**Análise fatorial múltipla aplicada
na obtenção de índices em
tabelas mistas**

Iván Robert Enríquez Guzmán

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO DE MESTRE
EM CIÊNCIAS

Área de Concentração: **Estatística**
Orientadora: **Profa. Dra. Lúcia Pereira Barroso**

Durante a elaboração deste trabalho o autor recebeu auxílio financeiro do CNPq
SÃO PAULO, JULHO DE 2006.

Análise fatorial múltipla aplicada na obtenção de índices em tabelas mistas

Este exemplar corresponde à redação final da
dissertação devidamente corrigida e defendida
por Iván Robert Enríquez Guzmán
e aprovada pela Comissão Julgadora.

SÃO PAULO, JULHO DE 2006.

Banca Examinadora:

Profa. Dra. Lúcia Pereira Barroso (Orientadora - IME - USP)

Profa. Dr. Rinaldo Artes (IBMEC)

Profa. Dra. Édina Shisue Miazaki (UnB)

Aos meus pais

Agradecimentos

Agradeço primeiro a Deus, por ter me permitido culminar o mestrado, além no país abençoado por Ele, aos meus Pais pela força que sempre me dão.

Agradecimento muito especial para minha Orientadora, Lúcia, ela se deu muito tempo pese a ser uma pessoa muito ocupada, obrigado por ter suportado meus erros no português e encaminhar minha dissertação e minha pesquisa. Aos membros da banca examinadora, Prof. Dr. Rinaldo Artes e Profa. Dra. Édina Shisue Miazaki, pelas suas importantes observações.

Não posso deixar passar a oportunidade de agradecer aos meus amigos do mestrado, Christian Clássico e Bayes, Joan, Juan Carlos, Pedrão camarada com quem já fiz muitas coisas que foram tão legais e até imprevisíveis, a pesar de ser ele probabilista, Fernando e Grazielle Yumi, que suportarão-me em Técnicas, cada quem com seu jeito, a Grã com esse jeito de transmitir paz para todos, o Fernando menos calmo mais vai lá e consegue. A Cillene Souza, por acordar-me os sábados cedo, para estudar Probabilidades, à galera em geral, Josivon, Fernando Cerecetti, Augusto, Mariana, Karina, Núbia, Juliana, Roberta, Eliane, Fabiola, Lincoln, Luis Fernando, Luis Affonso e Cecília, não posso deixar de mencionar à turma dos Cearenses, esses caras com os

que um deve de ficar estudando duro e esperto por que além de estudar muito, fazem as mais engraçadas piadas da língua portuguesa, Caio, Iesus e Juvêncio, fizeram mais agradável os dias de estudo intenso.

Devo agradecer aos Professores que de uma forma ou outra têm influenciado na formação deste novo mestre de sua prestigiosa instituição, faço extenso meus agradecimentos a todos aqueles que de uma e outra forma contribuíram a tornar os dias do mestrado uma etapa inesquecível da minha vida, obrigado.

Agradeço a CNPq pelo apoio financeiro sem o qual esta tese não teria sido possível, ou a teria tornado mais difícil.

Sumário

1	Introdução	1
2	Descrição dos Problemas	6
2.1	AFM para variáveis avaliadas no vestibular	7
2.2	AFM para confirmar os fatores indicadores de inteligência emocional .	9
3	Metodologia	13
3.1	Representação dos dados e espaços gerados	13
3.2	Ajuste da nuvem de indivíduos no espaço de variáveis	18
3.3	Métodos para o estudo de uma tabela de dados	21
3.3.1	Análise de componentes principais	21
3.4	Métodos para o estudo de múltiplas tabelas de dados	24
3.4.1	Algumas fundamentações da análise fatorial múltipla	25

Sumário

1	Introdução	1
2	Descrição dos Problemas	6
2.1	AFM para variáveis avaliadas no vestibular	7
2.2	AFM para confirmar os fatores indicadores de inteligência emocional .	9
3	Metodologia	13
3.1	Representação dos dados e espaços gerados	13
3.2	Ajuste da nuvem de indivíduos no espaço de variáveis	18
3.3	Métodos para o estudo de uma tabela de dados	21
3.3.1	Análise de componentes principais	21
3.4	Métodos para o estudo de múltiplas tabelas de dados	24
3.4.1	Algumas fundamentações da análise fatorial múltipla	25

3.4.2	A metodologia da análise fatorial múltipla	29
3.4.3	AFM em R^I : as variáveis	34
3.4.4	Representação das variáveis e CPs parciais nos eixos globais	34
3.4.5	AFM em R^p : os indivíduos	36
3.4.6	Influência da ponderação dos grupos sobre as K nuvens N_I^k	37
3.4.7	Influência da ponderação dos grupos sobre a nuvem N_I associada às variáveis	37
3.4.8	Qualidade de representação de cada nuvem N_I^k	38
4	Índices Propostos para Avaliar o Ingresso à Universidade	40
4.1	Análise de componentes principais das notas	41
4.2	Análise fatorial múltipla da tabela de dados	45
4.2.1	ACP em cada grupo de variáveis	45
4.2.2	Análise global da tabela completa	50
4.2.3	Descrição das componentes principais globais	52
4.2.4	Análise dos eixos parciais obtidos na ACP de cada grupo de gênero e idade	57
4.3	Conclusões	65
5	Comparação entre Sub-Grupos Populacionais via Análise Fatorial	

Múltipla	67
5.1 Análise dos eixos parciais da ACP em cada nuvem com os eixos globais da AFM	70
5.2 Representação superposta dos sub-grupos populacionais (grupos por idade e gênero)	74
5.3 Representação superposta da qualificação nos índices emocional do ponto de vista de grupo etário	77
5.4 Conclusões	83
6 Conclusões Gerais	85
A Análise nos Espaços Gerados	87
A.1 Análise dos indivíduos no espaço de variáveis	87
A.1.1 Transformação proposta	90
A.1.2 Os indivíduos projetados em eixos fatoriais	92
A.2 Análises das variáveis no espaço de indivíduos	92
A.2.1 Análise das variáveis	92
A.2.2 As variáveis projetadas nos eixos fatoriais da AFM	96
A.2.3 Representação superposta de nuvens dos grupos de variáveis .	98
B Tabelas	99

C Programas	106
C.1 AFM da Aplicação 1	106

Lista de Figuras

3.1	Tabela X	15
3.2	Projeção de um ponto sobre o eixo H	19
4.1	Primeiro plano das componentes principais	43
4.2	Segundo plano das componentes principais	44
4.3	Projeção das variáveis qualitativas no primeiro plano global. Grupo: F= Feminino, M= Masculino, 1=< 17, 2 =18, 3 =19, 4=20 a 24, 5=25 a 29, 6=30 anos a mais.	56
5.1	Projeções dos eixos parciais nos dois primeiros eixos principais da AFM global	78
5.2	Representação das variáveis e dos grupos etários.1:15 a 24 anos, 2:25 a 34 anos, 3:35 a 44 anos, 4:45 a 54 anos, 5:55 ou mais anos	81
5.3	Representação das CPs da AFM e das variáveis categorizadas	82
A.1	Projeção em H_1	89

Lista de Tabelas

2.1	Variáveis avaliadas em postulantes à Universidade Federal de Minas Gerais	7
2.1	Variáveis avaliadas em postulantes à Universidade Federal de Minas Gerais	8
2.2	Distribuição da amostra por sexo e faixa etária.	10
4.1	ACP global - proporção da variância explicada por CP	42
4.2	ACP global - coeficientes das componentes principais	42
4.3	Proporção da variância explicada por CP do grupo Ciências	46
4.4	Coeficientes das variáveis para cada CP do grupo Ciências	46
4.5	Correlações entre as variáveis do grupo Ciências	48
4.6	Proporção da variância explicada por CP do grupo Humanas	48
4.7	Coeficientes das variáveis para cada CP do grupo Humanas	48
4.8	Correlações entre as variáveis do grupo Humanas	49

4.9	Correlações entre os eixos parciais dos grupos de variáveis	49
4.10	ACP global ponderada: proporção da variância explicada por CP . .	51
4.11	ACP global ponderada: coeficientes das componentes principais . . .	51
4.12	Coordenadas e distâncias dos eixos projetados no primeiro plano das CPs globais.	54
4.13	Coordenadas e distâncias das variáveis projetadas no primeiro plano das CPs globais.	55
4.14	Proporção da variância explicada por CP	60
4.15	Coeficientes das variáveis para cada CP	61
4.16	Correlações das variáveis dos diferentes grupos	62
4.17	Coeficientes das variáveis para cada CP na faixa etária I (≤ 17 anos)	62
4.18	Coeficientes das variáveis para cada CP na faixa etária II (18 anos) .	63
4.19	Coeficientes das variáveis para cada CP na faixa etária III (19 anos)	63
4.20	Coeficientes das variáveis para cada CP na faixa etária IV (20 a 24 anos)	63
4.21	Coeficientes das variáveis para cada CP na faixa etária V (25 a 29 anos)	64
4.22	Coeficientes das variáveis para cada CP na faixa etária VI (≥ 30 anos) .	64
4.23	Médias por sexo das notas obtidas no vestibular na UFMG - 1997 . .	64

5.1	Coordenadas e contribuições de cada primeira componente parcial aos eixos principais	77
B.1	Proporção da variância explicada por CP do grupo intrapessoais . . .	99
B.2	Coeficientes das CPs das variáveis intrapessoais	99
B.3	Proporção da variância explicada por CP do grupo interpessoais . . .	100
B.4	Coeficientes das CPs das variáveis interpessoais	100
B.5	Proporção da variância explicada por CP do grupo adaptabilidade	100
B.6	Coeficientes das CPs das variáveis de adaptabilidade	100
B.7	Proporção da variância explicada por CP do grupo manejo do estresse . .	101
B.8	Coeficientes das CPs das variáveis de manejo do estresse	101
B.9	Proporção da variância explicada por CP do grupo estado do ânimo e motivação	101
B.10	Coeficienttes das CPs das variáveis de estado do ânimo e motivação	101
B.11	Proporção de variância explicada na ACP global do grupo de 15 a 24 anos	102
B.12	Coeficientes das CPs do grupo de 15 a 24 anos	102
B.13	Proporção de variância explicada na ACP global do grupo de 25 a 34 anos	102
B.14	Coeficientes das CPs do grupo de 25 a 34 anos	103

B.15 Proporção de variância explicada na ACP global do Grupo de 35 a 44 anos	103
B.16 Coeficientes das CPs do grupo de 35 a 44 anos	103
B.17 Proporção de variância explicada na ACP global do grupo de 45 a 54 anos	104
B.18 Coeficientes das CPs do grupo de 45 a 54 anos	104
B.19 Proporção de variância explicada na ACP global do grupo de 55 ou mais anos	104
B.20 Coeficientes das CPs do grupo de 55 ou mais anos	105

Capítulo 1

Introdução

Nas últimas décadas, temos observado a grande importância da Estatística como ciência, e como é sua influência na tomada de decisões. Dentro das diferentes metodologias que são utilizadas pela Estatística encontram-se os *Métodos de Análise Multivariada*, que surgem como uma necessidade de análise em pesquisas com base em um conjunto de variáveis, apresentando uma alternativa diferente à análise por separado de cada uma das variáveis. A Análise Multivariada permite a confrontação entre numerosas informações contidas nestas variáveis, o que pode ser muito mais conveniente que uma análise de cada uma delas, separadamente.

No entanto, o entendimento na obtenção dos resultados e nas operações matemáticas que estes envolvem, às vezes, é um tanto complicado. Assim sendo, pode ser muito mais conveniente apresentar os resultados graficamente. Poder observar resultados graficamente implica em visualizar a tendência de conjunto de dados, tendência que muitas vezes encontra-se escondida dentro de **grandes tabelas de dados**.

A capacidade que métodos gráficos têm de proporcionar ferramentas que mostram

de forma clara as tendências, as aglomerações, os pontos discrepantes, a hierarquização e a eliminação de efeitos pontuais ou marginais, que podem perturbar a percepção global, tem permitido o desenvolvimento e interesse sobre os mesmos, e estes vêm sendo aplicados a uma diversidade de campos como a Ecologia, Lingüística, Economia, Marketing, Setor Bancário e de Seguros, Sociologia, Medicina, Biologia, etc, (Escofier e Pagès, 1990).

Dentre as diversas técnicas multivariadas, a Análise Fatorial ou o que se entende por Análise Fatorial na escola francesa (Análise de Componentes Principais, Análise de Correspondência Múltipla, Análise Fatorial propriamente dita) ocupa um lugar primordial, e pode ser utilizada individualmente ou em conjunto com outros métodos de classificação. Prioritariamente é escolhida por dois motivos:

- permite visualizar em sentido literal, agrupamentos, tendências, oposições e aproximações, dificilmente discernidos diretamente de uma tabela de números, inclusive depois de testes exaustivos;
- mediante gráficos, o estatístico pode comunicar os resultados de forma clara e coerente, sem necessidade de se fazer demasiado tedioso, como no caso de uma exposição de tabelas e números, a pessoas que podem não ser estatísticos, mas precisam da interpretação e compreensão dessa informação, sem necessidade de entender a formalização matemática por trás dela.

Os desenvolvedores do conceito da Análise Fatorial Múltipla (AFM), Brigitte Escofier e Jérôme Pagès, justificam a terminologia Análise Fatorial (AF) empregada, uma vez que existem diferentes métodos adaptados a diferentes tipos de dados, mas o princípio deles é único. Matematicamente constituem-se em gráficos de nuvens de pontos correspondentes a linhas ou colunas de uma tabela em estudo. As representa-

ções de linhas e colunas relacionam-se freqüentemente e tratam de expressar os dados em duas dimensões.

A nomenclatura francesa de Análise Fatorial considera que a Análise de Componentes Principais (ACP) utilizada em variáveis numéricas, a Análise de Correspondência (AC) e a Análise de Correspondência Múltipla (ACM), aplicadas a variáveis categorizadas, são os diferentes métodos que apresentam esse tipo de tabelas, e dependem do tipo de variável ou medida usada. A metodologia se apóia principalmente na procura de reduzir a dimensionalidade dos dados sem perda significativa de variabilidade, buscando uma relação de fatores com indivíduos e com as variáveis e tentando encontrar tipologias de variáveis de indivíduos e tendências entre os dados. Neste trabalho, os termos *fator* e *componente principal* serão utilizados sem distinção alguma, tendo em consideração a terminologia utilizada pelos autores de AFM.

O objetivo principal desta dissertação é destacar as vantagens da AFM frente aos métodos clássicos de redução de dimensionalidade mais usados, como ACP ou AC. A AFM nos permite incluir na análise novas variáveis que representam grupos de variáveis. A AFM permite reponderar a importância das variáveis iniciais contidas em todos os grupos nos casos em que somente um deles explicaria quase toda variabilidade dos dados. Além disso a AFM permite analisar variáveis numéricas e categorizadas conjuntamente, permite analisar conjuntamente grupos de indivíduos e comparar os distintos resultados, podendo assim estabelecer que grupos de indivíduos têm comportamento similar ou distinto, de forma global e em cada grupo de variáveis, encontrando diferenças e proximidades dentro de sub-grupos de interesse, dentro de um mesmo conjunto de dados.

Como foi mencionado, a AFM permite a inclusão de variáveis qualitativas. A

pergunta natural é: existe alguma relação entre as CPs gerais e as dos sub-grupos de variáveis? Isto de fato implicaria em obter comparações das CPs obtidas para as diversas combinações de níveis das variáveis mencionadas, tendo assim, várias tabelas simultaneamente. Uma vez obtidas estas tabelas por nível, ou por grupo de variáveis, perguntamos: *é possível de alguma forma analisar em conjunto estes resultados? É possível descrever o conjunto de dados não só separadamente por grupo de variáveis senão também, conjuntamente, pelos grupos formados? É possível estabelecer relações entre as diferentes componentes encontradas em cada grupo? Se for possível, como desenvolver isso?* A idéia não é apenas sobrepor cada plano fatorial encontrado um sobre outro; deve-se ter em mente que a variabilidade das diferentes componentes principais e das variáveis dentro de cada grupo são diferentes, porém não estão dentro de um plano comum. Lembrando que cada componente é a combinação linear das variáveis em uma certa dimensão, torna-se mais complexo se a quantidade de variáveis dentro de cada grupo é diferente. É possível encontrar tipologias de variáveis e de indivíduos que melhorem a análise dos dados?

A Análise Fatorial Múltipla proporciona tais tipologias e permite trabalhar com dados que podem ser encontrados em múltiplas tabelas, seja porque dentro de tabelas de variáveis percebe-se sub-tabelas formadas por diferentes grupos de variáveis ou tabelas mistas, formadas por variáveis quantitativas e qualitativas. Pode-se também considerar o caso em que se tem o mesmo conjunto de variáveis, avaliadas em diferentes conjuntos de indivíduos.

Este trabalho está organizado da seguinte forma: no Capítulo 2 apresentamos a descrição dos problemas práticos a serem desenvolvidos, nos quais se considera a AFM como o método de análise; no Capítulo 3 apresentamos os métodos utilizados em nosso exemplo de aplicação, isto é, a ACP e AFM, como mencionamos algumas

metodologias de tratamentos de dados multivariados; no Capítulo 4 apresentamos uma aplicação referente aos dados das notas obtidas pelas pessoas que prestaram vestibular na UFMG no ano 1997, tendo por objetivo mostrar como a AFM trabalha com grupos de variáveis obtendo resultados melhores do que os da ACP, e no Capítulo 5 apresentamos uma outra aplicação tendo como objetivo mostrar uma comparação entre a ACP e a AFM onde destaca-se a importância de fazer uma análise considerando a totalidade dos indivíduos, e por sub-grupos de indivíduos, avaliados por variáveis e grupos de variáveis.

Capítulo 2

Descrição dos Problemas

Esta dissertação foi motivada pelo fato de que existem muitos problemas de análise de dados que apresentam variáveis quantitativas e qualitativas (categorizadas). Além disso, dentro do conjunto de variáveis quantitativas pode-se identificar conjuntos de variáveis que medem aspectos específicos, que de forma natural se agrupam em conjuntos que às vezes são definidos pelo pesquisador. De fato, pode ser um erro tratar de resolver o mencionado problema por ACP, pois pode-se ter componentes influenciadas por um só desses grupos. A necessidade de ter alguma forma de analisar simultaneamente esses conjuntos de dados foi que motivou a procura por um método que nos permita solucionar o problema, optando pela Análise Fatorial Múltipla.

Neste trabalho, serão tratados dois problemas práticos descritos a seguir.

2.1 AFM para variáveis avaliadas no vestibular

Atualmente a grande demanda de jovens pelo ingresso em uma universidade e as diferentes provas aplicadas para avaliar os alunos, nos fazem questionar quais são os conhecimentos necessários, que matérias devem ser dominadas pelos alunos, quais são os tipos de alunos que estão ingressando nas universidades e que matérias foram as que dominaram. Na busca de indicadores ou índices que possam avaliar o conhecimento dos alunos, optamos pelos métodos multivariados que procuram a obtenção de diversos fatores que possam explicar de forma adequada como é o desempenho dos jovens nos vestibulares. Para isso consideramos as notas de 12093 alunos no vestibular na Universidade Federal de Minas Gerais no ano de 1997. A base de dados contém grande número de variáveis qualitativas e quantitativas e a título de ilustração usaremos as variáveis descritas na Tabela 2.1:

Tabela 2.1: Variáveis avaliadas em postulantes à Universidade Federal de Minas Gerais

# variável	Nome	Descrição
1	Sexo	Masculino, feminino
2	Idade	Faixa etária (grupos de 9 anos)
3	Português	Nota obtida em português
4	Matemática	Nota obtida em matemática
5	Geografia	Nota obtida em geografia
6	História	Nota obtida em história
7	Inglês-Francês	Nota obtida em inglês/francês
<i>continua na próxima página</i>		

continuação

Tabela 2.1: Variáveis avaliadas em postulantes à Universidade Federal de Minas Gerais

8	Biologia	Nota obtida em biologia
9	Física	Nota obtida em física
10	Química	Nota obtida em química

Encontrar fatores baseados nas variáveis observadas nos possibilita avaliar os alunos não só pelas notas obtidas nas disciplinas, mas relacionar esses fatores com as variáveis categorizadas sexo e faixa etária. Assim, estamos interessados em saber quais são os fatores que mais influem para que alunos ingressem na faculdade, como por exemplo, em quais matérias necessitaram ter melhor desempenho. Dependendo da faculdade em que querem ingressar, os alunos devem dominar certas matérias mais do que outras. Para isso as variáveis estudadas são as consideradas para avaliar o conhecimento das diversas matérias, assim como, algumas variáveis de estratificação, como idade e sexo. De antemão pode-se classificar os alunos em grupos; os que têm melhor desempenho em matérias de ciências e aqueles que se saem melhor em humanas, podendo então, identificar grupos de indivíduos e grupos de variáveis que avaliem um respectivo conhecimento. Ou de forma mais geral, podemos perguntar se as componentes principais encontradas para toda a população de alunos são válidas para sub-grupos dentro da população, definidos por faixas etárias ou por sexo, por exemplo. Se os fatores da população como um todo não são válidos para os sub-grupos, pode ser de interesse investigar como os construtos ou fatores encontrados

nos diferentes grupos se relacionam entre si. Na presente pesquisa, damos resposta às inquietudes para as populações por faixas etárias e por gênero.

2.2 AFM para confirmar os fatores indicadores de inteligência emocional

Nosso segundo problema prático é relacionado a validar os fatores obtidos para uma amostra na população de Lima - Perú, tanto para as diferentes faixas etárias, como por gênero, ou seja, estamos interessados em descobrir se os fatores propostos por BarOn (1997) e BarOn (1996) em Toronto-Canadá, aplicados nos E.E.U.U., e confirmados para a população peruana no ano 2001, são validados para os grupos definidos por faixa etária, como para grupos de homens e mulheres. Se os fatores não são os mesmos, então surge a necessidade de saber como se relacionam interna e externamente, que grupos estão mais próximos, quais são mais diferenciados dos outros, etc.

A Dra. Nelly Ugarriza Chávez fez uma adaptação da pesquisa EQ-I BarOn Emotional Quotient Inventory do Dr. Reuven Bar-On realizada nos Estados Unidos, em 1997. Para isso foi considerada uma amostra final de 1996 pessoas de 16 anos ou mais, distribuída de acordo com a Tabela 2.2.

A inteligência emocional como a define BarOn (1997) "é um conjunto de capacidades, competências e habilidades não cognitivas que influem na habilidade própria de ter êxito ao afrontar as demandas e pressões do meio ambiente". Como tal, a inteligência emocional é um fator importante para determinar a habilidade própria de ter êxito na vida, e influencia diretamente o bem-estar psicológico geral. Detalhes

de como os fatores e as variáveis são obtidos encontram-se em Ugarriza (2001).

Os objetivos da pesquisa foram adaptar o inventário do quociente emocional considerando as características sócio-culturais da população, descrever os efeitos dos fatores sexo e idade, determinar a confiabilidade e validade do construto do inventário e desenvolver normas de administração, qualificação e interpretação dos resultados.

A amostra foi tomada com o objetivo de representar a população de Lima Metropolitana, segundo sua distribuição porcentual por idade. Esta população difere substancialmente da população americana, no contexto de grupos etários com predominância de pessoas de maior idade. Para lograr a extrapolação à população considerou-se o último censo do 1999 ("Encuesta nacional de Hogares ENAHO, INEI, 1999") e adotou-se a amostra proporcional.

Tabela 2.2: Distribuição da amostra por sexo e faixa etária.

Idade	Masculino	Feminino	Total	Porcentagem
15-24	263	323	586	29,36
25-34	178	289	467	23,39
35-44	145	243	388	19,44
45-54	109	172	281	14,08
55 >=	128	146	274	13,73
Total	823	1173	1996	100,00

É considerável destacar que 41,2% da amostra é masculina, e 58,8% é feminina. Dos casos avaliados 114 são de escola de ensino médio (5,7%), 656 são universitários (32,9%) e 1226 (61,4%) são entre profissionais, técnicos, empregados, donas de casa e desempregados.

O modelo de inteligência emocional de BarOn (1997) considera os seguintes fatores

e as correspondentes variáveis relacionadas:

- Componentes Fatoriais Intrapessoais: Compreensão de si mesmo (CM), Assertividade (AS), Auto-conceito (AC), Auto-realização (AR), Independência (IN);
- Componentes Fatoriais Interpessoais: Empatia (EM), Relações interpessoais (RI), Responsabilidade social (RS);
- Componentes Fatoriais de Adaptabilidade: Solução de problemas (SP), Prova da realidade (PR), Flexibilidade (FL);
- Componentes Fatoriais da Condução do Estresse: Tolerância a estresse (TE), Controle dos impulsos (CI);
- Componentes Fatoriais de Estado de Ânimo e Motivação: Felicidade (FE), Otimismo (OP).

O autor considerou a escala Likert para a avaliação de cada pergunta, escala de resposta de autoclassificação de cinco pontos: 1= Raramente ou nunca, 2= Poucas vezes, 3= Às vezes, 4= Muitas vezes e 5= Com muita frequência ou sempre.

O problema que logo surge é poder avaliar se os fatores obtidos pelo inventário de BarOn têm validade nas diversas sub-populações formadas pelas diferentes faixas etárias e por sexo. Uma vez que se tem uma tabela com sub-tabelas identificáveis, neste caso por indivíduo, desenvolveremos a AFM com o fim de avaliar os sub-grupos de indivíduos com base nas diversas variáveis propostas no teste de BarOn. Assim, se os resultados dos fatores por grupos são os mesmos, os resultados projetados nos planos das CPs finais devem ser muito próximos, caso contrário pode-se dizer que os diversos fatores obtidos em cada grupo etário são diferentes.

A AFM proporciona a solução de representar as diversas CPs em uma região onde se possa comparar os diferentes resultados, o que faz com que a análise seja muito mais rica em informação, pois nos permite encontrar não só tipologias baseadas nas variáveis, mas além disso, encontrar tipologias de grupo de variáveis, e descrever os resultados por ambas tipologias encontradas.

A existência de grupos de variáveis conduz a querer descrever os indivíduos não só por meio do conjunto das variáveis, mas também através de cada um dos grupos. Para isto, é sempre possível realizar uma análise separada para cada grupo; no entanto os resultados obtidos independentemente são dificilmente comparáveis entre si. Por exemplo, a identidade de sub-espacos fatoriais pode ser mascarada pelas rotações. Para comparar as tipologias é necessário situá-las no espaço referencial comum (Escofier e Pagès, 1990).

Capítulo 3

Metodologia

Neste capítulo apresentamos diferentes métodos empregados para a construção de fatores, baseados em tabelas de dados multivariados, com o propósito de encontrar construtos que expliquem adequadamente os dados. Antes de descrever esses métodos, apresentamos nas duas próximas seções um resumo sobre a representação de dados em espaços de dimensão menor que a original.

3.1 Representação dos dados e espaços gerados

A tabela de dados padronizada dispõe a massa de informação em uma tabela retangular. Para fixar as idéias, seja R a matriz de respostas iniciais, as linhas ($i = 1, \dots, n$) podem representar os n indivíduos ou observações, as colunas ($j = 1, \dots, p$) serão as p variáveis, que podem ser medidas numericamente ou ser atributos ou qualidades observadas nos indivíduos. Com o objetivo de utilizar os métodos de estatística exploratória multidimensional, é útil representar geometricamente as n linhas e as p

colunas da tabela de dados por pontos. As nuvens de pontos podem ser:

- a nuvem de n indivíduos (nuvem de pontos por linha, denotada por N_I) situada no espaço p dimensional R^p de variáveis (colunas); cada uma das n linhas é representada por um ponto de p coordenadas;
- a nuvem de p variáveis (nuvem de pontos por coluna, denotada por N_V) situada no espaço de n dimensões R^n de indivíduos (linhas); cada uma das p colunas é representada por um ponto de n coordenadas.

A tabela de dados padronizada denotada por $\mathbf{X}$, é uma matriz na qual cada vetor, linha ou coluna, representa um ponto em R^p ou em R^n , isto é, uma nuvem de indivíduos denotada por N_I ou uma nuvem de variáveis denotada por N_V . A tabela de dados é formada por um conjunto de indivíduos descrito por vários grupos de variáveis. Cada grupo de variáveis transformadas constitui uma tabela $\mathbf{X}_k$, que pode ser representada pela nuvem N_k^I ou N_k^V .

Todas as variáveis são definidas para o mesmo conjunto de indivíduos, de forma que se possa dispor todos os dados em uma só tabela, colocando as tabelas lado a lado e formando uma tabela geral de indivíduos e variáveis. Denotaremos por:

- $\mathbf{X}$: tabela completa padronizada;
- n : número de indivíduos;
- K : número de grupos de variáveis;
- $\mathbf{X}_k$:sub-tabela formada pelo k -ésimo grupo de variáveis das p variáveis;
- p : número de variáveis;

Figura 3.1: Tabela $\mathbf{X}$

- p_k : número de variáveis do grupo k ;

Podemos observar graficamente como podem estar dispostas as variáveis na tabela de dados $\mathbf{X}$, na Figura 3.1.

Considerando como exemplo os dados da pesquisa realizada com os 12093 que prestaram vestibular na UFMG no ano 1997, temos 10 variáveis das quais 8 são quantitativas, agrupadas naturalmente em dois grupos de variáveis *Ciências* e *Humanas*. Esses grupos estão constituídos por quatro variáveis quantitativas cada um. As variáveis restantes são de classificação (qualitativas). Então, identificando temos que $n = 12093$ estudantes, $K = 2$ grupos de variáveis, $p_1=4$, e $p_2=4$ variáveis em cada grupo.

Em cada uma das dimensões da tabela de dados precisamos observar qual é o comportamento dos dados. Por isso, definimos duas distâncias (ou duas proximidades) entre seus elementos linha ou coluna. O conjunto das colunas permite definir, com ajuda de fórmulas convenientes, as distâncias entre as linhas. Da mesma forma, o conjunto de linhas permite calcular as distâncias entre as colunas. Os pontos e as distâncias entre eles permitem definir a nuvem de pontos.

Uma forma ótima de visualizar a nuvem é projetar os pontos sobre planos ou sobre direções que minimizem as deformações que a projeção implica. Observando dois pontos em particular (i, i') , se a distância entre eles é $d_{(i, i')}$, quando são projetados em uma direção H qualquer (ver Figura A.1 do Apêndice A), $d_{(i, i')}^2$ pode ser decomposta como a soma dos quadrados dos catetos gerados, sendo a hipotenusa a distância $d_{(i, i')}^2$. Então o cateto que é a distância entre as projeções dos pontos na direção H , é menor ou igual a $d_{(i, i')}^2$, mas o objetivo é encontrar componentes que expliquem a maior variabilidade, então busca-se a maximização das novas distâncias de todos os pontos. Para isso, pode-se buscar o sub-espço unidimensional H que maximiza a soma dos quadrados das distâncias entre as projeções sobre H de todos os pares de pontos (i, i') , isto é,

$$Max \left[\sum_{i=1}^n \sum_{i'>i}^n d^2(i; i') \right]. \quad (3.1)$$

Se cada ponto recebe um peso w_i , a soma ponderada que se pretende maximizar é

$$\left[\sum_{i=1}^n \sum_{i'>i}^n w_i w_{i'} d^2((i; i')) \right].$$

Desta forma calcula-se o sub-espço vetorial que ajusta o melhor possível a nuvem de pontos; com respeito à análise de componentes principais, este último critério é

equivalente a obter as projeções correspondentes a

$$Max \left[\sum_{i=1}^n w_i d^2(i; G) \right],$$

em que G designa o ponto médio ou centro de gravidade das projeções.

Não obstante, nem sempre o interesse é pela forma da nuvem, algumas vezes sua posição com relação à origem é mais relevante. Assim, ao analisarmos as componentes principais, estamos interessados no início pela forma da nuvem de pontos, mas é a posição com relação à origem das variáveis que terá sentido no outro espaço, a projeção sob o sub-espaço H que maximiza as distâncias dos pontos a seu centro de gravidade (ver Seção A.1).

O modelo de análise com relação à origem, designado aqui sob o nome de análise geral, permite dar conta de distintas situações. Nada mais é do que uma representação geométrica da decomposição nos valores singulares, representados pela primeira vez por Eckart e Young (1936) e Eckart e Young (1939) para tabelas retangulares, generalizando os trabalhos relativos às matrizes quadradas. Gifi (1990) menciona também os trabalhos prévios e independentes de Beltrami (1873) e Jordan, Vinzi, Chatelin e Lauro (1874) (ver também Gower (1966) e Gabriel (1971)). O problema que se propõe solucionar é um problema de redução puramente numérico, isto é, um problema de se compactar os dados (Piron, Lebart e Morineau, 1995) ou de redução da dimensionalidade dos dados em fatores, componentes principais, construtos ou traços latentes, como vêm sendo conhecidos.

3.2 Ajuste da nuvem de indivíduos no espaço de variáveis

Para simplificar a formulação consideramos primeiramente p variáveis observadas para n indivíduos e que x_{ij} denota a j -ésima variável observada para o i -ésimo indivíduo, padronizada de forma que

$$x_{ij} = \frac{r_{ij} - \bar{r}_j}{s_j \sqrt{n}}, \quad (3.2)$$

onde r_{ij} é a resposta do i -ésimo indivíduo para a j -ésima questão e $\bar{r}_j$ e s_j são respectivamente a média e o desvio padrão amostrais da j -ésima questão.

Sob uma nuvem de indivíduos (n indivíduos) definidos no espaço de variáveis R^p , cada uma das n linhas da tabela $\mathbf{X}$ é considerada como um vetor ou um ponto em R^p . Se a nuvem está contida no sub-espaço vetorial de q dimensões de R^p , isto é, $q < p$, o problema de redução de dimensionalidade está praticamente resolvido com um conjunto de novas variáveis em R^q que representem também a nuvem de indivíduos ou se aproxime a ela o mais possível. Estamos à procura de um sub-espaço vetorial de q dimensões, ou seja uma direção linear passando pela origem, que apresente o melhor ajuste possível da nuvem de pontos.

É necessário para isso definir o vetor unitário nessa direção. Seja $\mathbf{u}$ este vetor coluna associado à matriz $\mathbf{X}$, e por $\mathbf{u}^\top$ a sua transposta. Considera-se que $\mathbf{u}$ é ortonormal, isto é, $\mathbf{u}^\top \mathbf{u} = 1$. Define-se também a matriz de dados padronizados como $\mathbf{X}$, onde cada $\mathbf{x}_i$ (i -ésima linha de $\mathbf{X}$) representa um indivíduo em R^p .

Observando a Figura 3.2, a norma ou longitude da projeção OH_i de um vetor OM_i que representa a posição de um indivíduo $\mathbf{x}_i$ no sub-espaço contido na direção de $\mathbf{u}$, não é outro que o produto escalar de OM_i por $\mathbf{u}$, a soma dos produtos termo

Figura 3.2: Projeção de um ponto sobre o eixo H

a termo dos componentes de OM_i e de $\mathbf{u}$, isto é,

$$OH_i = \langle \mathbf{x}_i, \mathbf{u} \rangle = \sum_{j=1}^p x_{ij} u_j.$$

Cada uma das n linhas da tabela $\mathbf{X}$ é um vetor individual $\mathbf{x}_i$ em R^p . O produto matricial $\mathbf{X}\mathbf{u}$ é um vetor coluna de n elementos, onde cada termo é o produto escalar de uma linha de $\mathbf{X}$ por $\mathbf{u}$:

$$\mathbf{X}\mathbf{u} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ \cdots & x_{ij} & \cdots \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_j \\ \vdots \\ u_p \end{bmatrix} = \begin{bmatrix} \cdots \\ \vdots \\ \sum_j^p x_{ij} u_j \\ \vdots \\ \cdots \end{bmatrix}. \quad (3.3)$$

São os n componentes do vetor $\mathbf{X}\mathbf{u}$ que situam sobre $\mathbf{u}$ as n projeções OH_i dos pontos da nuvem. Entre os critérios de ajuste de um sub-espço a uma nuvem, o

que se preserva e que conduz seguramente aos cálculos analíticos mais simples, é o critério clássico dos mínimos quadrados. Ele consiste em encontrar a direção máxima da nuvem de pontos e em conseqüência tornar mínima a soma dos quadrados das diferenças, devido a que M_i e H_i representam o mesmo indivíduo e é desejável que o indivíduo e sua projeção sejam o mais próximos possível, ou seja, deseja-se minimizar

$$\sum_{i=1}^n (M_i H_i)^2. \quad (3.4)$$

Pelo Teorema de Pitágoras aplicado a cada um dos n triângulos $H_i O M_i$ tem-se a relação

$$\sum_{i=1}^n (M_i H_i)^2 = \sum_{i=1}^n (O M_i)^2 - \sum_{i=1}^n (O H_i)^2. \quad (3.5)$$

Como a quantidade $\sum_{i=1}^n (O M_i)^2$ é uma quantidade fixa, independentemente do vetor $\mathbf{u}$, então minimizar (3.4) é equivalente a maximizar a quantidade:

$$\sum_{i=1}^n (O H_i)^2.$$

Mas

$$(O H_i)^\top = \langle \mathbf{x}_i, \mathbf{u}^\top \rangle = \sum_{j=1}^p x_{ij} u_j$$

e, generalizando para os n pontos, $i = 1 \dots n$,

$$OH = \mathbf{X} \mathbf{u} = \langle \mathbf{X}, \mathbf{u} \rangle,$$

o que conduz a encontrar $\mathbf{u}$ que maximiza a seguinte forma quadrática:

$$\sum_{i=1}^n (O H_i)^2 = (OH)^\top OH = (\mathbf{X} \mathbf{u})^\top \mathbf{X} \mathbf{u} = \mathbf{u}^\top \mathbf{X}^\top \mathbf{X} \mathbf{u},$$

sujeito à condição $\mathbf{u}^\top \mathbf{u} = 1$. Seja $\mathbf{u}_1$ o vetor que maximiza a forma quadrática acima. O subespaço bidimensional que melhor representa a nuvem contém necessariamente os subespaços gerados por $\mathbf{u}_1$. Em seguida procura-se $\mathbf{u}_2$, o segundo vetor básico deste subespaço, ortogonal a $\mathbf{u}_1$ que torna $\mathbf{u}_2^\top \mathbf{X}^\top \mathbf{X} \mathbf{u}_2$ máximo. Prossegue-se desta forma até obter o sub-espaco de q dimensões.

3.3 Métodos para o estudo de uma tabela de dados

Dentre os diferentes métodos que a análise multivariada de dados tem para desenvolver ferramentas que permitam a análise como um todo de diversas variáveis, os métodos fatoriais têm um lugar especial. O objetivo primordial destes é a redução da dimensionalidade dos dados e a obtenção de novas variáveis (componentes, fatores, variáveis latentes, construtos ou índices) que contenham informação relevante dos dados e que expliquem de forma adequada a estrutura de correlação existente.

Quando se dispõe de uma única tabela de dados sobre a qual se deseja obter resultados, a análise multivariada nos apresenta diferentes métodos. Entre eles temos a Análise de Componentes Principais, Análise de Correspondência Simples e Múltipla, Análise Fatorial Exploratória e Confirmatória, entre os mais destacados. Desenvolveremos brevemente a ACP, o método que está ligado à Análise Fatorial Múltipla, metodologia proposta para a análise dos problemas descritos nesta dissertação.

3.3.1 Análise de componentes principais

É uma técnica estatística que transforma linearmente um conjunto de p variáveis quantitativas em um conjunto de fatores com um número menor, digamos q , variáveis não correlacionadas, que explicam uma parcela substancial da variabilidade do conjunto de informação disponível. As p variáveis originais $(X_1, \dots, X_p)$ são transformadas em $(Y_1, \dots, Y_p)$ denominadas componentes principais, de modo que Y_1 é aquela que explica maior parcela de variabilidade total, Y_2 explica a segunda maior parcela e assim por diante (Barroso e Artes, 2003).

Esta técnica que foi desenvolvida por (Pearson, 1901) e por (Hotelling, 1933), tem

como primeiro resultado combinações lineares interpretáveis das variáveis, mas geralmente, o analista conserva só algumas destas variáveis, reduzindo a dimensionalidade dos dados.

A Análise de Componentes Principais depende só da matriz de covariâncias (Σ) ou da matriz de correlação (ρ), de $(X_1, \dots, X_p)$. Não requer qualquer suposição sobre a forma da distribuição multivariada dessas variáveis; assim procura a descrição e entendimento da correlação das variáveis.

Dada uma matriz $\mathbf{X}$ representando n observações em p variáveis $(X_1, \dots, X_p)$ o propósito principal é determinar as novas variáveis $(Y_1, \dots, Y_p)$, tal que Y_i é dada pela combinação linear das p variáveis de $\mathbf{X}$.

Definição 3.3.1 (Mardia, Kent e Bibby, 1989) *Seja $\mathbf{X}$ a matriz dos dados de dimensão $n \times p$, composta por p variáveis, $(X_1, \dots, X_p)$ e n indivíduos ou observações com vetor de média μ e matriz de covariâncias Σ , seja $\mathbf{x}$ vetor coluna $(X_1, \dots, X_p)^\top$. Então a transformação*

$$\mathbf{x} \rightarrow \mathbf{Y} = \mathbf{\Gamma}^\top (\mathbf{x} - \mu) \quad (3.6)$$

permite obter novas variáveis chamadas componentes principais, onde $\mathbf{\Gamma}$ é ortogonal, composta pelos autovetores, correspondentes aos autovalores da matriz de variância-covariância $\Sigma = \mathbf{X}^\top \mathbf{X}$, e $\mathbf{\Gamma}^\top \Sigma \mathbf{\Gamma} = \mathbf{\Lambda}$ é a matriz diagonal dos autovalores $\lambda_1 \geq \lambda_2 \dots \geq \lambda_p \geq 0$.

Assim, desde que Σ seja definida positiva, os autovalores $\lambda_1, \lambda_2, \dots, \lambda_p$ são positivos. A representação de Σ mediante a decomposição espectral permite encontrar as CPs e definir suas propriedades. A i -ésima CP de $\mathbf{X}$ é definida pelo i -ésimo elemento de $\mathbf{Y}$, dado por

$$\mathbf{y}_i = \gamma_i^\top (\mathbf{x} - \boldsymbol{\mu}), \quad (3.7)$$

onde γ_i é a i -ésima coluna de Γ , chamado i -ésimo vetor de cargas das CPs. A função y_i é a i -ésima componente principal de x .

Dando um enfoque de estimação pelas primeiras CPs, o resultado para r componentes principais será:

$$\begin{aligned} Y_1 &= v_{11}X_1 + v_{21}X_2 + \dots + v_{p1}X_p \\ Y_2 &= v_{12}X_1 + v_{22}X_2 + \dots + v_{p2}X_p \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ Y_r &= v_{1r}X_1 + v_{2r}X_2 + \dots + v_{pr}X_p, \end{aligned} \tag{3.8}$$

em que v_{ij} é o j -ésimo elemento do vetor γ_i .

Matricialmente pode-se denotar $\mathbf{Y} = \mathbf{V}^\top \mathbf{x}$, onde $\mathbf{V}$ contém as r colunas correspondentes aos primeiros r autovetores de $\mathbf{X}^\top \mathbf{X}$. A aproximação de $\mathbf{X}$ é dada por $\hat{\mathbf{X}} = \mathbf{V}\mathbf{Y}$ onde $\mathbf{Y}$ é uma matriz de $(r \times n)$ das r primeiras componentes principais e $\mathbf{V}$ é a matriz com os r primeiros autovetores de $\mathbf{X}^\top \mathbf{X}$ (Jobson, 1989).

Diminuir a dimensionalidade dos dados procurando capturar a maior variabilidade de $\mathbf{X}$, é fazer com que as diferenças entre $\mathbf{X}$ e $\widehat{\mathbf{X}}$ sejam mínimas, no sentido de ter o menor $\text{tr}(\mathbf{X} - \widehat{\mathbf{X}})^\top (\mathbf{X} - \widehat{\mathbf{X}})$.

Logo, a quantidade a ser minimizada é

$$\text{tr}[(\mathbf{X} - \widehat{\mathbf{X}})(\mathbf{X} - \widehat{\mathbf{X}})^\top] = \text{tr}(\mathbf{X}^\top \mathbf{X}) - \sum_{i=1}^r \lambda_i. \quad (3.9)$$

Uma vez que $\mathbf{X}^\top \mathbf{X} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^\top$, então $\text{tr}(\mathbf{X}^\top \mathbf{X}) = \text{tr}(\mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^\top) = \text{tr}(\mathbf{\Gamma}^\top \mathbf{\Gamma} \mathbf{\Lambda}) = \text{tr}(\mathbf{\Lambda}) = \sum_{k=1}^p \lambda_k$. Isto faz com que o critério para obter as r primeiras CPs deve ser

aquele que escolhe uma quantidade tal que os não escolhidos expliquem uma quantidade menor ou não significativa de variabilidade dos dados. Para maiores detalhes do método pode-se ver Mardia et al. (1989).

3.4 Métodos para o estudo de múltiplas tabelas de dados

A bibliografia nos permite encontrar muitos métodos para a análise das tabelas múltiplas, desenvolvidos através dos anos, entre os quais a análise canônica generalizada de Horst (1961), a de Carroll (1968), a AFM de Escofier e Pagès (1994), a Split Análise de componentes principais de Lohmöller (1989) e o algoritmo de máxima variância de Horst (1965). O enlace entre PLS (Partial Least Squares) e esses métodos tem sido estudado na prática por Guinot, Latreille e Tenenhaus (2001) e por Pagès e Tenenhaus (2001).

Descrevemos a seguir a Análise Fatorial Múltipla, foco deste trabalho.

Análise Fatorial Múltipla (AFM) é uma técnica multivariada usada para a redução de dimensionalidade dos dados. Esta é empregada quando se tem um conjunto de indivíduos (linhas) que são avaliados por diversas variáveis (colunas) onde pode-se identificar grupos de variáveis ou um conjunto de variáveis avaliadas para grupos de indivíduos.

Consideremos a primeira situação, a grande quantidade de variáveis em geral encontradas faz com que naturalmente não seja possível identificar padrões de comportamento entre indivíduos. Além disso, há a necessidade de avaliar os dados como

um todo, considerando todas as variáveis de uma vez. Isso dificulta ainda mais a procura desses padrões. Tentar visualizar tendências, relações, semelhanças e oposições entre variáveis, procurando explicar melhor a variabilidade dos dados faz imediatamente pensar em uma Análise Fatorial (AF), mas isto é dificultado devido à presença de grupos de variáveis influenciando na obtenção dos fatores ou eixos principais.

De acordo com este perfil, há várias sub-tabelas com as mesmas linhas, (geralmente indivíduos) avaliadas para diversos grupos de variáveis (formando sub-tabelas). A análise dos dados pode ser enriquecida pois pode-se obter em cada sub-tabela, fatores representativos interpretáveis considerando que todas as variáveis dentro de cada grupo devem ser do mesmo tipo (quantitativas ou qualitativas). Assim, com iguais objetivos dentro de cada grupo de variáveis, se aplica uma ACP, mas se tem a dificuldade de que a variância entre os diferentes grupos tenham grandes diferenças na construção dos eixos; daí surge a necessidade de procurar uma transformação na sub-matriz que não altere a estrutura interna da nuvem gerada por cada sub-grupo de variáveis, mas padronize de certa forma as influências dos diferentes grupos na construção dos eixos fatoriais gerais. Foi provado que para que se cumpram estes requisitos, a transformação deve ser isométrica, homotética e isomorfa. Uma das transformações propostas consiste em ponderar a matriz de variâncias-covariâncias por seu maior autovalor.

3.4.1 Algumas fundamentações da análise fatorial múltipla

Desde que foram criados os primeiros fatores, índices ou construtos, os psicólogos sempre têm a necessidade de tentar comparar se esses índices ou fatores obtidos para um determinado grupo de pessoas ou indivíduos, servem para outro grupo distinto.

Surgem então, técnicas como a Análise Fatorial Confirmatória, a Análise Procrustes, ou os diferentes métodos citados na seção anterior, que procuram responder a estas perguntas. Por exemplo, num determinado país, mediante a AF ou ACP pode-se encontrar um fator ou construto que meça o índice de inteligência para crianças de certa idade. A pergunta é: será que este índice serve para avaliar as crianças da mesma faixa etária no Brasil? Os resultados que podem ser encontrados nestes países podem ser comparados mediante uma Análise Procrustes. No entanto, dentro deste conjunto de crianças, podem existir diversos estratos, determinados como por exemplo por condições sócio-econômicas, gênero, procedências de habitação (rural ou urbana), etc. Surge a pergunta: globalmente valendo os mesmos fatores para os diferentes países, valerão também para cada sub-grupo dentro de cada população, isto é, valem também para os diferentes grupos populacionais? Como comparar estes diversos grupos? Se os grupos são mais de dois, então a aplicação da Análise Procrustes é inadequada.

Como se tem mais de dois grupos que devem ser comparados e estudados simultaneamente, devemos procurar aquelas transformações que permitam, aplicadas às diversas tabelas, preservar as distâncias que existem entre os diferentes indivíduos e variáveis. Estes tipos de transformações se denominam na Álgebra, transformações homotéticas e isomorfas.

Dentre todas as transformações sobre matrizes simétricas positivas definidas com as quais trabalhamos na análise multivariada, as que são de nosso interesse são aquelas que têm a propriedade de preservar o produto interno, ou seja as que conservam a propriedade do produto interno e da norma. Entre estas se encontram a proposta feita por Escofier e Pagès (1982). A importância dessa transformação está baseada na interpretabilidade dos resultados, pois tem a propriedade de poder comparar as

diferentes nuvens de pontos, que são representadas por matrizes. Todas elas podem ser comparadas sendo levadas a um sub-espço de menor dimensão, por meio de projeções, de forma que pode-se explicar a maior variabilidade dos pontos. Para conservar a forma da nuvem, esta transformação precisa ser isomorfa, e para conservar a propriedade das distâncias entre os pontos precisa ser isométrica e homotética. O fato de ser isométrica faz com que as propriedades do produto interno de uma matriz por um vetor se conservem, e desde que para avaliar as distâncias entre os pontos utiliza-se a norma euclideana, precisa-se que as propriedades da projeção euclideana prevaleçam.

Cabe mencionar que, matematicamente, todas aquelas transformações em que prevaleçam as propriedades da projeção euclideana serão de grande utilidade se o que se deseja é medir distâncias, similaridades, proximidades, ou dissimilaridades dos dados. No entanto, uma projeção depende do produto interno e da definição de norma. Isto nos leva finalmente a trabalhar com aquelas normas euclidianas mencionadas a seguir.

Seja $\mathbf{x}$ um elemento do espaço euclidiano de R^s . Definimos as normas mais conhecidas matematicamente e que englobam a qualquer outra como:

- o máximo dos elementos do vetor $\mathbf{x}$

$$\|\mathbf{x}\|_{\infty} = \max |x_i|_{1 \leq i \leq s};$$

- a soma dos elementos do vetor $\mathbf{x}$

$$\|\mathbf{x}\|_1 = \sum_{i=1}^s |x_i|;$$

- a soma do quadrado dos elementos do vetor $\mathbf{x}$

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^s |x_i|^2 \right)^{1/2} \quad (\text{norma - euclideana}).$$

Pode-se concluir então, que aquelas transformações que possibilitam projetar de um sub-espço a outro, são aquelas que conservam as propriedades do produto escalar de uma transformação, e para o caso de se avaliar uma medida nestes espaços, seria adequada uma norma euclideana.

Sendo uma matriz um elemento do sub-espço das matrizes de posto s ($\mathbf{X} \in M_{s \times s}$), com os critérios acima mencionados, definimos $\mathbf{a}^j$ como a coluna j da matriz $\mathbf{A}$ e a_{ij} o elemento da matriz $\mathbf{A}$ na posição (i, j) , as transformações neste caso são:

$$\begin{aligned} \|\mathbf{A}\|_\infty &= \max_{1 \leq j \leq s} |\mathbf{a}^j| = \max_{1 \leq i \leq s, \max_{1 \leq j \leq s} |a_{ij}|}, \\ \|\mathbf{A}\|_1 &= \sum_{i=1}^s |\mathbf{a}^i| = \sum_{i=1}^s |a_{ij}|, \\ \|\mathbf{A}\|_2 &= \left(\sum_{i=1}^s |\mathbf{a}^i|^2 \right)^{1/2} = \left(\sum_{i=1}^s \sum_{j=1}^s |a_{ij}|^2 \right)^{1/2} \quad (\text{norma - euclideana}), \end{aligned}$$

onde $|\mathbf{a}^j|$ representa a norma respectiva do vetor coluna $\mathbf{a}^j$.

Desde que a matriz $\mathbf{A}$ seja simétrica e positiva definida, pode ser decomposta em uma matriz diagonal de autovalores multiplicada pelas matrizes ortonormais de autovetores (decomposição espectral). Conclui-se que a primeira norma das acima citadas, concorda com o maior autovalor, a segunda com a soma dos autovalores, e a terceira com a raiz da soma dos autovalores.

Estas três transformações permitem comparar as diferentes nuvens de pontos e variáveis, só que cada uma tem diferente propósito. Dividir pelo maior autovalor

significa levar (mediante projeções) todas as diferentes nuvens a um único plano, o espaço onde a variância interna de cada nuvem não influi na construção das variáveis latentes ou componentes principais, pois todas elas se encontrarão dentro de uma hiperesfera de raio 1, e mantém as propriedades de comparação entre variáveis e componentes principais obtidas em cada sub-nuvem (ACP por grupo), que dão uma maior interpretação à obtenção das componentes da nuvem total (ACP total).

No caso em que se quer ter componentes, sem considerar a formação de possíveis grupos entre variáveis, mas logrando que a variância total seja 1, isto é, quando as novas variâncias indicarão proporção de variabilidade com relação à variabilidade total, deve-se utilizar a transformação da soma de autovalores; se fosse com respeito ao desvio padrão, utilizar-se-ia a raiz quadrada da soma dos autovalores. Estes valores têm diferentes interesses ao proposto por AFM (primeira transformação proposta). Deve-se aplicar esta última transformação só se deseja-se comparar várias tabelas colocando-as todas em uma super-tabela (tabelas justapostas), e analisando a proporção explicada de variabilidade de cada uma das variáveis. Assim, variáveis com baixa variância ficarão fora de nosso modelo, ou seja, a variância das variáveis influem na construção dos eixos das componentes principais.

3.4.2 A metodologia da análise fatorial múltipla

A AFM é uma análise fatorial na qual cada grupo de variáveis é ponderada, com a finalidade de reduzir o efeito dos grupos de variáveis na formação das componentes principais finais. AFM permite o estudo de vários grupos de variáveis definidas sobre um conjunto de indivíduos sob várias perspectivas, já usadas em outros métodos, como a análise canônica (Carroll, 1968) e (Horst, 1961), e análise procrustes (para

duas tabelas), STATIS (Structuration des Tableaux À Trois Indices de la Statistique) (Lavit, 1989) e INDSCAL de (Carroll e Chang, 1970).

A AFM é utilizada para explorar características na população por várias tabelas de dados retangulares $\mathbf{X}_k$, ou grupo de variáveis. Cada grupo pode ser de diferente número de variáveis, assim como também podem ser grupos heterogêneos (quantitativas ou qualitativas) no entanto, cada grupo é composto por um mesmo tipo de variável.

Nesta dissertação os exemplos desenvolvidos são de grupos de variáveis quantitativas; no caso de estar frente a grupos de variáveis categorizadas, desenvolve-se uma Análise de Correspondência (AC), e os dados são ponderados de igual forma, pela raiz do maior autovalor associado à matriz de distâncias entre as variáveis. Uma vez ponderada, as variáveis categorizadas são analisadas seguindo o mesmo procedimento que as quantitativas, e suas interpretações e projeções são feitas de forma similar a qualquer variável quantitativa.

Princípios

A AFM trabalha com os seguintes princípios.

Com os dados disponíveis e avaliados por grupo de variáveis, podemos realizar a ACP para cada grupo. Assim, a dispersão dos indivíduos no primeiro conjunto de variáveis está representada em uma nuvem que denotamos N_1^I , a qual está num espaço R^{p_1} com dimensão p_1 . Da mesma forma, a dispersão de N_k^I está num espaço R^{p_k} com p_k dimensões. Todas estas nuvens juntam-se num sub-espço total R^p que é a soma direta de K sub-espços ortogonais. A nuvem total que apresenta os indivíduos (nuvem N_I) pertence ao espaço R^p .

Uma vez que $R^{p_1}, \dots, R^{p_K}$ são sub-espacos ortogonais, a projeção da nuvem de indivíduos N_I no sub-espaco R^{p_1} é a nuvem N_1^I , e mais geralmente, a projeção de N_I no sub-espaco R^{p_k} é a nuvem N_k^I .

Todas as dispersões de N_k^I estão contidas em R^p ; é possível então ter as simultâneas (ou superpostas) representações projetadas sobre um mesmo espaco. As projeções são feitas sobre o sub-espaco gerado pelas componentes principais da dispersão média (direção $\mathbf{u}_s$).

A nuvem N_k^I , que pertence ao sub-espaco R^{p_k} , é projetada sobre um vetor $\mathbf{u}_s$, de R^s que não pertence a R^{p_k} . A projeção de N_k^I sobre $\mathbf{u}_s$ consiste em realizar sucessivamente uma projeção sobre um vetor $\mathbf{u}_s^k$ e logo sobre $\mathbf{u}_s$ que contrai a nuvem multiplicando as coordenadas por $\cos(\theta)_k^s$, onde θ_k^s é o angulo entre $\mathbf{u}_s$ e $\mathbf{u}_s^k$. Tenha-se presente que os vetores $\mathbf{u}_s$ são ortogonais, mas não os $\mathbf{u}_s^k$, de onde vem a necessidade de projetar sobre essas direções na representação simultânea.

A AFM é realizada em duas etapas:

1. *Análise de cada grupo de variáveis (ACP das tabelas $\mathbf{X}_k$)* : cada tabela é analisada separadamente; realiza-se uma Análise de Componentes Principais por tabela $\mathbf{X}_k$ em cada caso e selecionam-se os primeiros dos autovalores resultantes, denotados por λ_k^1 . Usualmente se costuma considerar todas as variáveis com igual peso, no entanto, é possível designar às variáveis distintos pesos. Em um caso em que as variáveis são os fatores de outra análise, seria dada a cada variável-fator peso igual ao da inércia ou desvio padrão;
2. *Análise Global (Análise de Componentes Principais da tabela $\mathbf{X}$)* : na ACP da matriz $\mathbf{X}$ formada por todas as variáveis, isto é, por todas as tabelas justapostas;

cada tabela $\mathbf{X}_k$ é ponderada pelo inverso da raiz do maior autovalor ($\frac{1}{\sqrt{\lambda_k^1}}$) resultante da ACP de cada tabela, realizada no passo anterior.

Pode-se identificar outros usos da AFM, como por exemplo no caso de os mesmos indivíduos avaliados pelo mesmo conjunto de variáveis, em diversos níveis de algumas variáveis categorizadas ou em diversos tempos. Isto enriquece tanto as possibilidades de uso que permite utilizar em tabelas de variáveis mistas (quantitativas e qualitativas) como também em diferentes tempos. Assim, o método é indicado para a análise de enquetes realizadas nos mesmos indivíduos em diferentes tempos (estudos de painel) e as mesmas variáveis, para poder visualizar a evolução das mesmas ao longo do tempo. O motivo pelo qual esses grupos de variáveis são afetados por seus pesos deve-se essencialmente a:

- o número de variáveis do grupo (quanto maior é o número, maior é a influência do grupo);
- a correlação das variáveis que formam o grupo: quanto existir uma forte correlação entre as variáveis dentro de cada grupo, a influência do grupo será maior. Isto é determinante para a construção dos principais fatores de variabilidade.

Com o propósito de equilibrar a influência dos grupos de variáveis, a AFM designa a cada variável um peso. Este peso é o mesmo para todas as variáveis dentro de um grupo com a finalidade de conservar a estrutura interna de cada grupo. Ao multiplicar cada uma das variáveis do grupo pelo peso associado ao grupo, a variabilidade da nuvem associada é multiplicada pelo quadrado desse peso em cada direção de seu espaço.

Outro fato importante que cabe destacar, refere-se à multiplicação pelo fator

$(\frac{1}{\sqrt{\lambda_k}})$. Esta transformação permite que as diferentes variâncias existentes no conjunto de dados sejam maximizadas pela unidade, ou seja tenham 1 como maior autovalor (ver Seção A.2). O objetivo desta ponderação é equilibrar o papel dos grupos. Sua influência se equilibra no sentido de que nenhum pode determinar por si mesmo a primeira componente principal do conjunto. Segundo Escofier e Pagès (1990), a ponderação tem duas propriedades associadas:

- existe uma influência balanceada para todos os grupos de variáveis na análise global, porque a variância do primeiro eixo fatorial de cada grupo de variáveis é igual a 1. Porém, a máxima contribuição de um grupo de variáveis à formação de um eixo é limitada por um;
- todas as variáveis de um dado grupo têm um mesmo peso e portanto a estrutura do grupo essencialmente não muda.

A AFM tem sido comparada com a Análise de Correlação Canônica de Carroll (1968) e STATIS (Lavit, 1989). Foi mostrado que a AFM apresenta melhores propriedades relacionadas à comparação global entre grupos, em comparação à análise de correlação canônica (Escofier e Pagès, 1990) e a simultânea representação de indivíduos em planos fatoriais é melhor se comparada a STATIS (Pagès, 1996). A AFM considera outros métodos similares para a análise de tabelas de dados mistos. Em outras palavras, grupo de categorias e grupos de variáveis quantitativas podem ser analisados simultaneamente (Escofier e Pagès, 1986).

3.4.3 AFM em R^I : as variáveis

O espaço R^I contém as representações das variáveis. A cada uma de suas dimensões associa-se um indivíduo. Esse espaço contém elementos que precisamos estudar:

- nesse espaço estão situadas as variáveis;
- as CPs obtidas dentro de cada grupo de variáveis também podem ser visualizadas nesse espaço. É útil ver as posições relativas destas CPs na mesma escala que as variáveis.

A transformação proposta para as variáveis originais é dada por:

$$x_{ij} = \left(\frac{r_{ij} - \bar{r}_j}{s_j \sqrt{n}} \right). \quad (3.10)$$

3.4.4 Representação das variáveis e CPs parciais nos eixos globais

A representação se obtém diretamente da ACP da tabela completa $\mathbf{X}$, isso permite visualizar a nuvem de indivíduos e representar de forma ótima as correlações entre as variáveis. Um dos destaques da AFM é que nos permite fazer um estudo das correlações que os eixos globais têm com as componentes de cada grupo de variáveis. Isso dá um aporte significativo à análise das CPs globais, que se faz última etapa da AFM.

A busca dos fatores comuns aos grupos de variáveis foi estudada por Hotelling (1933) com a ajuda da análise canônica. Nesta análise se busca simultaneamente uma combinação linear das variáveis do primeiro grupo e uma combinação linear das

variáveis do segundo grupo, tal que o coeficiente de correlação entre esses grupos seja máximo; encontrando um par se procede em seguida com a busca de segundo par. Tem-se proposto numerosas generalizações para grupos de variáveis. O objetivo é procurar K combinações lineares de variáveis (cada combinação linear de um grupo) tais que essas combinações estejam relacionadas o mais possível. Existem diversas generalizações da Análise Canônica Generalizada (análise multicanônica) baseadas em diversas formas de definir a relação entre grupos de variáveis. Assim, pode-se provar que a AFM é um caso particular da Análise Canônica de Carroll.

Assim, a projeção das variáveis sobre os eixos fatoriais globais é dada por,

$$\varphi_{\alpha} = \mathbf{u}_{\alpha} \sqrt{\lambda_{\alpha}} = \frac{1}{\sqrt{\lambda_{\alpha}}} \mathbf{X}' \boldsymbol{\psi}_{\alpha},$$

onde φ_{α} é a projeção das variáveis na α -ésima CP, $\mathbf{u}_{\alpha}$ é o autovetor na direção da α -ésima CP, correspondente ao autovalor λ_{α} da matriz $\boldsymbol{\rho} = \mathbf{X}'\mathbf{X}$ e $\boldsymbol{\psi}_{\alpha}$ é α -ésima CP global. Denotemos o j -ésimo elemento do vetor φ_{α} por $\varphi_{\alpha j}$, $j = 1, 2, \dots, p$. Daí, temos

$$\varphi_{\alpha j} = \sum_{i=1}^n \frac{(r_{ij} - \bar{r}_j)}{s_j \sqrt{n}} \left(\frac{\psi_{\alpha j}}{\sqrt{\lambda_{\alpha}}} \right).$$

Então

$$\varphi_{\alpha j} = \text{cor}(\mathbf{x}_j, \boldsymbol{\psi}_{\alpha}).$$

A coordenada da j -ésima variável no α -ésimo eixo principal é o coeficiente de correlação desta variável com o fator $\boldsymbol{\psi}_{\alpha}$ (combinação linear das variáveis iniciais), consideradas depois como variáveis artificiais, cujas coordenadas são constituídas pelas n projeções de indivíduos sobre este eixo.

Do mesmo modo as componentes parciais representadas nos eixos principais são dadas por

$$\psi_{\alpha i} = \frac{1}{\lambda_1^k} \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{j \in J_k} \lambda_{ij} \varphi_{\alpha j}.$$

As deduções são apresentadas no Apêndice A.

3.4.5 AFM em R^p : os indivíduos

O espaço R^p contém as representações dos indivíduos; a cada uma de suas dimensões associa-se uma variável. Neste espaço buscamos:

- uma representação gráfica da nuvem de indivíduos caracterizados pelo conjunto de variáveis. Pelo exemplo, cada pessoa que prestou vestibular na UFMG no ano 1997, pode ser representada no primeiro plano formado pelas primeiras CPs obtidas na ACP das 8 variáveis previamente ponderadas;
- uma representação das K nuvens de indivíduos caracterizados por CP obtida em cada grupo de variáveis; estas são as projecções das CPs sobre o primeiro plano formado pelas duas primeiras CPs da ACP global.

A coordenada ou projecção de um indivíduo i sobre o α -ésimo eixo principal é escrita explicitamente como

$$\psi_{i\alpha} = \sum_{j=1}^p \mathbf{u}_{\alpha j} x_{ij} = \sum_{j=1}^p \mathbf{u}_{\alpha j} \left(\frac{r_{ij} - \bar{r}_{ij}}{s_j \sqrt{n}} \right) \quad \forall i = 1, 2, \dots, n,$$

que não é outra que a correlação do vetor indivíduo i e o autovetor $\mathbf{u}_{\alpha}$ da direcção da α -ésima CP, onde $\psi_{i\alpha}$ é a projecção do i -ésimo indivíduo na α -ésima CP global, $\mathbf{u}_{\alpha}$ é o vetor direcção da α -ésima CP global e s_j o desvio padrão da j -ésima variável.

Os indivíduos suplementares (aqueles que representam um grupo de indivíduos) serão representados mediante projecções, como é indicado acima; eles geralmente são

calculados como a média do respectivo grupo e são valores representativos do conjunto de indivíduos. Em nossa aplicação serão calculados como do grupo a que pertencem.

3.4.6 Influência da ponderação dos grupos sobre as K nuvens

$$N_I^k$$

Ao k -ésimo grupo de variáveis corresponde uma nuvem que representa os indivíduos, denotada por N_I^k situada num espaço de dimensão p_k , denotado por R^{p_k} . Lembremos que a ponderação trata de equilibrar o papel dos grupos de variáveis, a qual consiste na divisão pela raiz do maior autovalor, a cada uma das variáveis do grupo k respectivamente. Esse coeficiente a ser idêntico para todas as variáveis do grupo, não modifica a forma da nuvem N_I^k , no entanto, normaliza as nuvens no sentido de que a máxima variabilidade de toda nuvem N_I^k em qualquer direção vale 1. Note que com esta ponderação, duas nuvens homotéticas se fazem iguais.

3.4.7 Influência da ponderação dos grupos sobre a nuvem N_I associada às variáveis

Ao conjunto de todas as variáveis corresponde uma nuvem situada no espaço R^p . Nesta nuvem, denotada por N_I , o quadrado das distâncias entre os pontos i e l é a soma dos quadrados de suas distâncias em N_I^j (Escofier e Pagès, 1990).

A distância entre dois elementos da nuvem N_I está influenciada pelos diferentes grupos. Multiplicar os pesos iniciais das variáveis do grupo k por um coeficiente α_k

é um meio de equilibrar a influência dos grupos, pois a distância se calcula como

$$d_{(i,l)}^2 = \sum_{j \in K} \alpha_j d^2(i^j, l^j).$$

Com a ponderação proposta, nenhum dos grupos pode ser mais preponderante na formação na primeira direção da inércia ou variância.

As CPs maximizam a inércia das projeções de todas as variáveis. A inércia projetada sobre a nuvem N_I^k pode ser interpretada como a contribuição do grupo, que ao ser dividido pela raiz de λ_k^1 equilibra a influência do grupo.

3.4.8 Qualidade de representação de cada nuvem N_I^k

A qualidade da representação de cada nuvem N_I^k é medida da forma clássica, pela razão entre a inércia projetada e a inércia total da nuvem. Lembremos que o vetor $\mathbf{u}_s$ de R^s , sobre o qual é projetada N_I^k , não pertence ao sub-espço no qual é situada a nuvem, ele forma com sua projeção $\mathbf{u}_s^k$ sobre esse espaço, um ângulo denotado por θ_s^k . Logo,

Qualidade da representação de N_I^k sobre $\mathbf{u}_s = [\cos(\theta_s^k)]^2 \times (\text{Qualidade da representação sobre } \mathbf{u}_s^k)$.

Como em geral os valores de $\cos(\theta_s^k)$ são pequenos, pois a soma é igual a 1, esta medida é mais débil do que a obtida por ACP da nuvem N_I^k . A análise trata de fazer com que a distância entre os valores dos indivíduos avaliados pelos grupos de variáveis seja baixa, uma vez que um indivíduo é representado tantas vezes quantos grupos de variáveis se têm (K). No entanto esses pontos representam um mesmo indivíduo e devem estar próximos (chamados pontos homólogos). É lógico tomar a

inércia intra nuvens como medida de semelhança entre as projeções das nuvens sob um eixo. Esse valor toma significado ao ser comparado com a inércia total. Assim, calcula-se para cada eixo e para cada plano a razão entre a inércia inter grupos (variabilidade dentro de cada grupo) e a inércia total. Essa razão constitui um bom indicador da qualidade da representação superposta das nuvens N_I^k . Se a razão é próxima de 1, todas as nuvens têm suficientes características comuns para justificar um estudo detalhado de seus diferenciais, mas se é próxima de 0 as diferenças de forma das nuvens são grandes e justificam um estudo por grupo.

Capítulo 4

Índices Propostos para Avaliar o Ingresso à Universidade

Neste capítulo da dissertação, apresentamos um exemplo prático para a aplicação das técnicas de AFM e a comparamos com a ACP normalmente utilizada, ressaltando as virtudes da AFM no tratamento dos dados. A base corresponde ao conjunto de variáveis avaliadas no vestibular de 1997 da Universidade Federal de Minas Gerais, assim como a algumas das variáveis sócio-econômicas sobre os alunos. O problema é identificar fatores que estejam explicando como o ingresso à universidade se associa às diferentes disciplinas avaliadas, assim como a variáveis que caracterizam os alunos.

O fato das variáveis apresentarem grupos referentes às matérias relacionadas aos cursos de Ciências (Matemáticas, Biologia, Física e Química) e outro grupo às matérias correspondentes a Humanas (Português, Geografia, História e Inglês-Francês), nos faz optar por uma AFM. Para efeito de comparação aplicamos também a ACP. A tabela de dados (Tabela 2.1) apresenta também variáveis categorizadas, que serão

incorporadas na análise como variáveis suplementares.

4.1 Análise de componentes principais das notas

Para começar, fazemos um estudo de uma análise de componentes principais nos dados quantitativos da base, ou seja, sobre as notas obtidas pelos alunos nas matérias avaliadas no vestibular.

A análise foi baseada na matriz de correlação. Os resultados nos mostram que nas 4 primeiras componentes principais, a variabilidade dos dados é explicada em 80% (Tabela 4.1).

Como pode ser visto na Tabela 4.2, a primeira CP é explicada por quase todas as variáveis da mesma forma, dado que suas cargas estão compostas por números muito próximos. Isto é esperado, porque as componentes principais buscam obter novas variáveis que expliquem a maior variabilidade dos dados, e desde que as variáveis tenham correlações próximas e se esteja avaliando o conhecimento em geral, espera-se que a primeira CP seja um indicador do conhecimento global dos alunos. A segunda CP apresenta um contraste entre Física e Português, a terceira CP também apresenta um contraste de Geografia com Química e Inglês-Francês, e a quarta CP é basicamente associada a Inglês-Francês em contraste com Química. Em muitas situações práticas, como neste exemplo, a primeira componente principal é como uma média entre todas as variáveis.

Graficamente pode-se corroborar os resultados numéricos encontrados (ver Figuras 4.1 e 4.2). Na Figura 4.1, podemos ver como todas as variáveis são atraídas para o mesmo sentido da CP1. Estas variáveis ao serem projetadas no primeiro plano, nos

Tabela 4.1: ACP global - proporção da variância explicada por CP

	CP1	CP2	CP3	CP4	CP5	CP6
Desvio Padrão	0,0198	0,0072	0,0070	0,0066	0,0061	0,0058
Proporção de variância	0,5897	0,0787	0,0736	0,0650	0,0560	0,0513
Proporção acumulada	0,5897	0,6684	0,7421	0,8071	0,8631	0,9144

Tabela 4.2: ACP global - coeficientes das componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6
Matemática	0,3717	-0,3436	-0,0864	0,1344	0,0991	-0,1978
Biologia	0,3823	-0,0100	-0,2079	-0,2850	0,0251	0,1932
Física	0,3506	-0,5702	0,0554	0,1930	0,1235	-0,4409
Química	0,3571	-0,1084	-0,4770	-0,4512	-0,0078	0,3591
Português	0,3420	0,5190	0,2034	-0,0505	0,7329	-0,1379
Geografia	0,3407	-0,1459	0,6785	0,1533	-0,1326	0,5965
História	0,3498	0,3372	0,2582	-0,3445	-0,5962	-0,4646
Inglês-Francês	0,3316	0,3751	-0,3878	0,7177	-0,2530	0,0980

Figura 4.1: Primeiro plano das componentes principais

Figura 4.2: Segundo plano das componentes principais

dão idéia de como vêm sendo formados os eixos principais. Observa-se claramente que a primeira componente está relacionada com quase todas as variáveis, vemos também que a segunda componente resulta no contraste que apresentam as variáveis Português e Física.

Na Figura 4.2 pode-se observar que a terceira CP se vê explicada pelo contraste das variáveis de Geografia versus Química e Inglês-Francês.

4.2 Análise fatorial múltipla da tabela de dados

Veremos agora o que acontece se aplicarmos uma AFM aos dados. Dividimos a base de dados em três grupos de variáveis uma vez que sabemos que existem alunos que têm mais pré-disposição para o estudo de Humanas que de Ciências, como também que, ao contrário, alguns jovens têm mais pré-disposição para as Ciências que para Humanas. Esta estratificação formará nossos dois primeiros grupos, aos quais será incorporado um terceiro grupo de variáveis categorizadas (idade e sexo).

4.2.1 ACP em cada grupo de variáveis

Na tabela referente às disciplinas de Ciências
(Matemática, Biologia, Física e Química)

Primeiramente fizemos uma ACP na tabela formada pelos indivíduos avaliados pelas notas nas matérias de ciências. Os resultados mostrados na Tabela 4.3 indicam que as duas primeiras CPs explicam 82,5% da variabilidade e considerando uma terceira CP, a explicação chega a 91,5%. As componentes resultantes são descritas a seguir e podem ser observadas na Tabela 4.4.

1. **Índice de Conhecimento Geral em Ciências:** deve-se notar que a primeira CP explica a maior variabilidade dos dados e é explicada por todas as variáveis. Assim, ao ser influenciada por todas, nos indicará o conhecimento geral sobre elas;
2. **Componente Físico-Químico:** a segunda CP é explicada na sua maioria pelas variáveis Física e Química, é a diferença entre elas;

3. **Componente Matemático-Físico:** notamos também que a terceira CP é explicada pelas variáveis que dão as notas obtidas pelos alunos nas provas de Matemática e Física;
4. **Componente Bioquímico:** já a quarta CP está mais vinculada às variáveis Biologia e Química.

Tabela 4.3: Proporção da variância explicada por CP do grupo Ciências

	CP1	CP2	CP3	CP4
Autovalores	2,797	0,503	0,361	0,339
Proporção de variância	0,6992	0,1258	0,0902	0,0848
Proporção acumulada	0,6992	0,8250	0,9152	1,0000

Tabela 4.4: Coeficientes das variáveis para cada CP do grupo Ciências

	CP1	CP2	CP3	CP4
Matemática	-0,5068	0,3068	0,8055	0,0145
Biologia	-0,5098	-0,3624	-0,1690	-0,7617
Física	-0,4872	0,6566	-0,5591	0,1377
Química	-0,4959	-0,5860	-0,1002	0,6330

Neste contexto, conclui-se que dentro deste grupo de variáveis existe uma estrutura de correlação entre elas. As correlações entre todas elas podem ser consideradas significativas, como se pode observar na matriz de correlação das variáveis (ver Tabela 4.5).

**Na tabela referente às disciplinas de Humanas
(Português, Geografia, História e Inglês-Francês)**

As componentes são descritas a seguir com base na Tabela 4.7.

1. **Índice de Conhecimento Geral em Humanas:** as 4 variáveis explicam a primeira CP que pode ser interpretada como um indicador do conhecimento geral das matérias de ciências humanas;
2. **Componente Geografia - Línguas Estrangeiras:** a segunda CP apresenta um contraste entre a variável Geografia e Inglês-Francês;
3. **Componente Geografia - Português:** a terceira é o contraste entre Geografia e Português;
4. **Índice de Conhecimento de História:** é a quarta componente principal e está sendo explicada destacadamente pela variável História.

No grupo de variáveis humanas, a variabilidade explicada pelas duas primeiras componentes principais chega a 77%. Considerando três componentes principais explica-se 88,8% da variabilidade dos dados (Tabela 4.6).

Comparando-se as Tabelas 4.3 e 4.6, pode-se notar que a variância explicada por análise de componente principal é semelhante, nos dois grupos de variáveis, pois os autovalores não são muito diferentes.

A Tabela 4.9 mostra que a correlação entre as primeiras componentes principais de cada grupo é considerável, no entanto, isto não acontece entre as segundas componentes (0,777 e 0,0871 respectivamente), o que estaria indicando que as primeiras componentes principais estão próximas.

Tabela 4.5: Correlações entre as variáveis do grupo Ciências

	Matemática	Biologia	Física	Química
Matemática	1,00	0,61	0,63	0,59
Biologia	0,61	1,00	0,57	0,66
Física	0,63	0,57	1,00	0,53
Química	0,59	0,66	0,53	1,00

Tabela 4.6: Proporção da variância explicada por CP do grupo Humanas

	CP1	CP2	CP3	CP4
Autovalores	2,510	0,566	0,477	0,447
Proporção de variância	0,6276	0,1414	0,1193	0,1117
Proporção acumulada	0,6276	0,7690	0,8883	1,0000

Tabela 4.7: Coeficientes das variáveis para cada CP do grupo Humanas

	CP1	CP2	CP3	CP4
Português	0,5120	-0,0581	0,6220	0,5895
Geografia	0,4921	0,6037	-0,5785	0,2424
História	0,5140	0,2001	0,3240	-0,7686
Inglês-Francês	0,4811	-0,7695	-0,4165	-0,0542

Tabela 4.8: Correlações entre as variáveis do grupo Humanas

	Português	Geografia	História	Inglês-Francês
Português	1,00	0,50	0,55	0,51
Geografia	0,50	1,00	0,53	0,44
História	0,55	0,53	1,00	0,49
Inglês-Francês	0,51	0,44	0,49	1,00

Tabela 4.9: Correlações entre os eixos parciais dos grupos de variáveis

	CP1C	CP2C	CP3C	CP4C
CP1H	0,78	0,00	-0,05	-0,01
CP2H	0,02	0,09	0,13	-0,05
CP3H	0,02	0,03	-0,01	0,01
CP4H	0,08	0,02	-0,03	-0,01

Os resultados obtidos nos dão a informação dentro de cada sub-grupo de dados (notas referentes a Ciências e Humanas). Agora cada sub-tabela será ponderada da forma proposta para realizar a análise de componentes globais da tabela $\mathbf{X}$ ponderada.

4.2.2 Análise global da tabela completa

Como os maiores autovalores foram 2,797 e 2,510 para os grupos Ciências e Humanas respectivamente, procede-se a ponderar as sub-tabelas de dados dividindo cada sub-tabela pela raiz desses valores. Feita a ponderação, as maiores inércias dentro de cada nuvem de dados têm como maior valor a unidade.

Obtendo finalmente as componentes principais da tabela conjunta de dados, formada pelos dois grupos de variáveis, observamos que até a quarta componente principal, temos explicada 81% da variabilidade total dos dados e daí em diante a contribuição devida a cada CP é menor que 6% (Tabela 4.10). Os resultados da análise global na AFM são mostrados nas Tabelas 4.10 e 4.11. Ao compararmos esses resultados com os da Tabela 4.1, percebemos que ao subdividir as variáveis nos grupos Ciências e Humanas, a proporção de variabilidade explicada por grupo não altera-se significativamente na obtenção das componentes finais, mas nos permite descrever os eixos principais baseados nas relações existentes entre grupos de variáveis. Assim, sem que sejam afetados pelo efeito de grupo, encontramos fatores que têm interpretabilidade, e são explicados, por variáveis, indivíduos e grupos de variáveis mediante os fatores encontrados em cada ACP por grupo.

Tabela 4.10: ACP global ponderada: proporção da variância explicada por CP

	CP1	CP2	CP3	CP4	CP5	CP6
Desvio Padrão	1,333	0,490	0,478	0,447	0,422	0,394
Proporção de variância	0,588	0,079	0,076	0,066	0,059	0,051
Proporção acumulada	0,588	0,6670	0,743	0,809	0,868	0,919

Tabela 4.11: ACP global ponderada: coeficientes das componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6
Matemática	-0,3595	0,3055	-0,1845	0,0680	-0,0973	0,2479
Biologia	-0,3703	0,0204	-0,1642	-0,2883	-0,0378	-0,2482
Física	-0,3386	0,5173	-0,1209	0,1418	-0,1173	0,5116
Química	-0,3449	0,0758	-0,3848	-0,4689	-0,0250	-0,4454
Português	-0,3550	-0,4920	0,3134	-0,0633	-0,7046	0,1519
Geografia	-0,3529	0,3176	0,5993	0,3773	0,0878	-0,5109
História	-0,3627	-0,2589	0,3553	-0,3606	0,6409	0,3592
Inglês-Francês	-0,3435	-0,4721	-0,4390	0,6292	0,2448	-0,0602

4.2.3 Descrição das componentes principais globais

- Analisando a Tabela 4.11, é possível observar que: a primeira componente principal vem sendo explicada proporcionalmente por todas as variáveis (componente que mede o conhecimento total das matérias testadas); o sentido negativo das cargas de cada uma das variáveis nesta componente, nada mais é que referencial.
- Na segunda componente principal ressalta-se a influência da variável conhecimento em Física em contraste com conhecimento em Português e Inglês-Francês. Se um indivíduo tem o valor desta CP próximo de zero, será um indicador de que o aluno tem um desempenho parecido entre as variáveis mencionadas, quanto mais positivo seja o valor deste indicador, tem um desempenho melhor em Física que nas variáveis Português e Inglês-Francês, e quanto mais negativo ocorre o contrário.
- A terceira componente apresenta contraste entre as variáveis referentes a Humanas excluindo o Idioma Estrangeiro e a de Química mais Língua Estrangeira. Assim os alunos com valores próximos e altos dessa componente têm desempenho melhor nas matérias de Geografia, Português e História e não em idioma estrangeiro e Química. Quanto mais próxima de zero for a pontuação, indica um desempenho similar nas variáveis mencionadas, e se for negativa a CP, nos estaria indicando como as matérias com melhor desempenho as de Química e Línguas Estrangeiras.
- A quarta componente principal apresenta contraste entre as variáveis Química e idioma estrangeiro.

Análise pelos eixos parciais projetados sobre os eixos fatoriais principais

A Tabela 4.12 nos mostra as coordenadas que cada eixo parcial nos grupos Ciência e Humanas tem ao ser projetado no plano formado pelos dois primeiros eixos globais gerados na ACP global; além disso, nos indica as distâncias que os eixos parciais têm com cada eixo global, valores dados pelo quadrado dos cossenos dos ângulos formados pelos eixos.

Na Tabela 4.13 temos as mesmas medidas da Tabela 4.12 mas referentes às variáveis quando projetadas no primeiro plano fatorial das duas primeiras componentes principais globais.

Assim, o primeiro eixo parcial de Ciências têm coordenadas 0,9407 e 0,2465 e o primeiro de Humanas 0,9408 e $-0,2466$, o que indica que um dos eixos é o reflexo do outro com respeito ao primeiro eixo da ACP global, e localizam-se a distâncias similares dos eixos globais. Para o primeiro eixo do grupo Ciências, as distâncias aos eixos globais são 0,8848 e 0,0608 e para o primeiro eixo do grupo Humanas são 0,8851 e 0,0608 (Tabela 4.12).

Devido a que o eixo de Ciências é o reflexo do de Humanas, ambos formam a primeira CP Global, ou seja, apresentam um indicador conjunto de todas as variáveis.

Pode-se observar na Figura 4.3 que a primeira CP do grupo Ciências se aproxima das variáveis Matemática e Geografia e o primeiro eixo parcial de Humanas das variáveis de História, Idioma estrangeiro e Português. Já os segundos eixos estão quase alinhados com a segunda CP global.

Para os segundos eixos parciais, as coordenadas são 0,0137 e $-0,4291$ de Ciências e 0,0007 e $-0,4570$ para o eixo de Humanas, isto é, estão quase um sobre o outro, e

Tabela 4.12: Coordenadas e distâncias dos eixos projetados no primeiro plano das CPs globais.

Resumo da Análise Parcial por Grupo				
Eixo Parcial	Coordenadas		<i>Cosseno</i> ²	
Primeiro eixo grupo Ciências	0,9407	0,2465	0,8848	0,0608
Segundo eixo grupo Ciências	0,0137	-0,4291	0,0002	0,1841
Primeiro eixo grupo Humanas	0,9408	-0,2466	0,8851	0,0608
Segundo eixo grupo Humanas	0,0007	-0,4570	0,0000	0,2088

muito perto da segunda componente global, o que confirma os valores das distâncias entre esses eixos e os da CP global (ver Tabela 4.12).

Assim, a segunda componente principal global é explicada pelos dois segundos eixos parciais ou componentes de cada grupo (Componente Físico-Químico e Geografia-Línguas estrangeiras), sendo eles do mesmo sentido, como também pelas variáveis que, projetadas, exercem maior proporção de explicação para o eixo respectivo, neste caso, as variáveis Português, Inglês-Francês e Física. Logo, esta componente é um indicador de conhecimento de Idiomas (Português, Inglês-Francês), como também de Física. Novamente pode-se verificar este resultado já que as distâncias entre os segundos eixos de cada grupo são próximas a zero na coordenada do primeiro eixo e diferentes de zero para o segundo eixo 0,1841 e 0,2088, o que reafirma a relação (Tabela 4.12). Gráficamente, essas conclusões podem ser vistas na Figura 4.3.

Tabela 4.13: Coordenadas e distâncias das variáveis projetadas no primeiro plano das CPs globais.

Ciências					Humanas				
Variável	Coordenadas		<i>Cosseno</i> ²		Variável	Coordenadas		<i>Cosseno</i> ²	
Matemática	0,7979	-0,2601	0,6367	0,0676	Português	0,7467	-0,4001	0,5576	0,1601
Biologia	0,8181	0,0372	0,6694	0,0014	Geografia	0,7374	0,1991	0,5437	0,0396
Física	0,7450	0,4424	0,5555	0,1957	Historia	0,7557	-0,2463	0,5710	0,0606
Química	0,7639	0,0878	0,5835	0,0077	Ing-Fran	0,7239	-0,3205	0,5240	0,1027

Figura 4.3: Projeção das variáveis qualitativas no primeiro plano global. Grupo: F= Feminino, M= Masculino, 1=< 17, 2 =18, 3 =19, 4=20 a 24, 5=25 a 29, 6=30 anos a mais.

Cabe ressaltar que componentes obtidas por esse processo podem ser comparadas com outros fatores ou componentes que possam ser obtidos para outro conjunto de indivíduos testados pelas mesmas variáveis. Isso nos permite questionar:

- Os índices obtidos podem ser utilizados em outras populações, por exemplo, em outra universidade?
- Esses índices podem ser utilizados para diferentes épocas, como para alunos que postulem seu ingresso no próximo ano nesta universidade, ou para alunos de anos anteriores?

Para obter estas respostas basta aplicar a AFM aos dados incorporando as novas populações a comparar e obter as correlações com as componentes principais. Além disso, pode-se analisar pelas diferentes variáveis qualitativas, ou seja por sub-grupos que possam ser encontrados dentro na população. Então surge a pergunta: a estrutura ou fatores obtidos nas diferentes populações ou amostras tem certa semelhança? Como estão desenvolvendo-se internamente os sub-grupos? Podemos resolver o problema fazendo uma AFM em cada sub-amostra e obtendo os diferentes componentes ou construtos em cada amostra, para novamente poder relacionar com os eixos principais já obtidos. Como exemplo, incluímos as variáveis categorizadas faixa etária e gênero na análise a seguir.

4.2.4 Análise dos eixos parciais obtidos na ACP de cada grupo de gênero e idade

Os aspetos medidos por grupo de variáveis não são facilmente entendidos em conjunto, uma vez que são analisados independentemente. É necessário portanto, dispor de

uma ferramenta que sintetize a percepção dos indivíduos não só por variáveis, senão também por grupos de variáveis, e esta análise para cada sub-grupo populacional. Para isto, a AFM nos proporciona uma representação gráfica.

Deve-se ter presente que ao considerar uma ACP por grupo de variáveis, estamos implicitamente projetando os indivíduos de R^8 em dois sub-espacos de R^4 uma vez que em cada ACP parcial se trabalha independentemente, isto é, cada indivíduo foi considerado com 4 variáveis e completado por zeros no resto de variáveis não pertencentes ao grupo. Basicamente, o que se faz é projetar cada indivíduo no sub-espaco correspondente ao grupo de variáveis em análise. As componentes principais obtidas em cada ACP parcial são tratadas como variáveis suplementares e são projetadas sobre os eixos fatoriais globais, que são obtidos depois de agrupar todas as variáveis padronizadas pelo maior autovalor de cada grupo, obtendo-se assim uma representação superposta destas duas nuvens parciais na nuvem média global.

A AFM é uma transformação homotética (que não modifica estritamente a forma da nuvem) que aplica-se à nuvem global; ela representa cada ponto da nuvem no primeiro plano de eixos principais. Logo os pontos que representam os grupos de indivíduos são projetados de igual forma sobre esta nuvem e em qualquer sub-espaco desta, em particular, sobre o primeiro plano formado pelos dois primeiros eixos globais (Figura 4.3).

Se analisarmos as componentes principais obtidas para cada grupo de indivíduos, observa-se que não existem diferenças entre os percentuais de explicação das primeiras componentes e que as distribuições das influências das variáveis foram muito similares (ver Tabelas 4.14, 4.15 e 4.16).

Os grupos de indivíduos formados por gênero e idade são representados pelo centro

de gravidade dos indivíduos pertencentes ao grupo. Na Figura 4.3 pode-se observar que o grupo de indivíduos de 25 a 29 anos é considerado como o grupo de menor rendimento nas provas, seu centro de gravidade está localizado na região negativa do primeiro eixo global. No entanto, os pertencentes aos grupos de 17, 18 e 19 anos estão muito próximos da origem das coordenadas dos eixos globais e têm um rendimento balanceado ou similar em todas as matérias (primeiro eixo global) e nos contrastes relacionados especificamente ao segundo eixo global (Física versus Português e Inglês-Francês).

No que se refere ao sexo, os centros representativos de cada grupo (masculino e feminino) são projetados também no primeiro plano fatorial como se observa na Figura 4.3. Os homens (M) têm melhores resultados que as mulheres (F), porém seus centros de gravidade encontram-se em quadrantes diferentes. Assim, os homens, em geral, têm mais afinidade com as matérias referidas a Física e Matemática e baixas de Inglês-Francês e Português. No entanto, em todas as demais notas as médias dos homens têm um valor superior à média total dos indivíduos. As mulheres (M) obtiveram em média notas inferiores à média de todos os indivíduos no conjunto, com exceção do Português (ver Tabela 4.23). É por isso que elas se encontram próximas a esta variável, no sentido positivo da segunda CP global relacionada precisamente com Português, e no sentido negativo da primeira CP global, indicando esse baixo rendimento nesta avaliação. Cabe ressaltar que entre os 701 ingressantes à UFMG desse ano, 475 foram homens e 226 mulheres.

As tabelas das ACPs feitas para cada faixa etária nos mostram o comportamento dentro de cada grupo. Nelas pode-se observar que cada grupo de variáveis são representadas com duas CPs, nas quais, a primeira CP sempre é explicada pela totalidade das variáveis, enquanto as outras CPs vão se diferenciando notoriamente, de faixa a

Tabela 4.14: Proporção da variância explicada por CP

Masculino					Feminino			
Ciências								
	CP1	CP2	CP3	CP4	CP1	CP2	CP3	CP4
Desvio Padrão	0,0206	0,0100	0,0090	0,0087	0,0202	0,0092	0,0088	0,0085
Proporção de variância	0,625	0,146	0,118	0,111	0,6370	0,1323	0,1195	0,1112
Proporção acumulada	0,625	0,772	0,889	1,0000	0,6370	0,7693	0,8888	1,0000
Humanas								
	CP1	CP2	CP3	CP4	CP1	CP2	CP3	CP4
Desvio Padrão	0,0218	0,0095	0,0078	0,0074	0,0211	0,0090	0,0079	0,0075
Proporção de variância	0,6989	0,1313	0,0887	0,0811	0,6917	0,1248	0,0968	0,0866
Proporção acumulada	0,6989	0,8302	0,9189	1,0000	0,6917	0,8166	0,9134	1,0000

faixa etária. Os resultados são apresentados nas Tabelas 4.17 a 4.22.

Tabela 4.15: Coeficientes das variáveis para cada CP

Masculino					Feminino			
Ciências								
	CP1	CP2	CP3	CP4	CP1	CP2	CP3	CP4
Matemática	−0,500	0,460	0,695	−0,235	−0,510	0,039	−0,759	0,402
Biologia	−0,505	−0,462	−0,285	−0,671	−0,516	−0,234	−0,098	−0,818
Física	−0,495	0,542	−0,625	0,266	−0,477	0,794	0,376	0,029
Química	−0,500	−0,531	0,212	0,651	−0,496	−0,560	0,522	0,410
Humanas								
	CP1	CP2	CP3	CP4	CP1	CP2	CP3	CP4
Português	0,519	−0,051	0,325	0,789	−0,512	0,067	0,282	0,809
Geografia	0,494	−0,539	−0,677	−0,081	−0,494	0,585	−0,628	−0,142
História	0,513	−0,185	0,592	−0,593	−0,507	0,129	0,646	−0,557
Inglês-Francês	0,472	0,820	−0,292	−0,138	−0,487	−0,798	−0,332	−0,127

Tabela 4.16: Correlações das variáveis dos diferentes grupos

Masculino					Feminino			
Ciências								
	Mat	Bio	Fís	Quí	Mat	Bio	Fís	Quí
Matemática	1,00	0,58	0,65	0,57	1,00	0,64	0,58	0,59
Biologia	0,58	1,00	0,57	0,67	0,64	1,00	0,57	0,64
Física	0,65	0,57	1,00	0,55	0,58	0,57	1,00	0,51
Química	0,57	0,67	0,55	1,00	0,59	0,64	0,51	1,00
Humanas								
	Port	Geo	Hist	Ing-Fran	Port	Geo	Hist	Ing-Fran
Português	1,000	0,530	0,550	0,500	1,00	0,53	0,55	0,52
Geografia	0,530	1,000	0,520	0,420	0,53	1,00	0,52	0,47
História	0,550	0,520	1,000	0,470	0,55	0,52	1,00	0,50
Ing-Franc	0,500	0,420	0,470	1,000	0,52	0,47	0,50	1,00

Tabela 4.17: Coeficientes das variáveis para cada CP na faixa etária I (≤ 17 anos)

	CP1	CP2	CP3	CP4
Matemática	-0,5117	0,2911	-0,2709	-0,7616
Biologia	-0,5109	-0,2763	-0,6623	0,4732
Física	-0,4896	0,6381	0,4152	0,4252
Química	-0,4873	-0,6571	0,5617	-0,1235

Tabela 4.18: Coeficientes das variáveis para cada CP na faixa etária II (18 anos)

	CP1	CP2	CP3	CP4
Matemática	0,5100	-0,2806	0,7866	0,2061
Biologia	0,5126	0,2943	-0,0161	-0,8065
Física	0,4868	-0,6561	-0,5710	0,0814
Química	0,4901	0,6358	-0,2346	0,5482

Tabela 4.19: Coeficientes das variáveis para cada CP na faixa etária III (19 anos)

	CP1	CP2	CP3	CP4
Matemática	-0,5048	0,2684	0,8023	-0,1717
Biologia	-0,5085	-0,3706	-0,3450	-0,6965
Física	-0,4869	0,6875	-0,4869	0,2308
Química	-0,4996	-0,5639	0,0150	0,6574

Tabela 4.20: Coeficientes das variáveis para cada CP na faixa etária IV (20 a 24 anos)

	CP1	CP2	CP3	CP4
Matemática	-0,5038	-0,3404	0,7848	0,1204
Biologia	-0,5089	0,4346	-0,0243	-0,7427
Física	-0,4854	-0,6465	-0,5880	-0,0265
Química	-0,5016	0,5266	-0,1946	0,6582

Tabela 4.21: Coeficientes das variáveis para cada CP na faixa etária V (25 a 29 anos)

Matemática	0.4979	-0.4350	-0.7502	-0.0017
Biologia	0.5103	0.4557	0.0760	-0.7253
Física	0.4874	-0.5738	0.6561	0.0512
Química	0.5041	0.5232	0.0296	0.6865

Tabela 4.22: Coeficientes das variáveis para cada CP na faixa etária VI (>= 30 anos)

	PC1	PC2	PC3	PC4
Matemática	0.4756	-0.6914	0.4814	0.2529
Biologia	0.5137	0.5456	-0.0697	0.6584
Física	0.4885	-0.2953	-0.7910	-0.2201
Química	0.5209	0.3701	0.3710	-0.6738

Tabela 4.23: Médias por sexo das notas obtidas no vestibular na UFMG - 1997

	Português	Matemática	Geografia	História	Ing-Fran	Biologia	Física	Química
Homens	10,25	8,84	6,01	5,93	5,70	6,24	5,74	5,71
Mulheres	10,34	7,76	5,07	5,52	5,53	5,75	4,71	5,37
Total	10,29	8,29	5,52	5,72	5,51	5,99	5,21	5,54

4.3 Conclusões

- A ACP nos proporciona CPs que descrevem os dados baseados nas variáveis quantitativas de interesse. Nesta aplicação a primeira componente é composta por todas as variáveis, aquelas que avaliam o conhecimento em Ciências e aquelas que avaliam as matérias Humanas. Este é um índice que avalia o desempenho dos alunos no vestibular.
- A AFM nos proporciona CPs que além de serem descritas como as anteriores, podem ser explicadas pelas diferentes CPs que sejam necessárias, dos grupos de variáveis Ciências e Humanas (eixos parciais).
- A necessidade de obter resultados para diferentes sub-grupos de indivíduos (gênero ou faixa etária) pode ser resolvida realizando uma ACP em cada sub-grupo. Essas análises ficam isoladas, ou seja, são para cada grupo e não são comparáveis. Uma vez que cada nuvem de indivíduos, que pertencem a um determinado espaço, é representado em diferentes planos ou sub-espacos que representem a maior variabilidade dos dados, os quais não necessariamente coincidem, é desejável encontrar a forma de poder visualizar ou medir estas relações. A AFM nos permite representar os diferentes eixos dos diferentes sub-espacos (de maior variabilidade) gerados pelas ACPs prévias.
- Assim, no conjunto, os homens obtiveram melhores resultados que as mulheres. Eles apresentam médias maiores nas notas da maioria das matérias avaliadas, em particular nas matérias de Física e Matemática. Isto não acontece com as mulheres, elas têm melhor desempenho na matéria de Português (ver Tabela 4.23). Na Figura 4.3 o grupo das mulheres encontra-se localizado no eixo positivo da segunda componente global e no negativo da primeira e vale ressaltar

que a segunda CP é um contraste entre Português, Inglês-Francês e Física.

- Quanto às faixas etárias, as pessoas menores de 17 anos, 18 e 19 anos obtiveram um rendimento balanceado nas diversas matérias uma vez que elas estão muito próximas da origem dos eixos. No entanto, as pessoas que encontram-se na faixa etária de 20 a 24 anos e de 25 a 29 anos obtiveram as piores notas (ver Figura 4.3).
- A grande diferença que nos permite a AFM, é tornar possível a identificação destas componentes por gênero, que se relacionam não só com as variáveis mas também com os eixos gerados pelos grupos de variáveis. Se observarmos na primeira ACP por gênero, não será possível perceber as diferenças que podem estar presentes entre estas. No entanto, com a utilização da AFM é possível afirmar que as mulheres de 20 a 24 anos obtiveram menor nota que as mais novas (ver Figura 4.3).
- Finalmente é notório que a AFM nos permite obter CPs, eixos ou fatores que trazem consigo maior informação sobre os dados. Em comparação com a ACP, este exemplo nos mostra como os eixos globais que se têm no final não perdem porcentagem de explicação, e ainda apresentam maiores recursos para poder descrever melhor os dados.

Capítulo 5

Comparação entre Sub-Grupos Populacionais via Análise Fatorial Múltipla

BarOn (1996) fez uma Análise Fatorial Exploratória de 103 perguntas. Nessa análise encontrou 13 fatores que desviam-se ligeiramente das 15 componentes básicas teoricamente estabelecidas pela Psicologia. Para resolver as discrepâncias de alguns itens foi desenvolvida uma Análise Fatorial Confirmatória.

Posteriormente, BarOn (1997) assumiu que os 15 fatores podem ser explicados por 5 componentes teoricamente sustentáveis. Ele considera suficiente este modelo para estimar um só índice, chamado Índice de Quociente Emocional. Desde o início, o enfoque de BarOn foi multifatorial com a finalidade de ampliar as habilidades envolvidas na inteligência não cognitiva. O autor estudou sua proposta de 11 a 15 fatores, componentes da inteligência não cognitiva, durante duas décadas, encontrando que

esta última estrutura, com 15 variáveis, se comprova. Maiores detalhes podem ser encontrados em Ugarriza (2001).

Algumas questões, entretanto, permanecem sem resposta. Embora, dentro de cada população os fatores de BarOn fossem confirmados, como se comportam dentro de grupos de diferentes faixas etárias? Será que os fatores são independentes dentro das diferentes faixas etárias? Para dar resposta a essas perguntas, temos a Análise Fatorial Múltipla como opção.

Sob a metodologia da AFM, identificamos, neste caso, 5 grupos de indivíduos (faixas etárias), os quais devem ser comparados, tanto pelas variáveis referidas a comportamento emocional (15 fatores) como pelos diferentes grupos de variáveis (5 fatores resultantes da AF feita por BarOn). Primeiramente, lembremos que os fatores e as variáveis componentes são descritos a seguir (ver Capítulo 2):

- Intrapessoais:

1. CM: habilidade de entender-se, aceitar os sentimentos e emoções, distinguir e conhecer o porque dos mesmos;
2. AS: habilidade de expressar sentimentos, crenças, pensamentos e defender os direitos próprios de uma forma não destrutiva;
3. AC: habilidade para reconhecer e entender os sentimentos próprios, para saber o que sente e qual foi a causa, aceitando os aspetos positivo e negativo, como também as limitações e possibilidades;
4. AR: habilidade para realizar o que realmente podemos, queremos e desfrutarmos de fazer;
5. IND: habilidade para autodirigir-se, sentir-se seguro de si mesmo e ser

independente emocionalmente para tomar decisões;

- Interpessoais:

1. EM: habilidade de perceber, compreender e respeitar os sentimentos das pessoas;
2. RI: habilidade para estabelecer e manter relações mútuas satisfatórias caracterizadas por proximidade emocional e intimidade;
3. RS: habilidade para demonstrar-se como uma pessoa que coopera, contribui e que é membro construtivo do grupo social;

- Adaptabilidade:

1. SP: habilidade para identificar e definir os problemas como também para gerar e implementar soluções;
2. PR: habilidade para avaliar a correspondência entre o que experimentamos (subjetivo) e o que realmente existe (objetivo);
3. FL: habilidade para realizar um ajuste adequado de nossas emoções, pensamentos e condutas a situações e condições de mudanças;

- Manejo do estresse:

1. TE: habilidade para suportar eventos adversos, situações estressantes e emoções fortes sem "desmoronar", enfrentando ativa e positivamente o estresse;
2. CI: habilidade para resistir a um impulso ou tentações e controlar as emoções;

- Estado de ânimo e motivação:

1. FE: habilidade para sentir-se satisfeito com a vida, desfrutar de si mesmo e dos outros e para divertir-se e expressar sentimentos positivos;
2. OP: habilidade para ver o aspeto mais brilhante da vida e manter uma atitude positiva apesar dos sentimentos negativos.

Os grupos etários são descritos na Tabela 2.2.

Procedendo a AFM, realizamos uma ACP para cada grupo de variáveis e cada grupo é ponderado pelo maior autovalor correspondente de cada sub-matriz. Prosseguimos juntando as sub-tabelas já ponderadas e fazemos uma ACP geral.

Das ACPs parciais obtivemos as duas primeiras CPs que descrevem os grupos de variáveis, chamadas *Eixos Parciais* de cada grupo e os obtidos na ACP global, que representam as CPs globais chamadas *Eixos Principais* (ver Figura 5.1). As duas primeiras CPs parciais e globais explicam uma alta percentagem de variabilidade (ver Tabelas B.1 até B.20). Finalmente para a análise conjunta, as componentes parciais, as variáveis quantitativas e as variáveis qualitativas são projetadas sobre o primeiro plano fatorial da AFM global (ver Figuras 5.1 e 5.2).

5.1 Análise dos eixos parciais da ACP em cada nuvem com os eixos globais da AFM

Primeiramente descrevemos cada eixo parcial obtido por grupo de variáveis.

No primeiro grupo de variáveis (Intrapessoais) as duas primeiras CPs explicam 77% da variabilidade total. A primeira componente está relacionada com *AC ha-*

bilidade para reconhecer e entender os sentimentos próprios, para saber o que sente e qual foi a causa, aceitando os aspectos positivo e negativo, como também as limitações e possibilidades e AR habilidade para realizar o que realmente podemos, queremos e desfrutarmos de fazer; já a segunda CP é influenciada por CM habilidade de entender-se, aceitar os sentimentos e emoções, distinguir e conhecer o porque dos mesmos, assim como por AC já descrita, e a terceira CP é um contraste entre IND habilidade para autodirigir-se, sentir-se seguro de si mesmo e ser independente emocionalmente para tomar decisões, AR e AS (ver Tabelas B.1 e B.2).

Para o segundo grupo de variáveis (Interpessoais), tem-se que as duas primeiras CPs explicam 95,0% da variabilidade, a primeira CP é relacionada com **RI** *habilidade para estabelecer e manter relações mútuas satisfatórias caracterizadas por proximidade emocional e intimidade, RS habilidade para demonstrar-se como uma pessoa que coopera, contribui e que é membro construtivo do grupo social, assim como EM habilidade de perceber, compreender e respeitar os sentimentos das pessoas. A segunda CP é um contraste entre as variáveis RS e EM com RI no sentido oposto (ver Tabelas B.3 e B.4).*

No grupo de variáveis referentes à Adaptabilidade, a porcentagem de explicação do primeiro plano é de 83% da variabilidade. A primeira CP se relaciona com **SP** *habilidade para identificar e definir problemas como também para gerar e implementar soluções, PR habilidade para avaliar a correspondência entre o que experimentamos e o que realmente existe e FL habilidade para realizar um ajuste adequado de nossas emoções, pensamentos e condutas a situações e condições de mudanças. A segunda CP apresenta um contraste entre PR e FL (ver Tabelas B.5 e B.6).*

Nos dois últimos grupos, Manejo do estresse e Estado de ânimo e motivação, a

proporção de explicação da variabilidade da primeira CP é de 78% e 79%, respectivamente (ver Tabelas B.7 e B.9). A primeira CP do Manejo do estresse é associada a **CI** *habilidade para resistir a um impulso ou tentações e controlar as emoções* e a segunda com a variável **TE** *habilidade para suportar eventos adversos, situações estressantes e emoções fortes sem desmoronar, enfrentando ativa e positivamente o estresse* (ver Tabela B.8). Para as variáveis de Estado de ânimo e motivação, o primeiro eixo é relacionado a **OP** *habilidade para ver o aspecto mais brilhante da vida e manter uma atitude positiva apesar dos sentimentos negativos* e **FE** *habilidade para sentir-se satisfeito com a vida, desfrutar de si mesmo e dos outros para divertir-se e expressar sentimentos positivos*. Já a segunda CP é o contraste entre elas (ver Tabela B.10).

Lembremos que as projeções dos eixos parciais como das variáveis sobre o primeiro plano da ACP global, são as correlações que elas têm com os eixos globais. As correlações entre os dois primeiros fatores da *ACP parcial* com as duas primeiras CPs da AFM (ACP global) são apresentadas na Figura 5.1. Em cada caso as duas primeiras CPs da ACP parcial são apresentadas. Os resultados mostram que os primeiros fatores referentes a *Adaptabilidade*, *Manejo do estresse* e *Estado de ânimo e motivação* estão no mesmo sentido, em contraposição aos primeiros eixos dos *Intrapessoais* e *Interpessoais*.

Assim, observa-se também que a primeira CP geral é explicada tanto pelas variáveis consideradas, como também pelos primeiros eixos fatoriais do primeiro, segundo e quinto grupo (eixos parciais 1 GP1, 1 GP5 e 2 GP2), os quais são representados por vetores bastante prolongados, ao contrário dos segundos eixos parciais do quinto, primeiro e terceiro grupos de variáveis (ver Figura 5.1). Isto nos permitirá uma melhor interpretação, pois coloca em evidência que a primeira CP geral depende dos

eixos parciais 1 G1, 1 G5, 2 G2 e 1 G3 e das variáveis *AC*, *OP*, *AR*, *CM*, *FE* e *AS*, podendo-se concluir que não depende de todas as variáveis, como indicaria uma ACP simples. Em outro sentido, não só confirmaria ou não uma estrutura fatorial dentro do conjunto de dados se uma AFC é aplicada, mais que isso, permite-nos identificar melhor as variáveis que influem nas CPs finais.

Em nosso exemplo, percebe-se que o primeiro eixo principal depende das variáveis ligadas ao grupo Intrapessoal e Estado de ânimo e motivação, e ao segundo eixo parcial do grupo Interpessoal (que é relacionado a *RI*, *EM* e *RS*). Como foi dito, os primeiros eixos parciais são explicados por todas as variáveis pertencentes a cada grupo, logo, representam o respectivo grupo de variáveis. Assim, pode-se afirmar que o primeiro eixo global refere-se a conduta Intrapessoal, Estado de ânimo e motivação e conduta Interpessoal. Assim, é relacionado à pessoa e ao seu estado de ânimo, como a compreensão do eu, a capacidade de proximidade emocional e intimidade, como também a valorizar-se como pessoa social.

O segundo eixo global está ligado às variáveis **CI** *habilidade para resistir a um impulso ou tentações e controlar as emoções* e **EM** *habilidade de perceber, compreender e respeitar os sentimentos das pessoas*, apresentando um contraste entre elas, assim como aos primeiros eixos parciais do segundo e quarto grupo de variáveis, referidas a Manejo do estresse e ao fator Interpessoal, e também ao segundo eixo parcial do quarto grupo em menor proporção (ver Figura 5.1). O primeiro eixo do quarto grupo indica habilidade para o Manejo do estresse, habilidade que tem a pessoa para resistir a impulsos ou tentações e controlar suas emoções, e o segundo eixo do grupo quarto (2G4), que é descrito como habilidade para suportar eventos adversos ao qual soma-se a variável **CI** *habilidade para resistir a um impulso ou tentações e controlar as emoções*, que explica-se como habilidade para o controle de emoções. Este eixo

global está associado mais ao quarto grupo de variáveis, ligado ao Manejo do estresse e controle e solução das situações difíceis.

Sob esta visão, poder-se-ia ter dois índices gerais, um primeiro associado ao desempenho pessoal (Inter e Intra) e ao Estado do ânimo e motivação, e um segundo referente ao Manejo do estresse, condutas e adequação a situações mutantes, isto é, um fator ou uma CP de controle sobre si mesmo.

As primeiras componentes ou eixos parciais das análises por grupo de variáveis estão associadas com todas as variáveis que pertencem a cada grupo. No entanto, as segundas CPs marcam as diferenças; por exemplo o primeiro grupo está associado a *AC* e *CM*; o terceiro grupo a *FL*, o quarto grupo a *TE* e o quinto grupo a *OP* e *FE*. Tudo indica que estas segundas CPs destacam a diferença significativa com respeito a seus grupos, ao serem ortogonais às primeiras CPs de cada grupo respectivo.

5.2 Representação superposta dos sub-grupos populacionais (grupos por idade e gênero)

A Figura 5.3 é um gráfico do sub-espço em que os diversos grupos de variáveis e indivíduos podem ser comparados. Nesse gráfico pode-se interpretar tanto as posições como as tendências dos grupos de variáveis e de indivíduos dentro de um mesmo plano.

Em nosso caso, temos cinco grupos de variáveis, que formam cada um, um fator proposto por BarOn (1996), e grupos de indivíduos, dois formados por sexo e cinco formados por faixa de idade. Os grupos de variáveis são etiquetados por G1, G2, G3, G4, G5, correspondentes a cada fator de BarOn descritos anteriormente e estão

ligados por segmentos ao redor do centróide representativo do grupo etário ou de gênero a que pertencem, pois cada grupo de indivíduos é avaliado em todos os grupos de variáveis.

Deve-se levar em conta que as projeções dos eixos parciais e das variáveis nos eixos principais são sempre correlações entre eles e os eixos principais ou globais (ver Seções A.2.2 e A.2.3). Essas projeções sempre se encontram na hipersfera de raio 1 ao serem representadas em qualquer sub-espço de dimensão $s < p$ (p número de variáveis), como pode-se ver na Figura 5.1; a AFM nos proporciona ferramentas gráficas que permite-nos representar os grupos de indivíduos por projeções dos vetores que representam os valores médios ou centróides de cada grupo de variáveis, isto é, são novos indivíduos representativos dos grupos, projetados no espaço onde encontram-se as duas primeiras CPs globais. Eles não necessariamente estarão no círculo unitário de eixos e variáveis, como pode-se observar nas Figuras 5.2 e 5.3. O que deseja-se verificar precisamente é que os pontos pertencentes a um mesmo grupo de variáveis estejam superpostos, por exemplo, o ponto que representa os indivíduos de 25 a 34 anos, avaliados pelo grupo 1 de variáveis (Intrapessoais) está quase superposto ao de 25 a 44 anos. Isto nos indicará se a estrutura de correlação proposta por BarOn é a mesma para sub-grupos populacionais.

É evidente que se os grupos de idade tivessem um comportamento homogêneo, eles deveriam ter localização próxima nesse plano, quanto mais perto estiverem os grupos (15 a 24, 25 a 34 anos ou 55 ou mais, ou Masculino e Feminino) menor é a diferenciação entre eles e, de forma similar, pode-se considerar ao projetar eixos parciais dos diversos grupos de indivíduos avaliados nos grupos de variáveis (Figura 5.3).

Devido ao gráfico nos proporcionar múltiplas possibilidades, começamos por analisar os grupos de indivíduos, primero por sexo e depois por idade. Em seguida, passamos a realizar a análise por grupos de variáveis dentro de cada grupo de indivíduos, e finalmente comparamos as possíveis relações que os fatores obtidos apresentam.

Observando os resultados por sexo, na Figura 5.3, existe uma contraposição entre os vetores que representam os comportamentos masculino e feminino, em torno do segundo eixo principal. Isto se nota melhor ao realizarmos a análise por grupos de variáveis, observando-se que existe na maioria dos casos uma oposição dos vetores representativos dos grupos de variáveis dos homens e das mulheres, apresentando maior diferença para os grupos Interpessoais e Manejo do estresse. Assim, pode-se concluir que os fatores globais não são válidos para esses grupos de indivíduos, uma vez que têm comportamentos diferentes (sentidos opostos).

Das Figuras 5.2 e 5.3, temos que os grupos de idade 25 a 34 anos e de 35 a 44 anos estão muito próximos. Isto indica que o comportamento emocional é muito similar entre eles. Por outro lado, existe uma diferença notória entre aqueles que pertencem aos grupos de idade 15 a 24 anos e os outros todos. Os indivíduos de 15 a 24 anos têm um comportamento muito diferenciado e quase contraditório aos demais grupos no que se refere à primeira CP global. Já os indivíduos com 25 a 34 anos, 35 a 44 anos e 45 a 54 anos, estão relacionados positivamente com o primeiro eixo principal da AFM global que, como visto anteriormente, é um indicador relacionado ao conhecimento do próprio eu, aos sentimentos e emoções nos aspectos Interpessoais e Intrapessoais, assim como perceber-se como um integrante importante da sociedade, contrariamente aos indivíduos do primeiro grupo de 15 a 24 anos. Isto faz sentido uma vez que é provável que uma grande porcentagem dos mais jovens ainda não tenha desenvolvido uma postura frente ao aspeto pessoal (intra, inter e social), e maturação

dos sentimentos e emoções, oposto ao que ocorre com as pessoas de maior idade.

Tabela 5.1: Coordenadas e contribuições de cada primeira componente parcial aos eixos principais

Grupo Etário	Coordenadas		Contribuição	
	Eixo1	Eixo2	Eixo Principal 1	Eixo Principal 2
15 a 24 anos	0,886	0,584	22,6%	27,3%
25 a 34 anos	0,67	0,483	18,6%	18,7%
35 a 44 anos	-0,84	-0,063	20,3%	26,5%
45 a 54 anos	-0,71	-0,282	16,5%	6,4%
55 ou mais anos	-0,863	-0,173	21,5%	24,1%

5.3 Representação superposta da qualificação nos índices emocional do ponto de vista de grupo etário

Nesta seção tentamos dar resposta à seguinte pergunta: os fatores emocionais encontrados para as populações de Estados Unidos e Lima-Perú, são válidas para sub-grupos populacionais? Se existe diferença entre os grupos populacionais, será que podemos avaliar ou estabelecer esta diferença? Que grupos são os mais próximos e quais são os mais discrepantes? Daremos resposta a essas questões considerando algumas variáveis suplementares. Para isto, a cada indivíduo associamos uma sub-nuvem que corresponde a cada fator pré-definido por BarOn. Uma vez que o conceito do fator implica na independência, pontos representativos de um mesmo fator devem

Figura 5.1: Projeções dos eixos parciais nos dois primeiros eixos principais da AFM global

sobrepôr-se graficamente. Podemos observar na Figura 5.3, que a dispersão relativa aos valores de cada sub-grupo têm diferentes comportamentos pois a maioria dos pontos referidos a um mesmo fator estão localizados em diferentes regiões no primeiro plano fatorial. Isto indica que cada grupo etário têm comportamento diferenciado para cada grupo de variáveis.

Os grupos etários de 25 a 34 anos e de 35 a 44 anos são os mais semelhantes, tendo sua maior diferença com o fator referido a Adaptabilidade, por isso estão próximos. Quanto aos outros grupos, pode-se observar maiores diferenças salvo alguns pontos, como o caso entre os indivíduos de 55 ou mais anos e os de 45 a 54 anos no que se refere ao fator Interpessoal; com respeito aos outros fatores, são muito diferenciados.

Os indivíduos da primeira faixa etária são diferenciados dos demais já que o grupo está localizado do lado negativo do primeiro eixo global; já os demais grupos em conjunto se encontram próximos, destacando-se a proximidade entre o grupo de 25 a 34 anos e os de 35 a 44 anos de idade e 55 ou mais. Assim, pode-se assumir que as diferenças entre estes grupos são mínimas (ver Figura 5.2), mas ao fazer a análise por grupos de variáveis, encontra-se que a principal diferença está no que se refere ao segundo e quarto grupo de variáveis, e em menor proporção ao terceiro grupo. Na Figura 5.3, pode-se observar como os dois grupos de idade maior se diferenciam dos demais precisamente no segundo grupo de variáveis (G2) no sentido que tem menor valor quanto ao eixo positivo da segunda CP global. Isto indica que, no que se refere ao fator Interpessoal, existe um controle diferenciado. No que concerne ao quarto grupo de variáveis, G4 está mais próximo de zero para o grupo de 55 ou mais anos, que para as demais faixas etárias, entendendo-se que tem um menor Manejo do estresse e um comportamento menos controlado frente a situações sob pressão.

Quanto ao primeiro e terceiro grupos de variáveis (G1 e G3), salvo o quinto grupo etário, o valor da primeira CP aumenta à medida que o grupo etário aumenta em idade. No entanto, para o resto dos grupos suas nuvens não são independentes posto que se transpõem entre si e isto acontece para qualquer grupo. Assim, pode-se concluir que é pouco provável que os grupos etários tenham a mesma estrutura fatorial. Mediante a AFM pode-se obter respostas de como cada fator descreve cada grupo populacional e como pode ser diferenciado entre eles.

Por exemplo, os indivíduos da faixa etária 55 ou mais velhas são pessoas mais velhas que se identificam por um comportamento emocional relacionado com as variáveis *RS*, *RI*, *AS*, *CM*, *FE*, *AR* e *EM*, que se identificam pelos fatores *Interpessoais*, *Intrapessoais* e *Estado de ânimo*, dado pela variável *EM* (ver Figura 5.2). Tudo isso faz concluir que a estrutura de correlação neste grupo não se identifica com os fatores propostos por BarOn, mas pode ser uma combinação entre eles ou variáveis a eles associadas.

A Tabela 5.2 nos proporciona também a proporção de explicação de variância de cada grupo ou a contribuição do grupo na formação das duas primeiras CPs. Tudo isto faz concluir que cada grupo etário é avaliado de forma diferente pelos fatores dados por BarOn, chamados aqui grupo de variáveis.

Detalhadamente dentro de cada faixa etária temos que os grupos de 25 a 34 anos, 35 a 44 anos e de 45 a 54 anos com respeito aos fatores Adaptabilidade e Manejo do estresse têm comportamento similar, e um tanto menos com respeito ao fator de Estado de ânimo e motivação ao qual se soma os indivíduos de 55 ou mais anos. Como pode-se notar, em nenhum caso os indivíduos de 15 a 24 anos são próximos dos demais grupos (ver Figura 5.3).

Figura 5.2: Representação das variáveis e dos grupos etários. 1:15 a 24 anos, 2:25 a 34 anos, 3:35 a 44 anos, 4:45 a 54 anos, 5:55 ou mais anos

Figura 5.3: Representação das CPs da AFM e das variáveis categorizadas

5.4 Conclusões

Além de que os fatores propostos por BarOn para a população dos Estados Unidos foram confirmados para a população de Lima-Perú tem-se o interesse de saber se eles são os mesmos para os sub-grupos etários. Ao fazer uma Análise Fatorial Confirmatória (AFC) se põe em evidência 5 fatores, que foram confirmados por Ugarriza (2001), salvo pequenas diferenças entre as populações avaliadas.

- Da Análise Fatorial Múltipla pode-se obter como conclusão que dentro de cada grupo etário não são válidos os fatores propostos.
- Os grupos etários de 25 a 34 anos, de 35 a 44 anos e de 55 ou mais anos têm um comportamento emocional similar no conjunto. Os outros dois grupos etários (15 a 24 anos e 45 a 54 anos) são diferenciados entre si e dos anteriores. De igual forma, ao considerar os gêneros, existem diferenças entre os grupos ainda que sejam menos evidentes (Figura 5.2).
- Internamente, os grupos considerados, ao serem analisados pelos cinco fatores ou grupos de variáveis, apresentam comportamentos diferenciados. Por exemplo, os grupos formados por sexo apresentam uma avaliação emocional mais similar no que refere-se aos grupos de variáveis *Intrapessoais* e *Adaptabilidade*, oposições entre as avaliações feitas por *Manejo do estresse* e *Estado de ânimo e motivação*, e uma diferença notória quanto às variáveis *Interpessoais*. Pode-se notar que mulheres e homens são diferenciados pelo segundo, quarto e quinto grupo de variáveis (ligados à segunda CP global) do resto de grupos. Isso é indício de que os 5 fatores propostos são válidos para a estratificação por gênero, pois os fatores 2, 4 e 5 são refletidos pelo centro das coordenadas dos eixos

principais, indicando que os fatores podem ser os mesmos, mas a avaliação ou interpretação é distinta (contrária) de homens para mulheres.

- Com respeito aos grupos etários, o primeiro (15 a 24 anos) é mais homogêneo com respeito aos 5 grupos de variáveis. Este grupo etário está localizado ao redor do lado negativo da primeira CP global. No entanto, não podemos concluir que apresentem a mesma estrutura de correlação e atinja os mesmos fatores que a população geral. Isto é concluído uma vez que este grupo está correlacionado com a primeira CP, que é explicada por uma combinação de variáveis pertencentes a 4 dos fatores teóricos propostos (AC, OP, AR, AS, FE, SP, PR, TE, CM). Isto indica que se realizarmos a análise fatorial é possível não chegar aos mesmos fatores. Esta mesma conclusão é obtida para o resto dos grupos etários com maior justificativa uma vez que eles apresentam maior heterogeneidade em seus resultados.

Em nosso exemplo, a análise feita anteriormente por faixas etárias não tinha podido encontrar que os grupos correspondentes às faixas etárias 25 a 34, 35 a 44 e 44 a 55 anos têm comportamentos muito diferenciados no que se refere aos diferentes grupos de variáveis, e que eles se diferenciam muito com respeito às pessoas que pertencem à faixa etária 15 a 24 anos. Também foram representados distintos grupos de variáveis dentro de um mesmo plano fatorial para permitir realizar análises como a aqui desenvolvida.

Capítulo 6

Conclusões Gerais

A Análise Fatorial Múltipla:

- permite-nos identificar quais variáveis iniciais ou fatores que influem decididamente na formação dos eixos ou fatores globais;
- permite-nos analisar os eixos globais não só por variáveis de interesse mas também por CPs parciais (que representam grupos de variáveis), isto é, incorpora variáveis que são interpretáveis, que explicam a estrutura de correlação entre as variáveis pertencentes ao grupo. Isto permite explorar totalmente o comportamento entre os diferentes grupos de variáveis, entre eles e dentro deles. Neste caso é muito útil visualizar as posições relativas dos fatores parciais na mesma escala que as variáveis iniciais, de forma tal que enriquece a interpretação dos eixos finais;
- não só permite uma identificação visual, que de fato é um modo didático para mostrar os resultados, mas também nos permite quantificar, mediante as cor-

relações entre as variáveis iniciais, CPs parciais e globais, quais são aquelas variáveis que influem na formação de eixos globais e o modo como isto acontece. Esta abordagem, comparada com os métodos tradicionais como a ACP ou AC, nos dá uma maior flexibilidade no uso das variáveis pois pode tanto trabalhar com variáveis quantitativas quanto com variáveis categorizadas;

- permite-nos comparar as estruturas fatoriais de diversas sub-populações dentro de uma mesma estrutura global populacional, considerando as correlações dos elementos representativos dentro de cada sub-grupo de indivíduos, identificando que grupos são mais homogêneos entre si e que relações possuem com os demais grupos;
- permite-nos comparar de forma simples as relações existentes entre um grupo de indivíduos avaliados por diferentes grupos de variáveis, considerando a maior explicação da variabilidade dos dados e maximizando a relação entre os eixos parciais de cada grupo e os eixos globais da ACP global (correlação canônica), melhorando as análises feitas sob outras abordagens, como a análise Procrustes que resulta limitada ao caso de dois grupos de variáveis, e complexo para mais de dois, e ou STATIS no caso de uma análise no conjunto dos grupos.

Apêndice A

Análise nos Espaços Gerados

A.1 Análise dos indivíduos no espaço de variáveis

Seja a base de dados original (respostas iniciais) dada por:

$$\begin{bmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \vdots & \vdots \\ \cdots & r_{ij} & \cdots \\ \vdots & \vdots & \vdots \\ r_{n1} & \cdots & r_{np} \end{bmatrix} \quad (\text{A.1})$$

Assim, r_{ij} é o valor resultante do indivíduo i na variável j .

As interpretações entre as linhas e as colunas das tabelas de dados permitem representar visualmente as proximidades entre os indivíduos e variáveis. Para os n indivíduos, as $\frac{n(n-1)}{2}$ distâncias em R^p correspondentes aos pares de pontos que representam os indivíduos têm uma interpretação direta. Para ilustrar a distância

entre o indivíduo i e o indivíduo i' temos:

$$d^2(i; i') = \sum_{j=1}^p (r_{ij} - r_{i'j})^2 \quad \forall j = 1, \dots, p \text{ e } i, i' = 1, \dots, n. \quad (\text{A.2})$$

Trata-se da distância euclidiana clássica. Dois pontos são muito próximos se suas p coordenadas são muito próximas. Os dois indivíduos de interesse são então caracterizados pelos valores quase iguais para cada variável (Piron et al., 1995).

Em nossos exemplos, duas pessoas avaliadas pelas perguntas mencionadas em cada aplicação representam pontos próximos se têm os mesmos valores ou valores próximos para as mesmas variáveis avaliadas.

Desde que seja necessário encontrar as direções ou sentidos das maiores dispersões dos dados, não é necessário tornar máxima a soma dos quadrados das distâncias à origem da projeção, mas sim maximizar a soma dos quadrados das distâncias entre todos os pares de indivíduos (Figura A.1). Assim, se as projeções dos pontos i e i' são h_i e $h_{i'}$ respectivamente, temos

$$\begin{aligned} \sum_{i=1}^n \sum_{i'=1}^n d^2(i; i') &= \sum_{i=1}^n \sum_{i'=1}^n (h_i - h_{i'})^2 = n \sum_{i=1}^n h_i^2 + n \sum_{i'=1}^n h_{i'}^2 - 2 \sum_{i=1}^n h_i \sum_{i'=1}^n h_{i'} \\ &= 2n^2 \left(\frac{\sum_{i=1}^n h_i^2}{n} - \bar{h}^2 \right) = 2n \sum_{i=1}^n (h_i - \bar{h})^2, \end{aligned}$$

onde $\bar{h}$ designa a média das projeções dos n indivíduos:

$$\bar{h} = \frac{\sum_{i=1}^n h_i}{n}$$

e corresponde à projeção sobre H_1 do centro de gravidade G da nuvem. Então o

j -ésimo valor da coordenada é:

$$\bar{r}_j = \frac{\sum_{i=1}^n r_{ij}}{n}.$$

Figura A.1: Projeção em H_1

Conseqüentemente, temos:

$$\sum_{i,i'}^2 d^2(i; i') = 2n \sum_{i=1}^n d^2(i; G).$$

Voltar a maximizar a soma de quadrados das distâncias entre todos os pares de indivíduos no lugar de maximizar a soma de quadrados das distâncias entre os pontos e o centro de gravidade de nuvem G , isto é,

$$Max \left[\sum_{i=1} \sum_{i'=1} d^2((i; i')) \right]$$

é equivalente a encontrar

$$Max \left[\sum_{i=1} \sum_{i'=1} d^2((i; G)) \right].$$

Se a origem se toma em G , a quantidade a maximizar será a nova soma de quadrados das distâncias à origem, a qual corresponde ao problema de análise geral em R^p . O sub-espço procurado resulta da análise geral da tabela transformada $\mathbf{X}$, de termo geral

$$x_{ij} = r_{ij} - \bar{r}_j.$$

A.1.1 Transformação proposta

A distância entre dois indivíduos i e i' é a distância euclidiana usual dada pela expressão (A.2). Podem existir valores de j para os quais as variáveis correspondentes são de escalas muito distintas; por exemplo: notas obtidas em uma avaliação de 1 a 10 e a idade em anos da pessoa; deseja-se que a distância entre dois pontos, seja independente das unidades. Pode-se, às vezes, desejar, sobretudo quando as unidades de medidas não são as mesmas, aplicar a cada variável um papel idêntico na definição das proximidades entre indivíduos. Trata-se, então, de análise de componentes principais padronizada. Para isso se corrigem as escalas adotando a distância

$$d^2(i; i') = \sum_{j=1}^p \left(\frac{r_{ij} - r_{i'j}}{s_j \sqrt{n}} \right)^2,$$

onde s_j é o desvio padrão da variável j .

$$s_j^2 = \frac{\sum_{i=1}^n (r_{ij} - \bar{r}_j)^2}{n}.$$

Isto induz a concluir que a padronização em R^p da tabela inicial $\mathbf{X}$ deve ser

$$x_{ij} = \left(\frac{r_{ij} - \bar{r}_j}{s_j \sqrt{n}} \right), \quad (\text{A.3})$$

ou seja, cada elemento da matriz deve ter como termo geral

$$x_{ij} = \left(\frac{r_{ij} - \bar{r}_j}{s_j \sqrt{n}} \right). \quad (\text{A.4})$$

Todas as variáveis assim transformadas são comparáveis, pois elas possuem igual dispersão

$$s^2(x_j) = 1.$$

Em resumo, a análise da nuvem de pontos de indivíduos em R^p nos conduz a efetuar uma translação da origem ao centro de gravidade da nuvem e trocar, no caso da análise da norma, a escala nos diferentes eixos. A análise da tabela transformada $\mathbf{X}$ nos conduz a diagonalizar a matriz

$$\boldsymbol{\rho} = \mathbf{X}^\top \mathbf{X}.$$

O termo geral $\rho_{jj'}$ desta matriz é escrito como:

$$\rho_{jj'} = \sum_i^n x_{ij} x_{ij'},$$

ou seja,

$$\rho_{jj'} = \frac{1}{n} \sum_i^n \frac{(r_{ij} - \bar{r}_j)(r_{ij'} - \bar{r}_{j'})}{s_j s_{j'}}.$$

Isso resulta em

$$\rho_{jj'} = \text{cor}(R_j, R_{j'}). \quad (\text{A.5})$$

A função $\rho_{jj'}$ não é outra que o coeficiente de correlação amostral entre as variáveis com índices j e j' . Daí a utilidade do coeficiente $\sqrt{n}$ introduzido no denominador da relação (A.4). Em resumo, a matriz com a qual se deve trabalhar é a matriz de correlação. Esta será a matriz diagonalizável e sobre a qual posteriormente se poderá encontrar as componentes ou fatores.

A.1.2 Os indivíduos projetados em eixos fatoriais

As coordenadas dos n pontos de indivíduos nos eixos fatoriais μ_α (α -ésimo autovetor da matriz ρ associada ao autovalor λ_α são as n componentes do vetor

$$\psi_\alpha = X\mu_\alpha.$$

O fator próprio ψ_α já é uma combinação linear das variáveis iniciais. Por estar a nuvem de indivíduos situada em torno do centro de gravidade, a média do fator é nula, isto é,

$$\sum_i^n \psi_{\alpha i} = 0$$

e sua variância é o autovalor

$$\text{var}(\psi_\alpha) = \lambda_\alpha.$$

A coordenada de um indivíduo i sobre a α -ésima CP, é a projeção da linha $\mathbf{x}_i$ sobre o espaço de direção μ_α escrita explicitamente como

$$\psi_{i\alpha} = \sum_{j=1}^p \mu_{\alpha j} x_{ij} = \sum_{j=1}^p \mu_{\alpha j} \left(\frac{r_{ij} - r'_{ij}}{s_j \sqrt{n}} \right). \quad (\text{A.6})$$

A.2 Análises das variáveis no espaço de indivíduos

A.2.1 Análise das variáveis

Se os valores tomados por duas variáveis particulares são muito próximos para todos os indivíduos, estas variáveis serão representadas por dois pontos muito próximos em R^n . Isso quer dizer que estas variáveis medem um mesmo fenômeno ou também que são vinculadas por uma relação particular (Piron et al., 1995).

Consideramos primeiro a nuvem dos n indivíduos não ponderados. Queremos, no espaço das variáveis, representar a nuvem de pontos em um subespaço de uma e seguidamente de duas dimensões, de modo que se obtenha uma representação visual o mais fiel possível à representação exata, de forma tal que se possa captar as proximidades que existem entre os n indivíduos frente às p variáveis. A distância entre variáveis deriva-se da análise em R^p . Calculamos a distância euclidiana usual entre duas variáveis j e j' :

$$d^2(j, j') = \sum_{i=1}^n (x_{ij} - x_{ij'})^2,$$

ou seja,

$$d^2(j, j') = \sum_{i=1}^n (x_{ij})^2 + \sum_{i=1}^n (x_{ij'})^2 - 2 \sum_{i=1}^n x_{ij} x_{ij'}.$$

Substituindo x_{ij} por seu valor dado em (A.4) e sabendo de fato que

$$s_j^2 = \frac{\sum_{i=1}^n (r_j - \bar{r}_j)^2}{n},$$

obtemos $\sum_{i=1}^n x_{ij}^2 = \sum_{i=1}^n x_{ij'}^2 = 1$ e igualmente

$$\sum_{i=1}^n x_{ij} x_{ij'} = \rho_{jj'}.$$

Segue que a relação que vincula a distância em R^n entre duas variáveis x_j e $x_{j'}$ é função do coeficiente de correlação $\rho_{x_j, x_{j'}}$ entre essas variáveis,

$$d^2(x_j, x_{j'}) = 2(1 - \rho_{x_j, x_{j'}}).$$

Isto implica em que

$$0 \leq d^2(j, j') \leq 4.$$

No espaço de R^n , o cosseno do ângulo entre duas variáveis é o coeficiente de correlação entre essas duas variáveis dado por (A.5) e pode também ser obtido da definição:

- cada vetor é a representação de uma variável e tem como norma a unidade:

$$\|\text{variável}\|^2 = \sum_{i=1}^n \left(\frac{1}{n} \left(\frac{x_{ik} - x_k}{s_k} \right)^2 \right) = 1$$

Por isso a nuvem N^V está contida em uma hipersfera de raio um;

- o cosseno do ângulo formado pelos vetores que estão representando as duas variáveis j e j' coincidem com o coeficiente de correlação entre ambas.

$$\text{cosseno}(j, j') = \langle j, j' \rangle = \sum_i \left(\frac{1}{n} \left(\frac{x_{ij} - x_j}{s_j} \right) \right) \left(\frac{1}{n} \left(\frac{x_{ij'} - x_{j'}}{s_{j'}} \right) \right) = \text{cor}(R_j, R_{j'}).$$

A interpretação desta última relação é muito importante porque é geométrica, e conseqüentemente visual. Esta propriedade requer a centralização prévia. É importante ter em conta que a padronização escolhida (expressão (A.4)) implica que, *na representação das variáveis, o que nos está interessando principalmente são as direções determinadas pelas variáveis, isto é, mais como vetores que como pontos.*

Ao ser a longitude dos vetores que representam as variáveis igual à unidade, a coordenada da projeção de uma variável sobre a outra pode ser interpretada como a medida de correlação entre elas.

O sistema de proximidades entre dois pontos de variáveis dada pela relação acima é familiar para um estatístico:

- duas variáveis fortemente correlacionadas são muito próximas uma da outra ($c_{jj'} = 1$) ou pelo contrário elas podem distanciar-se o máximo possível ($c_{jj'} = -1$) dependendo da relação linear que elas têm, direta ou inversa;
- duas variáveis ortogonais estão a uma distância média. A proximidade entre os pontos variáveis é interpretada conseqüentemente em termos de correlações.

Conclusão:

Efetuar um balanço de coeficientes de correlação entre as variáveis supõe estudar os ângulos entre os vetores que definem esta nuvem N^V . É impossível realizar tal estudo diretamente, dada a dimensão de R^n . Segue-se o interesse na ACP a proporcionar-nos variáveis sintéticas que constituem um resumo do conjunto de variáveis iniciais e permitem uma representação plana aproximada das variáveis e os ângulos que formam entre si.

Distância das variáveis à origem

A análise em R^n não é feita em relação ao centro de gravidade da nuvem de pontos variáveis, contrariamente à nuvem dos pontos indivíduos, mais sim em relação à origem. A distância de uma variável j à origem O se expressa por

$$d^2(j, j') = \sum_{i=1}^n x_{ij}^2 = 1.$$

Todos os pontos variáveis estão dentro de uma esfera de raio 1 centrada na origem das coordenadas.

Os planos de ajuste cortam a esfera em grandes círculos (de raio 1), círculos de correlações. A inércia deles é a inércia das variáveis pertencentes ao plano. Enfatizamos que transformações analíticas simples da expressão (A.4) nos espaços R^p e R^n dão interpretações geométricas diferentes. Consideremos por exemplo a interpretação de centralização das variáveis:

- em R^p , esta transformação equivale a uma translação de origem dos eixos ao centro de gravidade (ou ponto médio) da nuvem;

- em R^n , é uma projeção paralela à primeira bissetriz dos eixos sobre os eixos do hiperplano que lhe é ortogonal.

A.2.2 As variáveis projetadas nos eixos fatoriais da AFM

Temos observado como os indivíduos são representados nos planos fatoriais por

$$\psi_\alpha = \mathbf{X} \mathbf{u}_\alpha \quad (\text{A.7})$$

para $\alpha = 1, \dots, p$. Sejam $\mathbf{u}_\alpha$ os autovetores e os λ_α respectivos autovalores da matriz $\boldsymbol{\rho} = \mathbf{X}'\mathbf{X}$ de ordem $n \times n$, onde os $\mathbf{u}_\alpha$ autovetores são unitários ortonormais, e estão na direção da α -ésima componente principal.

O vetor

$$\varphi_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{X} \mathbf{u}_\alpha \quad (\text{A.8})$$

será o autovetor unitário de $\mathbf{X} \mathbf{X}^\top$. Então a α -ésima componente principal em R^n (das variáveis) é escrita como

$$\varphi_\alpha = \mathbf{X}' \mathbf{v}_\alpha.$$

Da expressão (A.7) temos:

$$\frac{1}{\sqrt{\lambda_\alpha}} \mathbf{X}' \mathbf{X} \mathbf{u}_\alpha \quad (\text{A.9})$$

e pela descomposição espectral chegamos a,

$$\varphi_\alpha = \mathbf{u}_\alpha \sqrt{\lambda_\alpha}. \quad (\text{A.10})$$

Como $\psi_\alpha = \mathbf{X} \mathbf{u}_\alpha$ substituindo na expressão (A.9) temos,

$$\varphi_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{X}^\top \psi_\alpha,$$

onde $\varphi_{\alpha j} = \{\varphi_{\alpha j}\}$ para $j = 1, \dots, p$. Logo, os valores $\varphi_{\alpha j}$, $j = 1, \dots, p$ são as coordenadas fatoriais das variáveis nos eixos α que são obtidas por $\mathbf{u}_\alpha \sqrt{\lambda_\alpha}$.

Além disso, podemos observar que:

$$\varphi_{\alpha} = \mathbf{u}_{\alpha} \sqrt{\lambda_{\alpha}} = \frac{1}{\sqrt{\lambda_{\alpha}}} \mathbf{X}^{\top} \boldsymbol{\psi}_{\alpha} = \{\varphi_{\alpha j}\}$$

e daí, temos,

$$\varphi_{\alpha j} = \sum_{i=1}^n \frac{(r_j - \bar{r}_j)}{s_j \sqrt{n}} \left(\frac{\psi_{\alpha j}}{\sqrt{\lambda_{\alpha}}} \right).$$

Então:

$$\varphi_{\alpha j} = \text{cor}(R_j, \boldsymbol{\psi}_{\alpha}).$$

A coordenada de uma variável no eixo principal não é outra senão o coeficiente de correlação desta variável com o fator $\boldsymbol{\psi}_{\alpha}$ (combinação linear de variáveis iniciais) consideradas depois como variáveis artificiais, cujas coordenadas são constituídas pelas n projeções de indivíduos sobre este eixo.

Enfatizando que:

- as CPs são variáveis artificiais não correlacionadas duas a duas, que resumem as informações das variáveis iniciais;
- as CPs não se traduzem como relações lineares entre variáveis, um coeficiente baixo entre duas variáveis não significa que estas sejam independentes, porque pode existir uma relação de grau superior;
- a coordenada da variável no eixo α é necessariamente inferior a 1 em valor absoluto e a soma delas é 1.

A.2.3 Representação superposta de nuvens dos grupos de variáveis

Na AFM, as projeções das componentes principais obtidas nas nuvens parciais sobre os eixos principais fatorais da nuvem total têm a propriedade de não participar na construção das CPs da tabela total, isto é, são consideradas como variáveis suplementares.

Resulta então:

$$\psi_{\alpha i} = \frac{1}{p} \sum_k \psi_{\alpha i^k}.$$

Porém,

$$\psi_{\alpha i} = \frac{1}{\lambda_1^k} \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j \in J_k} \lambda_{ij} \varphi_{\alpha j}.$$

Apêndice B

Tabelas

Tabela B.1: Proporção da variância explicada por CP do grupo intrapessoais

	CP1	CP2	CP3	CP4	CP5
Desvio Padrão	8,4785	3,9272	3,1476	3,0610	2,6168
Proporção de variância	0,6337	0,1360	0,0873	0,0826	0,0604
Proporção acumulada	0,6337	0,7697	0,8570	0,9396	1,0000

Tabela B.2: Coeficientes das CPs das variáveis intrapessoais

	CP1	CP2	CP3	CP4	CP5
CM	0,3736	-0,8651	0,1367	-0,2313	-0,1994
AS	0,3677	-0,1670	-0,6013	0,3525	0,5925
AC	0,6305	0,4056	0,1310	-0,6105	0,2193
AR	0,4898	0,1452	0,5247	0,6688	-0,1285
IND	0,2963	0,1950	-0,5721	0,0477	-0,7379

Tabela B.3: Proporção da variância explicada por CP do grupo interpessoais

	CP1	CP2	CP3
Desvio Padrão	6,9820	4,0824	1,8581
Proporção de variância	0,7079	0,2420	0,0501
Proporção acumulada	0,7079	0,9499	1,0000

Tabela B.4: Coefficientes das CPs das variáveis interpessoais

	CP1	CP2	CP3
EM	0,4703	-0,4225	-0,7748
RI	0,7032	0,7098	0,0398
RS	0,5331	-0,5636	0,6310

Tabela B.5: Proporção da variância explicada por CP do grupo adaptabilidade

	CP1	CP2	CP3
Desvio Padrão	5,9546	3,3343	3,0394
Proporção de variância	0,6353	0,1992	0,1655
Proporção acumulada	0,6353	0,8345	1,0000

Tabela B.6: Coefficientes das CPs das variáveis de adaptabilidade

	CP1	CP2	CP3
SP	0,4620	-0,0196	-0,8867
PR	0,7774	-0,4723	0,4155
FL	0,4269	0,8812	0,2030

Tabela B.7: Proporção da variância explicada por CP do grupo manejo do estresse

	CP1	CP2
Desvio Padrão	7,5096	3,9903
Proporção de variância	0,7798	0,2202
Proporção acumulada	0,7798	1,0000

Tabela B.8: Coeficientes das CPs das variáveis de manejo do estresse

	CP1	CP2
TE	0,4040	-0,9148
CI	0,9148	0,4040

Tabela B.9: Proporção da variância explicada por CP do grupo estado do ânimo e motivação

	CP1	CP2
Desvio Padrão	5,5016	2,8173
Proporção de variância	0,7922	0,2078
Proporção acumulada	0,7922	1,0000

Tabela B.10: Coeficientes das CPs das variáveis de estado do ânimo e motivação

	CP1	CP2
FE	0,7926	-0,6098
OP	0,6098	0,7926

Tabela B.11: Proporção de variância explicada na ACP global do grupo de 15 a 24

anos	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Desvio Padrão	12,6583	7,2354	6,2843	4,9057	4,2957	3,8550	3,4170	3,3314	3,1213	2,9099
Propor. variância	0,4202	0,1373	0,1036	0,0631	0,0484	0,0390	0,0306	0,0291	0,0256	0,0222
Propor. acumulada	0,4202	0,5575	0,6611	0,7242	0,7725	0,8115	0,8421	0,8712	0,8968	0,9190

Tabela B.12: Coeficientes das CPs do grupo de 15 a 24 anos

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
CM	0,2415	-0,1100	-0,0122	-0,2382	0,4835	-0,3375	0,3811	-0,3742	0,0832	-0,4127
AS	0,2200	-0,1178	0,1395	0,1057	0,1699	-0,2382	0,2439	0,1881	0,0516	0,5117
AC	0,3881	0,0087	0,2795	0,0831	-0,1929	0,3213	0,2770	-0,0557	0,5247	-0,1465
AR	0,3098	-0,0701	0,1062	0,0381	-0,0951	0,3092	0,3173	0,1404	-0,5373	-0,0846
IND	0,1750	0,0376	0,1769	0,1779	0,0229	-0,0925	0,1889	0,3116	0,2215	0,3304
EM	0,1601	-0,2766	-0,5077	0,1611	-0,0982	-0,0339	-0,0451	0,0261	0,0418	-0,0413
RI	0,3254	-0,3568	-0,0895	-0,5337	0,0544	-0,1627	-0,2689	0,0090	0,0086	0,3341
RS	0,2041	-0,1975	-0,6367	0,2418	-0,0737	0,1532	0,0755	0,0772	0,1564	-0,0422
SP	0,1796	0,0717	-0,0377	0,2597	-0,0373	-0,2585	0,1910	-0,0811	-0,5548	0,0753
PR	0,2783	0,2395	-0,0132	0,2664	0,7082	0,3219	-0,4028	0,0763	-0,0403	-0,0070
FL	0,1174	0,1076	0,0120	-0,1203	-0,0320	-0,3412	-0,0945	0,7587	0,0208	-0,4876
TE	0,2986	0,2356	0,0894	0,2570	-0,2315	-0,4099	-0,4060	-0,2482	0,0846	0,0275
CI	0,1949	0,7571	-0,3630	-0,3930	-0,1158	0,0556	0,1833	-0,0377	0,0148	0,1738
FE	0,3240	-0,1477	0,1853	-0,3322	-0,1775	0,3128	-0,2770	-0,0017	-0,1804	-0,0658
OP	0,2954	0,0094	0,1128	0,1977	-0,2647	-0,1226	-0,1390	-0,2167	-0,0341	-0,1903

Tabela B.13: Proporção de variância explicada na ACP global do grupo de 25 a 34

anos	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Desvio Padrão	13,5180	7,1435	5,4730	4,7953	4,4128	3,8430	3,4412	3,2559	3,0379	2,9421
Propor. variância	0,4686	0,1309	0,0768	0,0590	0,0499	0,0379	0,0304	0,0272	0,0237	0,0222
Propor. acumulada	0,4686	0,5995	0,6763	0,7353	0,7852	0,8231	0,8534	0,8806	0,9043	

Tabela B.14: Coeficientes das CPs do grupo de 25 a 34 anos

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
CM	0,2571	-0,2146	-0,0176	0,4791	-0,2713	0,3831	-0,2134	0,1087	-0,2893	0,4180
AS	0,2137	-0,1694	0,2088	0,0482	-0,1517	0,1531	0,5358	-0,0203	-0,4048	-0,2361
AC	0,3503	-0,0182	0,2080	-0,1952	0,2209	-0,2321	0,1644	0,1193	-0,1336	-0,0047
AR	0,3084	-0,1103	0,1858	-0,0387	0,1204	-0,1549	0,1576	0,3947	0,4300	0,1540
IND	0,1678	0,0252	0,2502	-0,0453	-0,0795	0,0253	0,3673	-0,3001	-0,1124	-0,1018
EM	0,1389	-0,2725	-0,4371	-0,2757	-0,1813	0,0108	0,0346	-0,0241	-0,0436	0,0646
RI	0,3290	-0,2896	-0,2707	0,3482	0,3426	0,2160	-0,1079	-0,1327	0,2404	-0,5718
RS	0,1588	-0,2814	-0,5243	-0,3106	-0,2195	-0,2605	0,1147	-0,0967	-0,0556	0,1010
SP	0,1561	0,0109	0,0556	-0,1974	-0,1932	0,4116	0,1061	0,4249	0,2582	0,1194
PR	0,2683	0,1038	0,1266	0,2755	-0,5741	-0,5370	-0,2532	0,1263	0,0611	-0,2844
FL	0,1335	0,1048	-0,0047	0,2460	-0,1240	-0,0273	0,2951	-0,5606	0,5120	0,3744
TE	0,3372	0,2586	0,1839	-0,4058	-0,1264	0,2453	-0,4281	-0,3895	-0,0820	-0,0552
CI	0,2594	0,7637	-0,4603	0,1162	0,0659	0,0835	0,2065	0,1846	-0,1114	-0,0455
FE	0,3401	-0,0375	0,0190	0,0957	0,4808	-0,2968	-0,1719	-0,0588	-0,2833	0,3876
OP	0,2805	-0,0383	0,1061	-0,2648	0,0393	0,1514	-0,1918	0,0132	0,2056	-0,0585

Tabela B.15: Proporção de variância explicada na ACP global do Grupo de 35 a 44 anos

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Desvio Padrão	12,7018	7,2292	5,4069	4,4849	3,7510	3,6568	3,2216	2,9559	2,8445	2,5277
Propor. variância	0,4665	0,1511	0,0845	0,0582	0,0407	0,0387	0,0300	0,0253	0,0234	0,0185
Propor. acumulada	0,4665	0,6176	0,7021	0,7603	0,8010	0,8396	0,8697	0,8949	0,9183	0,9368

Tabela B.16: Coeficientes das CPs do grupo de 35 a 44 anos

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
CM	0,3025	-0,1311	0,0649	-0,2718	0,2090	-0,4554	0,0251	-0,0828	0,1192	-0,4523
AS	0,2268	-0,1782	-0,1098	0,0474	0,1127	-0,3750	0,1805	-0,0111	-0,5758	-0,0698
AC	0,3497	-0,0376	-0,3121	0,1792	0,1461	0,4271	0,2109	0,3851	-0,1336	-0,3532
AR	0,2555	-0,0799	-0,0782	0,0232	0,1333	0,2553	0,2224	-0,4784	0,1722	-0,2261
IND	0,1494	-0,0346	-0,1792	0,0508	-0,0103	-0,1328	0,2847	-0,1813	-0,3988	0,5557
EM	0,1555	-0,2092	0,5535	0,1542	-0,1602	0,0439	0,1043	0,0049	-0,0753	-0,0408
RI	0,3394	-0,2095	0,1069	-0,5994	-0,1519	0,0051	-0,3762	0,1428	-0,0647	0,0843
RS	0,1569	-0,1771	0,6210	0,2137	-0,0825	0,1507	0,1756	0,1281	-0,0471	0,0741
SP	0,1928	0,0376	0,0061	0,3183	-0,1551	-0,0868	-0,3314	-0,6255	0,0238	-0,0718
PR	0,2904	0,0814	0,1076	0,2638	0,7288	-0,0654	-0,3046	0,1314	0,2062	0,3355
FL	0,2013	0,0858	-0,0485	-0,0868	-0,1418	-0,3338	0,5130	0,1050	0,5611	0,2369
TE	0,3099	0,1726	-0,1687	0,3295	-0,4213	-0,2217	-0,3521	0,3161	0,0140	-0,0000
CI	0,2196	0,8755	0,2482	-0,1969	-0,0077	0,0691	0,1100	-0,0621	-0,2112	-0,0866
FE	0,3099	-0,0915	-0,1226	-0,3066	-0,0557	0,4237	-0,0617	-0,1510	0,0394	0,3320
OP	0,2843	-0,0539	-0,1601	0,1999	-0,3047	0,0740	0,0544	-0,0117	0,1769	0,0067

Tabela B.17: Proporção de variância explicada na ACP global do grupo de 45 a 54

anos

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Desvio Padrão	12,7018	7,2292	5,4069	4,4849	3,7510	3,6568	3,2216	2,9559	2,8445	2,5277
Propor. variância	0,4665	0,1511	0,0845	0,0582	0,0407	0,0387	0,0300	0,0253	0,0234	0,0185
Propor. acumulada	0,4665	0,6176	0,7021	0,7603	0,8010	0,8396	0,8697	0,8949	0,9183	0,9368

Tabela B.18: Coeficientes das CPs do grupo de 45 a 54 anos

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
CM	0,2658	0,1598	-0,0576	0,2755	-0,4490	0,0214	-0,1918	0,0632	-0,1073	0,5406
AS	0,2245	0,1904	0,3019	-0,0617	-0,2663	0,2001	-0,1006	0,3288	0,0910	-0,5815
AC	0,3226	-0,0184	0,1961	-0,1946	0,1032	0,1954	0,1557	-0,0611	0,1169	0,3681
AR	0,2526	0,1615	0,0929	0,1407	0,2027	-0,1093	0,5586	-0,3171	0,0928	0,0833
IND	0,1231	0,0273	0,3383	-0,0235	-0,0908	-0,0582	0,0564	-0,1063	0,7356	0,0249
EM	0,1551	0,2584	-0,4002	0,1370	0,3293	0,2404	-0,2919	-0,0281	0,1266	0,0900
RI	0,3156	0,2124	-0,3987	0,0325	-0,4776	-0,0003	0,0626	-0,0046	0,1124	-0,1037
RS	0,1866	0,1666	-0,2808	0,2186	0,3941	0,2252	-0,0765	-0,0780	0,1929	-0,2842
SP	0,1992	0,0199	0,1886	-0,1135	0,0812	0,5075	0,1725	0,3931	-0,1799	0,1490
PR	0,2845	0,0406	0,3348	0,6426	0,1710	-0,3608	-0,0377	0,1459	-0,2670	-0,1132
FL	0,1758	-0,1841	0,2506	0,0121	-0,1551	0,2492	-0,3977	-0,7128	-0,1718	-0,1390
TE	0,3091	-0,0967	0,0122	-0,3260	0,2039	-0,4779	-0,4757	0,2078	0,1912	0,1250
CI	0,3292	-0,8396	-0,2440	-0,1409	-0,0126	0,1063	0,1233	0,1253	0,0711	-0,0849
FE	0,3330	0,1008	-0,2363	-0,3621	-0,0971	-0,3206	0,2804	-0,1475	-0,2662	-0,2219
OP	0,2667	0,1049	0,1416	-0,3335	0,2541	0,0722	-0,0961	-0,0309	-0,3244	0,0224

Tabela B.19: Proporção de variância explicada na ACP global do grupo de 55 ou

mais anos

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Desvio Padrão	11,4979	7,2438	5,3480	5,1067	4,1035	3,7854	3,5245	3,0715	2,7154	2,4032
Propor. variância	0,4175	0,1657	0,0903	0,0824	0,0532	0,0452	0,0392	0,0298	0,0233	0,0182
Propor. acumulada	0,4175	0,5832	0,6735	0,7558	0,8090	0,8542	0,8935	0,9233	0,9466	0,9648

Tabela B.20: Coeficientes das CPs do grupo de 55 ou mais anos

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
CM	-0,0729	0,0057	-0,0518	0,4103	-0,3499	0,1588	-0,3980	-0,2224	0,0687	-0,1287
AS	-0,1411	0,3151	-0,2241	0,1665	-0,2762	0,2013	0,2512	0,6823	0,3595	-0,0125
AC	-0,4595	0,0420	0,1336	-0,3525	-0,5258	0,2218	0,0758	-0,0763	-0,2901	0,2356
AR	-0,2920	0,1080	0,4183	0,2143	0,3287	0,4556	0,2285	-0,1214	-0,0629	-0,0727
IND	-0,1671	0,1960	-0,1511	0,0490	-0,0363	-0,2394	0,5596	-0,4819	0,2463	0,0555
EM	-0,1589	0,0997	-0,2471	0,2673	0,4051	0,0807	-0,0266	-0,0066	-0,2324	0,0231
RI	-0,0447	0,2214	0,2231	0,5848	-0,2727	-0,2020	-0,1252	-0,1758	-0,0158	0,2304
RS	-0,1298	0,1960	-0,3050	0,1416	0,1945	-0,0684	-0,2866	0,0485	0,0979	0,0440
SP	-0,2026	0,1697	-0,1064	0,0381	-0,0309	0,0717	-0,0079	0,0699	-0,4848	-0,6267
PR	-0,3054	-0,0842	-0,4436	-0,0822	0,2042	0,3536	-0,1071	-0,1668	0,0942	0,3843
FL	-0,2649	-0,0139	-0,1004	0,1370	0,0802	-0,5402	0,2707	0,1934	-0,3409	0,1149
TE	-0,3536	0,1046	-0,0542	-0,2404	-0,0021	-0,2056	-0,1520	-0,2227	0,4401	-0,4959
CI	-0,4104	-0,7822	0,0449	0,2190	-0,0334	-0,1174	-0,0160	0,1692	0,1403	-0,0631
FE	-0,2394	0,1932	0,5518	-0,0745	0,2979	-0,1026	-0,1512	0,2097	0,2443	0,1062
OP	-0,2261	0,2352	-0,0076	-0,2438	0,0200	-0,2821	-0,4217	0,0808	-0,1425	0,2217

Apêndice C

Programas

O programa a seguir foi gerado no aplicativo R e corresponde à aplicação 1:

C.1 AFM da Aplicação 1

Chame o pacote `scatterplot3d` e salve em Rlib, `install.packages('scatterplot3d',lib='U:/Rlib')`

```
require(scatterplot3d)
require(xtable)
a<-read.csv("VEST97.csv",header = TRUE, dec=".",sep=";", fill = TRUE)
class(a)
Extraer as variáveis quantitativas de interesse
notas<-a[,44:52]
b<-a[,33:37]
dim(b)
```

```

socioeconom<-cbind(b[,1:5],a[,19])
socioeconom[1,]
dim(notas)
notas[1,]
class(notas)\newline
\texttt{Padronização da Base}\newline
dim(notas)
medk<-vector()
sk<-vector()
for(j in 1:9)(medk[j]<-mean(notas[,j]))
for(j in 1:9)(sk[j]<-sqrt(var(notas[,j])))
medk
sk
matrizX<-matrix(0,12093,9)
for (j in 1:9){
matrizX[,j]<-(notas[,j]-medk[j])/((sk[j])*sqrt(12093))
}
matrizX
dim(matrizX)
notasst<-cbind(matrizX[,1:8])
notasst[1,]
colnames(notasst)<-c("Portugues","Matematica","Geografia","Historia",
"InglesFra","Biologia","Fisica","Quimica")
dim(notasst)
princomp(cor(notasst))
prcomp(notasst)

```

```

summary(prcomp(notasst))
library('xtable',lib='U:/Rlib')
xtable(summary(prcomp(notasst)))
xtable(prcomp(notasst))
\subsection{Calculo das CP dos indivíduos}
evals<-eigen(cor(notasst))\$values
evect<-eigen(cor(notasst))\$vectors
evals
evect
diag<-diag(evals)
diag
G<-evect\%*\%diag\%*\%t(evect)(\#verifique que se obtem matriz de covariancias\#)
CP<-t(evect)\%*\%t(notasst)
dim(CP)
ComPrint<-CP
ComPrint
dim(ComPrint)
class(notasst)
class(ComPrint)
ComPrincipal<-as.data.frame(ComPrint)\#não precisa salvo que queira usar Splus\#
x<-ComPrint[1,1:10000]
y<-ComPrint[2,1:10000]
z<-ComPrint[3,1:10000]
Existem comandos no R; "prcom" e "princomp", que extraem as CP de uma tabela
de dados, ou da matriz de correlação, é uma forma alternativa de obter os resultados.
Mais nos optamos por a alternativa matricial, para desenvolver os resultados.

```

```

\subsection{Gráficas das CP}
\subsubsection{Para os indivíduos}
par(mfrow=c(2,3))
scatterplot3d(x,y,z, highlight.3d=TRUE,
               col.axis="blue", col.grid="lightblue",type="p",xlab="CP1",ylab="CP2",
               zlab="CP3", pch=20)
scatterplot3d(x,z,y, highlight.3d=TRUE,
               col.axis="blue", col.grid="lightblue",type="p",xlab="CP1",ylab="CP3",
               zlab="CP2", pch=20)
scatterplot3d(z,y,x, highlight.3d=TRUE,
               col.axis="blue", col.grid="lightblue",type="p",xlab="CP3",ylab="CP2",
               zlab="CP1",pch=20)
plot(x,y,main="Plano das Componentes 1 e 2",col.axis="blue",
     col.grid="lightblue",type="n",xlab="CP2",ylab="CP1",cex=1.5,pch=20)
points(x,y,lty=2,col="red",pch=20)
plot(x,z,main="Plano das Componentes 1 e 3",col.axis="blue",
     col.grid="lightblue",type="n",xlab="CP2",ylab="CP1",cex=1.5,pch=20)
points(x,z,lty=2,col="red",pch=20)
plot(z,y,main="Plano das Componentes 3 e 2",col.axis="blue",
     col.grid="lightblue",type="n",xlab="CP2",ylab="CP1",cex=1.5,pch=20)
points(z,y,lty=2,col="red",pch=20)
\subsubsection{Para as variáveis} \#calculo das CP\#\#das variáveis\#
phi<-(notasst)\%*\%(evect)
dim(phi)
phi1<-evect[,1]*sqrt(evals[1])
phi2<-evect[,2]*sqrt(evals[2])

```



```

phi3<-evect[,3]*sqrt(evals[3])
phi4<-evect[,4]*sqrt(evals[4])
op <- par(bg = "light blue")
par(op)
par(mfrow=c(1,1))
plot(c(-0.85,-0.65),c(-0.9,0.8), bg=par("bg"),main="Plano da
1ar e 2da Componente Principal",
col.axis="blue", type="n", xlab="CP2", ylab="CP1",cex=1.5,pch=2,lwd=2)
abline(0,0,0,0,lty=2,col="black",lwd=1)
points(phi1,phi2,lty=2,col="red",pch=17:25,cex=3)
lines(phi1,phi2,lty=3,col='blue')
text(phi1[1],phi2[1],labels=c("Mat"),col="blue",font=9,pos="3")
text(phi1[2],phi2[2],labels=c("Bio"),col="blue",font=9,pos="3")
text(phi1[3],phi2[3],labels=c("Fis"),col="blue",font=9,pos="3")
text(phi1[4],phi2[4],labels=c("Qui"),col="blue",font=9,pos="3")
text(phi1[5],phi2[5],labels=c("Port"),col="blue",font=9,pos="3")
text(phi1[6],phi2[6],labels=c("Geo"),col="blue",font=9,pos="3")
text(phi1[7],phi2[7],labels=c("His"),col="blue",font=9,pos="3")
text(phi1[8],phi2[8],labels=c("Ing-Fra"),col="blue",font=9,pos="3")
op <- par(bg = "light blue")
plot(c(-0.5,0.65),c(-0.9,0.8), bg=par("bg"),main="Plano da 3ra e
4ta Componente Principal",
col.axis="blue", type="n", xlab="CP3", ylab="CP4",cex=1.5,pch=2,lwd=2)
abline(0,0,0,0,lty=2,col="black",lwd=1)
points(phi3,phi4,lty=2,col="red",pch=17:25,cex=3)
lines(phi3,phi4,lty=3,col='blue')

```

```

par(op)\
text(phi3[1],phi4[1],labels=c("Mat"),col="blue",font=9,pos="3")
text(phi3[2],phi4[2],labels=c("Bio"),col="blue",font=9,pos="3")
text(phi3[3],phi4[3],labels=c("Fis"),col="blue",font=9,pos="3")
text(phi3[4],phi4[4],labels=c("Qui"),col="blue",font=9,pos="3")
text(phi3[5],phi4[5],labels=c("Port"),col="blue",font=9,pos="3")
text(phi3[6],phi4[6],labels=c("Geo"),col="blue",font=9,pos="3")
text(phi3[7],phi4[7],labels=c("His"),col="blue",font=9,pos="3")
text(phi3[8],phi4[8],labels=c("Ing-Fra"),col="blue",font=9,pos="3")

\section{Programa para a AFM}

\subsection{Divisão da Base nas Sub bases do Grupo de variáveis}

dados1<-cbind(notasst[,2],notasst[,6:8])
colnames(dados1)<-c("Matematica","Biologia","Fisica","Quimica")
dados2<-cbind(notasst[,1],notasst[,3:4],notasst[,5])
colnames(dados2)<-c("Portugues","Geografia","Historia","IngFranc")
dados3<-socioeconom
dim(dados1)
dim(dados2)
dim(dados3)

\section{Resultados em cada grupo de variáveis}
princomp(cor(dados1))
summary(prcomp(dados1))
xtable(summary(prcomp(dados1)))\#importância da CP\#
xtable(prcomp(dados1))
xtable(cor(dados1))

```

```

princomp(cor(dados2))
summary(prcomp(dados2))
xtable(summary(prcomp(dados2)))\#importância da CP\#
xtable(prcomp(dados2))
xtable(cor(dados2))
evals<-eigen(cor(dados1))\$values
evals
evals<-eigen(cor(dados2))\$values
evals
\section{Análise da CP em cada Grupo ponderados}
\#dividimos a bases por o maior autovalor da matriz Cov de cada grupo\#
evals<-eigen(cov(dados1))\$values
evals
novodados1<-dados1/sqrt(evals[1])
novodados1
prcomp(novodados1)
xtable(prcomp(novodados1))
evals<-eigen(cov(dados2))\$values
novodados2<-dados2/sqrt(evals[1])
prcomp(novodados2)
xtable(prcomp(novodados2))
colnames(novodados1)<-c("Matematica","Biologia","Fisica","Quimica")
colnames(novodados2)<-c("Portugues","Geografia","Historia","IngFranc")
xtable(prcomp(novodados1))
xtable(prcomp(novodados2))
xtable(summary(prcomp(novodados1)))

```

```

xtable(summary(prcomp(novodados2)))
summary(prcomp(novodados1))
summary(prcomp(novodados2))
prcomp(novodados1)
prcomp(novodados2)
\section{As CP dos individuos em cada Grupo}
princomp(cor(dados1))
prcomp(dados1)
summary(prcomp(dados1))\#importância de cada CP#
xtable(summary(prcomp(dados1)))\#importância de cada CP#
xtable(prcomp(dados1))
xtable(cor(dados1))
xtable(cor(dados2))
princomp(cor(dados2))
prcomp(dados2)
summary(prcomp(dados2))\#importância de cada CP\#
xtable(summary(prcomp(dados2)))\#importância de cada CP\#
xtable(prcomp(dados2))
xtable(cor(dados2))
evals<-eigen(cor(dados1))\ $values
evect<-eigen(cor(dados1))\ $vectors
evals
evect
diag<-diag(evals)
diag
G<-evect\%*\%diag\%*\%t(evect)

```

G

```
CP<-t(evect)\%*\%t(dados1)
dim(CP)
ComPrintdeGrup1<-CP
dim(ComPrint)
CPG1<-ComPrintdeGrup1[1,1:12000]
CPG2<-ComPrintdeGrup1[2,1:12000]
CPG3<-ComPrintdeGrup1[3,1:12000]
par(mfrow=c(2,3))
scatterplot3d(CPG1,CPG3,CPG2, highlight.3d=TRUE,
              col.axis="blue", col.grid="lightblue",type="p",xlab="CP1",ylab="CP3",
              zlab="CP2", pch=20)
scatterplot3d(CPG1,CPG2,CPG3, highlight.3d=TRUE,
              col.axis="blue", col.grid="lightblue",type="p",xlab="CP1",ylab="CP2",
              zlab="CP3", pch=20)\
scatterplot3d(CPG3,CPG2,CPG1, highlight.3d=TRUE,
              col.axis="blue", col.grid="lightblue",type="p",xlab="CP3",ylab="CP2",
              zlab="CP1",pch=20)
plot(CPG1,CPG2,main="Plano das Componentes 1 e 2",col.axis="blue",
     col.grid="lightblue",type="n",xlab="CP2",ylab="CP1",cex=1.5,pch=20)
points(CPG1,CPG2,lty=2,col="red",pch=20)
plot(CPG3,CPG2,main="Plano das Componentes 3 e 2",col.axis="blue",
     col.grid="lightblue",type="n",xlab="CP2",ylab="CP1",cex=1.5,pch=20)
points(CPG3,CPG2,lty=2,col="red",pch=20)
plot(CPG1,CPG3,main="Plano das Componentes 1 e 3",col.axis="blue",
     col.grid="lightblue",type="n",xlab="CP2",ylab="CP1",cex=1.5,pch=20)
```

```

points(CPG1,CPG3,lty=2,col="red",pch=20)

\section{AFM Global - Correlaciones con os eixos Principais }

basefin<-cbind(novodados1,novodados2)

colnames(basefin)<-c("Matematica","Biologia","Fisica","Quimica",
"Portugues","Geografia","Historia","IngFranc")

summary(prcomp(basefin))

prcomp(basefin)

xtable(prcomp(basefin))

xtable(summary(prcomp(basefin)))

prcomp(basefin)


\section{AFM da Aplicação 2}


#Chame o pacote scatterplot3d e salve em Rlib#
install.packages('scatterplot3d',lib='U:/Rlib')
require(scatterplot3d)
require(xtable)

##Programas para adequar e salvar bases##

colnames(basefinal)<-c("v1","v2","v3","v4","v5","v6","v7","v8",
"v9","v10","v11","v12","v13","v14")

basefinal<-round((baseponderada),3)

a<-read.csv("IEnelly3.csv",header = TRUE, dec=".",sep=";", fill = TRUE)

class(a)

summary(a)

a<-a[a[,32]=="1",]### escolhemos a amostra ####

dim(a)

```

```

sink("U:/IEnelly1996.txt")
sink()
list(names(a))
colnames(a)<-c("Suj", "Edad", "Sex", "Gedad", "Gedad2", "Gedad3", "GedadH",
"GedadInei", "IPCI", "INCI", "CM", "AS", "AC", "AR", "IND", "EM", "RI", "RS", "SP",
"PR", "FL", "TE", "CI", "FE", "OP", "IA", "IE", "AD", "ME", "AG", "CE", "Muestra")
### Escolher as variaveis quantitativas de interesse ###
notas<-a[,9:31]
princomp(cor(notas))
summary(princomp(cor(notas)))
prcomp(notas)
list(names(notas))
#Padronização da Base#
dim(notas)
medk<-vector()
sk<-vector()
for(j in 1:23)(medk[j]<-mean(notas[,j]))
for(j in 1:23)(sk[j]<-sqrt(var(notas[,j])))
medk
sk
matrizX<-matrix(0,1996,23)
for (j in 1:23){
matrizX[,j]<-(notas[,j]-medk[j])/((sk[j])*sqrt(1996))
}
matrizX
dim(matrizX)

```

```

notasst<-cbind(matrizX[,1:23])
notasst[1,]
#Standarizacion de la Base a#
a1<-a[,11:25]
dim(a1)
media<-vector()
sd<-vector()
for(j in 1:15)(media[j]<-mean(a1[,j]))
for(j in 1:15)(sd[j]<-sqrt(var(a1[,j])))
media
sd
matrizX<-matrix(0,1996,15)
for (j in 1:15){
matrizX[,j]<-(a[,j]-media[j])/((sd[j])*sqrt(1996))}
matrizX[1912,]
dim(matrizX)
astan<-cbind(matrizX[,1:15])
astan[1,]

##### dividimos a base por individuos #####
list(names(a))
aidade<-cbind(a[,2:3],a[,8],a[,11:25])
colnames(aidade)<-c("Edad","Sex","GedadInei","CM","AS","AC","AR","IND","EM",
"RI","RS","SP","PR","FL","TE","CI","FE","OP")
aidade[1,]
dim(aidade)

```



```

a1<-aidade[aidade[,3]=="1",]
dim(a1)
a2<-aidade[aidade[,3]=="2",]
dim(a2)
a3<-aidade[aidade[,3]=="3",]
dim(a3)
a4<-aidade[aidade[,3]=="4",]
dim(a4)
a5<-aidade[aidade[,3]=="5",]
dim(a5)

list(names(a1))
colnames(a1)<-c("Edad","Sex","GedadInei","CM","AS","AC","AR",
"IND","EM","RI","RS","SP","PR","FL","TE","CI","FE","OP")
colnames(a2)<-c("Edad","Sex","GedadInei","CM","AS","AC","AR",
"IND","EM","RI","RS","SP","PR","FL","TE","CI","FE","OP")
colnames(a3)<-c("Edad","Sex","GedadInei","CM","AS","AC","AR",
"IND","EM","RI","RS","SP","PR","FL","TE","CI","FE","OP")
colnames(a4)<-c("Edad","Sex","GedadInei","CM","AS","AC","AR",
"IND","EM","RI","RS","SP","PR","FL","TE","CI","FE","OP")
colnames(a5)<-c("Edad","Sex","GedadInei","CM","AS","AC","AR",
"IND","EM","RI","RS","SP","PR","FL","TE","CI","FE","OP")
class(a1)

##### Deve-se ter como Matriz #####
a1<-as.matrix(a1)

```

```

a2<-as.matrix(a2)
a3<-as.matrix(a3)
a4<-as.matrix(a4)
a5<-as.matrix(a5)
class(a1)

#### Análise de CP em cada Grupo sem ponderar ####
#### escolhemos as variaveis de interesse de cada sub tabela ####
vara1<-cbind(a1[,4:18])
vara2<-cbind(a2[,4:18])
vara3<-cbind(a3[,4:18])
vara4<-cbind(a4[,4:18])
vara5<-cbind(a5[,4:18])
xtable(summary(prcomp(vara1)))#importância da CP#
xtable(prcomp(vara1))
xtable(summary(prcomp(vara2)))#importância da CP#
xtable(prcomp(vara2))
xtable(summary(prcomp(vara3)))#importância da CP#
xtable(prcomp(vara3))
xtable(summary(prcomp(vara4)))#importância da CP#
xtable(prcomp(vara4))
xtable(summary(prcomp(vara5)))#importância da CP#
xtable(prcomp(vara5))

##### Análise em cada grupo ACP parcial #####
aidade<-cbind(a[,2:3],a[,8],astan[,1:15])

```

```
colnames(aidade)<-c("Edad","Sex","GedadInei","CM","AS","AC",
"AR","IND","EM","RI","RS","SP","PR","FL","TE","CI","FE","OP")
aidade[1,]
dim(aidade)
```

```
#### grupos de variáveis #####
```

```
vara11<-cbind(a1[,4:8])
vara21<-cbind(a2[,9:11])
vara31<-cbind(a3[,12:14])
vara41<-cbind(a4[,15:16])
vara51<-cbind(a5[,17:18])
```

```
xtable(summary(prcomp(vara11)))#importância da CP#
xtable(prcomp(vara11))
xtable(summary(prcomp(vara21)))#importância da CP#
xtable(prcomp(vara21))
xtable(summary(prcomp(vara31)))#importância da CP#
xtable(prcomp(vara31))
xtable(summary(prcomp(vara41)))#importância da CP#
xtable(prcomp(vara41))
xtable(summary(prcomp(vara51)))#importância da CP#
xtable(prcomp(vara51))
```

```
## Análise Descritivo###
```

```
for(i in 1:1996)(
hist(a[,"INCI"], labels = TRUE))
```

```

par(mfrow=c(2,3))
hist(a[,9], labels = TRUE,main="Histograma IPCI")
hist(a[,10], labels = TRUE,main="Histograma INCI")
hist(a[,11], labels = TRUE,main="Histograma CM")
hist(a[,12], labels = TRUE,main="Histograma AS")
hist(a[,13], labels = TRUE,main="Histograma AC")
hist(a[,14], labels = TRUE,main="Histograma AR")
par(mfrow=c(2,3))
hist(a[,15], labels = TRUE,main="Histograma INC")
hist(a[,16], labels = TRUE,main="Histograma EM")
hist(a[,17], labels = TRUE,main="Histograma RI")
hist(a[,18], labels = TRUE,main="Histograma RS")
hist(a[,19], labels = TRUE,main="Histograma SP")
hist(a[,20], labels = TRUE,main="Histograma PR")
par(mfrow=c(2,3))
hist(a[,21], labels = TRUE,main="Histograma FL")
hist(a[,22], labels = TRUE,main="Histograma TE")
hist(a[,23], labels = TRUE,main="Histograma CI")
hist(a[,24], labels = TRUE,main="Histograma FE")
hist(a[,25], labels = TRUE,main="Histograma OP")
hist(a[,26], labels = TRUE,main="Histograma IA")
par(mfrow=c(2,3))
hist(a[,27], labels = TRUE,main="Histograma IE")
hist(a[,28], labels = TRUE,main="Histograma AD")
hist(a[,29], labels = TRUE,main="Histograma ME")
hist(a[,30], labels = TRUE,main="Histograma AG")

```

```

hist(a[,31], labels = TRUE,main="Histograma CE")
hist(a[,8], labels = TRUE,main="Histograma EDADINEI")

### Extraemos as variáveis quantitativas de interesse ###
notas<-a[,9:31]
princomp(cor(notas))
summary(princomp(cor(notas)))
prcomp(notas)

#Padronização da Base#
dim(notas)
medk<-vector()
sk<-vector()
for(j in 1:23)(medk[j]<-mean(notas[,j]))
for(j in 1:23)(sk[j]<-sqrt(var(notas[,j])))
medk
sk
matrizX<-matrix(0,1996,23)
for (j in 1:23){
matrizX[,j]<-(notas[,j]-medk[j])/((sk[j])*sqrt(1996))}
matrizX
dim(matrizX)
notasst<-cbind(matrizX[,1:23])
notasst[1,]

##### dividimos a base por individuos #####

```

```

list(names(a))
asex<-cbind(a[,2:3],a[,8],a[,11:25])
colnames(asex)<-c("Edad","Sex","GedadInei","CM","AS","AC","AR","IND",
"EM","RI","RS","SP","PR","FL","TE","CI","FE","OP")
dim(asex)
am<-asex[asex[,2]=="1",]
dim(am)
af<-asex[asex[,2]=="2",]
dim(af)
list(names(af))
colnames(af)<-c("Edad","Sex","GedadInei","CM","AS","AC","AR","IND",
"EM","RI","RS","SP","PR","FL","TE","CI","FE","OP")
colnames(am)<-c("Edad","Sex","GedadInei","CM","AS","AC","AR","IND",
"EM","RI","RS","SP","PR","FL","TE","CI","FE","OP")

##### Salvamos como Matriz #####
am<-as.matrix(am)
af<-as.matrix(af)
class(am)

#### Análise de CP em cada Grupo sem ponderar ####
#### escolher as variáveis de interesse de cada sub-tabela ####
varam<-cbind(am[,4:18])
varaf<-cbind(af[,4:18])
princomp(cov(varam))#ihombres#
princomp(cov(varaf))#mujeres#

```

```

prcomp(varam)
prcomp(varaf)
summary(prcomp(varam))
summary(prcomp(varaf))
xtable(summary(prcomp(varam)))#importância da CP#
xtable(prcomp(varam))
xtable(cor(varam))
xtable(summary(prcomp(varaf)))#importância da CP#
xtable(prcomp(varaf))
xtable(cor(varaf))

##### Padronizamos por Raiz de lamda1 #####

evals<-eigen(cov(varam))$values
evals
am1<-varam/sqrt(evals[1])
prcomp(am1)
evals<-eigen(cov(varaf))$values
evals
af1<-varaf/sqrt(evals[1])
prcomp(af1)
dim(t(am1))
dim(t(af1))
transA<-cbind(t(am1),t(af1))
apadronizada<-t(cbind(t(am1),t(af1)))
dim(apadronizada)

```

```

colnames(apadronizada)<-c("CM","AS","AC","AR","IND","EM","RI","RS",
"SP","PR","FL","TE","CI","FE","OP")
list(names(apadronizada))
class(apadronizada)
apadronizada[1,]
sink("U:/IEnelly1996.txt")
basefinal
sink()

```

```

##### Calculo das CP Globais #####

```

```

evals<-eigen(cor(apadronizada))$values
evect<-eigen(cor(apadronizada))$vectors
evals
evect
diag<-diag(evals)
diag
G<-evect%*%diag%*%t(evect)
G
dim(G)
CP<-evect%*%t(apadronizada)
dim(CP)
CP
ComPrint<-CP
x<-ComPrint[1,1:1996]
x

```



```

y<-ComPrint[2,1:1996]

y
z<-ComPrint[3,1:1996]

z

### Calculo das CP na tabela conjunta de variáveis #####
evals
phi1<-evect[,1]*sqrt(evals[1])
phi1
phi2<-evect[,2]*sqrt(evals[2])
phi2
phi3<-evect[,3]*sqrt(evals[3])
phi3
phi4<-evect[,4]*sqrt(evals[4])
phi4
x<--phi1
y<-phi2
z<-phi3
w<-phi4
op <- par(bg = "light blue")
par(op)
par(mfrow=c(1,1))
plot(x,y,main="Plano da 1ar e 2da CP",col.axis="blue", col.grid="lightblue",type='l')
points(x,y,lty=2,col="red",pch=20)
text(x[1],y[1],labels=c("CM"),col="blue",font=9,pos="3")
text(x[2],y[2],labels=c("AS"),col="blue",font=9,pos="3")

```

```

text(x[3],y[3],labels=c("AC"),col="blue",font=9,pos="3")
text(x[4],y[4],labels=c("AR"),col="blue",font=9,pos="3")
text(x[5],y[5],labels=c("IND"),col="blue",font=9,pos="3")
text(x[6],y[6],labels=c("EM"),col="blue",font=9,pos="3")
text(x[7],y[7],labels=c("RI"),col="blue",font=9,pos="3")
text(x[8],y[8],labels=c("RS"),col="blue",font=9,pos="3")
text(x[9],y[9],labels=c("SP"),col="blue",font=9,pos="3")
text(x[10],y[10],labels=c("PR"),col="blue",font=9,pos="3")
text(x[11],y[11],labels=c("FL"),col="blue",font=9,pos="3")
text(x[12],y[12],labels=c("TE"),col="blue",font=9,pos="3")
text(x[13],y[13],labels=c("CI"),col="blue",font=9,pos="3")
text(x[14],y[14],labels=c("FE"),col="blue",font=9,pos="3")
text(x[15],y[15],labels=c("OP"),col="blue",font=9,pos="3")
par(op)
par(mfrow=c(1,1))
plot(z,w,main="Plano da 3ar e 4ta CP",col.axis="blue", col.grid="lightblue",type="
points(z,w,lty=2,col="red",pch=20)
text(z[1],w[1],labels=c("CM"),col="blue",font=9,pos="3")
text(z[2],w[2],labels=c("AS"),col="blue",font=9,pos="3")
text(z[3],w[3],labels=c("AC"),col="blue",font=9,pos="3")
text(z[4],w[4],labels=c("AR"),col="blue",font=9,pos="3")
text(z[5],w[5],labels=c("IND"),col="blue",font=9,pos="3")
text(z[6],w[6],labels=c("EM"),col="blue",font=9,pos="3")
text(z[7],w[7],labels=c("RI"),col="blue",font=9,pos="3")
text(z[8],w[8],labels=c("RS"),col="blue",font=9,pos="3")
text(z[9],w[9],labels=c("SP"),col="blue",font=9,pos="3")

```

```

text(z[10],w[10],labels=c("PR"),col="blue",font=9,pos="3")
text(z[11],w[11],labels=c("FL"),col="blue",font=9,pos="3")
text(z[12],w[12],labels=c("TE"),col="blue",font=9,pos="3")
text(z[13],w[13],labels=c("CI"),col="blue",font=9,pos="3")
text(z[14],w[14],labels=c("FE"),col="blue",font=9,pos="3")
text(z[15],w[15],labels=c("OP"),col="blue",font=9,pos="3")

dados1<-cbind(notasst[,3:7])
dados2<-cbind(notasst[,8:10])
dados3<-cbind(notasst[,11:13])
dados4<-cbind(notasst[,14:15])
dados5<-cbind(notasst[,16:17])

colnames(dados1)<-c("CM","AS","AC","AR","IND")
colnames(dados2)<-c("EM","RI","RS")
colnames(dados3)<-c("SP","PR","FL")
colnames(dados4)<-c("TE","CI")
colnames(dados5)<-c("FE","OP")

```

#Análise de CP em cada Grupo sem ponderar#

```

princomp(cor(dados1))#interpessoal#
princomp(cor(dados2))#intrapessoal#
princomp(cor(dados3))#adaptabilidade#
princomp(cor(dados4))#Manejo de estresse#
princomp(cor(dados5))#adaptabilidade no geral#

prcomp(dados1)
prcomp(dados2)

```

```

prcomp(dados3)
prcomp(dados4)
prcomp(dados5)
summary(prcomp(dados1))#importância de cada CP#
summary(prcomp(dados2))#importância de cada CP#
summary(prcomp(dados3))#importância de cada CP#
summary(prcomp(dados4))#importância de cada CP#
summary(prcomp(dados5))

##### Padronizamos por Raiz de lamda1 #####

evals<-eigen(cov(dados1))$values
novodados1<-dados1/sqrt(evals[1])
prcomp(novodados1)
evals<-eigen(cov(dados2))$values
novodados2<-dados2/sqrt(evals[1])
prcomp(novodados2)
evals<-eigen(cov(dados3))$values
novodados3<-dados3/sqrt(evals[1])
prcomp(novodados3)
evals<-eigen(cov(dados4))$values
novodados4<-dados4/sqrt(evals[1])
prcomp(novodados4)
evals<-eigen(cov(dados5))$values
novodados5<-dados5/sqrt(evals[1])
prcomp(novodados5)

```

```

# Calculo das CP FINAL #####
dadosfinal<-cbind(novodados1,novodados2,novodados3,novodados4,novodados5)
evals<-eigen(cor(dadosfinal))$values
evect<-eigen(cor(dadosfinal))$vectors
evals
evect
diag<-diag(evals)
diag
G<-evect%*%diag%*%t(evect)
G
CP<-evect%*%t(dadosfinal)
dim(CP)
CP
ComPrint<-CP
x<-ComPrint[1,1:2249]
x
y<-ComPrint[2,1:2249]
y
z<-ComPrint[3,1:2249]
z

#calculo das CPs das variáveis (final) #####
phi<-(dadosfinal)%*%(evect)
dim(phi)
x<-evect[,1]*sqrt(evals[1])

```

```
x
y<-evect[,2]*sqrt(evals[2])
y
z<-evect[,3]*sqrt(evals[3])
z
w<-evect[,4]*sqrt(evals[4])
w
summary(prcomp(dadosfinal))
xtable(summary(prcomp(dadosfinal)))
```

Referências Bibliográficas

- BarOn, R. (1996). *BarOn Emotional Quotient Inventory*, Multi-Health Systems, New York.
- BarOn, R. (1997). *Development of the BarOn EQ-i: A measure of emotional and social intelligence.*
- Barroso, L. e Artes, R. (2003). Tópicos de análise multivariada.
- Beltrami, E. (1873). Sulle funzioni bilineari, *Giorn, Math, Bataglin* **11**: 98–106.
- Carroll, J. D. (1968). A generalization of canonical correlation analysis to thress or more sets of variables., *Proceedings of the 76th Convention of the American Psychological Association* pp. 277–228.
- Carroll, J. D. e Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young' descomposition, *Psycometrika* **35**: 283–319.
- Eckart, C. e Young, G. (1936). The approximation of one matrix by another of lower rank, *Psychometrika* **1**: 211–218.
- Eckart, C. e Young, G. (1939). A principal axis transformation for non-hermitian matrices, *Bull.Amer.Math.Assoc.* **45**: 118–121.

- Escofier, B. e Pagès, J. (1982). Análisis factoriales simples y multiples. objetivos, métodos e interpretación, *Int. J. Soc. Psychiatry* **2**: 256–260.
- Escofier, B. e Pagès, J. (1986). Le traitement des variables qualitatives et des tableaux mistes par afm, *Data Analysis and Informatics, North Holland - Amsterdam* **V**: 179–191.
- Escofier, B. e Pagès, J. (1990). *Anályses factorielles simples et multiples*, DUNOD, París.
- Escofier, B. e Pagès, J. (1994). Multiple factor analysis - afmult package, *Computational Statistics and Data Analysis* **18**: 121–140.
- Gabriel, K. (1971). The biplot graphic display of matrices with application to principal component analysis, *Biometrika* **58**: 453–467.
- Gifi, A. (1990). *Non Linear Multivariate Analysis*, J. Wiley, Chichester.
- Gower, J. (1966). Some distance properties of latent and vector methods used in multivariate analysis., *Biometrika* **53**: 325–328.
- Guinot, C., Latreille, J. e Tenenhaus, M. (2001). Pls path modelling and multiple table analysis. application to the cosmetic habits of women in ile-de -france., *Chemometrics Intell* **58**: 247–259.
- Horst, P. (1961). Relations among m sets of variaveis., *Psychometrika* **26**: 126–149.
- Horst, P. (1965). Factor analysis of data matrices., *Holt Rinehart and Winston New York* .
- Hotelling, H. (1933). Analysis of complex of statistical variables with principal components, *Journal of Educational Psychology* **24**.

- Jobson, J. (1989). *Applied Multivariate Data Analysis, Volume II : Categorical and multivariate Methods*, Academic Press, London.
- Jordan, C., Vinzi, E. V., Chatelin, Y. M. e Lauro, C. (1874). Mémoire sur les formes bilinéaires, *J.Math. Pures et Appliquées* **19**: 35–54.
- Lavit, J. (1989). Statis-4: analyse conjointe de plusieurs multi-tableaux, *Laboratoire de Physique Industrielle et Traitement de l'Information, Faculte de Pharmacie* <http://www.pharma.univ-montp1.fr/agrostat/resume/vivien.pdf>.
- Lohmöller, J. (1989). Latent variables path modeling with partial least squares, *Physica Verlag Heildelberg*.
- Mardia, K. V., Kent, J. e Bibby, J. (1989). *Multivariate Analysis*, Academic Press, London.
- Pagès, J. (1996). Eléments ded comparaison entre l'analyse factorielle multiple et la méthodee statis, *Revue de Statistique Appliquée* **4**: 81–95.
- Pagès, J. e Tenenhaus, M. (2001). Multiple factor analysis combined with pls path modelling. appliation to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgements, *Chemometrics Intell. Lab. Systems* **58**: 261–273.
- Pearson, K. (1901). Principal component analysis, *Edinburgh and Philosophical Magazine and Journal* **6**: 2.
- Piron, M., Lebart, L. e Morineau, A. (1995). *Anályses factorielles simples et multiples*, DUNOD,París.

Ugarriza, N. (2001). La evaluación de la inteligencia emocional a través del inventario de baron (i-ce) en una muestra de lima metropolitana, *Revista de la Facultad de Psicología, Universidad de Lima* 4: 129–160.