatores que explicassem os traços e/ou habilidades humanas. Alguns psicólogos britânicos como Cyril Burt e P.E. Vernon, argumentavam que adicionalmente ao fator geral, haviam alguns fatores “menos gerais” para habilidades também “menos gerais”, de forma hierárquica. Ou seja, “*a mente é organizada hierarquicamente de uma habilidade geral até algumas extremamente específicas*” (Mulaik, 1972, p. 7-8).

Por volta de 1930, L.L. Thurstone, psicólogo americano, parte do raciocínio de que as habilidades podem ser explicadas por um grupo de fatores gerais não apenas um e de forma não hierárquica, tornando o modelo proposto mais parcimonioso. Começa o desenvolvimento da Análise Fatorial Múltipla (AFM). A idéia é encontrar fatores que fossem relacionados com as variáveis observáveis de tal forma que estas formassem subgrupos com a menor redundância possível (do inglês *nonoverlapping*). Quando esta estrutura era encontrada era atribuído o nome de *simple structure*. Para tal, matematicamente nos modelos propostos, é necessária a suposição de não correlação entre os fatores, ou geometricamente falando, fatores ortogonais.

Após este período a AFM passou a ser utilizada nas mais diversas áreas da ciência, muitas vezes criticada pela aplicação cega e sem fundamentação teórica da estrutura do problema analisado (Freeman, 1962). A AFM muitas vezes apresenta resultados não

---

interpretáveis, ou até mesmo contraditórios, frutos da má aplicação, inadequação da natureza dos dados ou mesmo de problemas matemáticos quando se alteram algumas das suposições do modelo (Johnson *et al.*, 1998).

No entanto até os dias atuais esta técnica é largamente aplicada em problemas de psicometria como validação e estudo de instrumentos psicométricos, muitas vezes baseados em respostas qualitativas ordinais, onde é utilizada uma métrica imposta no desenvolvimento do instrumento. Ao se utilizar as matrizes de correlação ou covariância como base para a AFM está se aceitando esta métrica como verdadeira, além do fato de que traços e estados psicológicos, muitas vezes “medidos” por estes instrumento serem inclusive teóricamente fortemente associados.

### A.3.2 O Modelo de Análise Fatorial Ortogonal

Nesta seção será apresentado o modelo de análise fatorial mais utilizado atualmente, como proposto em Johnson e Wichern (1998).

Seja a matriz das variáveis observáveis  $\mathbf{X}_{(I \times J)}$  com  $J$  componentes, com vetor de médias  $\boldsymbol{\mu}$  e matriz de covariância  $\boldsymbol{\Sigma}$ . O modelo fatorial ortogonal sobre a matriz de correlação  $\rho$  é dado pelo sistema

$$\begin{aligned} \frac{X_1 - \mu_1}{\sigma_1} &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1p^*}F_{p^*} + \varepsilon_1 \\ \frac{X_2 - \mu_2}{\sigma_2} &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2p^*}F_{p^*} + \varepsilon_2 \\ &\vdots \\ \frac{X_J - \mu_J}{\sigma_J} &= l_{J1}F_1 + l_{J2}F_2 + \dots + l_{Jp^*}F_{p^*} + \varepsilon_{p^*} \end{aligned} \quad (\text{A.3.1})$$

ou em sua forma matricial

$$(\mathbf{X} - \mathbf{1}^T \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1/2} = \mathbf{L} \mathbf{F} + \boldsymbol{\varepsilon} \quad (\text{A.3.2})$$

onde  $l_{ij}$  são chamadas cargas fatoriais da  $i$ -ésima variável no  $j$ -ésimo fator comum  $F_j$ , sendo  $\mathbf{L}$  a matriz das cargas. O componente  $\varepsilon_i$  refere-se ao fator específico da  $i$ -ésima variável. Maxwell (1977) aponta que em alguns estudos este fator, também chamado de erro, tende a ser uma combinação de erros de medição e fatores que são de exclusividade da variável.

Pode-se entender então o modelo como sendo uma regressão múltipla nas observações tendo como regressoras variáveis não observáveis, ou latentes, ou ainda construtos. Este modelo na forma apresentada em (2.2) seria algo quase indeterminado, fazendo-se necessárias algumas suposições a respeito de  $\mathbf{F}$  e  $\varepsilon$ . É assumido que:

$$\begin{aligned} E(\mathbf{F}) &= \mathbf{0} & Cov(\mathbf{F}) &= \mathbf{I} \\ E(\varepsilon) &= \mathbf{0} & Cov(\varepsilon) &= \Delta = \begin{bmatrix} \delta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \delta_p \end{bmatrix}_{p \times p} \end{aligned} \quad (\text{A.3.3})$$

Estas condições definem o *Modelo Fatorial Ortogonal*.

O modelo ortogonal, por consequência, impõe a estrutura de correlação de  $\mathbf{X}$  tal que

$$\rho(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \Delta \quad (\text{A.3.4})$$

ou seja

$\rho(\mathbf{X}_j) = 1 = l_{j1}^2 + l_{j2}^2 + \cdots + l_{jm}^2 + \delta_j$  e  $\rho(\mathbf{X}_j, \mathbf{X}_k) = l_{j1}l_{k1} + \cdots + l_{jm}l_{km}$ . E ainda que  $\rho(\mathbf{X}, \mathbf{F}) = \mathbf{L}$ .

Temos que  $c_j^2 = l_{j1}^2 + l_{j2}^2 + \cdots + l_{jm}^2$  é a chamada **comunalidade**, ou seja, a parte da variância das variáveis observáveis que contribui para a estrutura de covariância total, e  $\delta_j^2$  a chamada de **especificidade**, ou seja, a componente da variância específica da variável  $\mathbf{X}_j$ .

Logo, por construção, os fatores resultantes são ortogonais, o que pode ser visto como uma limitação imposta pelo modelo, podendo ser inadequada quanto as relações entre os fatores. Estas relações se existentes acabam sendo explicitadas através da representação dos indivíduos através dos “escores fatoriais”, ou seja, dos valores “preditos” pelo modelo, se fizermos analogia com a regressão múltipla como anteriormente.

Este seria o primeiro modelo ajustado em um problema de análise fatorial. Posteriormente, baseado nas cargas, escores e medidas apresentados, bem como de outros índices de diagnóstico de modelo, a análise pode partir para rotações de eixos que melhor expliquem os fatores, estimação de fatores oblíquos, etc... Estas técnicas complementares não serão detalhadas neste trabalho. Boas referências são Johnson e Wichern(1998) e Thurstone (1967).

A estrutura de correlação em que o modelo se baseia, bem como as cargas fatoriais, podem ser estimadas por diferentes métodos. No caso da estrutura de correlação o método mais comum é o dos momentos cruzados de Pearson no caso de variáveis contínuas ou

---

discretas com uma amplitude razoável. Há ainda estimadores para dados binários como a Correlação Tetracórica (Nunnally, 1967, 123-124), ou a correlação por postos de Spearman.

A AF admite varias forma de estimação das cargas. Métodos via Máxima Verossimilhança, Mínimos Quadrados Generalizados, Centróide, Fator Principal, são alguns dos mais conhecidos. No entanto o mais utilizado e encontrado na literatura de validações psicométricas é a solução por Componentes Principais. O próximo item dá uma introdução nesta técnica de acordo com a proposta do Cap. 2.

A escola francesa considera a Análise de Componentes Principais um método fatorial (Lebart *et al.*, 1995), interpretando os eixos como uma estrutura latente, e não apenas como uma combinação linear das variáveis. Mas não deve ser confundida com a AF apresentada neste item, onde há a suposição de um modelo.

---

### A.3.3- A solução por Componentes Principais

Utilizando a notação do Capítulo 2, os Componentes Principais são os eixos principais obtidos de forma mais simples, ou seja, as matrizes  $\Omega = \mathbf{I}$  e  $\Phi = \Lambda^{-1}$ , onde  $\Lambda$  é a matriz diagonal dos Desvios Padrão das colunas de  $\mathbf{H}$  e  $\mathbf{I}$  é identidade. Sem perda de generalidade seja  $\mathbf{X}$ , matriz  $I \times J$  tal que

$$\mathbf{X} = (\mathbf{H} - \mathbf{u}^T \mathbf{h}) \Lambda^{-1}$$

onde  $\mathbf{h}^T = [h_1, \dots, h_J]$  e  $h_j = \frac{1}{I} \sum_{i=1}^I h_{ij}$ , ou o vetor de médias.

Desta forma a matriz  $\rho = \mathbf{X}^T \mathbf{X}$  a ser decomposta é a matriz de correlação das  $J$  variáveis e pode ser decomposta como . Pelos teoremas 2.1 e 2.2 a matriz  $\rho$  pode ser aproximada por

$$\rho = \mathbf{M} \mathbf{D}_\lambda \mathbf{M}^T = (\mathbf{M} \mathbf{D}_\lambda^{1/2}) (\mathbf{M} \mathbf{D}_\lambda^{1/2})^T = \mathbf{L} \mathbf{L}^T + \mathbf{0} \quad (\text{A.3.5})$$

que é a expressão em (A.3.4) fazendo  $\Delta = \mathbf{0}$ .

Então as cargas fatoriais  $\mathbf{L}$  para os  $P^*$  fatores são obtidas por

$$\mathbf{L} = [\varphi_1, \dots, \varphi_{P^*}] \quad (\text{A.3.6})$$

onde  $\varphi_s$ ,  $s = 1, \dots, P^*$  é definido em (2.4.14).

A comunalidade da  $j$ -ésima variável é dada então por

$$c_j^2 = \sum_{s=1}^{P^*} \varphi_{(j)s}^2 \quad (\text{A.3.7})$$

e as especificidades em  $\Delta$  dadas por

$$\delta_j = 1 - h_j^2. \quad (\text{A.3.8})$$

Desta forma temos as soluções fatoriais através de Componentes Principais.

## Apêndice B

### B.1- Decomposição em Valores Singulares (DVS)

A abordagem feita sobre esta técnica segue a de Eckart-Young (1936) e prova semelhante é encontrada em Cardoso (1991). Antes algumas definições.

**Definição b.1:**  $\mathfrak{R}_P^{I \times J}$  é a classe das matrizes reais não nulas  $I \times J$  de posto  $P$ .

**Definição b.2:** A representação dos pontos linha no espaço das colunas definida pela transformação  $H(\mathbf{v}): \mathfrak{R}^I \rightarrow \mathfrak{R}^J$ ,  $H(\mathbf{v}) = \mathbf{H}\mathbf{v}$ , tem imagem  $R_{JI}(\mathbf{H}) = \{ \mathbf{u} \in \mathfrak{R}^J : \mathbf{u} = \mathbf{H}\mathbf{v}, \text{ para algum } \mathbf{v} \in \mathfrak{R}^I \}$ . A transformação simétrica a esta,  $H^T(\mathbf{u}): \mathfrak{R}^J \rightarrow \mathfrak{R}^I$  tem imagem definida por  $R_{JI}(\mathbf{H}) = \{ \mathbf{u} \in \mathfrak{R}^J : \mathbf{v} = \mathbf{H}^T \mathbf{u}, \text{ para algum } \mathbf{v} \in \mathfrak{R}^I \}$ .

**Definição b.3:** Seja o espaço nulo de  $H(\cdot)$ ,  $\mathbf{0}(\mathbf{H}) = \{ \mathbf{x} \in \mathfrak{R}^n : \mathbf{H}\mathbf{x} = \mathbf{0} \}$ , onde  $n$  é a dimensão da contra-imagem.

**Definição b.4:** Seja  $\mathbf{H} \in \mathfrak{R}_P^{I \times J}$  e  $\Lambda_P(\mathbf{H})$  matriz diagonal contendo os  $p$  autovalores de  $\mathbf{H}\mathbf{H}^T$  tal que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ . Então os Valores Singulares de  $\mathbf{H}$  são definidos por  $\sigma_p = \sqrt{\lambda_p}$  para  $p = 1, \dots, P$ . Então  $\Sigma_P(\mathbf{H}) = (\Lambda_P(\mathbf{H}))^{1/2}$  é a matriz diagonal dos Valores Singulares de  $\mathbf{H}$  ordenados de forma decrescente.

**Teorema 1:** (*Relação Baricêntrica*) Seja  $\mathbf{H} \in \mathfrak{R}_P^{I \times J}$  e  $\Sigma(\mathbf{H})$  como definida em b.4. Seja  $\mathbf{U}_P = [\mathbf{u}_1, \dots, \mathbf{u}_P]$  o conjunto dos autovetores de  $\mathbf{H}\mathbf{H}^T$  correspondendo a seus  $P$  autovalores não nulos tal que

$$\mathbf{H}\mathbf{H}^T \mathbf{u}_p = \lambda_p \mathbf{u}_p \quad (\text{b.1.1})$$

$$(\mathbf{u}_p, \mathbf{u}_q) = \mathbf{1}(p=q) \quad (\text{b.1.2})$$

onde  $p, q = 1, \dots, P$  e  $\mathbf{1}(p=q)$  é função indicadora para  $p = q$ .

Seja  $\mathbf{V}_P$  conjunto de  $P$  vetores tais que

$$\mathbf{v}_p = \frac{1}{\sigma_p} \mathbf{H}^T \mathbf{u}_p \quad (\text{b.1.3}).$$

Então  $\mathbf{V}_P$  é o conjunto de autovetores associados aos  $P$  autovalores não nulos de  $\mathbf{H}^T \mathbf{H}$  tais que

$$\mathbf{H}^T \mathbf{H} \mathbf{v}_p = \lambda_p \mathbf{v}_p \quad (\text{b.1.4})$$

$$(\mathbf{v}_p, \mathbf{v}_q) = \mathbf{1}(p = q) \quad (\text{b.1.5})$$

e ainda

$$\mathbf{u}_p = \frac{1}{\sigma_p} \mathbf{H}^T \mathbf{v}_p \quad (\text{b.1.6}).$$

Então de forma recíproca se  $\mathbf{V}_P$  satisfaz (b.1.4) e (b.1.5) e se  $\mathbf{U}_P$  é definido por (b.1.6), então satisfaz (b.1.1), (b.1.2) e (b.1.3).

**Prova:** Será demonstrada primeiramente a relação em (b.1.4).

Seja  $\mathbf{v}_p$  como definido anteriormente. Então pré-multiplicando ambos lados de (b.1.3) por  $\mathbf{H}^T \mathbf{H}$  temos

$$\mathbf{H}^T \mathbf{H} \mathbf{v}_p = \frac{1}{\sigma_p} \mathbf{H}^T \mathbf{H} \mathbf{H}^T \mathbf{u}_p = \sigma_p \mathbf{H}^T \mathbf{u}_p = \sigma_p^2 \mathbf{v}_p = \lambda_p \mathbf{v}_p.$$

Agora para (b.1.5) temos que

$$(\mathbf{v}_p, \mathbf{v}_q) = \frac{1}{\sigma_p \sigma_q} (\mathbf{H}^T \mathbf{u}_p, \mathbf{H}^T \mathbf{u}_q) = \frac{1}{\sigma_p \sigma_q} (\mathbf{H} \mathbf{H}^T \mathbf{u}_p, \mathbf{u}_q) = \frac{\sigma_p}{\sigma_q} (\mathbf{u}_p, \mathbf{u}_q) = \mathbf{1}(p = q).$$

A relação em (b.1.6) pode ser provada tomando-se (b.1.3) e pré-multiplicando-se ambos lados por  $\mathbf{H}$  tem-se:

$$\mathbf{H} \mathbf{v}_p = \frac{1}{\sigma_p} \mathbf{H} \mathbf{H}^T \mathbf{u}_p = \sigma_p \mathbf{u}_p \Rightarrow \mathbf{u}_p = \frac{1}{\sigma_p} \mathbf{H} \mathbf{v}_p$$

A prova das proposições recíprocas é análoga, substituindo-se  $\mathbf{H}$  por  $\mathbf{H}^T$  em cada passo.



**Teorema 2** (*Decomposição em Valores Singulares*)

Seja  $\mathbf{H} \in \mathfrak{R}_P^{I \times J}$ . Então existem matrizes unitárias  $\mathbf{U}$ ,  $I \times I$ , e  $\mathbf{V}$ ,  $J \times J$  tais que a matriz  $I \times J$

$$\Sigma(\mathbf{H}) = \mathbf{U}^T \mathbf{H} \mathbf{V} = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \sigma_P & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{0} \end{pmatrix} \quad (\text{b.2.7})$$

e conseqüentemente

$$\mathbf{H} = \mathbf{U} \Sigma(\mathbf{H}) \mathbf{V}^T \quad (\text{b.2.8}).$$

**Prova:**

Seja  $\mathbf{U} = [\mathbf{U}_P \ \mathbf{U}_0]$  com  $\mathbf{U}_P$  base ortonormal definida no Teorema 1 formando uma base de  $\mathbf{R}(\mathbf{H}\mathbf{H}^T) = \mathbf{R}(\mathbf{H})$ , e  $\mathbf{U}_0 = [\mathbf{u}_{P+1}, \dots, \mathbf{u}_I]$  base ortonormal de  $\mathbf{0}(\mathbf{H})$ . então  $\mathbf{U}$  é uma base ortonormal em  $\mathfrak{R}^I$  e

$$\mathbf{H}^T \mathbf{u}_p = 0 \quad \text{para} \quad p = P+1, \dots, I \quad (\text{b.2.9}).$$

De forma análoga seja  $\mathbf{V} = [\mathbf{V}_P \ \mathbf{V}_0]$  com  $\mathbf{V}_P$  base ortonormal definida no Teorema 1 formando uma base de  $\mathbf{R}(\mathbf{H}^T \mathbf{H}) = \mathbf{R}(\mathbf{H})$ , e  $\mathbf{V}_0 = [\mathbf{v}_{P+1}, \dots, \mathbf{v}_J]$  base ortonormal de  $\mathbf{0}(\mathbf{H}^T)$ . então  $\mathbf{V}$  é uma base ortonormal em  $\mathfrak{R}^J$  e

$$\mathbf{H} \mathbf{v}_p = 0 \quad \text{para} \quad p = P+1, \dots, J \quad (\text{b.2.10}).$$

Desta forma, considerando  $\mathbf{U}$  e  $\mathbf{V}$ , a matriz

$$\Sigma(\mathbf{H}) = \mathbf{U}^T \mathbf{H} \mathbf{V} = \{\sigma_{ij}\} \quad \text{para} \quad i = 1, \dots, I \quad \text{e} \quad j = 1, \dots, J$$

satisfaz

$\sigma_{ij} = \mathbf{u}_i^T \mathbf{H} \mathbf{v}_j$  se  $i > P$  ou se  $j > P$  por (b.2.9) e (b.2.10). E para  $i, j = 1, \dots, P$  temos

$$\begin{aligned} \sigma_{ij} &= \mathbf{u}_i^T \mathbf{H} \mathbf{v}_j \\ &= (1/\sigma_j) \mathbf{u}_i^T \mathbf{H} \mathbf{H}^T \mathbf{u}_j && \text{por (b.1.3)} \\ &= \sigma_j \mathbf{u}_i^T \mathbf{u}_j && \text{por (b.1.1)} \\ &= \sigma_j \mathbf{1}(i=j) = \sigma_j. \end{aligned}$$

Como  $\mathbf{U}$  e  $\mathbf{V}$  são unitárias temos que

$$\mathbf{H} = \mathbf{U} \Sigma(\mathbf{H}) \mathbf{V}^T .$$

**Teorema 3** (Eckart – Young)

Se os  $P - P^*$  termos de (b.2.8) são desprezados, então

$$\mathbf{S}^* = \mathbf{H}^* \mathbf{H}^{*T} = \mathbf{U}_{P^*} \Lambda_{P^*}(\mathbf{H}) \mathbf{V}_{P^*}^T = \sum_p^{P^*} \lambda_p \mathbf{u}_p \mathbf{v}_p^T \quad (\text{b.3.11})$$

onde  $\Lambda_{P^*}(\mathbf{H})$  é a matriz diagonal contendo os  $P^*$  maiores auto-valores de  $\mathbf{H}$  tal que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{P^*}$ , é a “aproximação por mínimos quadrados generalizados de ordem  $P^*$ ” que minimiza a função perda dada por

$$\sigma(\mathbf{H}^*) \equiv \text{tr}\{(\mathbf{S} - \mathbf{X}^*)^T (\mathbf{S} - \mathbf{X}^*)\} \quad (\text{b.3.12})$$

entre todas as matrizes  $\mathbf{X}^*$  de posto  $P^*$ .

**prova:**

Podemos escrever (b.3.12) como

$$\begin{aligned} \text{tr}\{\mathbf{U}\mathbf{U}^T (\mathbf{H} - \mathbf{X}^*) \mathbf{V}\mathbf{V}^T (\mathbf{H} - \mathbf{X}^*)^T\} &= \text{tr}\{(\mathbf{U}\mathbf{H}\mathbf{V}^T - \mathbf{U}\mathbf{X}\mathbf{V}^T)(\mathbf{U}\mathbf{H}\mathbf{V}^T - \mathbf{U}\mathbf{X}\mathbf{V}^T)^T\} = \\ &= \text{tr}\{(\Lambda(\mathbf{H}) - \mathbf{P}^*)(\Lambda(\mathbf{H}) - \mathbf{P}^*)^T\} = \sum (\lambda_p - \rho_{pp})^2 + \sum_{p=1}^P \sum_{p \neq q}^P \rho_{pq}^2 \end{aligned}$$

onde  $\mathbf{P}^*$  tem mesmo posto de  $\mathbf{X}^*$ . Logo é claro que  $\mathbf{P}^* \equiv \Lambda_{P^*}(\mathbf{H})$ , o que implica que  $\mathbf{X} = \mathbf{S}^*$  definido em (b.3.11) é ótimo.

□

## Apêndice C

### C.1- Programas em Ox 2.1 utilizados nas simulações

As simulações foram geradas em Ox 2.1 (Doornick, 1997) e seguiram todas a um mesmo algoritmo baseado em duas funções desenvolvidas para este intuito, *GeraConstruto* e *GeraMultInd*, que utilizam a função *runmultinomial* da própria linguagem.

A função *runmultinomial* retorna um vetor com a distribuição de frequência em  $n$  ensaios de uma distribuição Multinomial  $(n, \mathbf{p})$  e tem como parâmetros o número de ensaios  $n$  e o vetor de probabilidades  $\mathbf{p}$ .

A função desenvolvida *GeraConstruto* (quadro c.1) utiliza-se desta função para retornar o número do estado em que a cadeia de Markov se apresenta no  $i$ -ésimo passo. A cadeia se inicia com o sorteio do estado inicial, segundo uma uniforme discreta, e após este sorteio gera as  $J=passo$  observações seguintes segundo a matriz de transição definida pelos vetores  $\mathbf{v1}$  a  $\mathbf{v5}$  como linhas desta, onde  $\mathbf{v}_s = p_{(i, j+1)}(s) = P(x_{j+1}^{(g,c)} = s \mid x_j^{(g,c)} = s^*)$  definidos conforme o capítulo 4.

A função *GeraMultInd* gera  $J=passo$  realizações de uma multinomial  $(1, \mathbf{v})$  definida pelo vetor  $\mathbf{v}$  (quadro c.2) que corresponde a uma observação do grupo = *grupo*.

Desta forma cada simulação utilizou-se da mesma estrutura sendo alterados apenas as combinações entre vetores e funções. Os programas para a simulação 1 e da simulação 5 estão apresentados.

```

/*****
  GeraConstruto(const v1, const v2, const v3, const v4, const v5, const
passos, const grupo);
**          v1    in: vetor 1 x 5
**          v2    in: vetor 1 x 5
**          v3    in: vetor 1 x 5
**          v4    in: vetor 1 x 5
**          v5    in: vetor 1 x 5
**          passo in: inteiro
**          grupo in: inteiro
**
**Return Value
**  vector: 1 x passo
*****/
GeraConstruto(const v1, const v2, const v3, const v4, const v5, const
passos, const grupo)
{
  decl v_prob, v_mult, i_aux;
      decl i, v_indir, v_condenr;

inicial      v_prob=<0.2,0.2,0.2,0.2,0.2>; // Sorteio da categoria

      v_mult= ranmultinomial(1, v_prob);
      i_aux= vecindex(v_mult, 1)+1;

  for (i=0; i<passos; ++i)
  {
      if (i_aux==1)
      {
          v_prob= v1;
          v_mult= ranmultinomial(1, v_prob);
          i_aux= vecindex(v_mult, 1)+1;
      }
      else if (i_aux==2)
      {
          v_prob= v2;
          v_mult= ranmultinomial(1, v_prob);
          i_aux= vecindex(v_mult, 1)+1;
      }
      else if (i_aux==3)
      {
          v_prob= v3;
          v_mult= ranmultinomial(1, v_prob);
          i_aux= vecindex(v_mult, 1)+1;
      }
      else if (i_aux==4)
      {
          v_prob= v4;
          v_mult= ranmultinomial(1, v_prob);
          i_aux= vecindex(v_mult, 1)+1;
      }
      else if (i_aux==5)

```

```
        {
            v_prob= v5;
            v_mult= ranmultinomial(1, v_prob);
            i_aux= vecindex(v_mult, 1)+1;
        }

    if (i==0)// Gera as linhas das matrizes Indicadora
    {
        // e Condensada

        v_indir = v_mult;
        v_condenr = i_aux;
    }
    else
    {
        v_indir=v_indir~v_mult;
        v_condenr=v_condenr~i_aux;
    }
}
return v_condenr ;
```

Quadro c.1 – Programa da função *GeraConstruto* (continuação)

```

/*****
** GeraMultInd(const v1, const passos);
**           v1   in: vetor 1 x 5
**           passo in: inteiro
**
**
**Return Value
**   vector: 1 x passo
*****/
GeraMultInd(const v1, const passos)
{
    decl v_prob, v_mult, i_aux;
    decl i, v_indir, v_condenr;

    for(i=0; i<passos; ++i)
    {
        v_prob= v1;
        v_mult= ranmultinomial(1, v_prob);
        i_aux= vecindex(v_mult, 1)+ 1;

        if(i==0)
        {
            v_indir= v_mult;
            v_condenr= i_aux;
        }
        else
        {
            v_indir= v_indir-v_mult;
            v_condenr= v_condenr~i_aux;
        }
    }

    return v_condenr ;
}

```

Quadro c.2 – Programa da função *GeraMultInd*

*progsim5.ox - Programa em Ox 2.1 para simulação 5 (Capítulo 4)*

```

#include <oxstd.h>
#include <oxprob.h>
#include "GeraConstruto.ox"
#include "GeraMultiInd.ox"

main()
{
    decl    n,v_constr1,v_constr2,v_constr3,v_constr4,v_controle,    v_obs2,
    v_obs3,v_obs4,
           v_indep1,v_indep2, v_obs1;
    decl vi,vi, vu, vb1, vb2, vb3, vb4, vb5, vc1, vc2, vc3, vc4, vc5,
    passos, grupo ;
    decl mediac, medial, medial1, mediac1, medial2, mediac2, medial3,
    mediac3, mediac4;
    decl m_indic, m_conden, m_conden1, m_conden2, m_conden3, m_conden4,
    m_conden5 ;

    // Gera indivíduos do Grupo I

    vb1= <0.4,0.2,0.2,0.1,0.1>;
    vb2= <0.2,0.2,0.2,0.2,0.2>;
    vb3= <0.2,0.2,0.2,0.2,0.2>;
    vb4= <0.2,0.2,0.2,0.2,0.2>;
    vb5= <0.1,0.1,0.2,0.2,0.4>;

    vu =    <0.35,0.15,0.05,0.15,0.3>;
    vi =    <0.2,0.2,0.2,0.2,0.2>;
    vii= <0.7, 0.2, 0.05, 0.05, 0>;

    passos= 20; grupo= 1;

    for(n=0; n<1000; ++n)
    {
        v_constr1= GeraMultiInd(vu,passos);
        v_indep1= GeraMultiInd(vi, passos);
        v_obs1= grupo-v_constr1;           // Atribui o grupo ao
                                         indivíduo

        if (n==0)
        {
            m_conden1= v_obs1;           // Testa se o
            // Se sim, recebe
            // do primeiro
            // Se não, anexa o
            // indivíduo
        }
        m_conden1= m_conden1|v_obs1;
    }
}

```

quadro c.3- Programa para Simulação 5

```
mediac1= meanc(m_conden1);
println("Média das col", mediac1);

// Gera indivíduos do Grupo II

vc1= <0.05,0.1,0.15,0.30,0.40>;
vc2= <0,0.15,0.15,0.3,0.4>;
vc3= <0,0.05,0.2,0.35,0.4>;
vc4= <0,0,0.25,0.40,0.35>;
vc5= <0,0,0,0.4,0.6>;

passos= 20; grupo= 2;

for(n=0; n<1000; ++n)
{
    v_constr2=
GeraConstruto(vc1,vc2,vc3,vc4,vc5,passos,grupo);
    v_indep2= GeraMultInd(vii, passos);
    v_obs2= grupo-v_indep2-v_constr2;

    if (n==0)
    {
        m_conden2= v_obs2;
    }
    else
    {
        m_conden2= m_conden2|v_obs2;
    }
}

    mediac2= meanc(m_conden2);
    println("Média das col", mediac2);

// Gera indivíduos do Grupo III

grupo= 3; passos=20;

for(n=0; n<1000; ++n)
{
    v_constr3=
GeraConstruto(vb1,vb2,vb3,vb4,vb5,passos,grupo);
    v_constr4= v_indep2= GeraMultInd(vii, passos) ;
    v_obs3= grupo-v_constr3-v_constr4;

    if (n==0)
    {
        m_conden3= v_obs3;
    }
    else
    {
        m_conden3= m_conden3|v_obs3;
    }
}
```

quadro c.3- Programa para Simulação 5

```
}
        mediac3= meanc(m_conden3);
        println("Média das col", mediac3);

// Gera indivíduos do Grupo IV

        passos=40; grupo=4;

        for(n=0; n<1000; ++n)
        {
            v_controle= GeraMultInd(vi, passos);
            v_obs4= grupo-v_controle ;

            if (n==0)
            {
                m_conden4= v_obs4;
            }
            else
            {
                m_conden4= m_conden4|v_obs4;
            }
        }

        mediac4= meanc(m_conden4);
        println("Média das col", mediac4); */

//Anexa as matrizes, imprime perfis coluna, e grava matriz
condensada

        m_conden= m_conden1
        mediac= meanc(m_conden);
        println("Média das col", mediac);

}
```

quadro c.3- Programa para Simulação 5 (continuação)

## Referências

- Andrade, D.F., H.R. Tavares e R.C. Valle (2000) *Teoria de Resposta ao Item: Conceitos e Aplicações*, Livro texto de minicurso, 14º SINAPE/ABE, Caxambú
- Benzécri, J.-P.(1973) *L'analyse des Données. Tome 1: La Taxonomie, Tome 2: L'Analyse des Correspondences*, Dunod, Paris
- Binet, A. e T. Simon (1905) *Méthods nouvelles pour le diagnostic du niveau intellectuel des anormaux*, L'Année Psychologique. II, 191-244
- Burt, C. (1951) *Test construction and the scaling of itens*, British Journal of Psychology, 4, 85-129
- Burt, C. (1950)a *The influence of differential weighting*, British Journal of Psychology, 3, 105-128
- Burt, C. (1950)b *The factorial analysis of qualitative data*, British Journal of Psychology, 3, 166-185
- Cardoso, R.H. (1991) *Análise de Dados Multivariados Através de Técnicas Baseadas nas Decomposições em Valores Singulares*, Dissertação de Mestrado, IMECC/UNICAMP, Campinas
- Caroll, J.D. (1968) *Generalization of canonical correlation to three or more sets of variables*, Proceedings of American Psychological Association, 227-228
- Cattel, J.McK. (1890) *Mental tests and measurements*, Mind, 15, 373-381
- Cattel, R.B. (1966) *The scree test for the number of factors*, Multivariate Behavioral Research, 1, 245-276
- CISIA-CERESTA (1998), SPAD 3.5, CISIA-CERESTA, Saint-Mandé
- Crivisqui, E. e A. Castiglioni (1998) *Generalização da análise fatorial de correspondências: Análise Fatorial de Correspondências Múltiplas (trad. port.)*, In: Seminário de Métodos Estatísticos Multivariados Aplicado às Ciências Humanas. Material de apoio, IE/ UNICAMP, Agosto/1998
- De Leeuw, J. (1984) *The Gifi-system of non-linear multivariate analysis*, In: "Data Analysis and Mathematics, Vol. III"(ed. Diday, E.; Jambu, M.; Lebart, L.; Pages e Tomassone, R.), Amsterdam
- De Vellis (1999) *Theory and methods for scale development*
- Derogatis, L.R. (1994) *SCL-90-R (Symptom Check List-90-R): Administration, Scoring and Procedures Manual*, 3ª ed., National Computer Systems; Inc., Minneapolis
- Doornik, J.A. (1998) *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants, Londres
- Eckart, C. e Young, G. (1936) *The approximation of one matrix by another of lower rank*, Psychometrika, 1, 211-218
- Edgeworth, F.Y. (1888) *The statistics of examination*, Journal of the Royal Statistical Society, serie A(General), 51, 599-635

- Freeman, F.S. (1962) *Theory and Practice of Psychological Testing*, 3<sup>a</sup> ed., Holt, Rineart and Winston, Inc., Nova York
- Fuller, E.L. e Hemmerle (1966) *Robustness of the maximum likelihood estimation procedure in factor analysis*, *Psychometrika*, 31, 255-266
- Galton, F. (1888) *Co-relations and their measurement, chiefly from anthropometric*, *Proceedings of the Royal Society*, 45, 135-145
- Gaush, H.G. (1982) *Multivariate Analysis in Community Ecology*, Cambridge University Press, Cambridge
- Gifi, A. (1990) *Nonlinear Multivariate Analysis*, John Wiley & Sons, New York
- Greenacre, M.J. (1984) *Theory and Applications of Correspondence Analysis*, Academic Press, London
- Greenacre, M.J. e Hastie, T. (1987) *The geometric interpretation of Correspondence Analysis*, *Journal of American Statistical Association*, 82, #398, 437-447
- Guilford, J.P. (1954) *Psychometric Methods*, 2<sup>nd</sup> ed., McGraw-Hill, Nova York
- Guttman, L (1941) *The quantification of a class of atributes: A theory and method of scale construction*. In: "The Prediction of Personal Adjustment" (Horst P., ed.), Social Science Research Concl, Nova York, 319-348
- Guttman, L. (1944) *A basis for scaling qualitative data*, *American Sociol. Review*, 9, 139-150
- Guttman, L. (1950) *The principal components of scale analysis*, *Measurement and Prediction*, Princeton University Press, Princeton
- Hayashi, C. (1950) *On the quantification of qualitative data from the mathematico-statistical point of view*, *Ann. Inst. Statist. Math.*, 2, 35-47
- Hayashi, C. (1956) *Theory and examples of quantification (II)*, *Proceedings of the Institute of Stat. Math.*, 4 (2), 19-30
- Johnson, A.R. e Wichern D.W. (1998) *Applied Multivariate Statistical Analysis*, 4<sup>th</sup> ed., Prentice Hall, New Jersey
- Kaplan, D. (1991) *The behaviour of three weighted least squares estimators for structured means analysis with non-normal Likert variables*, *British Journal of Math. and Stat. Psychology*, 44, 333-346
- Kendall, D.G. (1971) *Seriation from abundance matrices*, In: "Mathematics in the Archeological and Historical Sciences (ed. Hodson, C.R.; Kendall, D.G. e Táutu, P.) Edinburgh University Press, Edinburgh, 215-252
- Kruskal, W.H. (1978) *Psychometrics*, In: "International encyclopedia of statistics" (ed. W.H. Kruskal and Judith M. Tanun), The Free Press, New York, Vol. 2, 764 – 782
- Lebart, M. , A. Morineau e M. Piron (1995) *Statistique Exploratoire Multidimensionnelle*, Dunod, Paris
- Likert (1932) *A technique for measurement of attitudes*, *Archives of Psychology*, 140, 55 ps

- 
- Muthén, B e Kaplan, D. (1985) *A comparison of some methodologies for the factor analysis of non-normal Likert variables*, British Journal of Math. and Stat. Psychology, 38, 171-189
- Muthén, B e Kaplan, D. (1992) *A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model*, British Journal of Math. and Stat. Psychology, 45, 19-30
- Nunnally, J.C. (1967) *Psychometric Theory*, McGraw-Hill, Nova York
- Pasquali, Luiz (1999) *Escalas psicométricas* In: "Instrumentos Psicológicos: Manual Prático de Elaboração (ed. Luiz Pasquali), LabPAM/IBAP, Brasília, 105-126
- Pearson, K. (1901) *On lines and planes of closest fit to points in space*, Phil. Magazine, 2, 559-572
- Pereira, J.C.R. (1998) *Análise de Dados Qualitativos: Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais*, Edusp, São Paulo
- Ross, S. (1996) *Stochastic Processes*, John Wiley & Sons, Inc., Nova York
- Schriever, B.F. (1983) *Scaling of order dependent categorical variables with Correspondence Analysis*, International Statistical Review, 51, 225-238
- Spearman, C. (1904) *General Inteligence objectively determined and measured*, American Journal of Psychology, 15, 201-293
- Spearman, C. (1913) *Correlation of sums and differences*, British Journal of Psychology, 5, 417-423
- StatSoft (1996), Statistica 5.0
- Stata (1995), Stata 5.0 intercooled, Texas
- Thorndike, R.L. e E. Hagen (1961) *Measurement and Evaluation in Psychology and Education*, 2ª ed., John Wiley & Sons, Inc., Nova York
- Thurstone, L.L. (1927) *A law of comparative judgement*, Psychological Review, 34, 273-286
- Thurstone, L.L. (1947) *Modern Factor Analysis*, The University of Chicago Press, Chicago
- Torgerson, W.S. (1958) *Theory and Methods of Scaling*, John Wiley & Sons, Nova York
- Weller, S.C. e Romney A.K. (1990) *Metric Scaling: Correspondence Analysis*, Sage University Papers in Social Sciences, series 07-075, Sage, Newbury Park, California
- vanRijckevorsel (1986) *About horseshoes in Multiple Correspondence Analysis*, In: "Classification as a Tool of Research (ed. W. Gaul e M. Schader), Elsevier Science Publishers B.V., North-Holland
- Yule, G.U. (1897) *On the theory of correlation*, Journal of the Royal Statistical Society, 60
- Zeller, R.A. e E. G. Carmines (1980) *Measurement in Social Sciences: The Link Between theory and Data*", Cambridge University Press, Nova York
-



## **ANEXOS**

### **Distribuições Marginais dos Itens**

**Instrumento Aplicado**

# Histogramas das marginais totais dos itens - SCL 90 R

	ABSOLU	%/TOTAL	%/EXPR.	HISTOGRAMME DES POIDS
3 . Q1				
AI_1 - 0	291	48.58	48.58	*****
AI_2 - 1	140	23.37	23.37	*****
AI_3 - 2	52	8.68	8.68	****
AI_4 - 3	62	10.35	10.35	****
AI_5 - 4	54	9.02	9.02	****
ENSEMBLE	599	100.00	100.00	
4 . Q2				
AJ_1 - 0	194	32.39	32.39	*****
AJ_2 - 1	161	26.88	26.88	*****
AJ_3 - 2	65	10.85	10.85	****
AJ_4 - 3	103	17.20	17.20	*****
AJ_5 - 4	76	12.69	12.69	*****
ENSEMBLE	599	100.00	100.00	
5 . Q3				
AK_1 - 0	284	47.41	47.41	*****
AK_2 - 1	93	15.53	15.53	*****
AK_3 - 2	50	8.35	8.35	****
AK_4 - 3	97	16.19	16.19	*****
AK_5 - 4	75	12.52	12.52	*****
ENSEMBLE	599	100.00	100.00	
6 . Q4				
AL_1 - 0	261	43.57	43.57	*****
AL_2 - 1	156	26.04	26.04	*****
AL_3 - 2	73	12.19	12.19	*****
AL_4 - 3	59	9.85	9.85	****
AL_5 - 4	50	8.35	8.35	****
ENSEMBLE	599	100.00	100.00	
7 . Q5				
AM_1 - 0	344	57.43	57.43	*****
AM_2 - 1	98	16.36	16.39	*****
AM_3 - 2	28	4.67	4.66	***
AM_4 - 3	57	9.52	9.53	****
AM_5 - 4	71	11.85	11.87	*****
ENSEMBLE	598	99.83	100.00	
8 . Q6				
AN_1 - 0	371	61.94	61.94	*****
AN_2 - 1	112	18.70	18.70	*****
AN_3 - 2	34	5.68	5.68	***
AN_4 - 3	44	7.35	7.35	****
AN_5 - 4	38	6.34	6.34	***
ENSEMBLE	599	100.00	100.00	
9 . Q7				
AO_1 - 0	411	68.51	68.73	*****
AO_2 - 1	83	13.86	13.88	*****
AO_3 - 2	32	5.34	5.35	***
AO_4 - 3	44	7.35	7.36	****
AO_5 - 4	28	4.67	4.68	***
ENSEMBLE	598	99.83	100.00	
10 . Q8				
AP_1 - 0	345	57.60	57.69	*****
AP_2 - 1	121	20.20	20.23	*****
AP_3 - 2	36	6.01	6.02	***
AP_4 - 3	55	9.18	9.20	****
AP_5 - 4	41	6.84	6.86	****
ENSEMBLE	598	99.83	100.00	
11 . Q9				
AQ_1 - 0	260	43.41	43.55	*****
AQ_2 - 1	147	24.54	24.62	*****
AQ_3 - 2	48	8.01	8.04	****
AQ_4 - 3	83	13.86	13.90	*****
AQ_5 - 4	59	9.85	9.88	*****
ENSEMBLE	597	99.67	100.00	
12 . Q10				
AR_1 - 0	276	46.08	46.08	*****
AR_2 - 1	150	25.04	25.04	*****
AR_3 - 2	52	8.68	8.68	****
AR_4 - 3	72	12.02	12.02	*****
AR_5 - 4	49	8.18	8.18	****
ENSEMBLE	599	100.00	100.00	
13 . Q11				
AS_1 - 0	184	30.72	30.72	*****
AS_2 - 1	136	22.70	22.70	*****
AS_3 - 2	86	14.36	14.36	*****
AS_4 - 3	98	16.36	16.36	*****
AS_5 - 4	95	15.86	15.86	*****
ENSEMBLE	599	100.00	100.00	
14 . Q12				
AT_1 - 0	331	55.26	55.26	*****
AT_2 - 1	109	18.20	18.20	*****
AT_3 - 2	50	8.35	8.35	****
AT_4 - 3	55	9.18	9.18	*****
AT_5 - 4	54	9.02	9.02	*****
ENSEMBLE	599	100.00	100.00	

15 - Q13					
AU_1 - 0	365	60.93	61.04	*****	
AU_2 - 1	85	14.19	14.21	*****	
AU_3 - 2	33	5.51	5.52	***	
AU_4 - 3	57	9.52	9.53	*****	
AU_5 - 4	58	9.68	9.70	*****	
ENSEMBLE	598	99.83	100.00		
-----					
16 - Q14					
AV_1 - 0	214	35.73	35.79	*****	
AV_2 - 1	175	29.22	29.26	*****	
AV_3 - 2	84	14.02	14.05	*****	
AV_4 - 3	73	12.19	12.21	*****	
AV_5 - 4	52	8.68	8.70	****	
ENSEMBLE	598	99.83	100.00		
-----					
17 - Q15					
AW_1 - 0	478	79.80	79.93	*****	
AW_2 - 1	47	7.85	7.86	****	
AW_3 - 2	15	2.50	2.51	**	
AW_4 - 3	26	4.34	4.35	**	
AW_5 - 4	32	5.34	5.35	***	
ENSEMBLE	598	99.83	100.00		
-----					
18 - Q16					
AX_1 - 0	469	78.30	78.43	*****	
AX_2 - 1	69	11.52	11.54	*****	
AX_3 - 2	18	3.01	3.01	**	
AX_4 - 3	26	4.34	4.35	**	
AX_5 - 4	16	2.67	2.68	**	
ENSEMBLE	598	99.83	100.00		
-----					
19 - Q17					
AY_1 - 0	334	55.76	55.85	*****	
AY_2 - 1	118	19.70	19.73	*****	
AY_3 - 2	49	8.18	8.19	****	
AY_4 - 3	56	9.35	9.36	*****	
AY_5 - 4	41	6.84	6.86	****	
ENSEMBLE	598	99.83	100.00		
-----					
20 - Q18					
AZ_1 - 0	156	26.04	26.22	*****	
AZ_2 - 1	134	22.37	22.52	*****	
AZ_3 - 2	75	12.52	12.61	*****	
AZ_4 - 3	119	19.87	20.00	*****	
AZ_5 - 4	111	18.53	18.66	*****	
ENSEMBLE	599	99.33	100.00		
-----					
21 - Q19					
BA_1 - 0	321	52.59	52.59	*****	
BA_2 - 1	112	18.70	18.70	*****	
BA_3 - 2	58	9.68	9.68	*****	
BA_4 - 3	63	10.52	10.52	*****	
BA_5 - 4	45	7.51	7.51	****	
ENSEMBLE	599	100.00	100.00		
-----					
22 - Q20					
BB_1 - 0	236	39.40	39.40	*****	
BB_2 - 1	94	15.69	15.69	*****	
BB_3 - 2	62	10.35	10.35	*****	
BB_4 - 3	95	15.86	15.86	*****	
BB_5 - 4	112	18.70	18.70	*****	
ENSEMBLE	599	100.00	100.00		
-----					
23 - Q21					
BC_1 - 0	350	58.43	58.63	*****	
BC_2 - 1	109	18.20	18.26	*****	
BC_3 - 2	40	6.68	6.70	****	
BC_4 - 3	52	8.68	8.71	*****	
BC_5 - 4	46	7.68	7.71	****	
ENSEMBLE	597	99.67	100.00		
-----					
24 - Q22					
BD_1 - 0	310	51.75	51.75	*****	
BD_2 - 1	112	18.70	18.70	*****	
BD_3 - 2	48	8.01	8.01	****	
BD_4 - 3	73	12.19	12.19	*****	
BD_5 - 4	56	9.35	9.35	*****	
ENSEMBLE	599	100.00	100.00		
-----					
25 - Q23					
BE_1 - 0	345	57.60	57.60	*****	
BE_2 - 1	119	19.87	19.87	*****	
BE_3 - 2	43	7.18	7.18	****	
BE_4 - 3	47	7.85	7.85	****	
BE_5 - 4	45	7.51	7.51	****	
ENSEMBLE	599	100.00	100.00		
-----					
26 - Q24					
BF_1 - 0	277	46.24	46.32	*****	
BF_2 - 1	117	19.53	19.57	*****	
BF_3 - 2	44	7.35	7.36	****	
BF_4 - 3	87	14.52	14.55	*****	
BF_5 - 4	73	12.19	12.21	*****	
ENSEMBLE	598	99.83	100.00		
-----					

27 . Q25					
BG_1 - 0	393	65.61	65.61	*****	
BG_2 - 1	76	12.69	12.69	*****	
BG_3 - 2	29	4.84	4.84	***	
BG_4 - 3	39	6.51	6.51	***	
BG_5 - 4	62	10.35	10.35	*****	
ENSEMBLE	599	100.00	100.00		
28 . Q26					
BH_1 - 0	334	55.76	55.95	*****	
BH_2 - 1	116	19.37	19.43	*****	
BH_3 - 2	46	7.68	7.71	****	
BH_4 - 3	57	9.52	9.55	*****	
BH_5 - 4	44	7.35	7.37	*****	
ENSEMBLE	597	99.67	100.00		
29 . Q27					
BI_1 - 0	236	39.40	39.46	*****	
BI_2 - 1	126	21.04	21.07	*****	
BI_3 - 2	49	8.18	8.19	****	
BI_4 - 3	96	16.03	16.05	*****	
BI_5 - 4	91	15.19	15.22	*****	
ENSEMBLE	598	99.83	100.00		
30 . Q28					
BJ_1 - 0	304	50.75	51.01	*****	
BJ_2 - 1	124	20.70	20.81	*****	
BJ_3 - 2	52	8.68	8.72	*****	
BJ_4 - 3	58	9.68	9.73	*****	
BJ_5 - 4	59	9.68	9.73	*****	
ENSEMBLE	596	99.50	100.00		
31 . Q29					
BK_1 - 0	245	40.90	40.97	*****	
BK_2 - 1	122	20.37	20.40	*****	
BK_3 - 2	56	9.35	9.36	*****	
BK_4 - 3	82	13.69	13.71	*****	
BK_5 - 4	93	15.53	15.55	*****	
ENSEMBLE	598	99.83	100.00		
32 . Q30					
BL_1 - 0	322	53.76	53.85	*****	
BL_2 - 1	107	17.86	17.89	*****	
BL_3 - 2	48	8.01	8.03	****	
BL_4 - 3	58	9.68	9.70	*****	
BL_5 - 4	63	10.52	10.54	*****	
ENSEMBLE	598	99.83	100.00		
33 . Q31					
BM_1 - 0	107	17.86	17.86	*****	
BM_2 - 1	114	19.03	19.03	*****	
BM_3 - 2	68	11.35	11.35	*****	
BM_4 - 3	164	27.38	27.38	*****	
BM_5 - 4	146	24.37	24.37	*****	
ENSEMBLE	599	100.00	100.00		
34 . Q32					
BN_1 - 0	346	57.76	57.86	*****	
BN_2 - 1	108	18.03	18.06	*****	
BN_3 - 2	52	8.68	8.70	****	
BN_4 - 3	49	8.18	8.19	****	
BN_5 - 4	43	7.18	7.19	****	
ENSEMBLE	598	99.83	100.00		
35 . Q33					
BO_1 - 0	292	48.75	48.75	*****	
BO_2 - 1	127	21.20	21.20	*****	
BO_3 - 2	52	8.68	8.69	****	
BO_4 - 3	68	11.35	11.35	*****	
BO_5 - 4	60	10.02	10.02	*****	
ENSEMBLE	599	100.00	100.00		
36 . Q34					
BP_1 - 0	193	32.22	32.22	*****	
BP_2 - 1	131	21.87	21.87	*****	
BP_3 - 2	67	11.19	11.19	*****	
BP_4 - 3	108	18.03	18.03	*****	
BP_5 - 4	100	16.69	16.69	*****	
ENSEMBLE	599	100.00	100.00		
37 . Q35					
BQ_1 - 0	382	63.77	63.77	*****	
BQ_2 - 1	86	14.36	14.36	*****	
BQ_3 - 2	42	7.01	7.01	****	
BQ_4 - 3	59	9.85	9.85	*****	
BQ_5 - 4	30	5.01	5.01	***	
ENSEMBLE	599	100.00	100.00		
38 . Q36					
BR_1 - 0	273	45.58	45.65	*****	
BR_2 - 1	145	24.21	24.25	*****	
BR_3 - 2	58	9.68	9.70	*****	
BR_4 - 3	82	13.69	13.71	*****	
BR_5 - 4	40	6.68	6.69	****	
ENSEMBLE	598	99.83	100.00		
39 . Q37					
BS_1 - 0	314	52.42	52.42	*****	
BS_2 - 1	109	18.20	18.20	*****	
BS_3 - 2	51	8.51	8.51	****	
BS_4 - 3	63	10.52	10.52	*****	
BS_5 - 4	62	10.35	10.35	*****	
ENSEMBLE	599	100.00	100.00		

40 . Q38				
BT_1 - 0	277	46.24	46.24	*****
BT_2 - 1	120	20.03	20.03	*****
BT_3 - 2	81	13.52	13.52	*****
BT_4 - 3	81	13.52	13.52	*****
BT_5 - 4	40	6.68	6.68	****
ENSEMBLE	599	100.00	100.00	
41 . Q39				
BU_1 - 0	293	48.91	48.91	*****
BU_2 - 1	115	19.20	19.20	*****
BU_3 - 2	61	10.18	10.18	****
BU_4 - 3	66	11.02	11.02	*****
BU_5 - 4	64	10.68	10.68	****
ENSEMBLE	599	100.00	100.00	
42 . Q40				
BV_1 - 0	294	49.08	49.16	*****
BV_2 - 1	127	21.20	21.24	*****
BV_3 - 2	52	8.68	8.70	****
BV_4 - 3	74	12.35	12.37	*****
BV_5 - 4	51	8.51	8.53	****
ENSEMBLE	598	99.83	100.00	
43 . Q41				
BW_1 - 0	371	51.94	62.04	*****
BW_2 - 1	93	15.53	15.55	*****
BW_3 - 2	41	6.84	6.86	****
BW_4 - 3	52	8.68	8.70	****
BW_5 - 4	41	6.84	6.86	****
ENSEMBLE	598	99.93	100.00	
44 . Q42				
BX_1 - 0	204	34.06	34.06	*****
BX_2 - 1	153	25.54	25.54	*****
BX_3 - 2	76	12.69	12.69	*****
BX_4 - 3	80	13.36	13.36	*****
BX_5 - 4	86	14.36	14.36	*****
ENSEMBLE	599	100.00	100.00	
45 . Q43				
BY_1 - 0	343	57.26	57.26	*****
BY_2 - 1	105	17.53	17.53	*****
BY_3 - 2	31	5.18	5.18	***
BY_4 - 3	59	9.85	9.85	*****
BY_5 - 4	61	10.18	10.18	*****
ENSEMBLE	599	100.00	100.00	
46 . Q44				
BZ_1 - 0	261	41.90	41.90	*****
BZ_2 - 1	113	18.86	18.86	*****
BZ_3 - 2	50	8.35	8.35	****
BZ_4 - 3	85	14.19	14.19	*****
BZ_5 - 4	100	16.69	16.69	*****
ENSEMBLE	599	100.00	100.00	
47 . Q45				
CA_1 - 0	318	53.09	53.09	*****
CA_2 - 1	121	20.20	20.20	*****
CA_3 - 2	52	8.68	8.68	****
CA_4 - 3	56	9.35	9.35	*****
CA_5 - 4	52	8.68	8.68	****
ENSEMBLE	599	100.00	100.00	
48 . Q46				
CB_1 - 0	284	47.41	47.49	*****
CB_2 - 1	130	21.70	21.74	*****
CB_3 - 2	51	8.51	8.53	****
CB_4 - 3	74	12.35	12.37	*****
CB_5 - 4	59	9.85	9.87	*****
ENSEMBLE	598	99.83	100.00	
49 . Q47				
CC_1 - 0	450	75.13	75.13	*****
CC_2 - 1	63	10.52	10.52	*****
CC_3 - 2	22	3.67	3.67	**
CC_4 - 3	39	6.51	6.51	***
CC_5 - 4	25	4.17	4.17	**
ENSEMBLE	599	100.00	100.00	
50 . Q48				
CD_1 - 0	359	58.93	59.03	*****
CD_2 - 1	105	17.53	17.56	*****
CD_3 - 2	44	7.35	7.36	****
CD_4 - 3	50	8.35	8.36	****
CD_5 - 4	46	7.68	7.69	****
ENSEMBLE	598	99.83	100.00	
51 . Q49				
CE_1 - 0	320	53.42	53.42	*****
CE_2 - 1	128	21.37	21.37	*****
CE_3 - 2	62	10.35	10.35	*****
CE_4 - 3	52	8.68	8.68	****
CE_5 - 4	37	6.18	6.18	***
ENSEMBLE	599	100.00	100.00	
52 . Q50				
CF_1 - 0	284	47.41	47.57	*****
CF_2 - 1	126	21.04	21.11	*****
CF_3 - 2	46	7.68	7.71	****
CF_4 - 3	91	15.19	15.24	*****
CF_5 - 4	50	8.35	8.38	****
ENSEMBLE	597	99.67	100.00	

53 . Q51				
CG_1 - 0	220	36.73	36.79	*****
CG_2 - 1	164	27.38	27.42	*****
CG_3 - 2	59	9.85	9.87	****
CG_4 - 3	83	13.86	13.88	*****
CG_5 - 4	72	12.02	12.04	*****
ENSEMBLE	598	99.83	100.00	
-----				
54 . Q52				
CH_1 - 0	270	45.08	45.15	*****
CH_2 - 1	148	24.71	24.75	*****
CH_3 - 2	53	8.85	8.86	****
CH_4 - 3	74	12.35	12.37	*****
CH_5 - 4	53	8.85	8.86	****
ENSEMBLE	598	99.83	100.00	
-----				
55 . Q53				
CI_1 - 0	332	55.43	55.43	*****
CI_2 - 1	106	17.70	17.70	*****
CI_3 - 2	51	8.51	8.51	****
CI_4 - 3	61	10.18	10.18	*****
CI_5 - 4	49	8.18	8.18	****
ENSEMBLE	599	100.00	100.00	
-----				
56 . Q54				
CJ_1 - 0	352	58.76	58.76	*****
CJ_2 - 1	89	14.86	14.86	*****
CJ_3 - 2	32	5.34	5.34	***
CJ_4 - 3	65	10.85	10.85	*****
CJ_5 - 4	61	10.18	10.18	*****
ENSEMBLE	599	100.00	100.00	
-----				
57 . Q55				
CK_1 - 0	296	49.42	49.58	*****
CK_2 - 1	136	22.70	22.78	*****
CK_3 - 2	45	7.51	7.54	****
CK_4 - 3	60	10.02	10.05	*****
CK_5 - 4	60	10.02	10.05	*****
ENSEMBLE	597	99.67	100.00	
-----				
58 . Q56				
CL_1 - 0	254	42.40	42.55	*****
CL_2 - 1	145	24.21	24.29	*****
CL_3 - 2	63	10.52	10.55	*****
CL_4 - 3	81	13.52	13.57	*****
CL_5 - 4	54	9.02	9.05	****
ENSEMBLE	597	99.67	100.00	
-----				
59 . Q57				
CM_1 - 0	255	42.57	43.07	*****
CM_2 - 1	123	20.53	20.78	*****
CM_3 - 2	70	11.69	11.82	*****
CM_4 - 3	81	13.52	13.68	*****
CM_5 - 4	63	10.52	10.64	*****
ENSEMBLE	592	98.83	100.00	
-----				
60 . Q58				
CN_1 - 0	287	47.91	48.15	*****
CN_2 - 1	115	19.20	19.30	*****
CN_3 - 2	60	10.02	10.07	*****
CN_4 - 3	80	13.36	13.42	*****
CN_5 - 4	54	9.02	9.06	*****
ENSEMBLE	596	99.50	100.00	
-----				
61 . Q59				
CO_1 - 0	339	56.59	56.69	*****
CO_2 - 1	104	17.36	17.39	*****
CO_3 - 2	32	5.34	5.35	***
CO_4 - 3	51	8.51	8.53	****
CO_5 - 4	72	12.02	12.04	*****
ENSEMBLE	598	99.83	100.00	
-----				
62 . Q60				
CP_1 - 0	390	65.11	65.11	*****
CP_2 - 1	60	10.02	10.02	*****
CP_3 - 2	48	8.01	8.01	****
CP_4 - 3	55	9.18	9.18	*****
CP_5 - 4	46	7.68	7.68	****
ENSEMBLE	599	100.00	100.00	
-----				
63 . Q61				
CQ_1 - 0	284	47.41	47.41	*****
CQ_2 - 1	121	20.20	20.20	*****
CQ_3 - 2	47	7.85	7.85	****
CQ_4 - 3	91	15.19	15.19	*****
CQ_5 - 4	56	9.35	9.35	*****
ENSEMBLE	599	100.00	100.00	
-----				
64 . Q62				
CR_1 - 0	385	64.27	64.38	*****
CR_2 - 1	82	13.69	13.71	*****
CR_3 - 2	29	4.84	4.85	***
CR_4 - 3	64	10.68	10.70	*****
CR_5 - 4	38	6.34	6.35	***
ENSEMBLE	598	99.83	100.00	
-----				
65 . Q63				
CS_1 - 0	445	74.29	74.29	*****
CS_2 - 1	58	9.68	9.68	*****
CS_3 - 2	28	4.67	4.67	***
CS_4 - 3	34	5.68	5.68	***
CS_5 - 4	34	5.68	5.68	***
ENSEMBLE	599	100.00	100.00	

-----					
66 . Q64					
CT_1 - 0	232	38.73	38.73	*****	
CT_2 - 1	122	20.37	20.37	*****	
CT_3 - 2	63	10.52	10.52	*****	
CT_4 - 3	89	14.86	14.86	*****	
CT_5 - 4	90	15.53	15.53	*****	
ENSEMBLE	599	100.00	100.00		
-----					
67 . Q65					
CU_1 - 0	392	65.44	65.55	*****	
CU_2 - 1	83	13.86	13.88	*****	
CU_3 - 2	39	6.51	6.52	***	
CU_4 - 3	56	9.35	9.36	*****	
CU_5 - 4	28	4.67	4.68	***	
ENSEMBLE	598	99.83	100.00		
-----					
68 . Q66					
CV_1 - 0	221	36.89	36.89	*****	
CV_2 - 1	128	21.37	21.37	*****	
CV_3 - 2	57	9.52	9.52	*****	
CV_4 - 3	99	16.53	16.53	*****	
CV_5 - 4	94	15.69	15.69	*****	
ENSEMBLE	599	100.00	100.00		
-----					
69 . Q67					
CW_1 - 0	462	77.13	77.13	*****	
CW_2 - 1	60	10.02	10.02	*****	
CW_3 - 2	20	3.34	3.34	**	
CW_4 - 3	31	5.18	5.18	***	
CW_5 - 4	26	4.34	4.34	**	
ENSEMBLE	599	100.00	100.00		
-----					
70 . Q68					
CK_1 - 0	385	64.27	64.27	*****	
CK_2 - 1	104	17.36	17.36	*****	
CK_3 - 2	34	5.68	5.68	***	
CK_4 - 3	39	6.51	6.51	***	
CK_5 - 4	37	6.18	6.18	***	
ENSEMBLE	599	100.00	100.00		
-----					
71 . Q69					
CY_1 - 0	231	38.56	38.63	*****	
CY_2 - 1	140	23.37	23.41	*****	
CY_3 - 2	56	9.35	9.36	*****	
CY_4 - 3	95	15.86	15.89	*****	
CY_5 - 4	76	12.69	12.71	*****	
ENSEMBLE	598	99.83	100.00		
-----					
72 . Q70					
CS_1 - 0	328	54.76	54.85	*****	
CS_2 - 1	110	18.36	18.39	*****	
CS_3 - 2	41	6.84	6.86	****	
CS_4 - 3	72	12.02	12.04	*****	
CS_5 - 4	47	7.85	7.86	*****	
ENSEMBLE	598	99.83	100.00		
-----					
73 . Q71					
DA_1 - 0	263	43.91	43.91	*****	
DA_2 - 1	142	23.71	23.71	*****	
DA_3 - 2	61	10.18	10.18	*****	
DA_4 - 3	77	12.85	12.85	*****	
DA_5 - 4	36	9.35	9.35	*****	
ENSEMBLE	599	100.00	100.00		
-----					
74 . Q72					
DB_1 - 0	404	67.45	67.45	*****	
DB_2 - 1	78	13.02	13.02	*****	
DB_3 - 2	23	3.84	3.84	**	
DB_4 - 3	50	8.35	8.35	****	
DB_5 - 4	44	7.35	7.35	****	
ENSEMBLE	599	100.00	100.00		
-----					
75 . Q73					
DC_1 - 0	393	65.61	65.61	*****	
DC_2 - 1	99	16.53	16.53	*****	
DC_3 - 2	31	5.18	5.18	***	
DC_4 - 3	49	8.18	8.18	****	
DC_5 - 4	27	4.51	4.51	***	
ENSEMBLE	599	100.00	100.00		
-----					
76 . Q74					
DD_1 - 0	404	67.45	67.45	*****	
DD_2 - 1	81	13.52	13.52	*****	
DD_3 - 2	25	4.17	4.17	**	
DD_4 - 3	51	8.51	8.51	****	
DD_5 - 4	38	6.34	6.34	***	
ENSEMBLE	599	100.00	100.00		
-----					
77 . Q75					
DE_1 - 0	315	52.59	52.59	*****	
DE_2 - 1	116	19.37	19.37	*****	
DE_3 - 2	33	5.51	5.51	***	
DE_4 - 3	80	13.36	13.36	*****	
DE_5 - 4	55	9.18	9.18	*****	
ENSEMBLE	599	100.00	100.00		
-----					

78 . Q76				
DF_1 - 0	278	46.41	46.41	*****
DF_2 - 1	139	23.21	23.21	*****
DF_3 - 2	44	7.35	7.35	****
DF_4 - 3	85	14.19	14.19	*****
DF_5 - 4	53	8.85	8.85	*****
ENSEMBLE	599	100.00	100.00	
-----				
79 . Q77				
DG_1 - 0	305	50.92	50.92	*****
DG_2 - 1	112	18.70	18.70	*****
DG_3 - 2	49	8.18	8.18	****
DG_4 - 3	68	11.35	11.35	*****
DG_5 - 4	65	10.85	10.85	*****
ENSEMBLE	599	100.00	100.00	
-----				
80 . Q78				
DH_1 - 0	320	53.42	53.42	*****
DH_2 - 1	90	15.03	15.03	*****
DH_3 - 2	54	9.02	9.02	*****
DH_4 - 3	78	13.02	13.02	*****
DH_5 - 4	57	9.52	9.52	*****
ENSEMBLE	599	100.00	100.00	
-----				
81 . Q79				
DI_1 - 0	364	60.77	60.87	*****
DI_2 - 1	98	16.36	16.39	*****
DI_3 - 2	39	6.51	6.52	***
DI_4 - 3	53	8.85	8.86	*****
DI_5 - 4	44	7.35	7.36	****
ENSEMBLE	598	99.83	100.00	
-----				
82 . Q80				
DJ_1 - 0	207	47.91	47.99	*****
DJ_2 - 1	104	17.36	17.39	*****
DJ_3 - 2	54	9.02	9.03	*****
DJ_4 - 3	77	12.85	12.88	*****
DJ_5 - 4	76	12.69	12.71	*****
ENSEMBLE	598	99.83	100.00	
-----				
83 . Q81				
DK_1 - 0	436	72.79	72.79	*****
DK_2 - 1	67	11.19	11.19	*****
DK_3 - 2	30	5.01	5.01	***
DK_4 - 3	38	6.34	6.34	***
DK_5 - 4	28	4.67	4.67	***
ENSEMBLE	599	100.00	100.00	
-----				
84 . Q82				
DL_1 - 0	379	63.27	63.38	*****
DL_2 - 1	85	14.19	14.21	*****
DL_3 - 2	28	4.67	4.68	***
DL_4 - 3	56	9.35	9.36	*****
DL_5 - 4	50	8.35	8.36	****
ENSEMBLE	598	99.83	100.00	
-----				
85 . Q83				
DM_1 - 0	301	50.25	50.33	*****
DM_2 - 1	110	18.36	18.39	*****
DM_3 - 2	48	8.01	8.03	****
DM_4 - 3	75	12.52	12.54	*****
DM_5 - 4	64	10.68	10.70	*****
ENSEMBLE	598	99.83	100.00	
-----				
86 . Q84				
DN_1 - 0	418	69.78	69.90	*****
DN_2 - 1	60	10.02	10.03	*****
DN_3 - 2	30	5.01	5.02	***
DN_4 - 3	50	8.35	8.36	****
DN_5 - 4	40	6.68	6.69	****
ENSEMBLE	598	99.83	100.00	
-----				
87 . Q85				
DO_1 - 0	332	55.43	55.52	*****
DO_2 - 1	118	19.70	19.73	*****
DO_3 - 2	48	8.01	8.03	****
DO_4 - 3	59	9.85	9.87	*****
DO_5 - 4	41	6.84	6.86	****
ENSEMBLE	598	99.83	100.00	
-----				
88 . Q86				
DP_1 - 0	402	67.11	67.11	*****
DP_2 - 1	77	12.85	12.85	*****
DP_3 - 2	35	5.84	5.84	***
DP_4 - 3	49	8.18	8.18	****
DP_5 - 4	36	6.01	6.01	***
ENSEMBLE	599	100.00	100.00	
-----				
89 . Q87				
DQ_1 - 0	263	43.91	43.98	*****
DQ_2 - 1	129	21.54	21.57	*****
DQ_3 - 2	39	6.51	6.52	***
DQ_4 - 3	92	15.36	15.38	*****
DQ_5 - 4	75	12.52	12.54	*****
ENSEMBLE	598	99.83	100.00	
-----				
90 . Q88				
DR_1 - 0	384	64.11	64.54	*****
DR_2 - 1	90	15.03	15.13	*****
DR_3 - 2	11	5.18	5.21	***
DR_4 - 3	55	9.18	9.24	*****
DR_5 - 4	35	5.84	5.88	***
ENSEMBLE	595	99.33	100.00	

-----

91 . Q89				
DS_1 - 0	319	53.26	53.26	*****
DS_2 - 1	133	22.20	22.20	*****
DS_3 - 2	42	7.01	7.01	****
DS_4 - 3	63	10.52	10.52	*****
DS_5 - 4	42	7.01	7.01	****
ENSEMBLE	599	100.00	100.00	

-----

92 . Q90				
DT_1 - 0	327	54.59	54.59	*****
DT_2 - 1	101	16.86	16.86	*****
DT_3 - 2	33	5.51	5.51	***
DT_4 - 3	72	12.02	12.02	*****
DT_5 - 4	66	11.02	11.02	*****
ENSEMBLE	599	100.00	100.00	

-----

# SCL - 90 - R

ESCALA DE AVALIAÇÃO DE SINTOMAS - 90 - R

Leonard R. Derogatis, PhD. (1994)

## PROTOCOLO PARA PESQUISA

Tradução e Adaptação do Original

Diana Tosello Laloni (1996)

Nome: \_\_\_\_\_

Idade: \_\_\_\_\_ Data de nascimento: \_\_\_\_\_

Sexo: M  F

Estado civil: solteiro  casado ou amigado  separado  viúvo

Grau de escolaridade: analfabeto  1ª. a 4ª.  5ª. a 8ª.  2º. Grau  3º. Grau

Data \_\_\_\_ / \_\_\_\_ / \_\_\_\_

### Orientações:

- 1- Preencha os dados da identificação na parte superior da folha.
- 2- Use um lápis preto para marcar a resposta.
- 3- Se você desejar alterar sua resposta, apague com cuidado a 1ª. marca e marque a nova resposta.
- 4- Não faça outras anotações fora dos círculos.

### INSTRUÇÕES:

Abaixo está uma lista de problemas que as pessoas algumas vezes têm. Por favor leia-os cuidadosamente e pinte o círculo que melhor descreve **o quanto aquele problema tem preocupado ou angustiado você durante os últimos 7 dias incluindo hoje**. Pinte o círculo em apenas 1 número para cada problema e não pule nenhum item. Se você mudar sua idéia apague sua primeira marca cuidadosamente. Leia o exemplo antes de começar e se você tiver alguma dúvida pergunte.

N E N H U M P O U C O	U M P O U C O	M O D E R A D A M E N T E	B A S T A N T E	E X T R E M A M E N T E MUITO	
					<b>EXEMPLO</b>  <b>O quanto você está preocupado com:</b>
0	①	②	③	④	<b>Dores no corpo</b>

	N E N H U M P O U C O	U M P O U C O	M O D E R A D A M E N T E	B A S T A N T E	E X T R E M A M E N T E M U L T O	O quanto você está preocupado com:
1	0	①	②	③	④	Dores de cabeça
2	0	①	②	③	④	Nervosismo ou tremores internos
3	0	①	②	③	④	Pensamentos ruins repetidos que não saem de sua mente (cabeça)
4	0	①	②	③	④	Fraqueza ou tonturas
5	0	①	②	③	④	Perda de interesse ou de prazer sexual
6	0	①	②	③	④	Sentir-se crítico em relação a outras pessoas
7	0	①	②	③	④	A idéia de que outra pessoa pode controlar seus pensamentos
8	0	①	②	③	④	Sentir que os outros são culpados de muitos dos seus problemas
9	0	①	②	③	④	Dificuldade para se lembrar das coisas
10	0	①	②	③	④	Preocupado com o desleixo ou falta de cuidado
11	0	①	②	③	④	Sentir-se facilmente incomodado ou irritado
12	0	①	②	③	④	Dores no coração ou no peito
13	0	①	②	③	④	Sentir medo em espaços abertos ou nas ruas
14	0	①	②	③	④	Sentir-se com pouca energia ou para baixo
15	0	①	②	③	④	Pensamentos de acabar com a própria vida
16	0	①	②	③	④	Escutar vozes que outras pessoas não escutam
17	0	①	②	③	④	Estar trêmulo (com tremedeira)
18	0	①	②	③	④	Sentir que não se pode confiar na maioria das pessoas
19	0	①	②	③	④	Ter pouco apetite
20	0	①	②	③	④	Chorar facilmente
21	0	①	②	③	④	Sentir-se envergonhado ou desconfortável em relação ao sexo oposto
22	0	①	②	③	④	Sentir-se trapaceado ou encurralado (sem saída)
23	0	①	②	③	④	Repentinamente sentir medo sem razão
24	0	①	②	③	④	Explosões que você não pode controlar (explosões de raiva)
25	0	①	②	③	④	Ter medo de sair de casa sozinho
26	0	①	②	③	④	Sentir-se culpado
27	0	①	②	③	④	Dores nas costas e quadris
28	0	①	②	③	④	Sentir-se bloqueado (com dificuldade) para terminar as coisas
29	0	①	②	③	④	Sentir-se sozinho
30	0	①	②	③	④	Sentir-se sem importância

	N E N H U M P O U C O	U M P O U C O	M O D E R A D A M E N T E	B A S T A N T E	E X T R E M A M E N T E <small>MUITO</small>	O quanto você está preocupado com:
31	0	①	②	③	④	Preocupar-se demais com as coisas
32	0	①	②	③	④	Sentir-se desinteressado pelas coisas
33	0	①	②	③	④	Sentir medo
34	0	①	②	③	④	Sentir-se facilmente ferido em seus sentimentos
35	0	①	②	③	④	Outras pessoas estarem a par de seus pensamentos íntimos
36	0	①	②	③	④	Sentir que os outros não o entendem ou são antipáticos
37	0	①	②	③	④	Sentir que os outros não são amigos ou não gostam de você
38	0	①	②	③	④	Ter que fazer as coisas muito devagar para ter certeza de que estão corretas
39	0	①	②	③	④	Coração palpitando ou disparando
40	0	①	②	③	④	Náuseas, enjôos ou estômago ruim
41	0	①	②	③	④	Sentir-se inferior aos outros
42	0	①	②	③	④	Dores musculares (dor no corpo)
43	0	①	②	③	④	Sentir-se vigiado e comentado pelos outros
44	0	①	②	③	④	Problemas para conciliar o sono
45	0	①	②	③	④	Ter que conferir e reconferir o que fez
46	0	①	②	③	④	Dificuldade para tomar decisões
47	0	①	②	③	④	Sentir medo de andar de ônibus, metrô ou trens
48	0	①	②	③	④	Problemas para respirar
49	0	①	②	③	④	Ondas de calor ou frio
50	0	①	②	③	④	Ter que evitar certas coisas, lugares ou atividades que o amedrontam (dão medo)
51	0	①	②	③	④	Um "branco" na cabeça (ter uma incapacidade momentânea de raciocinar ou lembrar-se de algo)
52	0	①	②	③	④	Dormência ou formigamento em partes do corpo
53	0	①	②	③	④	Um nó na garganta
54	0	①	②	③	④	Sentir-se sem esperança sobre o futuro
55	0	①	②	③	④	Dificuldade de concentração
56	0	①	②	③	④	Sentir fraqueza em partes do corpo
57	0	①	②	③	④	Sentir-se tenso ou travado
58	0	①	②	③	④	Sentir peso nos braços e pernas
59	0	①	②	③	④	Pensar sobre morte ou sobre estar morrendo
60	0	①	②	③	④	Comer demais

	N E N H U M P O U C O	U M P O U C O	M O D E R A D A M E N T E	B A S T A N T E	E X T R E M A M E N T E <small>MUITO</small>	O quanto você está preocupado com:
61	0	①	②	③	④	Sentir-se desconfortável quando as pessoas o observam ou falam de você
62	0	①	②	③	④	Ter pensamentos que não são seus
63	0	①	②	③	④	Ter desejos de bater, machucar ou ameaçar alguma pessoa
64	0	①	②	③	④	Acordar de madrugada
65	0	①	②	③	④	Ter que repetir as mesmas ações como tocar, contar ou lavar
66	0	①	②	③	④	Sono sem descanso ou perturbado
67	0	①	②	③	④	Ter desejos de quebrar ou destruir coisas
68	0	①	②	③	④	Ter idéias ou crenças que os outros não aceitam
69	0	①	②	③	④	Sentir-se muito acanhado ou preocupado com os outros
70	0	①	②	③	④	Sentir-se inquieto numa multidão, fazendo compras ou no cinema
71	0	①	②	③	④	Sentir que tudo é um esforço
72	0	①	②	③	④	Ondas de terror ou pânico
73	0	①	②	③	④	Sentir-se desconfortável por comer ou beber em público
74	0	①	②	③	④	Envolver-se freqüentemente em discussões
75	0	①	②	③	④	Sentir nervosismo quando é deixado sozinho
76	0	①	②	③	④	Os outros não dão o devido crédito (valor) às suas conquistas
77	0	①	②	③	④	Sentir-se solitário mesmo quando está acompanhado
78	0	①	②	③	④	Sentir-se tão agitado que não é capaz de parar quieto (de movimentar-se)
79	0	①	②	③	④	Sentimentos de inutilidade
80	0	①	②	③	④	O sentimento de que algo ruim está para acontecer com você
81	0	①	②	③	④	Gritar ou atirar coisas
82	0	①	②	③	④	Com medo de desmaiar em público
83	0	①	②	③	④	Sentir que as pessoas tirarão vantagens de você, caso permita
84	0	①	②	③	④	Ter pensamentos sobre sexo que o incomodam muito
85	0	①	②	③	④	A idéia de que você deveria ser punido pelos seus pecados
86	0	①	②	③	④	Pensamentos e imagens de natureza assustadora
87	0	①	②	③	④	A idéia de algo sério está errado com seu corpo
88	0	①	②	③	④	Nunca se sentir próximo a outra pessoa
89	0	①	②	③	④	Sentimentos de culpa
90	0	①	②	③	④	A idéia de que há algo errado com sua mente