

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas – ICEX

Programa de Pós-graduação em Estatística

Nanéia Guimarães Borges

**PERFIL DOS ALUNOS DOS CURSOS DE GRADUAÇÃO NA MODALIDADE À
DISTÂNCIA DE UMA INSTITUIÇÃO DE ENSINO SUPERIOR DA REGIÃO
METROPOLITANA DE BELO HORIZONTE: APLICAÇÃO DA ANÁLISE DE
CORRESPONDÊNCIA MÚLTIPLA**

Belo Horizonte

2016

Nanéia Guimarães Borges

**PERFIL DOS ALUNOS DOS CURSOS DE GRADUAÇÃO NA MODALIDADE À
DISTÂNCIA DE UMA INSTITUIÇÃO DE ENSINO SUPERIOR DA REGIÃO
METROPOLITANA DE BELO HORIZONTE: APLICAÇÃO DA ANÁLISE DE
CORRESPONDÊNCIA MÚLTIPLA**

Monografia apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Especialista em Estatística.

Orientador (a): Dra. Sueli Aparecida Mingoti.

Belo Horizonte

2016

2016, Nanéia Guimarães Borges.
Todos os direitos reservados.

Borges, Nanéia Guimarães

B732p Perfil dos alunos dos cursos de graduação na modalidade à distância de uma instituição de ensino superior da região metropolitana de Belo Horizonte [manuscrito]: aplicação da análise de correspondência múltipla / Nanéia Guimarães Borges —2016.
94f.. il.

Orientadora: Sueli Aparecida Mingoti.
Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.
Referências: 79-81 f.

1. Estatística. 2. Análise de correspondência múltipla 3. Educação à distância – Perfil – Aluno. 4. Variáveis categóricas I. Mingoti, Sueli Aparecida. II. Universidade Federal de Minas Gerais. I. Instituto de Ciências Exatas, Departamento de Estatística .III. Título.

CDU 519.2 (043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa CRB 6/1510
Universidade Federal de Minas Gerais – ICEX



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

ATA DO 150º. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DA ALUNA NANÉIA GUIMARÃES BORGES.

Aos onze dias do mês de julho de 2016, às 09:00 horas, na Sala 2076 do Instituto de Ciências Exatas, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso da aluna **Nanéia Guimarães Borges**, intitulado: **“Perfil dos Alunos dos Cursos de Graduação na Modalidade à Distância de Uma Instituição de Ensino Superior da Região Metropolitana de Belo Horizonte: Aplicação de Análise de Correspondência Múltipla”**, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, a Presidente da Comissão, Profa. Sueli Aparecida Mingoti - Orientadora, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra a candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Após a defesa, os membros da banca examinadora reuniram-se sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: a candidata foi considerada Aprovada condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente a candidata pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 11 de julho de 2016.


Profa. Sueli Aparecida Mingoti (Orientadora)
Departamento de Estatística / UFMG


Ela Mercedes Medrano de Toscano
Departamento de Estatística / UFMG


Profa. Ilka Afonso Reis
Departamento de Estatística / UFMG



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

DECLARAÇÃO DE CUMPRIMENTO DE REQUISITOS PARA CONCLUSÃO DO CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA.

Declaro para os devidos fins que Nanéia Guimarães Borges, número de registro 2012699671, cumpriu todos os requisitos necessários para conclusão do curso de Especialização em Estatística e que entregou para sua orientadora, a professora Sueli Aparecida Mingoti, que aprovou a versão final. O trabalho foi apresentado no dia 11 de julho de 2016 com o título “*Perfil dos Alunos dos Cursos de Graduação na Modalidade à Distância de Uma Instituição de Ensino Superior da Região Metropolitana de Belo Horizonte: Aplicação de Análise de Correspondência Múltipla*”.

Belo Horizonte, 19 de dezembro de 2022

Roberto da
Costa
Quinino: 8087
1291720

Assinado de forma
digital por Roberto da
Costa
Quinino: 80871291720
Data: 2022.12.19
11:00:00 -0200

Prof. Roberto da Costa Quinino
Coordenador da Comissão
do Curso de Especialização
em Estatística

Prof. Roberto da Costa Quinino

RESUMO

Este estudo teve como objetivo principal identificar o perfil dos alunos que estudam no ensino à distância, utilizando a técnica de Análise de Correspondência Múltipla para verificar as associações entre as variáveis observadas. Essa técnica multivariada permite analisar conjuntos grandes de dados categorizados e fazer análises individual e conjunta de todas as variáveis e categorias. Para formação do conjunto de dados desta pesquisa, utilizou-se os questionários respondidos por 204 alunos que cursaram o 2º período de uma graduação na modalidade de ensino à distância no ano de 2014. Esses dados foram tratados por meio do *software* R pacote *FactoMine* R. Através das análises e dos gráficos da ACM, foi possível identificar o perfil dos alunos e conhecer os fatores que influenciaram sua decisão na escolha por uma graduação na modalidade de ensino à distância. Com a pesquisa, constatou-se que o perfil do aluno do ensino à distância da instituição pesquisada é composto por um número maior de alunos do sexo feminino. Além disso, são alunos que trabalham e estudam, recebem até três salários mínimos, são casados e têm filhos. Portanto, conclui-se que a técnica de análise de correspondência múltipla facilita o tratamento e a interpretação dos dados categóricos, principalmente, quando aplicada em conjuntos de dados com grande número de variáveis e categorias.

Palavras-chave:

Análise de correspondência múltipla, Variáveis categóricas, Educação à distância, Perfil do aluno ingressante.

ABSTRACT

This study aimed to identify the profile of students studying in distance learning, using multiple correspondence analysis technique to verify the associations between the study variables. Applying this analysis technique in the database also enabled an individual and joint analysis of all categorical variables, resulting in the generation of graphs and charts that allow for quick viewing and understanding the behavior of research data. The data set of this research consists of 204 questionnaires answered by students who attended the 2nd period of undergraduate distance in the year 2014. To identify the profile of the students, the statistical technique Descriptive data was used. In the survey, students surveyed identified themselves have over 30 years of age, are married, have children, work and study and receive up to three minimum wages. In addition, it was noted that the largest number of students enrolled in 2nd period surveyed in 2014 are female. Therefore, it was concluded that multiple correspondence analysis coupled with computational tools facilitates handling large databases classified and without the use of mathematical calculations. Moreover, this exploratory data analysis technique allows to analyze the relationships between variables, and categorical variables and variables and objects.

Keywords:

Multiple correspondence analysis, Categorical variables, Distance learning, Profile of new students, Teaching mode from a distance.

LISTA DE FIGURAS

Figura 1 - Mapa de Correspondência do exemplo fictício	36
Figura 2 - Mapa de Correspondência do exemplo por meio da Matriz Indicadora	42
Figura 3 - Mapa de Correspondência do exemplo por meio da Matriz de Burt	44
Figura 4 - Mapa de Correspondência do exemplo por meio da Matriz de Burt Ajustado	47
Figura 5 - Análise de Correspondência Geral	61
Figura 6 - Mapa de Correspondência do Grupo 1	65
Figura 7 - Mapa de Correspondência do Grupo 2	69
Figura 8 - Mapa de Correspondência dos Grupos 3 e 4	73
Figura 9 - Mapa de Correspondência do Grupo 5	76

LISTA DE QUADROS

Quadro 1 - Codificações das categorias das variáveis na ACM do Grupo 1	62
Quadro 2 - Codificações das categorias das variáveis na ACM do Grupo 2	66
Quadro 3 - Codificações das categorias das variáveis na ACM dos Grupos 3 e Grupo 4	69
Quadro 4 - Codificações das categorias das variáveis na ACM do Grupo 5	73

LISTA DE TABELAS

Tabela 1 - Tabela de Contingência	21
Tabela 2 - Desempenho escolar de alunos do Ensino Médio em disciplinas cursadas no primeiro bimestre	23
Tabela 3 - Frequências esperadas do exemplo	23
Tabela 4 - Autovetores das linhas, valores singulares e perfil das linhas	30
Tabela 5 - Raiz dos valores do perfil das linhas	31
Tabela 6 - Coordenadas principais das linhas	31
Tabela 7 - Coordenadas principais das colunas	31
Tabela 8 - Principais inércias do exemplo	32
Tabela 9 - Perfil das linhas, coordenadas da linhas e coordenadas principais do exemplo	33
Tabela 10 - Contribuição relativa das linhas	33
Tabela 11 - Perfil das colunas, coordenadas das colunas e coordenadas principais	34
Tabela 12 - Contribuições relativas das colunas	34
Tabela 13 - Contribuições absolutas das linhas	35
Tabela 14 - Contribuições absolutas das colunas	36
Tabela 15 - Representação da Matriz Indicadora	38
Tabela 16 - Principais inércias do exemplo	40
Tabela 17 - Resultado da ACM do exemplo (x1000)	41
Tabela 18 - Representação da Matriz de Burt	42
Tabela 19 - As principais inércias do exemplo	43
Tabela 20 - Resultado da ACM do exemplo (x1000)	44
Tabela 21 - As principais inércias do exemplo	45
Tabela 22 - Resultado da ACM do exemplo (x1000)	45
Tabela 23 - Tamanho da Amostra por curso	48
Tabela 24 - Distribuição de frequência de alunos pesquisados de acordo com a cidade de residência	53
Tabela 25 - Distribuição de frequência de alunos pesquisados de acordo com o estado civil	53

Tabela 26 -	Distribuição de frequência de alunos pesquisados de acordo com o número de filhos	53
Tabela 27 -	Distribuição de frequência de alunos pesquisados por motivo da escolha de um curso na modalidade à distância	56
Tabela 28 -	Distribuição de frequência de alunos pesquisados por perspectiva profissional futura	57
Tabela 29 -	Probabilidade associada ao Qui-quadrado	58
Tabela 30 -	Comparativo das principais inércias e autovalores da Análise Geral	60
Tabela 31 -	Resultado das principais inércias e autovalores da análise geral	60
Tabela 32 -	Comparativo das principais inércias e autovalores do Grupo 1	63
Tabela 33 -	Resultado da inércia total do Grupo 1	63
Tabela 34 -	Resultado da ACM do Grupo 1 por meio da Burt Ajustado (x1000)	64
Tabela 35 -	Comparativo das principais inércias e autovalores do Grupo 2	66
Tabela 36 -	Resultado da inércia total do Grupo 2	67
Tabela 37	Resultado da ACM do Grupo 2 por meio da Burt Ajustado (x1000)	68
Tabela 38 -	Comparativo das principais inércias e autovalores dos Grupo 3 e Grupo 4	70
Tabela 39 -	Resultado da inércia total dos Grupo 3 e Grupo 4	70
Tabela 40 -	Resultado da ACM dos Grupo 3 e Grupo 4 por meio da Burt Ajustado (x1000)	71
Tabela 41 -	Comparativo das principais inércias e autovalores do Grupo 5	74
Tabela 42 -	Resultado da inércia total do Grupo 5	74
Tabela 43 -	Resultado da ACM do grupo 5 por meio da Burt Ajustado (x1000)	74

LISTA DE GRÁFICOS

Gráfico 1 -	Distribuição de questionários respondidos por curso	50
Gráfico 2 -	Distribuição de cursos por titulação	51
Gráfico 3 -	Distribuição de alunos pesquisados por sexo	51
Gráfico 4 -	Distribuição de alunos pesquisados por faixa etária	52
Gráfico 5 -	Distribuição de alunos pesquisados por raça	52
Gráfico 6 -	Distribuição de alunos pesquisados por faixa salarial	54
Gráfico 7 -	Distribuição de alunos pesquisados por curso de idioma	54
Gráfico 8 -	Distribuição de alunos pesquisados por primeira graduação	55

SUMÁRIO

1 INTRODUÇÃO	14
2 OBJETIVOS	15
2.1 Geral	15
2.2 Específico	15
3 REVISÃO DE LITERATURA	15
3.1 Conceito de Educação a Distância	15
3.2 Um breve histórico da Evolução da Educação a Distância	16
3.3 Educação a Distância no Brasil	17
3.4 O perfil do aluno EAD	18
4 METODOLOGIA DE PESQUISA	20
4.1 Estatística Descritiva	20
4.2 Teste Qui-quadrado	21
4.3 Análise de Correspondência	23
4.4 Breve histórico da Análise de Correspondência	38
4.5 Análise de Correspondência Múltipla	39
4.5.1 Matriz Indicadora	39
4.5.2 Matriz de Burt	43
5 CARACTERIZAÇÃO DA INSTUIÇÃO DE ENSINO SUPERIOR PESQUISADA	48
6 APRESENTAÇÃO DO BANCO DE DADOS	49
6.1 Questionário socioeconômico	49
6.2 Seleção e descrição da amostra	49
7 RESULTADOS E ANÁLISE DOS DADOS	52
7.1 Análises descritivas do banco de dados	52
7.2 Teste Qui-Quadrado do conjunto de dados	60
7.3 Aplicação da ACM	62
7.3.1 ACM do Grupo 1	64
7.3.2 ACM do Grupo 2	67
7.3.3 ACM da união dos Grupos 3 e 4	70
7.3.4 ACM do Grupo 5	74

8 CONCLUSÃO	77
9 REFERÊNCIAS	80
APÊNDICE A - QUESTIONÁRIO DE CARACTERIZAÇÃO SOCIOECONÔMICO	86
APÊNDICE B - PROBABILIDADES ASSOCIADAS AO QUI-QUADRADO CALCULADO DO CONJUNTO DE DADOS	87
ANEXO A - SCRIPT DO SOFTWARE R DA ANÁLISES DE CORRESPONDÊNCIA SIMPLES.....	88
ANEXO B - SCRIPT DO SOFTWARE R DA ANÁLISE DE CORRESPONDÊNCIA MÚLTIPLA	95

1 INTRODUÇÃO

No atual contexto, marcado por profundas mudanças sociais, a Educação a Distância vem crescendo de forma exponencial em muitos países do mundo, principalmente, em função da inserção de novas tecnologias de informação e comunicação, que aplicadas à educação, trazem várias possibilidades de acesso ao conhecimento.

No Brasil, esse movimento também não é diferente. De acordo com o Relatório Analítico da Aprendizagem a Distância no Brasil, divulgado em 2014, as matrículas para os cursos de nível superior chegaram a mais de um milhão de estudantes no ano de 2013. O crescimento do Ensino a Distância (EAD) no país se deve, principalmente, a criação de programas do governo que facilitam o acesso ao ensino superior. Além disso, outros fatores relacionados às vantagens que a EAD propicia aos estudantes, como por exemplo a flexibilidade de horários e mensalidades mais baratas têm atraído mais adeptos para essa modalidade de ensino.

Atentas a esse cenário, as Instituições de Ensino Superior (IES) investem cada vez mais em estratégias específicas de captação e retenção de alunos dos cursos EAD. É importante que as IES conheçam o perfil desse aluno para desenvolver cursos adequados ao seu público-alvo. Diferentemente do aluno do ensino presencial, o aluno do ensino a distância possui características próprias que devem ser consideradas na elaboração de metodologias didático-pedagógicas eficientes que propiciem condições de autoaprendizagem (BOLZAN, 1998).

Esta pesquisa tem como objetivo geral identificar o perfil dos alunos que cursam uma graduação na modalidade de ensino à distância, utilizando-se de técnicas estatísticas de Análise Descritiva e de Correspondência Múltipla. Além desse objetivo, a pesquisa contempla o seguinte objetivo específico : conhecer os fatores que influenciam na decisão do aluno a buscar um curso nessa modalidade de ensino.

Para alcançar os resultados, elaborou-se um questionário socioeconômico no qual abrangeu os seguintes eixos : econômico, educacional, social e autoavaliação. Esse questionário foi aplicado para os alunos que, na época desse estudo, cursavam o 2º período de uma graduação à distância. Com os questionários respondidos, construiu-se um banco de dados e foram aplicadas técnicas estatísticas de análise de correspondência a fim de se compreender as associações entre as perguntas (variáveis), bem como comparar a estrutura das respostas dos alunos.

Para facilitar a compreensão e leitura deste texto, ele está organizado em oito seções. Na seção três, é apresentada uma revisão da literatura sobre a Educação à Distância (EAD), a qual aborda aspectos de sua evolução, o desenvolvimento da EAD no Brasil, uma abordagem conceitual de EAD e o perfil do aluno EAD.

Na seção quatro, trata-se dos conceitos básicos, formulações matemáticas e estatísticas empregadas na elaboração desta pesquisa, como a Análise Descritiva, Análise de Correspondência Simples e Análise de Correspondência Múltipla.

Já na seção cinco, apresenta-se uma breve caracterização da Instituição de Ensino Superior pesquisada, enquanto que, na seção seis, trata-se da elaboração do questionário socioeconômico e da metodologia da criação do banco de dados em estudo.

Por fim, a apresentação dos resultados e da análise de dados é proposta na seção sete, e na seção oito, são apresentadas as considerações finais.

2 OBJETIVOS

2.1 Geral

O objetivo geral deste trabalho é identificar o perfil dos alunos que cursam uma graduação na modalidade de ensino à distância, utilizando técnicas estatísticas de Análise Descritiva e de Correspondência Múltipla.

2.2 Específico

Por conseguinte, o objetivo específico deste estudo é conhecer os fatores que induzem os alunos a buscar um curso de graduação na modalidade de ensino à distância, aplicando técnicas estatísticas apropriadas para análise de dados categóricos, como a Análise de Correspondência.

3 REVISÃO DE LITERATURA

3.1 Conceito de Educação à Distância

Desde o seu surgimento, a Educação à Distância (EAD) vem trazendo questionamentos quanto a sua definição. Na literatura, encontramos desde o conceito mais simples que define a EAD como qualquer forma de educação em que o professor se encontra distante do aluno, como também o conceito que incorpora o uso de novas tecnologias de informação e comunicação no processo de ensino-aprendizagem, propiciando a interação entre professores e alunos que estão separados, física ou temporalmente.

Atualmente, o conceito de EAD está associado à utilização de novas tecnologias no processo pedagógico (LANDIM, 1997). Na modalidade de ensino à distância, o Professor e o aluno estão em lugares distintos e, por isso utiliza os meios tecnológicos de informação e comunicação para interagir.

Com a publicação do Decreto nº 5.622, de 19 de dezembro de 2005, que regulamenta a Educação a Distância no Brasil, o conceito de EAD se torna mais abrangente e integra a variável tempo, permitindo que o aluno desenvolva as atividades em qualquer horário. Conforme o Decreto nº 5.622/2005, a EAD é definida como modalidade educacional na qual a mediação didático-pedagógico no processo de ensino e aprendizagem ocorre com a utilização das tecnologias de informação e comunicação, com alunos e professores desenvolvendo atividades educativas em lugares ou tempos diversos (BRASIL, 2005).

Em geral, as definições relacionam a educação à distância com os seguintes termos: separação física entre professor e o aluno e cursos mediados por tecnologias. Essa nova perspectiva de ensino aberto pelas tecnologias digitais e o uso da internet, mesmo que ainda não tenha um conceito definitivo para a EAD, passa a reconstruir virtualmente espaços de interação e de geração de conhecimento, permitindo o desenvolvimento de atividades de acordo com interesses e possibilidades dos participantes.

3.2 Um breve histórico da Evolução da Educação à Distância

A invenção da máquina de impressão tipográfica por Gutemberg, na Alemanha, no século XV, revolucionou a leitura e escrita e, também, teve um papel fundamental na história da Educação a Distância (MORINI,2006).

Com a Revolução Industrial ocorrida na Alemanha, no século XIX, alavancou o estudo por correspondência. Isso só foi possível com o aprimoramento das técnicas de impressão, com o melhoramento do sistema de transporte ferroviário e dos serviços postais (VILAÇA,2010). Tudo isso propiciou a produção e distribuição em grande escala de materiais didáticos para os mais diversos lugares do país.

Nas décadas de 1920 a 1960, o rádio, o telefone e a televisão foram introduzidos na sociedade. Eles se tornaram os novos protagonistas da disseminação da Educação a Distância. Com o aparecimento e o desenvolvimento das mídias de comunicação em massa, fomentou-se a criação e a transmissão de programas educativos para adultos e, assim, novas tecnologias foram inseridas no ensino a distância.

Diante do novo cenário econômico e social estabelecido durante o período da Revolução Industrial, a Educação a Distância tornou-se uma alternativa para capacitação profissional dos trabalhadores para o exercício de suas atribuições. O grande avanço na EAD se deu a partir dos anos 1960, quando diversas universidades do mundo e escolas do ensino básico começaram a oferecer o ensino à distância para atender a demanda de qualificação de profissional (VIDAL; MAIA, 2010).

Nas últimas décadas, as tecnologias da informação e comunicação têm promovido alterações profundas e permanentes na maneira de pensar e agir do homem nos âmbitos social, cultural e econômico. Na área educacional, essas mudanças fomentaram a criação novos formatos metodológicos de ensino-aprendizagem à distância que “variam quanto ao grau de presencialidade, nível de interatividade, tipos de recursos tecnológicos, área de atuação e escala de abrangência” (VIDAL; MAIA, 2010, p.14). Nos dias atuais, com essas mudanças, a adesão e expansão da EAD chega ao alcance mundial.

3.3 Educação a Distância no Brasil

A história da EAD no Brasil não difere muito dos fatos que ocorreram no mundo. Os primeiros registros que se tem sobre a Educação à Distância no Brasil é de um anúncio, publicado no Jornal do Brasil em 1904, que oferece curso por correspondência para datilógrafo (ALVES, 2011).

A popularização e a disseminação da EAD no Brasil ocorreu no início do século XX, com a utilização do rádio e materiais impressos. Em 1923, a Rádio Sociedade do Rio de Janeiro foi fundada com objetivo de levar cultura e educação à população (ALVES, 2007).

Depois, surgiram o Instituto Monitor, em 1939, e o Instituto Universal Brasileiro, em 1941, que ofertaram cursos profissionalizantes à distância por correspondência para todas as pessoas do território nacional.

Nas décadas de 1960 a 1990, introduz-se a televisão na área da educação, cuja finalidade era a de promover a inclusão social de adultos através dos programas educacionais como, por exemplo, Jornal da Educação - Edição do Professor, TV Escola e Telecurso 2000 (VIDAL; MAIA, 2010).

A utilização da televisão na área da educação contribuiu para o aumento da oferta de cursos na modalidade à distância, possibilitando a democratização do acesso à educação no país (VIDAL; MAIA, 2010).

As bases legais que nortearam o Ensino à Distância foram estabelecidas pela Lei de Diretrizes e Bases da Educação Nacional nº 9.394, de 20 de dezembro de 1996, porém, esta foi

regulamentada em 20 de dezembro de 2005 pelo Decreto nº 5.622, que revogou os Decretos nº 2.494 de 10/02/98, e nº 2.561 de 27/04/98, normatizada pela Portaria Ministerial nº 4.361 de 2004 (BRASIL, 2005).

Ainda no que tange ao Decreto nº 5.622, de 19 de dezembro de 2005, os cursos superiores na modalidade à distância são reconhecidos de forma igual aos presenciais, mas são impostas algumas regras de funcionamento, como, por exemplo, avaliações presenciais, estágios, apresentação de trabalho de conclusão de curso e, no que se refere à duração e à carga horária dos cursos à distância, devem também ser iguais às dos cursos presenciais.

Atualmente, a EAD cresce de forma acelerada no país devido à criação de programas do governo que facilitam o acesso de alunos ao ensino superior. De acordo com o Relatório Analítico da Aprendizagem à Distância no Brasil, divulgado em 2014, as matrículas para os cursos de nível superior chegaram a mais de um milhão de estudantes no ano de 2013. Outros fatores que contribuem para tal acontecimento estão relacionados às vantagens proporcionadas por essa modalidade de ensino como, por exemplo, comodidade, flexibilidade de horários e mensalidades mais baratas para o estudante.

3.4 O perfil do aluno EAD

Diferentemente do aluno do ensino presencial, que estuda em uma sala de aula com professores e diversos recursos pedagógicos ao seu alcance, o aluno do ensino à distância estuda em qualquer lugar, desde que tenha acesso à internet, não precisa de professores ao seu lado o tempo todo e pode desenvolver suas atividades acadêmicas em qualquer horário. Essas vantagens proporcionadas pelo ensino à distância tem atraído várias pessoas.

Diante da crescente procura pelo ensino à distância, as Instituições de Ensino Superior (IES) devem se atentar e conhecer o perfil desses alunos que apresentam aspectos culturais, financeiros e educacionais bastante heterogêneos, isto é, possuem características próprias que devem ser consideradas ao desenvolverem cursos mais adequados a diversidade em questão (BOLZAN, 1998).

Gomes, Leonardo e Mota (2014, p. 356) afirmam que é importante que as IES conheçam o perfil do aluno EAD, pois, isso “favorece a implantação de estratégias para aprendizagem, além de garantir maior eficiência no processo de ensino-aprendizagem”. A autora Belloni (2006) também corrobora com afirmação citada pelos autores ao afirmar que esta ação permite a integração do aluno na concepção de metodologias e estratégias pedagógicas, de modo que estas propiciam condições de autoaprendizagem.

Dessa forma, os cursos EAD atenderão às necessidades, expectativas e anseios dos estudantes que “veem na educação à distância uma alternativa para prosseguir nos seus estudos devido à flexibilidade de horários e a autonomia em desenvolver um cronograma de estudo de acordo com a sua disponibilidade de tempo” (AZEVEDO citado por SCHNITMAN, 2010, p. 3).

Um estudo realizado pelo Anuário Brasileiro Estatístico de Educação Aberta e a Distância (ABRAEAD), em sua edição do ano 2008, identificou as seguintes características do aluno EAD: tem idade acima de 30 anos, é casado, tem filhos, trabalha, recebe renda entre um e três salários mínimos, estudou em escola pública, tem menos acesso à *internet*, usa pouco o computador e tem pouco conhecimento em língua estrangeira. Outra característica identificada nesse estudo mostra que a EAD para cursos de nível superior é composta por uma população majoritariamente feminina.

Para o desenvolvimento e aprendizagem do aluno EAD, outras características relacionadas à competência e à habilidade de cada indivíduo são também consideradas pela autora Belloni (2006). Segundo a autora, o estudo mediado por novas tecnologias de informação e comunicação requer indivíduos proativos e autônomos, capazes de buscar e analisar as informações, de modo que possam administrar seu processo de aprendizagem. Esse perfil é também visto pelas empresas como essencial para os profissionais no mercado de trabalho de hoje, uma vez que tecnologias e processos produtivos mudam rapidamente, exigindo boa capacidade para acompanhar as mudanças.

Portanto, conhecer melhor as características dos alunos EAD estreita as relações entre as Instituições de Ensino Superior e sua comunidade acadêmica. As universidades podem identificar fatores que influenciam a escolha por cursos na modalidade de ensino à distância e desenvolver estratégias didático-pedagógicas mais adequadas ao perfil dos estudantes, bem como oferecer condições de aprendizado diferenciado e competitivo para o mercado de trabalho.

4 METODOLOGIA DE PESQUISA

4.1 Estatística Descritiva

Para organização, descrição e sintetização do conjunto de dados desta pesquisa, foram utilizadas técnicas de análise descritiva. Essas permitem que o pesquisador manuseie os dados de modo a extrair informações sobre o fenômeno em estudo. Antes da apresentação do método estatístico descritivo, é necessário ter clareza nos conceitos utilizados. A seguir, são apresentados alguns desses conceitos.

Na Estatística, os dados são obtidos de uma população ou de uma amostra. A população é definida como sendo um conjunto de elementos ou indivíduos que compartilham de, pelo menos,

uma característica comum, enquanto a amostra é um subconjunto de elementos selecionados da população segundo regras definidas (CRESPO, 2002). Essas regras possibilitam selecionar elementos da população aleatoriamente, de modo a garantir a representatividade da amostra em relação a população em estudo.

Existem várias técnicas de amostragem e a escolha por uma delas deverá ser feita pelo pesquisador de acordo com os objetivos propostos pela pesquisa. As técnicas de amostragem são subdivididas em amostragem probabilística e não probabilística. Na primeira, a probabilidade de um elemento da população ser escolhido é conhecida, enquanto que, na segunda não se conhece.

As principais técnicas de amostragem probabilística são: aleatória simples, estratificada, sistemática e por conglomerados. Com relação as técnicas de amostragem não probabilística, podemos citar amostragem aleatória por conveniência, por julgamento, por quotas e bola de neve e dentre outros.

Na pesquisa tratada nessa monografia, utilizou-se a Amostragem Estratificada com a Aleatória Simples. A Amostragem Aleatória Simples é fácil e pode ser feita por meio de um sorteio. Dessa forma, uma população com N elementos são numerados de 1 a N e, por meio de um sorteio, n elementos são escolhidos (FÁVERO *et al.* 2009).

Nessa técnica de amostragem, todos os elementos da população tem a mesma probabilidade de pertencer à amostra. Já na técnica de amostragem aleatória estratificada, a população é inicialmente dividida em subgrupos ou estratos. Depois, o pesquisador pode ou não aplicar a técnica de amostragem aleatória simples em cada estrato para selecionar os elementos da amostra de forma independente.

A amostra final é obtida, de modo que os elementos selecionados de cada subpopulação estão representados na amostra. Nesse tipo de amostragem, é necessário especificar quantos elementos serão selecionados aleatoriamente de cada estrato (FÁVERO *et al.* 2009).

Na análise descritiva dos dados, esses podem ser resumidos em tabelas de distribuição de frequências, de acordo com o tipo de variável. Podemos também representá-los de forma gráfica, permitindo que o pesquisador visualize as características do conjunto de dados estudado de forma rápida e eficiente (FÁVERO *et al.* 2009).

A utilização de ferramentas estatísticas adequadas, para o processamento de grande quantidade de dados, ajuda na extração de informações úteis e torna mais consistentes os resultados da análise. Visando atender a este propósito, o método de análise descritiva possui extrema importância na organização e descrição dos dados desta monografia. Nesta pesquisa, o conjunto de dados é apresentado por meio de tabelas de frequências e seus respectivos gráficos, uma vez que trata-se de dados categóricos.

4.2 Teste Qui-quadrado

O teste Qui-Quadrado não é pré-requisito para realização da Análise de Correspondência, mas auxilia o pesquisador a verificar se existe alguma associação ou dependência entre duas variáveis categóricas dispostas em uma tabela de contingência bidimensional (Tabela 1). Ele compara as frequências observadas em cada uma das células da tabela de contingência com as frequências esperadas.

A estatística qui-quadrado (χ^2) mede a diferença entre uma tabela de contingência observada e uma tabela de contingência esperada, partindo da hipótese que não há associação entre as variáveis estudadas. Se a distribuição de frequências observadas for igual à distribuição de frequências esperadas, o resultado da estatística qui-quadrado é zero. Então, um valor baixo de χ^2 indica independência entre as variáveis.

Tabela 1 - Tabela de Contingência

		<i>Variável W</i>				Total
		1	2	...	<i>q</i>	
<i>Variável X</i>	1	n_{11}	n_{12}	...	n_{1q}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2q}	$n_{2.}$

	<i>P</i>	n_{p1}	n_{p2}	...	n_{pq}	$n_{p.}$
Total	$n_{.1}$	$n_{.2}$...	$n_{.q}$	n	

onde:

p representa o número de categorias da variável X ;

q representa o número de categorias da variável W ;

n_{ij} representa o número de elementos que pertencem à categoria i da variável X e à categoria j da variável W ;

n representa o número total de elementos observados.

A estatística de teste é dada por:

$$\chi^2 = \sum_{(i,j)} \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

onde n_{ij} representa a frequência observada e E_{ij} representa a frequência esperada na casela (i,j) da tabela de contingência. O E_{ij} é dado por:

$$E_{ij} = n p_i p_j = n \frac{n_i}{n} \frac{n_j}{n} = \frac{n_i n_j}{n} \quad (2)$$

sendo $p_{i.} = \frac{n_{i.}}{n}$ e $p_{.j} = \frac{n_{.j}}{n}$.

O Teste Qui-quadrado de independência para duas variáveis assume as seguintes hipóteses nula (H_0) e alternativa (H_a) : (H_0), as variáveis são independentes e H_a (alternativa), as variáveis são dependentes.

Após definir as hipóteses, seguiremos com a especificação do nível de significância, o qual é denotado por α (letra grega minúscula alfa). Ele indica a probabilidade máxima para cometer um erro tipo I, ou seja, rejeitar a hipótese nula quando ela é de fato verdadeira. A probabilidade de ocorrer o erro do tipo II é denotada por β (a letra grega minúscula Beta), $0 < \beta < 1$. Esse tipo de erro ocorre quando a hipótese nula não é rejeitada, mas, no entanto, ela é falsa.

Já o *P-value* representa a probabilidade associada ao valor observado da amostra, indicando o menor nível de significância que levaria a rejeição de H_0 . A maioria dos *softwares* estatísticos calculam o correspondente *P-value*. Assim, para um nível de significância de 5%, se $p < \alpha$, a hipótese nula é rejeitada. Se $p > \alpha$, a hipótese nula não é rejeitada.

A distribuição da estatística de teste sob a hipótese nula é aproximada por uma distribuição qui-quadrado χ^2 com $k = (p - 1) \cdot (q - 1)$ graus de liberdade, em que p é o número de linhas da tabela de contingência e q é o número de colunas da mesma tabela. Para que a hipótese nula seja rejeitada, o valor observado da estatística de teste deve ser maior que o valor crítico do teste, ou seja, o valor $\chi^2_{c,\alpha}$ tal que $P[\chi^2_k \geq \chi^2_{c,\alpha}] = \alpha$, $0 < \alpha < 1$, sendo α o nível de significância do teste.

Na próxima seção, apresentaremos os conceitos e as formulações matemáticas da Análise de Correspondência. Para facilitar a compreensão da AC, será demonstrado sua aplicação por meio de um exemplo com dados fictícios.

4.3 Análise de Correspondência

Como exemplo, utilizaremos dados fictícios de uma pesquisa que visa avaliar o desempenho dos alunos do Ensino Médio em disciplinas cursadas no primeiro bimestre. Primeiramente, foi realizado o teste Qui-Quadrado para verificação da associação entre as variáveis do exemplo da Tabela 2.

Tabela 2 – Desempenho escolar de alunos do Ensino Médio em disciplinas cursadas no primeiro bimestre

Disciplina (X)	Desempenho escolar (W)			Total
	Baixa	Regular	Alta	
Matemática	156	14	2	172
Português	20	124	5	149
Física	7	12	77	96
Biologia	36	82	15	133
Total	219	232	99	550

onde:

Variável X = Disciplina e $p = 4$ é o número de categorias da variável Disciplina.

Variável W = Desempenho escolar e $q = 3$ é o número de categorias da variável Desempenho escolar.

$n = 550$ (número total da amostra).

Para identificar se existe uma relação de dependência entre as categorias das duas variáveis da Tabela 2, deve-se calcular a estatística χ^2 .

Esta é obtida com base na seguinte fórmula:

$$\chi^2 = \sum_{(i,j)} \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (3)$$

A fim de que seja calculada a estatística χ^2 , primeiramente, deve-se definir os valores esperados de cada casela da Tabela 2. Para isso, multiplica-se as frequências totais das linhas com as frequências totais das colunas e, depois, divide os resultados pelo total da amostra, como mostra o exemplo:

$$E_{ij} = \frac{n_i \cdot n_j}{n} = \frac{219 \times 172}{550} = 68,49$$

Dessa forma, obtemos os valores esperados de cada casela do exemplo, conforme a Tabela 3.

Tabela 3 - Frequências esperadas do exemplo

Disciplina	Baixa	Regular	Alta
Matemática	68,49	72,55	30,96
Português	59,33	62,85	26,82

Física	38,23	40,49	17,28
Biologia	52,96	56,10	23,94

Então, aplica-se a fórmula do χ^2 para o cálculo do teste. Assim, temos que:

$$\chi^2 = \frac{(156-68,49)^2}{68,49} + \frac{(14-72,55)^2}{72,55} + \frac{(2-30,96)^2}{30,96} + \frac{(20-59,33)^2}{59,33} + \frac{(124-62,85)^2}{62,85} + \frac{(5-26,82)^2}{26,82} +$$

$$\frac{(7-38,23)^2}{38,23} + \frac{(12-40,49)^2}{40,49} + \frac{(77-17,28)^2}{17,28} + \frac{(36-52,96)^2}{52,96} + \frac{(82-56,10)^2}{56,10} + \frac{(15-23,94)^2}{23,94} = 111,8119 +$$

$$47,2516 + 27,0892 + 26,0720 + 59,4960 + 17,7521 + 25,5117 + 20,0464 + 206,3934 +$$

$$5,4313 + 11,9574 + 3,3385 = 562,1515.$$

Na próxima etapa, calcula-se o grau de liberdade (gl) para o nível de significância de 5%.

O cálculo do grau de liberdade é dado por :

$$\text{g.l.} = (p - 1) \times (q - 1) \quad (4)$$

O resultado referente ao cálculo do grau de liberdade do exemplo da Tabela 2 é dado a seguir:

$$\text{g.l.} = (4-1) \times (3-1) = 6$$

Para 6 graus de liberdade, temos na Tabela Qui-Quadrado o valor χ^2 crítico = 12,59 (para o nível de significância de 5%). Dessa forma, como o χ^2 calculado 562, 1515 é maior que o valor χ^2 crítico = 12,59, podemos rejeitar a hipótese nula de que as duas variáveis categóricas são independentes, ou seja, existe indícios de associação estatisticamente significativa, ao nível de significância de 5%, entre a disciplina e o desempenho escolar.

A seguir, construiremos a Matriz de Correspondência (**P**) de dimensão 4×3 tomando como base os dados da Tabela de Contingência. Essa matriz é constituída por valores das proporções entre linhas e colunas da Tabela 2, calculada através da fórmula (5) :

$$p_{ij} = \frac{n_{ij}}{n} \quad (5)$$

sendo n_{ij} o número de observações da casela (i, j) da tabela de contingência e n o número total de observações da amostra. Cada frequência relativa calculada representa uma proporção da frequência n_{ij} em relação ao total n .

Os valores da Matriz P foram obtidos conforme o seguinte cálculo:

$$p_{ij} = \frac{n_{ij}}{n} = \frac{156}{550} = 0,2836$$

Logo, a Matriz de Correspondência (P) para o exemplo é dada por:

$$P_{4 \times 3} = \begin{bmatrix} 0,2836 & 0,0255 & 0,0036 \\ 0,0364 & 0,2255 & 0,0091 \\ 0,0127 & 0,0218 & 0,1400 \\ 0,0655 & 0,1491 & 0,0273 \end{bmatrix}$$

No próximo passo, deve-se elaborar as Matrizes Diagonal D_r e Diagonal D_c referentes aos perfis linha e coluna, respectivamente. Para o cálculo dessas matrizes, são considerados os totais das frequências marginais das linhas e das colunas da tabela de contingência, dividido pelo total da amostra. Esses perfis são denominados de massa (peso) dos elementos constituintes da matriz de correspondência (MENDES, 2009).

A Matriz de Perfil de linhas é a matriz diagonal D_r , cujos elementos da sua diagonal é o vetor dado por $r = \left(\frac{n_{1.}}{n} \frac{n_{2.}}{n} \dots \frac{n_{p.}}{n} \right)$. O valor de cada perfil linha é calculado com base na seguinte fórmula:

$$\frac{n_{1.}}{n} = \frac{172}{550} = 0,3127$$

No exemplo da Tabela 2, a matriz de perfil de linha é dada por:

$$D_r = \begin{bmatrix} 0,3127 & 0 & 0 & 0 \\ 0 & 0,2709 & 0 & 0 \\ 0 & 0 & 0,1745 & 0 \\ 0 & 0 & 0 & 0,2418 \end{bmatrix}$$

De forma análoga, a Matriz de Perfil de Colunas é dada pela matriz diagonal D_c cujos elementos da sua diagonal é calculado por $c = \left(\frac{n_{.1}}{n} \frac{n_{.2}}{n} \dots \frac{n_{.q}}{n} \right)$. Como exemplo, o valor do perfil coluna 0,3981 foi obtido a partir do seguinte cálculo:

$$\frac{n_{.1}}{n} = \frac{219}{550} = 0,3981$$

No exemplo da Tabela 02, a matriz de perfil das colunas é dada por:

$$\mathbf{D}_c = \begin{bmatrix} 0,3981 & 0 & 0 \\ 0 & 0,4218 & 0 \\ 0 & 0 & 0,1800 \end{bmatrix}$$

Cada perfil de linha ou coluna pode ser representado como pontos em um mapa bidimensional (PAMPLONA, 1998). Esses pontos têm um formato de nuvem e podem aparecer próximos ou isolados, podendo indicar a existência ou não de associação entre eles.

Para apresentar as associações dos pontos gerados através de gráficos com dimensões reduzidas e propiciar uma melhor visualização do comportamento dos dados estudados em um espaço ótimo, é necessário utilizar a metodologia de decomposição em valores singulares (JELIHOVSKI; FERRAZ, 2010).

A metodologia de Decomposição em Valores Singulares permite determinar as coordenadas principais dos perfis de linha e coluna em relação a todos os eixos principais, de maneira que cada coluna represente as coordenadas de uma dimensão (GREENACRE citado por VIDIGAL, 2010).

De acordo com Mingoti (2005), seja $\tilde{\mathbf{P}} = P - rc'$ uma matriz que tem entradas do tipo $\left(p_{ij} - \frac{n_i \cdot n_j}{n}\right)$, ou seja, os termos representam uma comparação da proporção observada no interior de cada casela da tabela, com aquela esperada sob um modelo na qual as variáveis X e W são independentes. O posto (rank) da matriz $\tilde{\mathbf{P}}_{pxq}$ é dado por $k = \min(p-1, q-1)$.

A matriz $\tilde{\mathbf{P}}_{pxq}$ pode ser decomposta em valores singulares. Dessa decomposição são extraídas as coordenadas principais que resumem a informação contida nas linhas da tabela de contingência e as coordenadas principais que resumem a informação contida nas colunas da tabela.

Para realização dessa decomposição, é aplicado o teorema de decomposição de matrizes em seus valores singulares (JOHNSON, WICHERN; JOBSON; SEBER citados por MINGOTI, 2005), obtendo-se o seguinte resultado:

$$\tilde{\mathbf{P}} = \mathbf{A} \mathbf{\Lambda} \mathbf{B}' \quad (6)$$

onde $\mathbf{A} = \mathbf{D}_r^{1/2} \mathbf{U}_{pxk}$ é uma matriz de dimensão pxk , $\mathbf{B} = \mathbf{D}_c^{1/2} \mathbf{V}_{qk}$ é uma matriz de dimensão qk , \mathbf{U} e \mathbf{V} são matrizes ortogonais e $\mathbf{\Lambda}$ é a matriz de dimensão kxk contendo os valores singulares da matriz $\tilde{\mathbf{P}}$ ordenados em ordem decrescente. A matriz \mathbf{U} contém os autovetores da matriz $\tilde{\mathbf{P}}\tilde{\mathbf{P}}'$, e a matriz \mathbf{V} contém os autovetores da matriz $\tilde{\mathbf{P}}'\tilde{\mathbf{P}}$.

Assim, para ilustrar a decomposição em valores singulares, realizaremos um exemplo a partir de uma matriz $E = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$. O procedimento para resolução dessa matriz também pode ser aplicada para outras matrizes não quadradas.

Então, para encontrarmos as soluções da matriz E , deve-se aplicar a equação característica $\det(\lambda I - E) = 0$. Ao desenvolvermos essa equação, nos deparamos com um polinômio chamado de polinômio característico, do qual são extraídas as raízes. Essas raízes são os valores singulares da respectiva matriz.

Primeiramente, devemos encontrar os valores singulares da matriz E , a seguir:

$$E = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

Desenvolvendo a equação característica $\det(\lambda I - E) = 0$, teremos:

$$\det \left(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \right) = 0$$

$$\det \left(\begin{bmatrix} \lambda - 2 & -2 \\ -2 & \lambda - 2 \end{bmatrix} \right)$$

$$(\lambda - 2)^2 - 4 = 0$$

$$\lambda^2 - 4\lambda = 0$$

$$\lambda(\lambda - 4) = 0$$

Dessa equação, têm-se os seguintes valores singulares em ordem decrescente: $\lambda_1 = 4$ e $\lambda_2 = 0$.

Para encontrarmos os autovetores da matriz E , utilizaremos os valores singulares calculados no exemplo, a saber: $\lambda_1 = 4$ e $\lambda_2 = 0$. Então, soluciona-se a equação $(\lambda I - E) \mathbf{v} = \mathbf{0}$ para $\lambda = 0$ e $\lambda = 4$.

Para $\lambda = 0$

$$(\lambda I - E) \mathbf{v} = \mathbf{0}$$

$$\begin{bmatrix} \lambda - 2 & -2 \\ -2 & \lambda - 2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 - 2 & -2 \\ -2 & 0 - 2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -2 & -2 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} -2x_1 - 2y_1 = 0 \\ -2x_1 - 2y_1 = 0 \end{cases}$$

$$x_1 = -y_1$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} -y_1 \\ y_1 \end{bmatrix} = y_1 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

O vetor de $\lambda_1 = 0$ é $\mathbf{v}_1 = (-1, 1)$.

De forma análoga, deve-se solucionar a equação para $\lambda = 4$.

Para $\lambda = 4$

$$(\lambda \mathbf{I} - \mathbf{E}) \mathbf{v} = \mathbf{0}$$

$$\begin{bmatrix} \lambda - 2 & -2 \\ -2 & \lambda - 2 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 4 - 2 & -2 \\ -2 & 4 - 2 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} 2x_2 - 2y_2 = 0 \\ -2x_2 + 2y_2 = 0 \end{cases}$$

$$x_2 = y_2$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_2 \\ y_2 \end{bmatrix} = y_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

O vetor de $\lambda_2 = 4$ é $v_2 = (1, 1)$.

Logo, a matriz dos autovetores de E é a matriz $X = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$.

O exemplo da matriz E , exposto anteriormente, foi apenas para demonstrar como é realizado o cálculo para obtenção dos valores singulares e dos autovetores de uma matriz. Quando se fala em valor singular da matriz \tilde{P} , o cálculo é feito usando as matrizes $\tilde{P}\tilde{P}'$ e $\tilde{P}'\tilde{P}$. Os valores singulares não nulos dessas matrizes são iguais e elas aparecem ao quadrado, sendo necessário extrair as raízes. As raízes quadradas dos valores singulares positivos de $\tilde{P}\tilde{P}'$ e $\tilde{P}'\tilde{P}$ são chamadas de Valores Singulares de \tilde{P} , sendo denotada por δ .

Os cálculos dessas matrizes também podem ser realizados por meio de programas computacionais. Portanto, visando facilitar os cálculos da Decomposição em valores singulares do exemplo da Tabela 2, utilizou-se o *software* estatístico R.

Abaixo, seguem as matrizes geradas na Decomposição em valores singulares da Tabela 2, sendo elas : U é a matriz que contém os autovetores das linhas ; V é a matriz que contém os autovetores da colunas, δ é a matriz dos valores singulares e Λ é a matriz dos valores singulares ao quadrado.

$$U = \begin{bmatrix} -0,5571 & 0,5992 \\ -0,0496 & -0,6537 \\ 0,8288 & 0,3573 \\ -0,0181 & -0,2931 \end{bmatrix}$$

$$V = \begin{bmatrix} -0,5449 & 0,5522 \\ -0,0182 & -0,7602 \\ 0,8383 & 0,3424 \end{bmatrix}$$

$$\delta = \begin{bmatrix} 0,7654 & 0 \\ 0 & 0,6605 \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} 0,5859 & 0 \\ 0 & 0,4362 \end{bmatrix}$$

Definidas as matrizes de valores singulares e os autovetores, as coordenadas principais das linhas e das colunas podem ser calculadas. Para determinar o número máximo de coordenadas principais, utiliza-se a fórmula $k = \min(p - 1, q - 1)$.

No exemplo da Tabela 2, o número de coordenadas principais possível de ser obtido é igual a $k = 2 = \min(4 - 1, 3 - 1)$. As duas primeiras coordenadas principais das linhas e das colunas são mais representativas em termos da associação total que existe entre as variáveis X e W , pois estão relacionadas aos maiores valores singulares da matriz \tilde{P}_{pq} .

A fim de auxiliar os cálculos das coordenadas principais das linhas, apresentam-se os valores das matrizes autovetores das linhas, valores singulares e perfil das linhas calculados anteriormente na Tabela 4.

Tabela 4 – Autovetores das linhas, valores singulares e perfil das linhas

Valores da Matriz Autovetores das Linhas (U)		Valores Singulares (δ)		Perfil das linhas	
-0,5571	0,5992	0,7654	0,6605	Matemática	0,3127
-0,0496	-0,6537			Português	0,2709
0,8288	0,3573			Física	0,1745
-0,0181	-0,2931			Biologia	0,2418

Para o cálculo das coordenadas das linhas, utiliza-se a seguinte fórmula:

$$\text{Coord. Linhas} = \frac{U * \delta}{\sqrt{\text{Perfil das Linhas}}} \quad (7)$$

Inicialmente, deve-se multiplicar os valores da Matriz de autovetores das linhas pelos valores singulares, como segue:

$$U * \delta = \begin{bmatrix} (-0,5571 * 0,7654) & (0,5992 * 0,6605) \\ (-0,0496 * 0,7654) & (-0,6537 * 0,6605) \\ (-0,8288 * 0,7654) & (0,3573 * 0,6605) \\ (-0,0181 * 0,7654) & (-0,2931 * 0,6605) \end{bmatrix} = \begin{bmatrix} -0,4264 & 0,3958 \\ -0,0380 & -0,4318 \\ 0,6344 & 0,2360 \\ -0,0139 & -0,1936 \end{bmatrix}$$

Com o resultado da multiplicação dessas matrizes, realiza-se a divisão dos valores obtidos pela raiz quadrada do perfil de linhas de cada categoria (Tabela 5).

Tabela 5 – Raiz dos valores do perfil das linhas

Perfil das linhas		$\sqrt{\text{Perfil. Linhas}}$
Matemática	0,3127	0,5592
Português	0,2709	0,5205
Física	0,1745	0,4177
Biologia	0,2418	0,4917

Dessa forma, obtêm-se os valores das coordenadas de linhas, como segue na Tabela 6:

Tabela 6 - Coordenadas principais das linhas

Variável X	Coordenada 1	Coordenada 2
Matemática	-0,763	0,708
Português	-0,073	-0,830
Física	1,518	0,565
Biologia	-0,028	-0,394

O cálculo das coordenadas das colunas, deve ser realizado de forma análoga, porém utilizando a matriz de autovetores de colunas e os valores do perfil das colunas na fórmula dada por:

$$\text{Coord. Colunas} = \frac{v * \delta}{\sqrt{\text{Perfil das Colunas}}} \quad (8)$$

Dessa forma, obtêm-se as coordenadas principais das colunas, como mostrado na Tabela 7:

Tabela 7 - Coordenadas principais das colunas

Variável (W)	Coordenada 1	Coordenada 2
Baixa	-0,661	0,578
Regular	-0,021	-0,773
Alta	1,512	0,533

Para auxiliar na interpretação do mapa de correspondência, podemos utilizar alguns conceitos, como por exemplo, a inércia total e a contribuição relativa. A inércia total é uma medida utilizada para indicar a variabilidade dos dados no espaço determinado pelas dimensões.

A forma de se calcular a inércia total é a partir dos valores singulares não-nulos da diagonal da matriz \mathbf{A} . Para esse cálculo, utiliza-se a seguinte fórmula:

$$\sum_{i=1}^k \lambda_i^2 \quad (9)$$

onde λ_i são os valores singulares não-nulos da diagonal da matriz \mathbf{A} .

Logo, temos que :

$$\sum_{i=1}^K \lambda_i^2 = \lambda_1^2 + \lambda_2^2 = 0,5859 + 0,4362 = 1,022 .$$

Os valores singulares ao quadrado da diagonal da matriz \mathbf{A} também são chamados de inércias principais parciais. Esses valores, quando divididos pelo valor da inércia total, indicam o percentual de contribuição de explicação de determinada linha ou coluna para construção de cada dimensão do mapa de correspondência.

Na análise de correspondência, quanto maior a inércia principal total, maior será o nível de associação entre as categorias em linha e em coluna, possibilitando maior dispersão dos pontos no mapa de correspondência e melhor visualização da associação das variáveis e suas categorias.

Assim, para calcular a proporção de explicação da i -ésima coordenada principal em relação à inércia total, deve-se utilizar a seguinte fórmula:

$$\frac{\lambda_i^2}{\sum_{i=1}^k \lambda_i^2} \quad (10)$$

Assim, temos a proporção de explicação da primeira coordenada principal do exemplo dado por:

$$\frac{\lambda_1^2}{\sum_{i=1}^k \lambda_i^2} = \frac{0,5859}{1,0221} = 0,5732 . 100 = 57,32\%$$

Por meio da análise da Tabela 8, pode-se afirmar que as dimensões 1 e 2 explicam, respectivamente, 57,32% e 42,68% da inércia principal total.

Tabela 8 – Principais inércias do exemplo

Dimensão	Valor singular (λ^2)	Inércia parcial (%)	Inércia Acumulada (%)
1	0,5859	57,319	57,319
2	0,4362	42,681	100,000

No que tange à contribuição relativa de uma linha ou coluna, esta pode ser calculada por meio da decomposição da inércia total. Essa decomposição permite identificar quais pontos têm maior contribuição para as dimensões principais. Esse valor é chamando de contribuição de um ponto a uma dimensão e é denotada por Ctr .

Assim, por meio da equação (11), determina-se a contribuição relativa das linhas:

$$\text{Contrib. Linhas} = \frac{\text{Perfil de Linhas} * (\text{Coord. Linhas})^2}{\text{Valor singular}} \quad (11)$$

Tabela 9 – Perfil das linhas, coordenadas da linhas e coordenadas principais do exemplo

Perfil da Linha		Coordenadas das Linhas			Coordenadas principais	Valores singulares
		Coordenada 1	Coordenada 2			
Matemática	0,3127	Matemática	-0,763	0,708	1	0,5859
Português	0,2709	Português	-0,073	-0,830		
Física	0,1745	Física	1,518	0,565	2	0,4362
Biologia	0,2418	Biologia	-0,028	-0,394		

Então, aplica-se os dados da Tabela 9 à fórmula (11) e obtemos os seguintes resultados:

$$\text{Contrib. Linhas (Matemática), coord. 1} = \frac{0,3127 * (0,763)^2}{0,5859} \times 100 = 31,04$$

$$\text{Contrib. Linhas (Matemática), coord. 2} = \frac{0,3127 * (0,708)^2}{0,4362} \times 100 = 35,91$$

Na Tabela 10, apresentamos as contribuições relativas das linhas do exemplo:

Tabela 10 – Contribuição relativa das linhas

Variável (X)	Contribuição para coordenada 1	Contribuição para coordenada 2
	(%)	(%)
Matemática	31,04	35,91
Português	0,25	42,73
Física	68,68	12,77
Biologia	0,03	8,59

Para o cálculo das contribuições das colunas, deve-se seguir o modelo apresentado acima.

O cálculo da contribuição das colunas é dado por:

$$\text{Contrib. Colunas} = \frac{\text{Perfil Colunas} * (\text{Coord. Colunas})^2}{\text{Valor singular}} \quad (12)$$

Na Tabela 11, são apresentados os perfis das colunas, as coordenadas das colunas e as coordenadas principais 1 e 2 do exemplo fictício.

Tabela 11 – Perfil das colunas, coordenadas das colunas e coordenadas principais

Perfil das Colunas		Coordenadas das colunas			Coordenadas principais	Valores singulares
		Coordenada 1	Coordenada 2			
Baixa	0,3982	Baixa	-0,661	0,578	1	0,5859
Regular	0,4218	Regular	-0,021	-0,773	2	0,4362
Alta	0,1800	Alta	1,512	0,533		

Logo, temos:

$$\text{Contrib. Coluna (Baixa), coord. 1} = \frac{0,3982 * (-0,661)^2}{0,5859} \times 100 = 29,69\%$$

$$\text{Contrib. Coluna (Baixa), coord. 2} = \frac{0,3982 * (0,578)^2}{0,4362} \times 100 = 30,49\%$$

Na Tabela 12, apresentamos as contribuições das colunas para cada coordenada principal.

Tabela 12 – Contribuição relativa das colunas

Variável (W)	Contribuição para coordenada 1 (%)	Contribuição para coordenada 2 (%)
Baixa	29,69	30,49
Regular	0,03	57,79
Alta	70,27	11,73

Outro valor que também auxilia na interpretação dos dados nas coordenadas principais é a contribuição absoluta. A contribuição absoluta das linhas e das colunas exprime o quanto um ponto contribui para a inércia principal da coordenada em estudo (PAMPLONA, 1998). Esta é denotada por *Cor*.

Para o cálculo da contribuição absoluta das linhas, utilizamos a fórmula a seguir:

$$\text{Cor} = \frac{(\text{Coord.Linhas na dimensão 1})^2}{(\text{Coord.Linhas na dimensão 1})^2 + (\text{Coord.Linhas na dimensão 2})^2} \quad (13)$$

Assim, obtemos os seguintes valores das contribuições da linha 1 para as dimensões 1 e 2:

$$\text{Cor (dimensão 1)} = \frac{(-0,763)^2}{(-0,763)^2 + (0,708)^2} = \frac{0,582}{1,083} = 0,537$$

Para a dimensão 2, temos :

$$Cor = \frac{(Coord. Linhas na dimensão 2)^2}{(Coord. Linhas na dimensão 1)^2 + (Coord. Linhas na dimensão 2)^2}$$

$$Cor (dimensão 2) = \frac{(0,708)^2}{(-0,763)^2 + (0,708)^2} = \frac{0,501}{1,083} = 0,463$$

Na Tabela 13, apresentamos os valores das contribuições absolutas das colunas referentes ao exemplo estudado. Pode-se observar que os valores de contribuições absolutas das linhas para formação das coordenadas foram: Coordenada 1 (0,573) e (0,878) ; Coordenada 2 (0,463) e (0,992) , (0,995).

Tabela 13 – Contribuições absolutas das linhas

Disciplina	Contribuição para coordenada 1	Contribuição para coordenada 2
Matemática	0,537	0,463
Português	0,008	0,992
Física	0,878	0,122
Biologia	0,005	0,995

O cálculo da contribuição absoluta da coluna 1 da dimensão 1 é dado por:

$$Cor = \frac{(Coord.colunas na dimensão 1)^2}{(Coord.colunas na dimensão 1)^2 + (Coord.colunas na dimensão 2)^2} \quad (14)$$

Logo, temos:

$$Cor (dimensão 1) = \frac{(-0,661)^2}{(-0,661)^2 + (0,578)^2} = \frac{0,437}{0,771} = 0,567$$

O cálculo da contribuição absoluta da coluna 1 da dimensão 2 é dado por:

$$Cor = \frac{(Coord.colunas na dimensão 2)^2}{(Coord.colunas na dimensão 1)^2 + (Coord.colunas na dimensão 2)^2} \quad (15)$$

$$Cor (dimensão 2) = \frac{(0,578)^2}{(-0,661)^2 + (0,578)^2} = \frac{0,334}{0,771} = 0,433$$

Na Tabela 14, apresentam-se os valores das contribuições absolutas das colunas referente ao exemplo estudado. Pode-se observar que os valores das contribuições absolutas das colunas para formação das coordenadas foram : Coordenada 1 (0,567) e (0,889) ; Coordenada 2 (0,433) e (0,999).

Tabela 14 – Contribuições absolutas das colunas

Desempenho	Contribuição para coordenada 1	Contribuição para coordenada 2
Baixa	0,567	0,433
Regular	0,001	0,999
Alta	0,889	0,111

Após a aplicação da Análise de Correspondência no conjunto de dados do exemplo da Tabela 2, gera-se o Mapa de Correspondência, como apresentado na Figura 1. A análise pode ser feita observando as proximidades dos pontos do Mapa de Correspondência, podendo indicar uma associação entre as categorias da variável linha com as categorias da variável coluna.

Para seguir com a análise do Mapa de Correspondência da Figura 1, observa-se uma forte associação entre as categorias Física e Alta, pontos localizados no lado direito superior do mapa. Já no lado esquerdo, verificamos uma associação entre as categorias Biologia, Português e Regular. Quanto aos pontos relacionados às categorias Matemática e Baixa, que estão no canto superior da figura, percebe-se também uma associação entre eles.

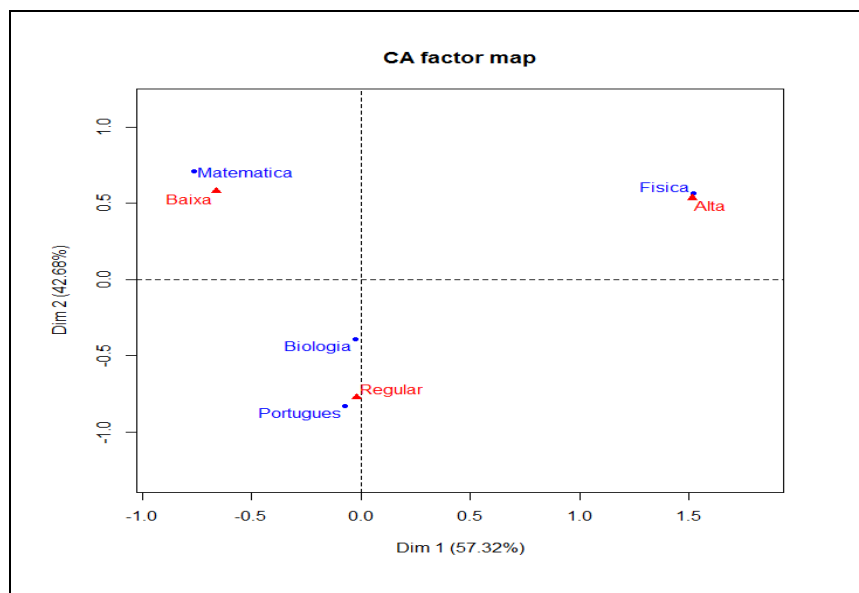


Figura 1 – Mapa de Correspondência do exemplo fictício

Definidos os conceitos da AC, a seguir, apresentamos um breve histórico da análise de correspondência para auxiliar na compreensão e na aplicação dessa técnica estatística.

4.4 Breve histórico da Análise de Correspondência

A partir de 1933 a 1960, foram publicados vários trabalhos sobre a técnica de análise multivariada de correspondência, porém, em meados da década de 60, Jean-Paul Benzécri trouxe uma nova proposta para o estudo de tabelas de contingência em linguística (MINGOTI, 2005).

Essa ferramenta se tornou popular a partir da década de 1975, sendo utilizada hoje em várias áreas do conhecimento como, por exemplo, da Ciências Sociais, da Biológicas e da Saúde, devido a sua facilidade de aplicação e à possibilidade de interpretação de dados através dos gráficos (CZERMAINSKI, 2004).

A Análise de Correspondência (AC) é definida como uma técnica estatística multivariada que permite estudar associações entre variáveis categóricas através da conversão de uma matriz de dados não negativos em gráficos, conhecida por tabela de contingência, que exibem as linhas e colunas da matriz como pontos em uma dimensão reduzida, de modo que estabeleçam relações entre linhas, entre colunas e entre linhas e colunas, permitindo uma rápida interpretação dos dados (GREENACRE; HASTIE citados por PAMPLONA, 1998).

Dessa forma, a representação gráfica obtida através da AC possibilita visualizar a distribuição das variáveis na sua relação com todas as outras, bem como sua intensidade de associação. Cada categoria de cada variável é representada por um ponto, e as distâncias entre os pontos representam as relações entre as variáveis que se deseja analisar.

Já a Análise de Correspondência Múltipla (ACM), pode ser considerada uma extensão da Análise de Correspondência, diferindo apenas na apresentação dos dados. Essa técnica de análise exploratória é aplicada no estudo de mais duas variáveis categóricas.

Inicialmente, utiliza-se a Matriz Indicadora e de Burt que exploram a relação entre um conjunto de variáveis categóricas. Logo, as informações contidas na Matriz Indicadora, que é formada por variáveis codificadas com valores 1 ou 0, permite gerar a Matriz de Burt. Esta é representada pela tabulação cruzada de todas as categorias e variáveis, cuja matriz é simétrica.

Na próxima seção, apresentamos os conceitos da ACM, assim como sua aplicação prática por meio de um exemplo com dados fictícios.

4.5 Análise de Correspondência Múltipla

Assim como a AC, a Análise de Correspondência Múltipla (ACM) é também uma técnica de análise exploratória de dados categóricos, porém aplicada a tabelas multidimensionais. A ACM

pode ser entendida como uma extensão da AC devido à aplicação do algoritmo da AC a tabelas de múltiplas entradas, chamadas de Matriz Indicadora e Matriz de Burt (NASCIMENTO, 2011).

Na ACM, as tabelas “são compostas de linhas por indivíduos e colunas por características observadas, onde cada linha contém todos os códigos correspondentes às modalidades atribuídas a um indivíduo ou elemento para cada uma das características observadas” (PAMPLONA, 1998, p.5).

A Matriz Indicadora (Z) é formada por variáveis codificadas como variáveis *dummy*, recebendo valores 1 ou 0. Na Matriz de Burt, é realizado o cruzamento entre várias variáveis categóricas duas a duas, sendo a diagonal principal formada de matrizes diagonais de frequências marginais de coluna (PAMPLONA, 1998). Essa matriz é calculada a partir do produto da transposta da Matriz Indicadora por ela mesma.

4.5.1 Matriz Indicadora

Nesta seção, apresentaremos a aplicação da ACM em um banco de dados fictício, conforme Tabela 2. Para darmos continuidade à explicação da Análise de Correspondência Múltipla, utilizaremos uma parte da Tabela de Desempenho escolar de alunos do Ensino Médio das disciplinas cursadas acrescida da variável idade para representar a Matriz Indicadora (Tabela 15).

Uma Matriz Indicadora apresenta um banco de dados com N observações e Q variáveis ($Q > 2$), e cada variável Q_q ($q = 1, 2, \dots, Q$) possuindo J_q categorias. Logo, o número total de categorias envolvidas em uma análise de correspondências múltipla é :

$$J = \sum_{q=1}^Q J_q \quad (17)$$

A Tabela 15 representa uma parte da Matriz Indicadora do banco de dados do exemplo fictício:

Tabela 15 – Representação da Matriz Indicadora

Indivíduo (Observação)	Disciplina (Q_1)				Desempenho (Q_2)			Idade (Q_3)		
	Matemática	Português	Física	Biologia	Baixa	Regular	Alta	Até 24 anos	De 25 a 34 anos	Acima de 35 anos
1	1	0	0	0	1	0	0	1	0	0
157	0	1	0	0	0	1	0	0	0	1
173	0	1	0	0	1	0	0	0	0	1
242	1	0	0	0	0	1	0	0	0	1

306	0	1	0	0	0	1	0	0	0	1
386	0	0	1	0	0	0	1	1	0	0
422	0	0	0	1	1	0	0	0	1	0
550	0	0	0	1	0	0	1	1	0	0
Total de indivíduos (N_{qj})	172	149	96	133	234	232	84	198	205	147
Número de categorias J_q	4				3			3		

onde:

Q_q representa a variável ($q = 1, \dots, Q$).

J_q representa o número de categorias da variável Q_q , sendo $q = 1, 2, \dots, Q$, onde Q é o número total de variáveis.

N_{qj} é o número total de indivíduos que escolheram a categoria J da variável Q_q , sendo $q = 1, 2, 3, \dots, Q$, $j = 1, 2, \dots, J_q$.

J é o número total de categorias, sendo $\sum_{q=1}^Q J_q$.

Na ACM, formam-se duas nuvens de pontos no mapa de correspondência. Essas nuvens referem-se aos elementos ou indivíduos (linhas) e às categorias (colunas). As nuvens de pontos são interpretadas também de acordo com a proximidade e distanciamento dos pontos, podendo existir ou não uma associação entre as variáveis e categorias em estudo.

O objetivo da ACM é reduzir o espaço multidimensional em um subespaço ótimo para que seja possível o estudo das nuvens de pontos. Portanto, a dimensionalidade de uma nuvem de pontos, na ACM, é determinada pelo número total de categorias menos o número total de variáveis, dado por:

$$L < J - Q \quad (18)$$

onde L é o número de coordenadas, J é o número total de categorias e Q é o número de variáveis. No caso do exemplo, $L < (10 - 3) = 7$. Dessa forma, podemos identificar quantas dimensões serão geradas, quando for aplicada a ACM em um conjunto de dados.

O algoritmo da AC fornece as regras básicas para interpretação da ACM. Por isso, pode ser aplicado em tabelas de contingência de múltiplas entradas (Matriz Indicadora ou Matriz de Burt). Para a ACM do exemplo da Tabela 15, utilizamos o programa R-project, um pacote do *software* estatístico R, a fim de facilitar os cálculos e fazer a demonstração dos resultados. Nessa análise do exemplo, o número máximo de dimensões geradas para explicar a variância total da nuvem de pontos é 7.

Na Tabela 16, observamos que as duas primeiras coordenadas são as mais representativas por obterem os maiores valores singulares (em destaque). Para o conjunto de dados do exemplo, as duas primeiras coordenadas acumulam 60,2% da variabilidade dos dados.

Tabela 16– Principais inércias do exemplo

Coordenadas	Valor singular	Inércia%	Inércia acumulada%
1	0,788755	33,8	33,8
2	0,616438	26,4	60,2
3	0,508462	21,8	82,0
4	0,160349	6,9	88,9
5	0,118468	5,1	94,0
6	0,091442	3,9	97,9
7	0,049419	2,1	100,0
Total	2,333333	100	

A apresentação dos resultados da ACM por meio da Matriz indicadora estão apresentados na Tabela 17, e ela exibe os seguintes valores:

- Massa (mass) é o peso relativo de cada categoria dentro da respectiva variável.
- Qualidade (qlt) de representação de um ponto em um subespaço, sendo calculada pela soma das contribuições absolutas.
- Inércia (inr) representa a proporção de contribuição de cada ponto na inércia total.
- k denota a coordenada principal de ordem k , $k = 1, 2, \dots, L$. Cada variável tem um valor numérico para cada coordenada de ordem k .
- Contribuição absoluta (cor) indica o quanto um ponto contribui para a inércia principal do eixo em estudo.
- Contribuição relativa (ctr) permite encontrar quais são os pontos mais importantes na orientação da coordenada principal.

Para interpretar os resultados da ACM e o Mapa de Correspondência, é necessário levar em conta as contribuições absolutas e relativas de cada coordenada. Os valores que apresentam maior contribuição absoluta e relativa para formação das coordenadas principais estão em negrito na Tabela 17, bem como suas respectivas coordenadas principais. São elas : Coordenada 1 ($k=1$) e Coordenada 2 ($k=2$).

Tabela 17 - Resultado da ACM do exemplo (x1000)

name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
Matéria:Biologia	81	120	87	-319	33	10	522	87	36
Matéria:Física	58	736	121	1438	437	152	-1188	299	133
Matéria:Matemática	104	533	92	392	70	20	1009	463	172
Matéria:Português	90	723	111	-1094	445	137	-865	278	110
Nota:Alta	60	856	136	1580	548	190	-1184	308	136
Nota:Baixa	133	621	79	243	39	10	938	582	189
Nota:Regular	141	702	89	-904	596	146	-380	105	33
Idade:Acima35anos	89	598	107	-1011	373	116	-785	225	89
Idade:Ate24anos	120	738	102	1137	727	197	-137	11	4
Idade:De25a34anos	124	370	77	-373	83	22	695	287	97

Outra forma de interpretar a ACM é por meio da qualidade de representação (qr) de um ponto. O valor da qualidade de representação de um ponto é dado pela soma das contribuições absolutas de um ponto em cada coordenada. Ele mostra o quanto cada ponto está ou não bem representado no subespaço ótimo. Na interpretação da tabela dos resultados da ACM, devemos considerar as categorias com os valores mais altos e próximos ou iguais a 1.

Outras duas formas com as quais um ponto pode contribuir para aumentar a inércia principal são: quando está localizado a uma grande distância do centro das coordenadas principais; e quando tem uma grande massa (JELIHOVSCHI; FERRAZ, 2010).

Para finalizar as análises da ACM, apresentamos o Mapa de Correspondência do exemplo fictício (Figura 2). Na figura, observa-se que os pontos relacionados às categorias Idade (De 25 a 34 anos), Matéria (Biologia), Matéria (Matemática) e Nota (Regular) estão localizados na parte superior do mapa de correspondência. No lado direito do mapa, os pontos pertencentes às categorias Idade (Até 24 anos), Matéria (Física) e Nota (Alta) estão mais próximos. No lado oposto, estão os pontos referentes às categorias Idade (Acima 35 anos), Matéria (Português) e Nota (Regular).

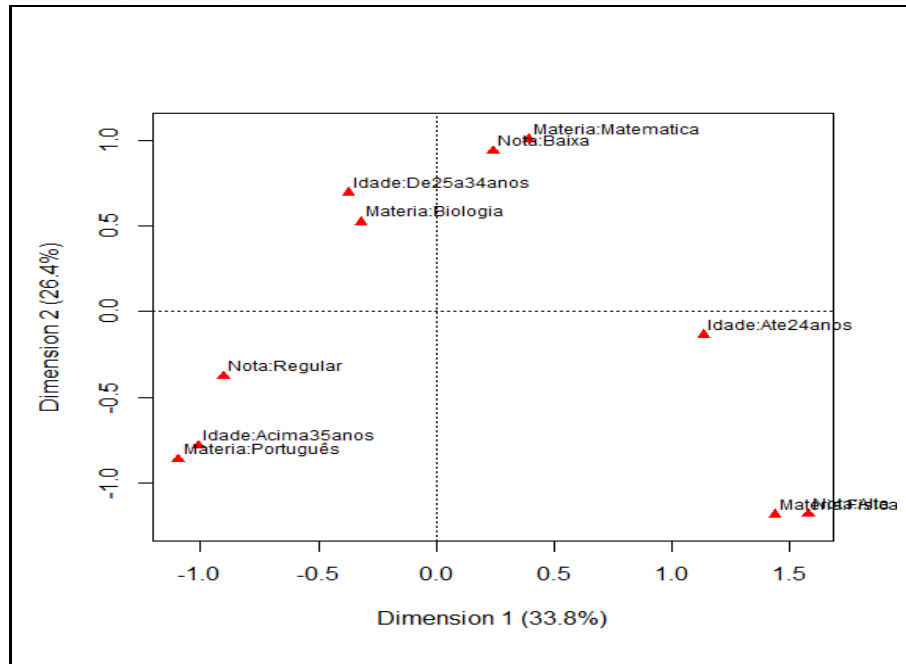


Figura 2– Mapa de Correspondência do exemplo por meio da Matriz Indicadora

4.5.2 Matriz de Burt

Uma alternativa de análise de correspondência para tabelas multidimensionais é a Matriz de Burt. Segundo Greenacre (2008) a Matriz de Burt é uma matriz $\mathbf{B} = \mathbf{Z}^T \mathbf{Z}$ simétrica quadrada de categorias por categorias, sendo formada por tabelas de contingência de pares de variáveis, incluindo na diagonal principal a tabulação cruzada de uma variável com ela mesma. Dado que \mathbf{B} é uma matriz simétrica, a análise de correspondência das linhas e das colunas desta matriz são iguais.

Assim como fizemos nas seções anteriores, apresentaremos a aplicação de um exemplo da ACM por meio da Matriz de Burt, utilizando para esse propósito os mesmos dados do exemplo da seção anterior.

Na Tabela 18, apresentamos a Matriz de Burt do exemplo:

Tabela 18 – Representação da Matriz de Burt

	Matemática	Português	Física	Biologia	Baixa	Regular	Alta	Até 24 anos	De 25 a 34 anos	Acima de 35 anos
Matemática	172	0	0	0	156	14	2	94	56	22
Português	0	149	0	0	20	124	5	5	26	118

Física	0	0	96	0	7	12	77	77	12	7
Biologia	0	0	0	133	36	82	15	22	111	0
Baixa	156	20	7	36	234	0	0	107	92	35
Regular	14	124	12	82	0	232	0	7	113	112
Alta	2	5	77	15	0	0	84	84	0	0
Até 24 anos	94	5	77	22	107	7	84	198	0	0
De 25 a 34 anos	56	26	12	111	92	113	0	0	205	0
Acima de 35 anos	22	118	7	0	35	112	0	0	0	147

Essa matriz possui uma coluna e uma linha para cada categoria de cada variável categórica. Ela é composta de todos os pares da tabulação cruzada entre as variáveis, ou seja, nas interseções são colocadas as frequências observadas simultaneamente em ambas as variáveis.

A Matriz de Burt é uma matriz simétrica, onde no triângulo inferior encontra-se a tabulação cruzada de cada variável com todas as outras variáveis e, no triângulo superior, encontra-se a sua contraparte transposta. Na diagonal dessa matriz, aparece o número total de observações em cada categoria das diversas variáveis.

A seguir, destacamos as propriedades desta análise e sua relação com a AC por meio de uma Matriz Indicadora (GREENACRE, 2008, p.190):

- A Análise de Correspondência sobre a Matriz de Burt produz coordenadas idênticas as da produzida pela Matriz Indicadora;
- As coordenadas principais são as coordenadas padrão multiplicadas pela raiz quadrada das inércias principais;
- As porcentagens das inércias principais da Matriz de Burt são o quadrado das inércias da Matriz Indicadora.

Na Tabela 19, apresentamos os resultados das principais inércias do exemplo fictício por meio da Matriz de Burt, bem como o seu dimensionamento. Como pode-se observar, o número de dimensões geradas é o mesmo do resultado da Tabela 15. Porém, ao contrário dessa tabela, a Matriz de Burt considera todas as informações dos 550 indivíduos.

No que tange as inércias de cada dimensão apresentada na Tabela 19, observa-se que as duas primeiras dimensões representam juntas 76,43% de variância explicada. As inércias de cada dimensão gerada por meio da aplicação da ACM via Matriz de Burt são maiores que as geradas pela

Matriz Indicadora (exemplo da seção anterior). Isso ocorre por que os valores singulares gerados pela Matriz de Burt são o quadrado dos valores da Matriz Indicadora, elevando as porcentagens das inércias da análise de Burt (GREENACRE,2008).

Tabela 19 - Principais inércias do exemplo

Dimensão	Valor singular	Inércia %	Inércia Acumulada %
1	0,622134	47,4	47,4
2	0,379995	29,0	76,4

Na Tabela 20, apresentamos os resultados obtidos da ACM sobre a Matriz de Burt deste exemplo. Essa tabela mostra que a massa (mass), a inércia (inr) e a contribuição relativa (ctr) de cada ponto são idênticas as da Matriz Indicadora (seção anterior).

Tabela 20 - Resultado da ACM do exemplo (x1000)

name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
Matéria:Biologia	81	176	87	-284	57	10	410	119	36
Matéria:Física	58	915	121	1277	596	152	-933	318	133
Matéria:Matemática	104	650	92	348	105	20	792	545	172
Matéria:Português	90	876	111	-972	588	137	-680	288	110
Nota:Alta	60	955	136	1404	664	190	-930	291	136
Nota:Baixa	133	755	79	216	60	10	737	695	189
Nota:Regular	141	888	89	-803	780	146	-298	108	33
Idade:Acima35anos	89	754	107	-898	512	116	-616	241	89
Idade:Ate24anos	120	923	102	1010	913	197	-108	10	4
Idade:De25a34anos	124	502	77	-331	135	22	546	367	97

Além disso, o mapa de correspondência resultante da ACM por meio da Matriz de Burt (Figura 3) também é semelhante ao gráfico produzido por meio da Matriz Indicadora, diferindo apenas nas dimensões. Isso se deve ao fato das inércias principais da Matriz de Burt serem o quadrado dos valores da Matriz Indicadora, conseqüentemente, as dimensões 1 e 2 do gráfico gerado pela Matriz de Burt sempre serão mais elevados que as dimensões do gráfico da Matriz Indicadora.

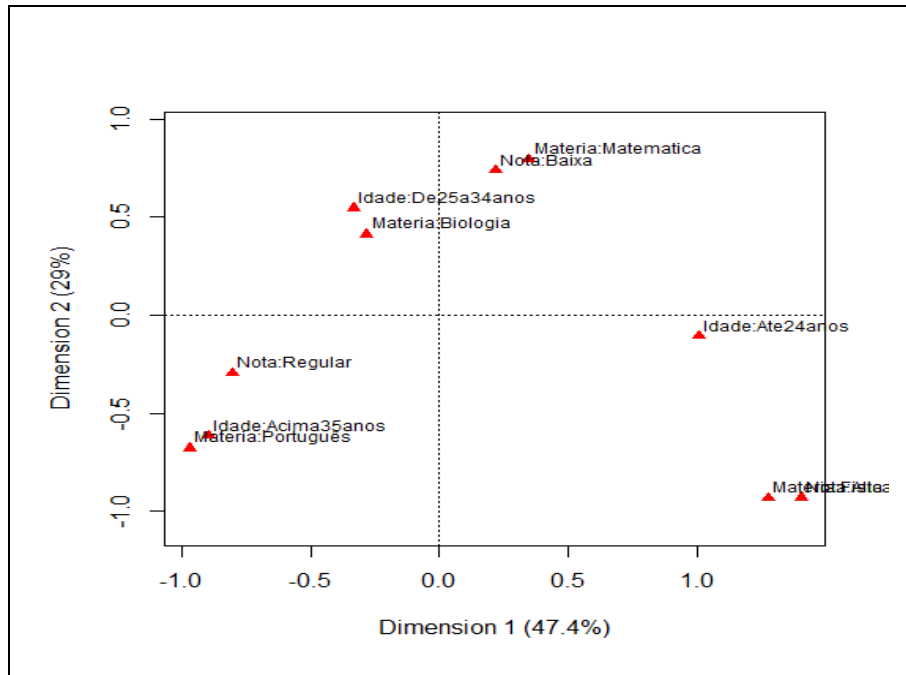


Figura 3– Mapa de Correspondência do exemplo por meio da Matriz de Burt

Nos exemplos da aplicação da ACM por meio das Matrizes Indicadora e de Burt, é possível constatar que a qualidade da representação gráfica relacionada à percentagem de inércia associada às duas primeiras coordenadas é muito reduzida em comparação com as inércias do mesmo exemplo gerada pela AC simples por meio da tabela de contingência de dupla entrada.

Essa diferença na qualidade de explicação dos dados deve-se ao fato de, na ACM, cada variável é codificada em diversas colunas, criando dimensões artificiais que inflacionam a inércia total e subestimam seus respectivos valores nas diversas dimensões, que são construídas com maior impacto nas primeiras (GRENACRE, 2008).

Assim, para contornar essa desvantagem, é utilizada uma fórmula de Bénécri que ajusta os valores das inércias (GRENACRE, 2008). Esta fórmula é dada por:

$$\lambda_k^{ajustada} = \left(\frac{Q}{Q-1}\right)^2 \times \left(\sqrt{\lambda_k} - \frac{1}{Q}\right)^2, k = 1, 2, \dots \quad (19)$$

onde λ_k indica a k -ésima inércia principal da Matriz de Burt e Q é o número de variáveis.

Quando aplicamos o método do Burt Ajustado à base de dados do exemplo, observa-se um melhoramento significativo nas inércias das duas coordenadas principais. Na Tabela 21, observa-se que a primeira dimensão explica 58,3% dos dados e a segunda, 22,5%.

Tabela 21 - As principais inércias do exemplo

Dimensão	Valor singular	Inércia (%)	Inércia Acumulada(%)
1	0,466669	58,3	58,3
2	0,180333	22,5	80,8

Com o ajustamento das inércias, há uma redução na quantidade de dimensões geradas pela análise devido ao aumento no percentual das inércias das duas coordenadas principais. O método do Burt Ajustado ajusta as inércias, de forma a acrescentar qualidade na representação dos dados, e eliminam as contribuições que inflacionam e subestimam seus valores nas dimensões (Tabela 22).

Tabela 22 - Resultado da ACM do exemplo (x1000)

Name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
Matéria:Biologia	81	255	87	-246	110	10	-282	145	36
Matéria:Física	58	942	121	1106	704	152	643	238	133
Matéria:Matemática	104	619	92	302	145	20	-546	474	172
Matéria:Português	90	873	111	-841	666	137	468	206	110
Nota:Alta	60	868	136	1216	680	190	641	189	136
Nota:Baixa	133	705	79	187	84	10	-507	620	189
Nota:Regular	141	948	89	-695	872	146	206	76	33
Idade:Acima35anos	89	793	107	-778	611	116	425	182	89
Idade:Ate24anos	120	979	102	875	972	197	74	7	4
Idade:De25a34anos	124	592	77	-287	218	22	-376	374	97

Na Figura 4, apresentamos o Mapa de Correspondência gerado por meio da Matriz de Burt Ajustado.

Nele, observa-se que as dimensões 1 e 2 são maiores que as dimensões apresentadas na Figura 3. Além disso, houve deslocamento dos pontos, tornando-o diferente da figura anterior.

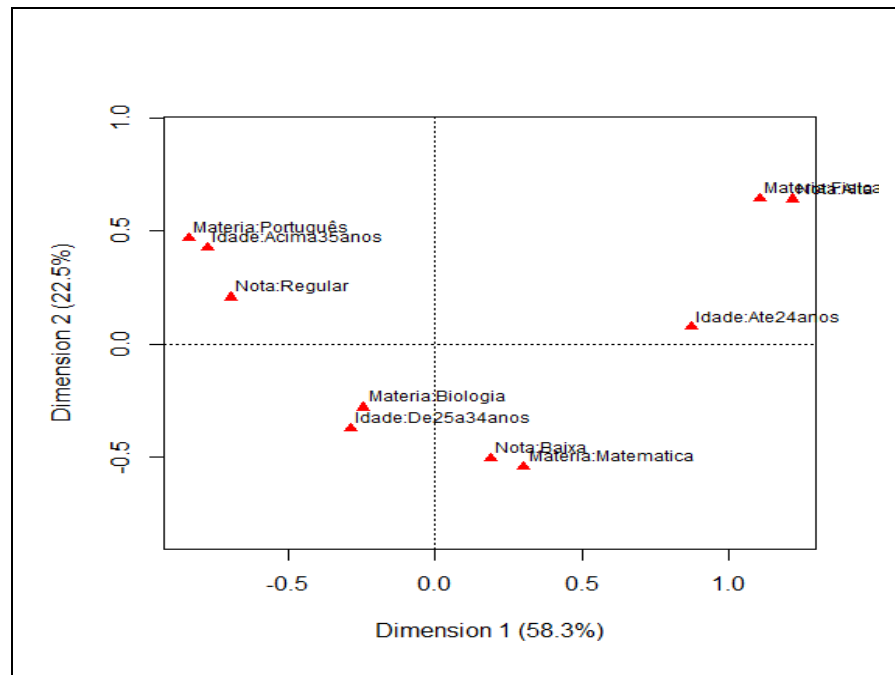


Figura 4 – Mapa de Correspondência do exemplo por meio da Matriz de Burt Ajustado

5 CARACTERIZAÇÃO DA INSTITUIÇÃO DE ENSINO SUPERIOR PESQUISADA

A Instituição de Ensino Superior pesquisada oferece cursos de graduação, pós-graduação e cursos livres na modalidade de Educação à Distância. Seu sistema de ensino semipresencial permite que os alunos assistam às aulas transmitidas via satélite, ao vivo dos estúdios da sede da universidade, para as salas de aulas do polo de apoio da universidade situado na região metropolitana de Belo Horizonte, Minas Gerais.

No polo, o aluno participa de atividades em grupo e assiste às aulas uma vez por semana, com a presença de um tutor em sala. Nos outros dias, o aluno estuda online, por meio dos inúmeros recursos e conteúdos disponibilizados no Ambiente Virtual de Aprendizagem – AVA e conta com o suporte de tutores eletrônicos e professores.

Além de todo o material disponível para estudo no AVA, os alunos recebem também materiais didáticos impressos com conteúdos preparados por professores capacitados, que os organizam pedagogicamente, tornando esses materiais mais interativos e didáticos.

Dessa forma, a instituição promove a interação entre estudantes, tutores eletrônicos e professores através das ferramentas como *chat* e fóruns de discussão, os quais podem ser acessados por meio da internet. Nesse momento de interatividade, os alunos podem esclarecer dúvidas, debater coletivamente sobre temas discutidos em aula e, ainda, receber orientações sobre atividades e trabalhos acadêmicos.

6 APRESENTAÇÃO DO BANCO DE DADOS

6.1 Questionário socioeconômico

Para alcançar os objetivos desta pesquisa, foi elaborado um questionário socioeconômico para o levantamento dos dados e posterior análise. O questionário socioeconômico apresentado no Apêndice A desta monografia é composto por vinte perguntas fechadas que abrangem aspectos social, econômico, educacional e de auto avaliação.

As perguntas de caráter social visam identificar algumas características dos entrevistados como, por exemplo, o sexo, faixa etária, raça, estado civil, local de residência, acesso à internet e se tem filhos. No que tange ao estado econômico, as perguntas do questionário tem como objetivo conhecer a ocupação atual dos entrevistados e sua renda mensal. Quanto às perguntas relacionadas ao aspecto educacional, estas têm como finalidade conhecer a formação prévia dos entrevistados: tipo de escola do ensino médio, tipo de curso realizado no ensino médio, período cursado no ensino médio e conhecimento em língua estrangeira.

Por fim, as perguntas relacionadas à auto avaliação fornecem informações sobre o conhecimento do entrevistado sobre si mesmo nos quesitos relacionados ao seu desempenho ou sua percepção sobre o ensino à distância. Nesse eixo do questionário, abordam-se perguntas sobre os motivos que fizeram o entrevistado escolher um curso na modalidade à distância, grau de conhecimento na utilização do computador, tempo de dedicação aos estudos, perspectiva profissional futura, conhecimento da Instituição de Ensino Superior pesquisada e o grau de satisfação com o curso de graduação.

Na seção 6.2 , apresentaremos, com mais detalhes, os dados sobre a população em estudo e o processo de seleção da amostra para composição do banco de dados desta pesquisa.

6.2 Seleção e descrição da amostra

A população desta pesquisa foi composta por 509 estudantes do 2º período distribuídos entre os 16 cursos de graduação oferecidos pela Instituição de Ensino Superior pesquisada, durante os meses de outubro e novembro, no ano de 2014 (Tabela 23). A escolha dessa população deve-se ao fato de que estes alunos, com a vivência do curso à distância no 1º período cursado, possivelmente, já tivessem adquirido uma percepção sobre o ensino à distância e conseguiriam responder a todos os eixos de abrangência do questionário socioeconômico.

Para aplicação do questionário, optou-se por estudar parte dessa população, ou seja, foi selecionada uma amostra aleatória. O procedimento de amostragem utilizado na seleção da amostra foi a da amostragem estratificada com o uso da amostragem aleatória simples sem reposição.

Assim, assumindo o nível de significância de 5% e a margem de erro também de 5% para estimação de proporções populacionais, selecionou-se 225 estudantes para compor a amostra desta pesquisa. Por conseguinte, a estratificação da população foi efetuada considerando Curso como fator de estratificação e o número de estudantes selecionados para compor a amostra foi calculado considerando-se o número de estudantes matriculados no 2º período de cada curso, como pode ser observado na Tabela 23.

Nessa tabela, mostra-se também o tamanho de cada amostra previamente calculado, o tamanho da amostra observada após a aplicação dos questionários, bem como a quantidade de questionários não respondidos em relação ao número estipulado previamente.

Tabela 23 - Tamanho da Amostra por Curso

Curso	Número de alunos matriculados no 2º período por curso – ano 2014	Número de questionários previstos	Número de questionários respondidos	Número de questionários não respondidos
Administração	37	17	17	0
Ciências Contábeis	21	9	9	0
Ciências Biológicas	12	7	6	1
Educação Física	51	22	21	1
História	18	8	4	4
Letras	4	2	4	0
Matemática	13	6	5	1
Pedagogia	133	59	52	7
Serviço Social	48	21	20	1
Tecnologia em Análise e Desenvolvimento de Sistemas	4	2	4	0
Tecnologia em Estética e Imagem Pessoal	53	23	20	3
Tecnologia em Gestão Ambiental	14	7	6	1
Tecnologia em Gestão de Recursos Humanos	20	8	6	2
Tecnologia em Gestão Pública	23	10	9	1
Tecnologia em Logística	20	8	6	2
Tecnologia em Processos Gerenciais	38	16	15	1
Total	509	225	204	25

Fonte: Dados da instituição de ensino

Inicialmente, o número de questionários previstos para aplicação era de 225. No entanto, durante a aplicação dos questionários, optou-se por aplicar o questionário para todos os estudantes dos cursos de Letras e de Tecnologia em Análise e Desenvolvimento de Sistemas, que tiveram um menor número de matriculados no 2º semestre de 2014. Com isso, gerou-se um acréscimo de 4 questionários ao número total de questionários previstos, ou seja, no final, foram aplicados 229 questionários. Esse fato explica o aumento de 4 questionários na coluna “Número de questionários respondidos” da Tabela 23.

Os questionários foram entregues de forma aleatória para os estudantes do 2º período que participaram das aulas nos dias da pesquisa. Assim, obtivemos 204 questionários respondidos e, apenas 25 não respondidos. Neste estudo, consideramos “Questionários não respondidos” sendo aqueles questionários que foram entregues aos alunos presentes e estes entregaram-no sem responder; recusa do aluno presente a receber o questionário e; a quantidade de alunos insuficiente na sala de aula durante a aplicação dos questionários, não sendo possível o retorno à sala do aluno faltante por indisponibilidade de tempo.

Como pode-se observar na Tabela 23, os cursos de História, Tecnologia em Gestão de Recursos Humanos e Tecnologia em Logística obtiveram os menores índices de questionários respondidos 50%, 60% e 60%, respectivamente. É importante destacar que a modalidade de ensino à distância possibilita que o aluno estude em outros locais e em horários flexibilizados. Desse modo, não é necessário que ele sempre compareça ao polo de apoio presencial para assistir as aulas. Esse fato pode ter contribuído para obtenção de 11% dos questionários não respondidos dentre os 225 previstos inicialmente para coleta, de acordo com o plano amostral.

Portanto, o banco de dados desta pesquisa é formado por 204 questionários socioeconômicos respondidos. Esses dados foram analisados por meio do programa *R-project*, um pacote do *software* R, que permite o tratamento de dados multivariados de diferentes tipos de variáveis (quantitativa ou qualitativa) e de estrutura de dados, além de gerar relatórios, gráficos e outros, de acordo com a necessidade do pesquisador.

7 RESULTADOS E ANÁLISE DOS DADOS

7.1 Análises descritivas do banco de dados

Do total de 225 alunos selecionados por meio da amostragem estratificada, 204 alunos responderam o questionário socioeconômico. Esse número representa 90,67% de questionários respondidos. No Gráfico 1, apresentamos a distribuição de questionários respondidos por curso.

Como pode-se observar no Gráfico 1, o índice percentual de questionários respondidos pelos alunos selecionados na amostra é de 26%, seguido pelos cursos de Educação Física, com 10,3%, Estética e Imagem Pessoal e Serviço Social, com 9,8%. Por outro lado, os cursos de História, Letras e Análise de Desenvolvimentos de Sistemas representam, respectivamente, 2,0%, 2,0% e 1,5% dos questionários respondidos pelos alunos entrevistados.

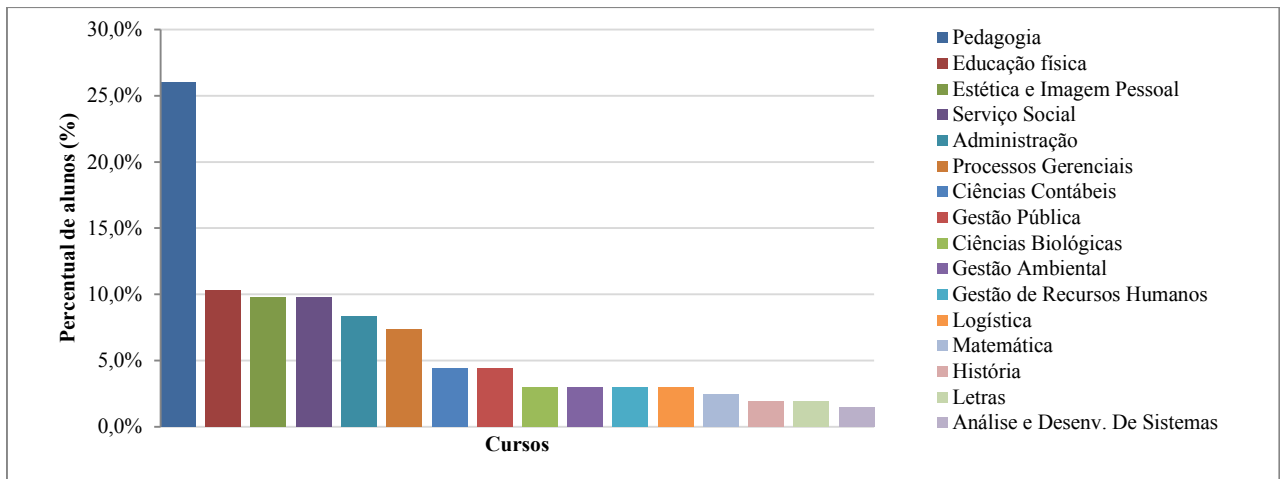


Gráfico 1 – Distribuição de questionários respondidos por curso

Fonte: Dados da pesquisa

No Gráfico 2, observa-se que 45,6% dos alunos entrevistados cursam a graduação voltada para o título de Licenciatura. Com relação aos graus de Bacharelado e Tecnológico têm-se 22,5% e 31,9% respectivamente, dos entrevistados.

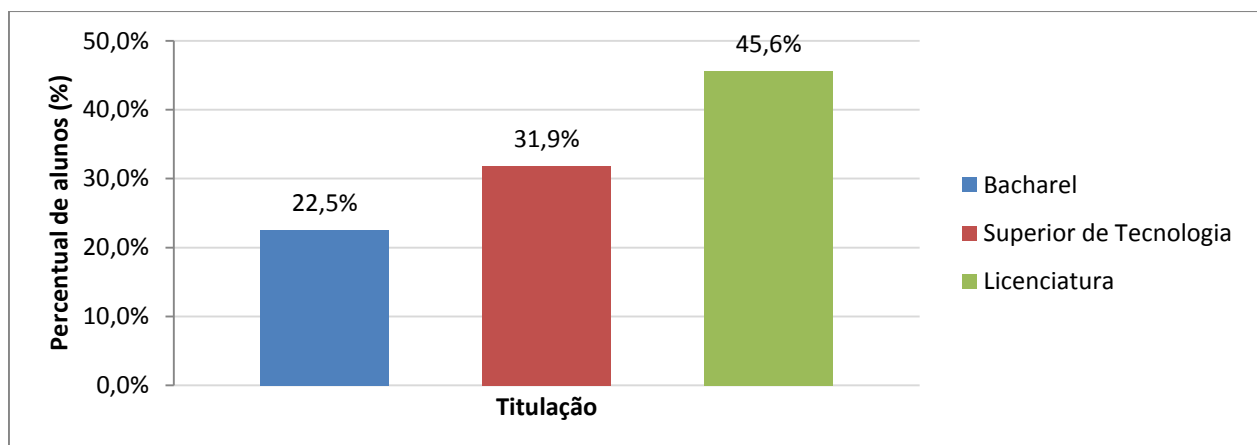


Gráfico 2 – Distribuição de cursos por titulação

Fonte: Dados da pesquisa

Na pesquisa, observa-se também um grande público feminino estudando na modalidade de educação à distância. Dos 204 entrevistados, 70,1% são do sexo feminino, como pode-se observar no Gráfico 3.

Esse dado corrobora com o dado do Anuário Brasileiro Estatístico de Educação Aberta e a Distância (ABRAEAD), edição 2008, ao qual identifica grande participação feminina em cursos de nível superior na modalidade à distância.

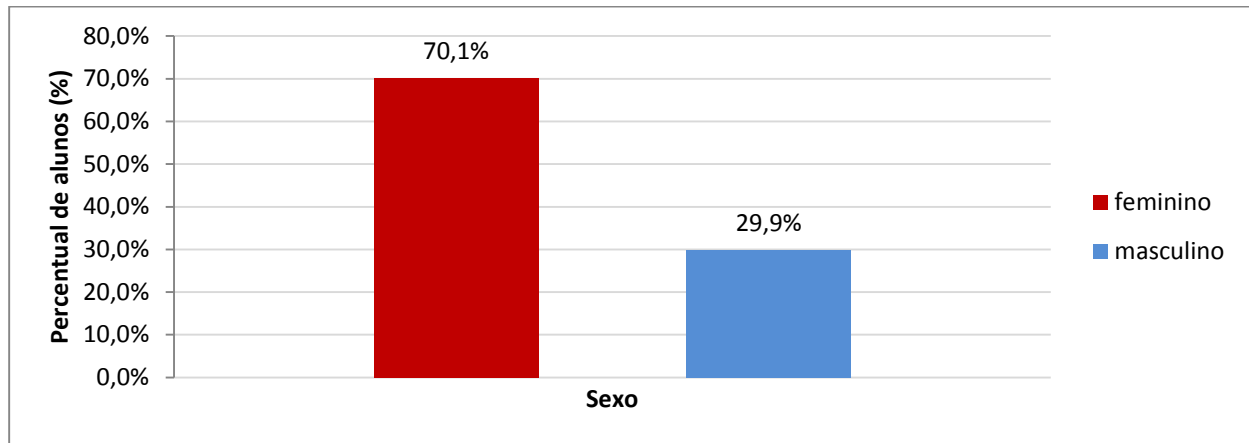


Gráfico 3 – Distribuição de alunos pesquisados por sexo

Fonte: Dados da pesquisa

Quanto à faixa etária dos alunos entrevistados, constata-se que a maioria tem idade acima de 33 anos. A autora Schnitman (2010) caracteriza o aluno que estuda na modalidade à distância como sendo um ser adulto, mais maduro, que percebe na educação à distância uma possibilidade de dar continuidade nos seus estudos. Portanto, no Gráfico 4, nota-se que a maior parte dos alunos entrevistados tem mais de 33 anos, com 49,5%, e uma menor proporção de alunos está na faixa etária entre 18 a 22 anos (15,7%).

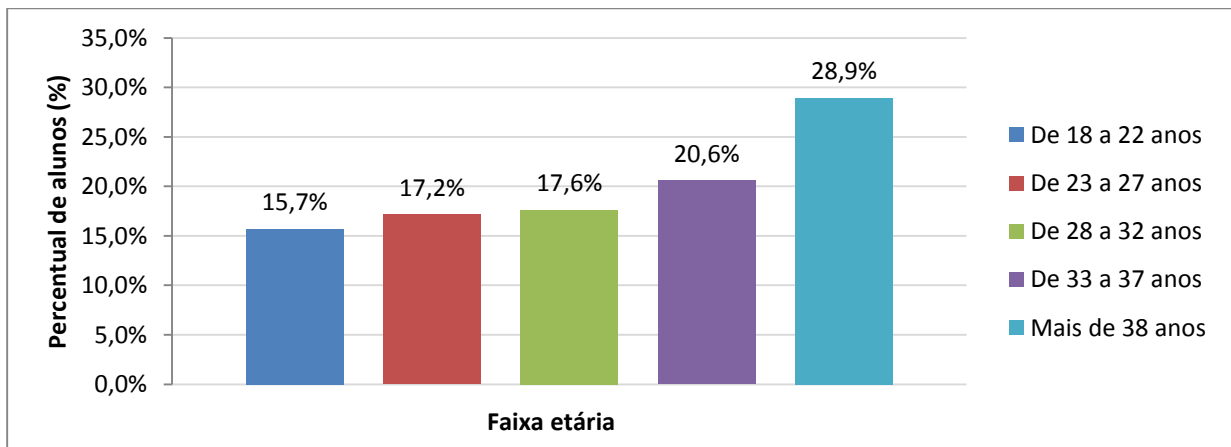


Gráfico 4 – Distribuição de alunos pesquisados por faixa etária

Fonte: Dados da pesquisa

Com relação à raça declarada pelos estudantes entrevistados, pode-se observar, no Gráfico 5, que 52% dos alunos se declararam pardos, seguido de alunos que se autodeclararam brancos (32,4%). Observa-se também menor participação de alunos entrevistados que se declararam das raças Indígenas e Amarelas, com 1,0% e 0,5%, respectivamente.

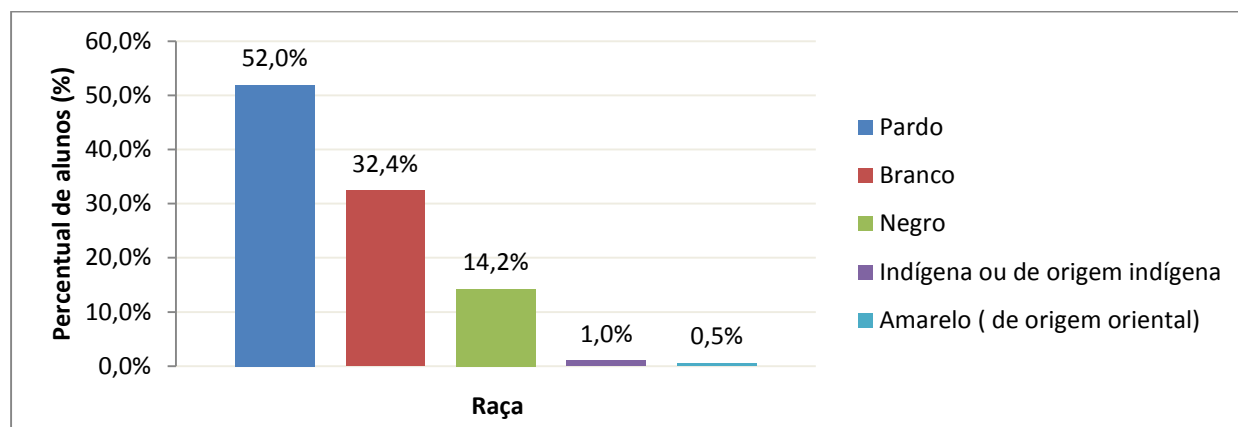


Gráfico 5 – Distribuição de alunos pesquisados por raça

Fonte: Dados da pesquisa

Com base na análise dos dados referente à cidade de residência dos alunos entrevistados, observa-se que 60,8% dos estudantes residem na cidade de Betim, onde localiza-se o polo de apoio presencial. Entretanto, a abrangência de atuação do polo não se limita ao município-sede, pois nota-se também alunos oriundos das cidades circunvizinhas, como Igarapé, Brumadinho e Contagem, que obtiveram índices consideráveis de estudantes entrevistados dessas cidades, enquanto que, o restante teve números menos expressivos, como mostra a Tabela 24.

Tabela 24 – Distribuição de frequência de alunos pesquisados de acordo com a cidade de residência

Cidade	Frequência	%
Belo Horizonte	1	0,50
Buenópolis	1	0,50
Divinópolis	1	0,50
Raposos	1	0,50
Mário Campos	2	1,00
Sarzedo	3	1,50
Mateus Leme	4	2,00
Rio Manso	4	2,00
Juatuba	5	2,50
São Joaquim de Bicas	5	2,50
Florestal	7	3,40
Contagem	8	3,90
Esmeraldas	9	4,40
Brumadinho	11	5,40
Igarapé	18	8,80
Betim	124	60,80
Total	204	100,00

Fonte: Dados da pesquisa

Já no que tange à estrutura familiar dos estudantes entrevistados, na Tabela 25, observa-se que 53,4% são casados ou em união estável. Observa-se também que 35,8% dos alunos entrevistados são solteiros, enquanto que, os divorciados e separados são 10,8%.

Tabela 25 - Distribuição de frequência de alunos pesquisados de acordo com o estado civil

Estado civil	Frequência	%
Divorciado(a)/Separado(a)	22	10,80%
Solteiro(a)	73	35,80%
Casado(a) ou União estável	109	53,40%
Total	204	100,00%

Fonte: Dados da pesquisa

Na Tabela 26, verifica-se também que cerca de 51,5 % dos entrevistados têm de 1 a 4 filhos, tendo como maior índice a família formada por no máximo 2 filhos, com 42,6%.

Tabela 26 - Distribuição de alunos pesquisados de acordo com o número de filhos

Número de Filhos	Frequência	%
0	99	48,5%
1	50	24,5%
2	37	18,1%
3	15	7,4%
4	3	1,5%
Total	204	100,0%

Fonte: Dados da pesquisa

Quando observado a ocupação dos estudantes entrevistados, dos 204 alunos entrevistados, 87% desenvolvem atividades remuneradas. Desse total, 39,2% trabalham em empresas privadas e 27,5% no serviço público. Como pode-se observar no Gráfico 6, o salário dos estudantes entrevistados que trabalham, verifica-se que a maioria recebe até 03 salários mínimos (R\$724,00 corresponde ao salário mínimo do ano de 2014).

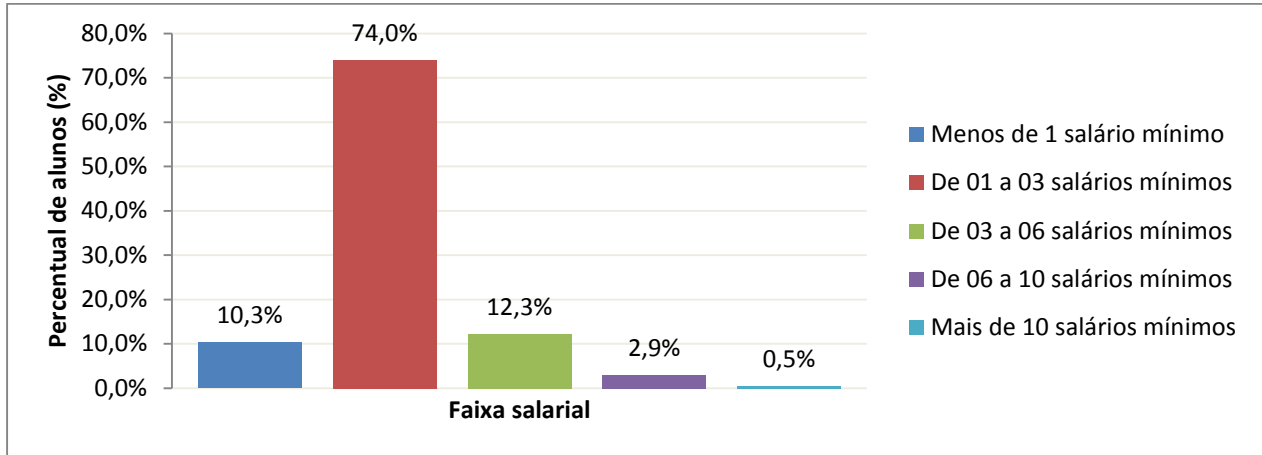


Gráfico 6 – Distribuição de alunos pesquisados por faixa salarial

Fonte: Dados da pesquisa

No que diz respeito ao ensino médio dos alunos entrevistados, na pesquisa, observa-se que 91,7% estudaram em escola pública e 73,5% fizeram o ensino médio tradicional, bem como 38,7% frequentaram o curso no período noturno.

Com relação ao conhecimento em língua estrangeira, constata-se que 92,6% desses alunos não fazem algum curso de idioma e, dos 7,4% que cursam ou já cursaram algum idioma, há maior predominância na língua inglesa, como pode-se verificar no Gráfico 07.

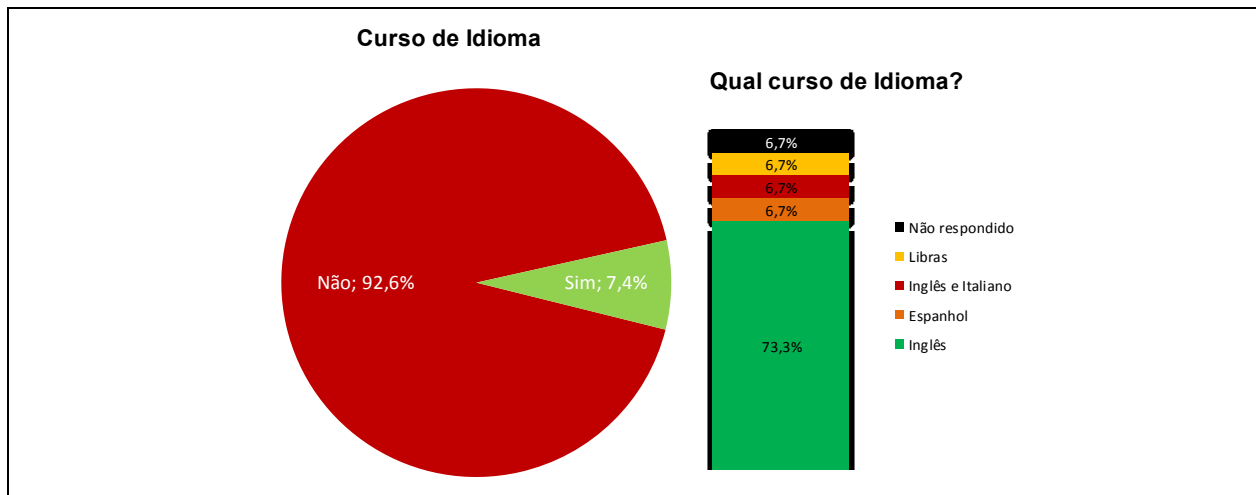


Gráfico 7 – Distribuição de alunos pesquisados por curso de idioma

Fonte: Dados da pesquisa

Na pesquisa, observa-se também um grande número de estudantes entrevistados cursando sua primeira graduação na modalidade de ensino à distância, cerca de 94,1% do total de 204 alunos da amostra (Gráfico 8).

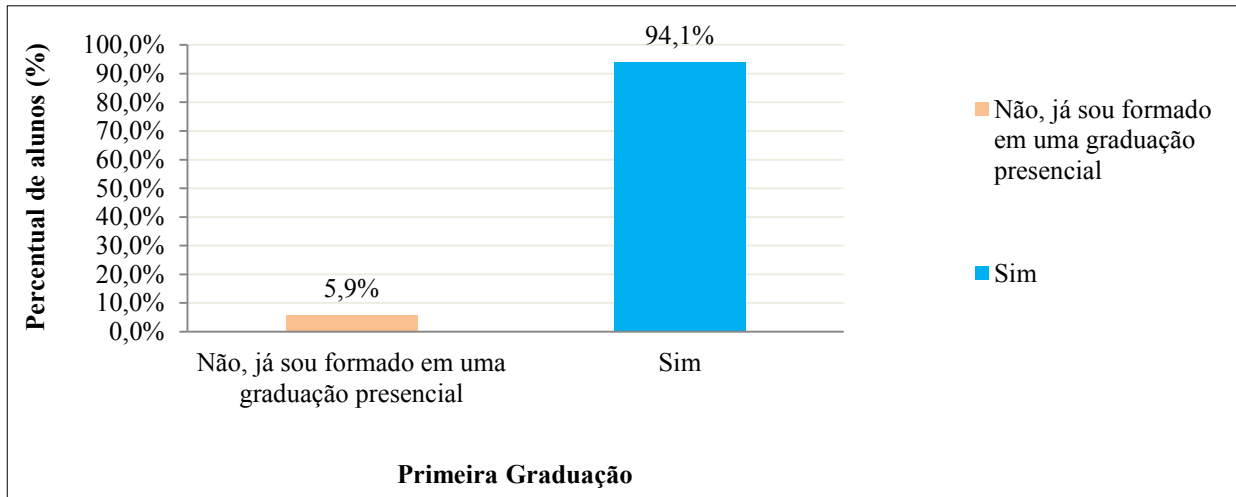


Gráfico 8 - Distribuição de alunos pesquisados por primeira graduação

Fonte: Dados da pesquisa

Contudo, mesmo que pequeno o índice de incidência de alunos entrevistados já formados em algum curso presencial (5,9%), nota-se que a modalidade de ensino à distância pode ser uma opção para as pessoas que tem graduação e desejam fazer outro curso. Para os alunos que já concluíram algum curso superior, na pesquisa, constatou-se que os cursos de maior predominância são: Administração; Direito; Pedagogia ; e Tecnologia em Logística.

Outro ponto abordado no questionário foi sobre o motivo que levou o estudante pesquisado a optar por um curso de graduação à distância. Como pode-se observar na Tabela 27, a falta de tempo é um dos motivos que mais influencia o aluno na hora de escolher a modalidade de ensino à distância para cursar uma graduação (77,5%). Vale ressaltar que motivos relacionados às questões financeiras, à distância das Instituições de Ensino Superior e à facilidade de estudar em um curso EAD também foram mencionados pelos alunos como sendo fatores de escolha para fazer um curso de graduação na modalidade à distância.

Além desses, outros fatores citados pelos alunos pesquisados como, por exemplo, sentir constrangimento em frequentar um curso com pessoas mais jovens, adquirir conhecimento, ganhar uma bolsa do Programa Universidade para Todos e por não ter pessoas para cuidar dos filhos todas as noites, mostra que EAD tem sido uma opção para as pessoas que possuem limitações de tempo ou por algum motivo pessoal não conseguem frequentar um curso presencial.

Tabela 27 – Distribuição de alunos pesquisados por motivo da escolha de um curso na modalidade à distância

	Frequência	%
Facilidade de poder estudar em qualquer lugar do Brasil	2	0,01
Não respondida	3	0,01

Outros	4	0,02
Achei que fosse mais fácil que o curso presencial	10	0,05
Por questões financeiras	13	00,1
Distância das instituições de Ensino Superior	14	6,90
Falta de tempo para frequentar diariamente um curso	158	77,5
Total	204	100,0%

Fonte: Dados da pesquisa

Nesta pesquisa, percebe-se também que a maioria dos alunos entrevistados tem familiaridade com o uso do computador e tem acesso à internet para desenvolver as atividades acadêmicas. Com base nos dados, constatou-se que 76,9% dos alunos entrevistados afirmaram ter entre bom e ótimo nível de conhecimento na utilização do computador. No que diz respeito ao acesso à internet para desenvolvimento das atividades acadêmicas, 77,5% dos entrevistados acessam em casa e, apenas 14,7%, no trabalho.

O fato dos alunos entrevistados acessarem mais a internet de casa vem de encontro com a afirmação dos autores Demarco e Honorato (2013). Para os autores, a chegada de nova tecnologia e a expansão do acesso à internet propiciou a interatividade entre professor e aluno e, além disso, trouxe mais comodidade aos usuários, assim, o aluno pode conciliar seu trabalho, seus estudos e suas atividades familiares.

Na EAD, é imprescindível que o aluno tenha acesso à internet para desenvolver trabalhos e as atividades no ambiente virtual de aprendizagem. Percebe-se que a maioria dos alunos entrevistados dedica uma parte de seu tempo, fora da instituição de ensino, aos estudos. Verifica-se que 81,4% desses alunos estudam até 5 horas semanais, enquanto que apenas 5% não dedicam nenhuma hora além das aulas assistidas no polo.

Com relação à perspectiva profissional futura dos alunos entrevistados, na Tabela 28, pode-se constatar que 66,2% dos estudantes almejam trabalhar na área a qual está cursando a graduação, bem como manter-se no emprego. Isso mostra que a Educação à distância possibilita o desenvolvimento de estudantes que buscam à formação profissional e à qualificação no trabalho (PALLOF; PRATT, PAUL citados por BELLONI,2006).

Tabela 28 – Distribuição de alunos pesquisados por perspectiva profissional futura

	Frequência	%
Trabalho em outra área, mas pretendo buscar uma atividade na minha área de graduação	76	37,3
Já tenho trabalho na área e pretendo continuar nele	59	28,9
Vou me dedicar à atividade acadêmica e buscar um curso de pós-graduação	35	17,2
Vou prestar concurso para atividade em empresa pública	26	12,7

Pretendo trabalhar em empresa privada	5	2,5
Outras	3	1,5
Total	204	100,0%

Fonte: Dados da pesquisa

Além disso, percebe-se também que os alunos pesquisados tem como perspectiva futura profissional de seguir à atividade acadêmica e fazer um curso de uma pós-graduação (17,2%), prestar concursos público (12,7%) e trabalhar em uma empresa privada (2,5%).

Na pesquisa, observou-se também que 1,5% dos estudantes tem outras perspectivas quanto a profissão no futuro. Enquanto que alguns visam utilizar a graduação para agregar à profissão, outros, ainda não decidiram se atuarão em outras áreas.

Outro ponto tratado nesta pesquisa foi a forma como o aluno conheceu a Instituição de Ensino Superior. Essa informação é muito importante para a Instituição de Ensino identificar os tipos de mídias de divulgação mais eficientes e, assim, focar suas estratégias de marketing.

Na pesquisa, identificou-se que, além das mídias de massa (televisão, rádio) utilizadas para divulgar os cursos de graduação da Instituição de Ensino Superior pesquisada, a indicação de amigos ou parentes mostra-se como uma possível ferramenta estratégica de divulgação. Cerca de 48,5% dos alunos entrevistados conheceram o Instituição de Ensino por meio de anúncios em internet, *outdoor*, panfletos, televisão e entre outros meios de comunicação. Entretanto, a propaganda “boca a boca” também foi um ótimo influenciador, tanto através de parentes como de alunos da própria instituição (44,7%). No quesito de grau de satisfação dos alunos entrevistados sobre o curso de graduação, 75% dos alunos alegaram estarem satisfeitos com o curso e apenas 8,3 % alegaram insatisfação com o curso.

De maneira geral, o perfil do aluno entrevistado corrobora com a pesquisa realizada pela ABRAEAD, edição 2008, sobre o perfil do aluno EAD no Brasil. No presente trabalho, constatou-se que o aluno típico estudou o ensino médio em escola pública durante o período noturno, tem idade acima de 33 anos, trabalha e estuda, é casado, tem filhos, recebe até 3 salários mínimos e não tem conhecimento em outro idioma. Outro ponto interessante na pesquisa que também confirma o levantamento da ABRAED sobre o perfil do aluno EAD, é a predominância do público do sexo feminino nos cursos de graduação à distância na Instituição de Ensino Superior pesquisada.

7.2 Teste Qui-Quadrado do conjunto de dados

Para avaliar a existência ou não de dependência das variáveis estudadas, assumimos o nível de significância de 5% e foram formuladas as seguintes hipóteses:

H_0 : Não há associação entre as variáveis estudadas, ou seja, essas são independentes.

H_a : Há associação entre as variáveis, ou seja, essas são dependentes.

Assim, o cálculo da estatística χ^2 foi realizado entre as categorias de cada par de variáveis do conjunto de dados gerado a partir do questionário socioeconômico (Apêndice A). Portanto, em vez de analisarmos se $\chi_{cal}^2 > \chi_c^2$, verificamos se o nível de significância do χ_{cal}^2 é menor que 0,05 (5%) a fim de darmos continuidade à análise de correspondência. Dessa forma, se o $P\text{-value} < 0,05$ pode-se considerar que existe associação entre as duas variáveis categóricas, rejeitando a hipótese nula.

As probabilidades associadas à estatística χ_{cal}^2 foram apresentadas no quadro referido no Apêndice B, destacando-se os valores inferiores à 0,05 (nível de significância), para demonstrar as variáveis que apontam uma dependência entre elas. A partir dessa análise, resultou a Tabelas 29 que contém as variáveis associadas e seus respectivos $P\text{-value}$, a seguir.

Tabela 29 – Probabilidades associadas ao Qui-Quadrado calculado

Variável (linha)	Variável (coluna)	<i>P-value</i>
Curso	Titulação	0,000
	Sexo	0,000
	Faixa etária	0,002
	Estado civil	0,027
	Ocupação	0,005
	Renda	0,028
	Primeira Graduação	0,015
	Conhecimento computador	0,009
	Perspectiva profissional	0,029
Titulação	Perspectiva profissional	0,004
Sexo	Faixa etária	0,041
	Estado civil	0,033
	Número de filhos	0,033
	Renda	0,001
	Tipo de curso - Ensino Médio	0,009
Idade	Estado civil	0,000
	Número de filhos	0,000
	Tipo de curso - Ensino Médio	0,000
	Período do curso	0,012
	Conhecimento computador	0,001
Raça	Tipo de escola - Ensino Médio	0,036
Estado civil	Número de filhos	0,000
	Tipo de curso - Ensino Médio	0,000

Número de filhos	Tipo de curso - Ensino Médio	0,000
	Local onde acessa a internet	0,019
	Conhecimento da instituição de ensino	0,003
Período do curso - Ensino Médio	Primeira graduação	0,018
	Perspectiva profissional	0,000
	Satisfação com o curso de graduação	0,000
Ocupação	Renda	0,001
	Perspectiva profissional	0,009
	Satisfação com o curso de graduação	0,000
Renda	Curso de Idioma	0,013
	Primeira Graduação	0,000
	Horas de estudo	0,000
	Conhecimento da instituição de ensino	0,002
	Perspectiva profissional	0,000
	Satisfação com o curso de graduação	0,002
Tipo de escola - Ensino Médio	Período do curso - Ensino Médio	0,036
	Curso de idioma	0,000
	Perspectiva profissional	0,000
Tipo de curso - Ensino Médio	Período do curso - Ensino Médio	0,001
Curso de idioma	Primeira graduação	0,037
	Conhecimento da instituição de ensino	0,010
	Horas de estudo	0,000
Primeira graduação	Motivo que escolheu um curso na EAD	0,035
	Conhecimento computador	0,013
	Perspectiva profissional	0,000
	Conhecimento da instituição de ensino	0,000
Motivo que escolheu um curso na EAD	Local onde acessa a internet	0,000
	Horas de estudo	0,026
	Conhecimento da instituição de ensino	0,000
	Perspectiva profissional	0,000
	Satisfação com o curso de graduação	0,000
Local onde acessa a internet	Satisfação com o curso de graduação	0,000
Horas de estudos semanais	Perspectiva profissional	0,003
	Conhecimento da instituição de ensino	0,008
Perspectiva profissional	Conhecimento da instituição de ensino	0,000

7.3 Aplicação da ACM

Nesta seção, apresentamos os resultados obtidos pela ACM por meio das Matrizes Indicadora, de Burt e Burt Ajustado. O objetivo dessa apresentação é fazer um comparativo das principais inércias e verificar qual matriz melhor explica o conjunto de dados desta pesquisa.

A ACM foi aplicada ao conjunto de variáveis que foram consideradas estatisticamente associadas pelo teste qui-quadrado na seção 7.2. Os resultados da ACM obtidos por meio das Matrizes Indicadora, de Burt e Burt Ajustado foram apresentados na Tabela 30.

Nessa tabela, identificamos que a percentagem das inércias acumuladas nas primeiras dimensões das Matrizes Indicadora e de Burt são muito reduzidas em relação aos valores obtidos quando feito o ajustamento das inércias (Burt Ajustado). Nesse último, o percentual acumulado das principais inércias é de 29,4% de explicação dos dados no espaço determinado pelas duas primeiras dimensões do mapa de correspondência, enquanto que, nas Matrizes Indicadora e de Burt, têm-se apenas 8,7% e 19,9%, respectivamente.

Tabela 30 – Comparativo das principais inércias e autovalores da Análise Geral

Principais Inércias (autovalores) : Matriz Indicadora				Principais Inércias (autovalores) : Matriz de Burt				Principais Inércias (autovalores) : Burt Ajustado			
Dimensão	Autovalor	%	acumulado %	Dimensão	Autovalor	%	acumulado %	Dimensão	Autovalor	%	acumulado %
1	0,214601	4,6	4,6	1	0,046054	11,0	11,0	1	0,028359	16,9	16,9
2	0,191876	4,1	8,7	2	0,036816	8,8	19,9	2	0,020834	12,4	29,4

Na Tabela 31, mostra-se as dimensões geradas pela matriz para as análises dos dados.

Tabela 31 – Resultado das principais inércias e autovalores da análise geral

Matriz Indicadora				Matriz de Burt				Burt Ajustado			
Dimensão	Autovalor	%	Acumulado %	Dimensão	Autovalor	%	Acumulado %	Dimensão	Autovalor	%	Acumulado %
1	0,215000	4,6	4,60	1	0,046054	11	11,00	1	0,028000	16,90	16,90
2	0,192000	4,1	8,70	2	0,036816	8,8	19,90	2	0,021000	12,40	29,40
3	0,168000	3,6	12,30	3	0,028217	6,8	26,60	3	0,014000	8,50	37,90
4	0,151000	3,2	15,50	4	0,022671	5,4	32,00	4	0,010000	6,00	43,90
5	0,137000	2,9	18,50	5	0,018881	4,5	36,60	5	0,008000	4,50	48,40
6	0,123000	2,6	21,10	6	0,015136	3,6	40,20	6	0,005000	3,00	51,40
7	0,117000	2,5	23,60	7	0,013754	3,3	43,50	7	0,004000	2,60	54,00
8	0,113000	2,4	26,10	8	0,004000	2,20	56,20
9	0,109000	2,3	28,40	84	0	0	100,00	9	0,003000	1,90	58,10
10	0,106000	2,3	30,70					10	0,003000	1,70	59,80
...
84	0,000000	0	100,00					36	0,000000	0,00	70,40
Total	4,666667			Total	0,417357			Total	0,167397		

Para facilitar a visualização das associações do conjunto de dados estudado, não foram colocados os pontos referentes às categorias de cada variável. Dessa forma, pôde-se verificar as

associações das variáveis pela proximidade dos pontos gerados na Figura 5. Vale ressaltar que, nessa figura, destacamos apenas as variáveis que foram consideradas associadas pelo teste qui-quadrado.

Assim, no lado direito da Figura 5, percebe-se que existe uma associação entre as variáveis categóricas Motivo de escolha curso EAD, Perspectiva profissional e Conhecimento da Instituição de Ensino. Já no lado superior à esquerda da figura, observamos uma relação entre as variáveis Idade e Curso. Na parte inferior à direita, identificamos uma associação entre as variáveis Renda, Horas de estudo, Nível de conhecimento em computador e Primeira graduação.

E por fim, verificamos também associações entre as variáveis Estado civil, Ocupação, Número de filhos, Sexo, Raça, Titulação, Tipo de curso do Ensino Médio, Período de curso do Ensino médio e Local de acesso à internet.

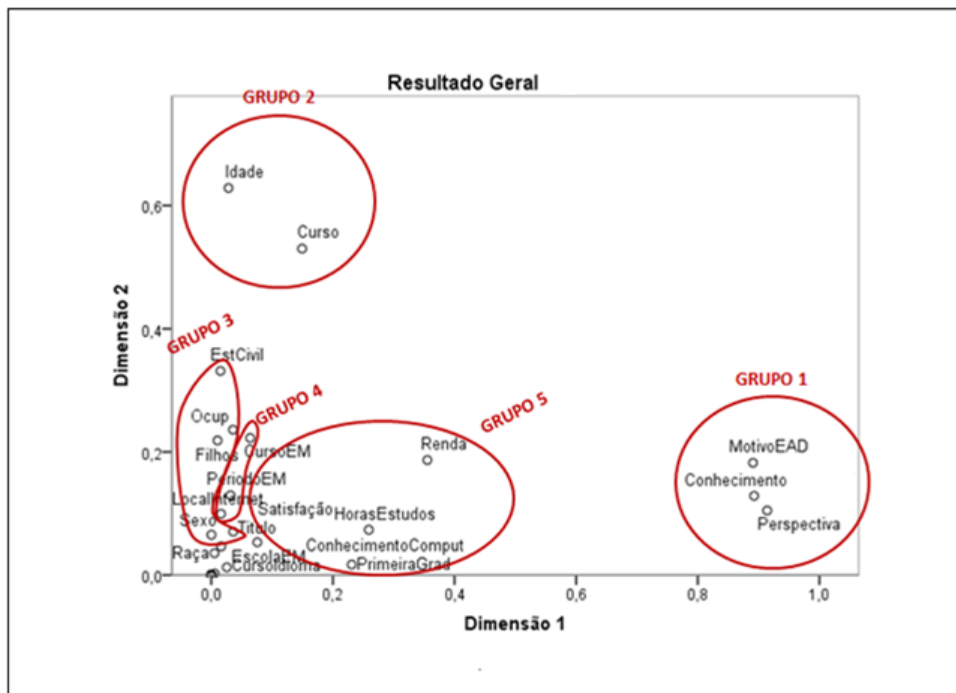


Figura 5 – Análise de Correspondência Geral

Diante do exposto e para continuar com a aplicação da ACM, considerou-se a formação de cinco grupos, a seguir : Grupo 1 (Motivo de escolha curso EAD, Perspectiva profissional e Conhecimento da Instituição de Ensino; Grupo 2 (Idade e Curso); Grupo 3 (Renda, Horas de estudo, Nível de conhecimento em computador e Primeira graduação; Grupo 4 (Estado civil, Ocupação, Número de filhos, Sexo, Titulação, Tipo de curso do Ensino Médio, Período de curso do Ensino médio, Local de acesso à internet) e Grupo 5 (Satisfação, Horas de estudo, Nível de conhecimento em computador, Primeira graduação e Renda).

Nas análises dos dados dos grupos, optou-se unir os Grupos 3 e 4 para verificar o comportamento, devido à proximidade dos pontos desses dois grupos, como observado na Figura 5.

Na próxima seção, apresentaremos a ACM dos grupos formados por meio do método de Burt Ajustado, bem como seus gráficos.

7.3.1 ACM do Grupo 1

Para análise das variáveis do Grupo 1, as categorias foram codificadas e estão apresentadas no Quadro 1.

Quadro 1 - Codificações das categorias das variáveis consideradas na ACM do Grupo 1

Variável	Categoria	Código
Motivo da escolha curso EAD	Achei que fosse mais fácil que o curso presencial	MOT1
	Constrangimento em frequentar o curso junto com pessoas mais jovens	MOT2
	Distância das instituições de Ensino Superior	MOT3
	Facilidade de poder estudar em qualquer lugar do Brasil	MOT4
	Falta de tempo para frequentar diariamente um curso	MOT5
	Faz o curso só para adquirir conhecimento	MOT6
	Por questões financeiras	MOT7
	Porque ganhou uma bolsa do PROUNI	MOT8
	Porque não tem pessoas para cuidar dos filhos todas as noites	MOT9
	Não respondeu	MOT10
Perspectiva profissional futura	Agregar a atual graduação à minha profissão de advogado	PP1
	Ainda não decidi se vou atuar em outra área	PP2
	Desempregada	PP3
	Já tenho trabalho na área e pretendo continuar nele	PP4
	Pretendo trabalhar em empresa privada	PP5
	Trabalho em outra área, mas pretendo buscar uma atividade na minha área de graduação	PP6
	Vou prestar concurso para atividade em empresa pública	PP7
Conhecimento da Instituição de Ensino Superior	Alunos da instituição	U1
	Divulgação em minha escola	U2
	Empresa em que trabalha	U3
	Indicação de amigos	U4
	Internet, Outdoor, Panfletos, Televisão e outros meios de comunicação	U5
	Parentes	U6
	Não respondeu	U7

Como pode-se observar na Tabela 32, o método do Burt Ajustado tem maior o percentual acumulado de variação de explicação nas duas primeiras dimensões, de 92,4%, enquanto que, nas soluções sem ajuste tem-se apenas 20,8% e 39,2%, respectivamente.

Tabela 32 – Comparativo das principais inércias e autovalores do Grupo 1

Principais Inércias (autovalores) : Matriz Indicadora				Principais Inércias (autovalores) : Matriz de Burt				Principais Inércias (autovalores) : Burt Ajustado			
Dimensão	Valor	%	acumulado %	Dimensão	Valor	%	acumulado %	Dimensão	Valor	%	acumulado %
1	10000000	14,3	14,3	1	10000000	32,5	32,5	1	10000000	89,5	89,5
2	0,453543	6,5	20,8	2	0,205701	6,7	39,2	2	0,032513	2,9	92,40

Na Tabela 33, são apresentados os resultados das inércias gerados nas dimensões mais relevantes das análises e seus respectivos totais.

Tabela 33 – Resultado da inércia total do Grupo 1

Matriz Indicadora				Matriz de Burt				Burt Ajustado			
Dimensão	Valor	%	Acumulado %	Dimensão	Valor	%	Acumulado %	Dimensão	Valor	%	Acumulado %
1	10000000	14,3	14,3	1	10000000	32,5	32,5	1	10000000	89,5	89,5
2	0,453543	6,50	20,8	2	0,205701	6,70	39,2	2	0,032513	2,90	92,4
3	0,417611	6,00	26,7	3	0,174399	5,70	44,8	3	0,015981	1,40	93,9
4	0,412657	5,90	32,6	4	0,170286	5,50	50,4	4	0,014157	1,30	95,1
5	0,400951	5,70	38,4	5	0,160762	5,20	55,6	5	0,010287	0,90	96
6	0,387505	5,50	43,9	6	0,15016	4,90	60,5	6	0,006603	0,60	96,6
7	0,367927	5,30	49,1	7	0,135371	4,40	64,9	7	0,002693	0,20	96,9
8	0,35017	5,00	54,1	8	0,122619	4,00	68,9	8	0,000638	0,10	96,9
9	0,340844	4,90	59,0	9	0,116175	3,80	72,6	9	0,000127	0,00	96,9
10	0,333333	4,80	63,8	10	0,111111	3,60	76,2				
11	0,333333	4,80	68,5	11	0,111111	3,60	79,8				
12	0,333333	4,80	73,3	12	0,111111	3,60	83,5				
13	0,329118	4,70	78,0	13	0,108319	3,50	87				
14	0,305411	4,40	82,4	14	0,093276	3,00	90				
15	0,277027	4,00	86,3	15	0,076744	2,50	92,5				
16	0,264533	3,80	90,1	16	0,069978	2,30	94,8				
17	0,252517	3,60	93,7	17	0,063765	2,10	96,8				
18	0,232239	3,30	97,0	18	0,053935	1,80	98,6				
19	0,207946	3,00	100	19	0,043242	1,40	100				
Total:	7,000000	100,0		Total	3,078063	100,0		Total	1,117095	96,9	

Os resultados das análises gerados por meio da Burt Ajustado estão representados na Tabela 34. Nessa tabela, exibimos os seguintes os valores:

- Massa (mass) é o peso relativo de cada categoria dentro da respectiva variável.
- Qualidade (qlt) de representação de um ponto em um subespaço, sendo calculada pela soma das contribuições absolutas.
- Inércia (inr) representa a proporção de contribuição de cada ponto na inércia total.
- k é a coordenada de cada categoria na coordenada principal.
- Contribuição absoluta (cor) indica o quanto um ponto contribui para a inércia principal do eixo em estudo.
- Contribuição relativa (ctr) permite encontrar quais são os pontos mais importantes na orientação da coordenada principal.

Tabela 34 – Resultado da ACM do grupo 1 por meio da Burt Ajustado (x1000)

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	PP1	2	507	36	-70	7	0	-581	500	17
2	PP2	2	1.000	108	14.248	1000	332	0	0	0
3	PP3	2	507	36	-70	7	0	-581	500	17
4	PP4	96	254	27	-70	84	0	-99	169	29
5	PP5	8	75	36	-70	9	0	-189	66	9
6	PP6	181	561	17	-70	215	1	89	346	44
7	PP7	42	54	33	-70	26	0	-74	28	7
8	MOT1	16	422	37	-70	8	0	521	415	136
9	MOT2	2	26	36	-70	3	0	182	23	2
10	MOT3	23	124	36	-70	12	0	217	112	33
11	MOT4	3	19	36	-70	5	0	-115	14	1
12	MOT5	258	659	9	-70	290	1	-79	369	50
13	MOT6	2	1.000	108	14.248	1000	332	0	0	0
14	MOT7	21	17	36	-70	11	0	51	6	2
15	MOT8	2	19	36	-70	5	0	-115	14	1
16	MOT9	2	714	37	-70	1	0	1688	713	143
17	MOT10	5	446	37	-70	5	0	684	441	70
18	U1	108	105	26	-70	95	1	-24	11	2
19	U2	3	4	37	-70	3	0	-53	2	0
20	U3	8	86	36	-70	13	0	166	73	7
21	U4	10	30	37	-70	7	0	128	23	5
22	U5	162	451	20	-70	101	1	-131	350	85
23	U6	41	737	35	-70	13	0	520	723	339
24	U7	2	1.000	108	14.248	1000	332	0	0	0

Para interpretação da Tabela 34 e análise do Mapa de Correspondência, foram considerados os pontos com maior contribuição absoluta, pontos com a contribuição relativa alta, bem como os pontos mais bem representados no subespaço ótimo e os pontos com maior massa.

Dessa forma, os valores destacados nessa tabela estão relacionados aos pontos que mais contribuíram para explicação dos dados e para formação das coordenadas principais (CP). As

coordenadas principais formadas foram as seguintes na CP1 (PP2, MOT6, U7). Na CP2, as combinações foram (PP1, PP3, MOT1, MOT5, MOT9, MOT10, U5, U6).

Na Figura 6, observa-se as relações entre as categorias PP2 (Ainda não decidi se vou atuar em outra área); MOT6 (Faz o curso só para adquirir conhecimento); U7 (Não respondeu, pontos localizados à direita do gráfico). Esses pontos estão associados às respostas dos alunos pesquisados que escolheram fazer uma graduação à distância para adquirir conhecimento, ainda não decidiram se atuarão em outra área e também não responderam por qual meio de comunicação que obteve conhecimento sobre a Instituição de Ensino Superior que estuda.

Verificamos também uma associação entre os pontos MOT1 (Achei que fosse mais fácil que o curso presencial) e U6 (Parentes). Diante disso, pode-se constatar que os alunos pesquisados, que escolheram um curso na modalidade à distância por acharem mais fácil que o presencial, conheceram a Instituição de Ensino Superior através de indicação de parentes.

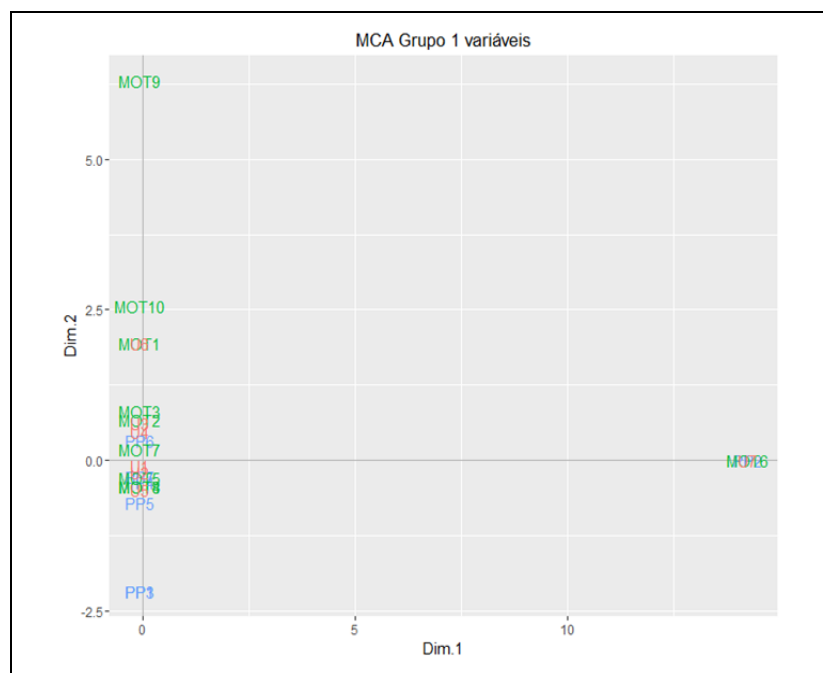


Figura 6 – Mapa de Correspondência do Grupo 1

Nas próximas seções, apresentaremos as análises dos Grupo 2, Grupos 3 e 4, Grupo 5 de forma análoga do Grupo 1, uma vez que aplicou-se a mesma técnica de Análise de Correspondência Múltipla por meio da Matriz Ajustada para todos os grupos.

7.3.2 ACM do Grupo 2

Para análise das variáveis do grupo 2, as categorias foram também codificadas e estão apresentadas no Quadro 2.

Quadro 2 - Codificações das categorias das variáveis consideradas na ACM do Grupo 2

Variável	Categoria	Código
Curso	Administração	C1
	Análise e Desenvolvimento De Sistemas	C2
	Ciências Biológicas	C3
	Ciências Contábeis	C4
	Educação física	C5
	Estética e Imagem Pessoal	C6
	Gestão Ambiental	C7
	Gestão de Recursos Humanos	C8
	Gestão Pública	C9
	História	C10
	Letras	C11
	Logística	C12
	Matemática	C13
	Pedagogia	C14
	Processos Gerenciais	C15
	Serviço Social	C16
Faixa etária	De 18 a 22 anos	ID1
	De 23 a 27 anos	ID2
	De 28 a 32 anos	ID3
	De 33 a 37 anos	ID4
	Mais de 38 anos	ID5

Conforme os dados da Tabela 35, a análise do Grupo 2 com duas coordenadas principais é melhor explicada pelo Burt Ajustado, com percentual acumulado de 76,10% da inércia total.

Tabela 35 – Comparativo das principais inércias e autovalores do Grupo 2

Principais Inércias (autovalores) : Matriz Indicadora				Principais Inércias (autovalores): Matriz de Burt				Principais Inércias (autovalores): Burt Ajustado			
Dimensão	Autovalor	%	acumulado %	Dimensão	Autovalor	%	acumulado %	Dimensão	Autovalor	%	acumulado %
1	0,746700	7,9	7,9	1	0,557561	11,2	11,2	1	0,243443	51,4	51,4
2	0,670833	7,1	14,9	2	0,450017	9,0	20,2	2	0,116736	24,7	76,10

Na Tabela 36, são apresentados os resultados das inércias gerados nas dimensões mais relevantes das análises e seus respectivos totais.

Tabela 36 - Resultado da inércia total do Grupo 2

Matriz Indicadora				Matriz de Burt				Burt Ajustado			
Dimensão	Valor	%	Acumulad o %	Dimensão	Valor	%	Acumulad o %	Dimensã o	Valor	%	Acumulad o %
1	0,746700	7,9	7,9	1	0,557561	11,2	11,2	1	0,243443	51,4	51,4
2	0,670833	7,1	14,9	2	0,450017	9	20,2	2	0,116736	24,7	76,1
3	0,634266	6,7	21,6	3	0,402294	8,1	28,3	3	0,07211	15,2	91,3
4	0,601451	6,3	27,9	4	0,361744	7,3	35,5	4	0,041169	8,7	100
5	0,500000	5,3	33,2	5	0,250000	5	40,5				
6	0,500000	5,3	38,5	6	0,250000	5	45,6				
7	0,500000	5,3	43,7	7	0,250000	5	50,6				
8	0,500000	5,3	49	8	0,250000	5	55,6				
9	0,500000	5,3	54,2	9	0,250000	5	60,6				
10	0,500000	5,3	59,5	10	0,250000	5	65,6				
11	0,500000	5,3	64,8	11	0,250000	5	70,6				
12	0,500000	5,3	70	12	0,250000	5	75,6				
13	0,500000	5,3	75,3	13	0,250000	5	80,6				
14	0,500000	5,3	80,6	14	0,250000	5	85,7				
15	0,500000	5,3	85,8	15	0,250000	5	90,7				
16	0,398549	4,2	90	16	0,158841	3,2	93,9				
17	0,365734	3,8	93,9	17	0,133761	2,7	96,5				
18	0,329167	3,5	97,3	18	0,108351	2,2	98,7				
19	0,253300	2,7	100	19	0,064161	1,3	100				
Total:	9,500000	100		Total:	4,986729	100		Total:	0,473458	100	

As coordenadas principais foram formadas pelas combinações na CP1 (C1,C5, C14, C16, ID1, ID5). Na CP2, temos (C3, C2, C10, ID2) conforme destaque no Tabela 37.

Tabela 37 – Resultado da ACM do Grupo 2 por meio da Burt Ajustado (x1000)

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	C1	42	563	49	684	561	80	-42	2	1
2	C2	7	962	51	976	414	29	-1123	-1123	79
3	C3	15	899	50	-86	7	0	-988	-988	123
4	C4	22	880	49	472	647	20	-284	-284	15
5	C5	51	700	48	588	699	73	22	22	0
6	C6	49	707	46	196	225	8	287	287	35
7	C7	15	543	50	-591	462	21	247	247	8
8	C8	15	702	50	523	311	17	586	586	43
9	C9	22	411	49	-390	342	14	175	175	6
10	C10	10	990	50	-133	18	1	-974	-974	80
11	C11	10	891	50	843	568	29	635	635	34
12	C12	15	877	50	-365	233	8	-607	-607	46
13	C13	12	573	51	972	567	48	97	97	1
14	C14	130	952	39	-360	922	69	65	65	5
15	C15	37	923	47	-475	801	34	-185	-185	11

16	C16	49	956	47	-499	844	50	182	182	14
17	ID1	78	901	50	890	826	255	268	75	48
18	ID2	86	976	47	277	129	27	-709	847	370
19	ID3	88	323	45	152	60	8	316	262	76
20	ID4	103	291	42	-268	290	30	16	1	0
21	ID5	145	862	41	-549	848	179	71	14	6

Quanto às associações dos pontos, na Figura 7, percebe-se que existem relações entre as variáveis “faixa etária” e “cursos”. Do lado direito no canto inferior da figura, observamos uma associação entre os pontos ID1 (De 18 a 22 anos) com os pontos C1 (Administração), C5 (Educação Física), C13 (Matemática). Isto significa que os alunos pesquisados com faixa etária de 18 a 22 anos tendem a cursar Matemática, Educação Física e Administração. Ainda na parte inferior da figura, percebe-se também uma associação entre os pontos ID3 (De 28 a 32 anos) e C6 (Estética e Imagem Pessoal), ou seja, os alunos pesquisados com faixa etária de 28 a 32 anos tendem a cursa Estética e Imagem Pessoal.

Na parte superior do lado direito da figura, tem-se associação dos seguintes pontos: ID2 (De 23 a 27 anos), C4 (Ciências Contábeis), C2 (Análise e Desenvolvimento de Sistemas). Neste caso, os alunos pesquisados com faixa etária de 23 a 27 anos cursam Ciências Contábeis e Análise de Desenvolvimento de Sistemas. Do lado esquerdo, na parte inferior da figura, observa-se uma associação forte entre as categorias ID4 (De 33 a 37 anos) e ID5 (Mais de 38 anos) com as categorias C7 (Gestão Ambiental), C9 (Gestão Pública), C14 (Pedagogia), C15 (Processos Gerenciais), C16 (Serviço Social). Percebe-se que alunos pesquisados com idade acima de 33 anos cursam Gestão Ambiental, Gestão Pública, Pedagogia, Processos Gerenciais e Serviço Social.

Podemos concluir nesta análise que os alunos pesquisados mais jovens tendem a buscar cursos com duração de 3 ou 4 anos, ou seja, com titulação de Bacharel, enquanto que, os que têm acima de 33 anos tendem a cursar uma graduação com titulação em Superior de tecnologia e sua duração é de 2 anos.

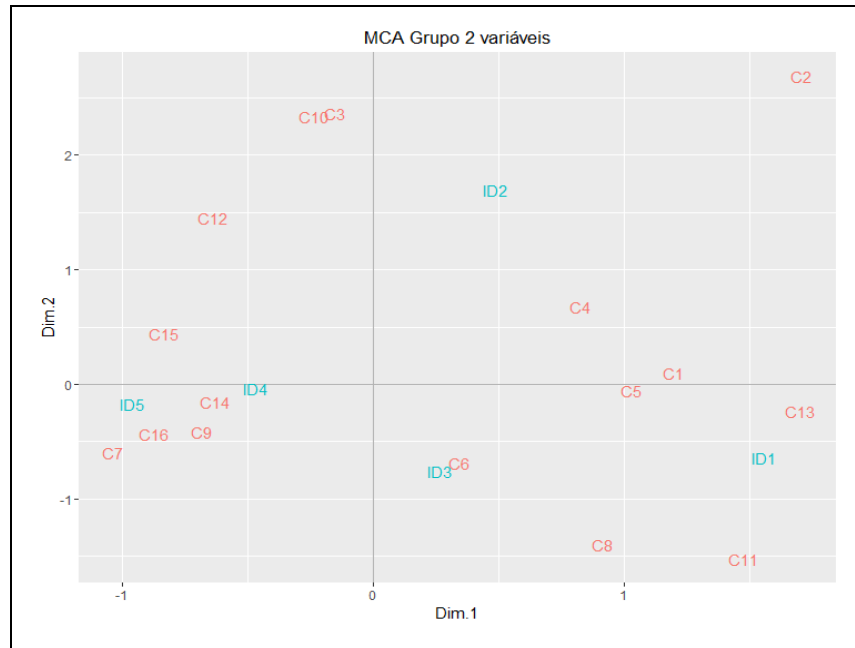


Figura 7 – Mapa de Correspondência do Grupo 2

7.3.3 ACM da união dos Grupos 3 e 4

Apresentamos as codificações dos Grupos 3 e Grupo 4, conforme o Quadro 3:

Quadro 3 - Codificações das categorias das variáveis consideradas na ACM dos Grupos 3 e 4

Variável	Categoria	Código
Sexo	Feminino	F
	Masculino	M
Estado civil	Casado(a) ou União estável	EC1
	Divorciado(a)/Separado(a)	EC2
	Solteiro(a)	EC3
Número de filhos	Nenhum	F0
	Um filho	F1
	Dois filhos	F2
	Três filhos	F3
	Quatro filhos ou mais	F4
Tipo de curso - Ensino Médio	EJA	C1
	Tradicional	C2
	Profissionalizante	C3
Período do curso - Ensino Médio	Maior parte diurno	P1
	Maior parte noturno	P2
	Todo diurno	P3
	Todo noturno	P4
Local de acesso à internet	Casas dos parentes	L1
	Em casa	L2
	Em uma <i>lan house</i>	L3

	Na instituição de ensino	L4
	No trabalho	L5
	Pelo celular	L6
Ocupação	Cooperativa	O1
	Creche comunitária	O2
	Creche conveniada com o município	O3
	Empregado em Empresa Privada	O4
	Empresário	O5
	Estagiário	O6
	Monitora Educacional Infantil e fundamental.I	O7
	Não trabalho, apenas estudando	O8
	Profissional liberal	O9
	Serviços domésticos	O10
	Servidor público	O11
	Trabalhador autônomo	O12
Titulação	Bacharel	T1
	Licenciatura	T2
	Superior de Tecnologia	T3

Conforme os dados da Tabela 38, a análise do Grupo 3 e 4, com duas coordenadas principais é melhor explicada pelo Burt Ajustado, com percentual acumulado de 57,9% da inércia total.

Tabela 38 – Comparativo das principais inércias e autovalores dos Grupos 3 e 4

Principais Inércias (autovalores): Matriz Indicadora				Principais Inércias (autovalores) : Matriz de Burt				Principais Inércias (autovalores): Burt Ajustado			
Dimensão	Valor	%	acumulado %	Dimensão	Valor	%	acumulado %	Dimensão	Valor	%	acumulado %
1	0,27651	7,4	7,4	1	0,07646	14	14	1	0,02998	34,7	34,7
2	0,21544	5,75	13,1	2	0,04641	8,5	22,6	2	0,01068	12,4	47

Na Tabela 39, são apresentados os resultados das inércias gerados nas dimensões mais relevantes das análises e seus respectivos totais.

Tabela 39 - Resultado da inércia total dos Grupos 3 e 4

Matriz Indicadora				Matriz de Burt				Burt Ajustado			
Dimensão	Valor	%	Acumulado %	Dimensão	Valor	%	Acumulado %	Dimensão	Valor	%	Acumulado %
1	0,276515	7,4	7,4	1	0,07646	14,00	14,00	1	0,029984	34,70	34,70
2	0,215443	5,7	13,1	2	0,046416	8,50	22,60	2	0,010684	12,40	47,00
3	0,199004	5,3	18,4	3	0,039603	7,30	29,80	3	0,007153	8,30	55,30
4	0,1791	4,8	23,2	4	0,032077	5,90	35,70	4	0,003823	4,40	59,70
5	0,173235	4,6	27,8	5	0,030011	5,50	41,20	5	0,003039	3,50	63,20
6	0,163664	4,4	32,2	6	0,026786	4,90	46,20	6	0,001953	2,30	65,50
7	0,159572	4,3	36,4	7	0,025463	4,70	50,80	7	0,001561	1,80	67,30
8	0,149526	4	40,4	8	0,022358	4,10	55,00	8	0,000786	0,90	68,20
9	0,148232	4	44,4	9	0,021973	4,00	59,00	9	0,000705	0,80	69,00
10	0,140645	3,8	48,1	10	0,019781	3,60	62,60	10	0,00032	0,40	69,40
11	0,13331	3,6	51,7	11	0,017772	3,30	65,90	11	9e-05000	0,10	69,50

12	0,129949	3,5	55,2	12	0,016887	3,10	69,00	12	3,20E-50	0,00	69,50
13	0,128243	3,4	58,6	13	0,016446	3,00	72,00	13	1,40E-50	0,00	69,50
14	0,126393	3,4	61,9	14	0,015975	2,90	74,90	14	3e-06000	0,00	69,50
15	0,124362	3,3	65,3	15	0,015466	2,8	77,8				
16	0,122801	3,3	68,5	16	0,01508	2,8	80,6				
17	0,11518	3,1	71,6	17	0,013267	2,4	83				
18	0,113538	3	74,6	18	0,012891	2,4	85,4				
19	0,105689	2,8	77,5	19	0,01117	2,1	87,4				
20	0,100668	2,7	80,1	20	0,010134	1,9	89,3				
21	0,099414	2,7	82,8	21	0,009883	1,8	91,1				
22	0,092431	2,5	85,3	22	0,008544	1,6	92,7				
23	0,085045	2,3	87,5	23	0,007233	1,3	94				
24	0,084739	2,3	89,8	24	0,007181	1,3	95,3				
25	0,079722	2,1	91,9	25	0,006356	1,2	96,5				
26	0,073583	2	93,9	26	0,005414	1	97,5				
27	0,068731	1,8	95,7	27	0,004724	0,9	98,3				
28	0,063597	1,7	97,4	28	0,004045	0,7	99,1				
29	0,060417	1,6	99	29	0,00365	0,7	99,7				
30	0,037252	1	100	30	0,001388	0,3	100				
Total	3,75	100		Total	0,544431	100		Total	0,086492		

Conforme a Tabela 40, as coordenadas principais foram formadas pelas seguintes combinações: CP1 (F, M, EC3, F0) ; CP2 (T2, O09).

Tabela 40 – Resultado da ACM do Grupos 3 e 4 por meio da Burt Ajustado (x1000)

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	TITULO:T1	28	415	26	39	19	1	-176	396	82
2	TITULO:T2	57	532	18	-59	143	7	97	389	51
3	TITULO:T3	40	80	22	57	75	4	-14	5	1
4	SEXO:F	88	661	10	91	655	24	9	6	1
5	SEXO:M	37	661	24	-212	655	56	-20	6	1
6	ESTADO:EC1	67	605	18	154	501	53	-70	104	31
7	ESTADO:EC2	13	560	30	333	500	50	116	60	17
8	ESTADO:EC3	45	741	30	-331	709	163	70	32	21
9	OCUP:O1	1	731	30	-987	718	20	134	13	1
10	OCUP:O10	1	87	29	-76	7	0	258	80	4
11	OCUP:O11	34	342	23	95	252	10	-57	90	10
12	OCUP:O12	17	71	27	-49	38	1	45	33	3
13	OCUP:O2	1	164	30	176	18	1	-508	146	15
14	OCUP:O3	1	114	29	-224	101	2	80	13	1
15	OCUP:O4	49	392	19	-77	230	10	-65	162	19
16	OCUP:O5	1	286	31	225	40	2	-556	246	35
17	OCUP:O6	1	295	30	109	8	0	-670	287	26
18	OCUP:O7	1	113	30	357	96	3	149	17	1
19	OCUP:O8	16	384	29	19	2	0	237	382	84
20	OCUP:O9	4	415	32	352	168	15	426	247	62

21	FILHOS:F0	61	723	24	-264	704	141	43	19	11
22	FILHOS:F1	31	548	27	178	318	33	-152	230	66
23	FILHOS:F2	23	444	29	260	436	51	-35	8	3
24	FILHOS:F3	9	542	34	470	471	68	182	70	28
25	FILHOS:F4	2	254	33	200	24	2	619	230	66
26	CURSO:C1	12	687	36	555	603	120	208	85	47
27	CURSO:C2	92	681	11	-115	653	41	-24	28	5
28	CURSO:C3	21	295	28	193	294	27	-11	1	0
29	PERIODO:P1	20	92	26	1	0	0	-72	92	9
30	PERIODO:P2	21	189	27	26	7	0	-133	182	35
31	PERIODO:P3	36	476	25	-188	424	43	-66	53	15
32	PERIODO:P4	48	683	22	129	323	27	136	359	84
33	LOCAL:L1	1	427	30	-439	102	4	786	326	35
34	LOCAL:L2	97	183	7	-27	178	2	-4	5	0
35	LOCAL:L3	2	130	33	-41	1	0	429	129	32
36	LOCAL:L4	6	196	32	-35	2	0	332	193	57
37	LOCAL:L5	18	527	27	173	295	18	-154	233	41
38	LOCAL:L6	2	11	30	-26	1	0	83	10	1

Na Figura 5, observamos quatro pontos que contribuíram também para a explicação dos dados e formação das coordenadas principais, são eles: O1, O2 , O5, O6 e O7. Estes pontos estão localizados bem longe do centro do Mapa de Correspondência da ACM do grupos 3 e 4, por isso tem participação significativa na contribuição da variabilidade dos dados.

Observamos também uma forte associação entre as categorias O1 (Cooperativa), O4 (Empregado em empresa privada), EC3 (Solteiro), M (sexo masculino) e F0 (Nenhum filho). Percebemos que os alunos pesquisados do sexo masculino trabalham em empresas privadas e em cooperativas, são solteiros e não têm filhos.

Na mesma figura, verificamos uma associação entre as seguintes categorias: O2 (Creche comunitária), O3 (Creche conveniada com o município), O9 (Profissional liberal), O10 (Serviços domésticos) , O11 (Servidor público), EC2 (Divorciado/Separado), F2 (Dois filhos), F3 (Três filhos), F4 (Quatro filhos ou mais), T2 (Licenciatura). Isto significa que os alunos pesquisados do sexo feminino que trabalham como profissionais liberais, em creches ou são servidoras públicas, cursam uma graduação na área da Licenciatura, são divorciadas e têm entre 2 e 4 filhos.

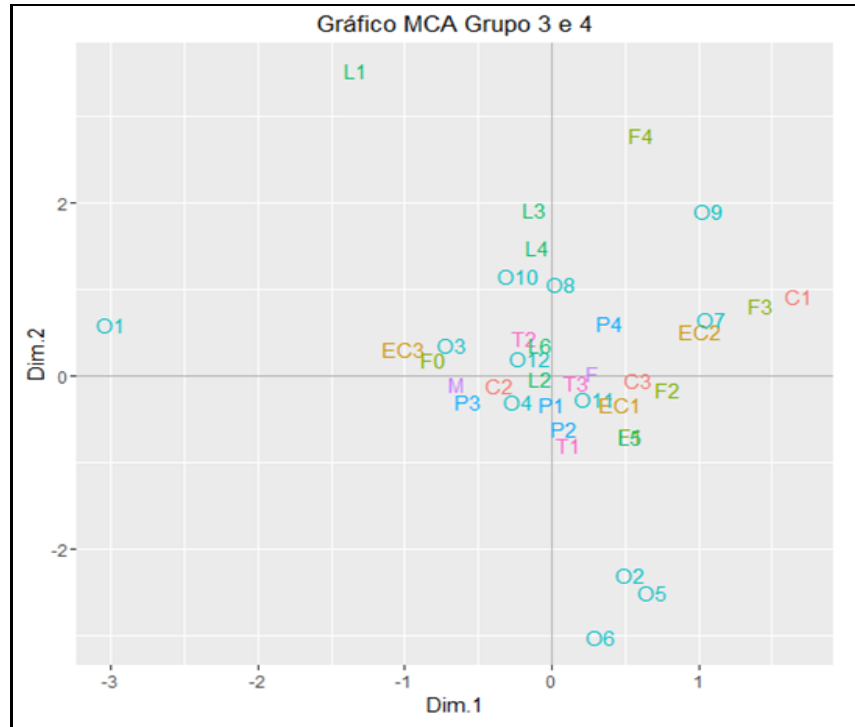


Figura 8 - Mapa de Correspondência dos Grupos 3 e 4

7.3.4 ACM do Grupo 5

No Quadro 4, apresentamos a codificação do Grupo 5 :

Quadro 4 - Codificações das categorias das variáveis consideradas na ACM do Grupo 5

Variável	Categoria	Código
Renda	Até 01 SM	RD1
	De 01 a 03 SM	RD2
	De 03 a 06 SM	RD3
	De 06 a 10 SM	RD4
	Acima de 10 SM	RD5
Primeira graduação	Sim	G1
	Não	G2
Nível de conhecimento em computador	Pouco	PC1
	Bom	PC2
	Ótimo	PC3
Horas de estudo	Nenhuma, apenas assisto às aulas	H1
	De 1 a 2 horas	H2
	De 3 a 5 horas	H3
	De 6 a 8 horas	H4
	Acima de 8 horas	H5
Satisfação	Muito insatisfeito	S1
	Insatisfeito	S2
	Pouco satisfeito	S3
	Satisfeito	S4

Como pode-se observar na Tabela 41, a análise do Grupo 5 é melhor explicada pelo Burt Ajustado, com percentual acumulado de 60% da inércia total.

Tabela 41 – Comparativo das principais inércias e autovalores do Grupo 5

Principais Inércias (autovalores) : Matriz Indicadora				Principais Inércias (autovalores) : Matriz de Burt				Principais Inércias (autovalores) : Burt Ajustado			
Dimensão	Valor	%	acumulado %	Dimensão	Valor	%	acumulado %	Dimensão	Valor	%	acumulado %
1	0,33648	12	12	1	0,11322	18,4	18,4	1	0,029104	41,1	41,1
2	0,29261	10,5	22,5	2	0,08562	13,9	32,2	2	0,013401	18,9	60,0

Na Tabela 42, são apresentados os resultados das inércias gerados nas dimensões mais relevantes das análises e seus respectivos totais.

Tabela 42 - Resultado da inércia total do Grupo 5

Matriz Indicadora				Matriz de Burt				Burt Ajustado			
Dimensão	Valor	%	Acumulado %	Dimensão	Valor	%	Acumulado %	Dimensão	Valor	%	Acumulado %
1	0,33648	12	12	1	0,11322	18,4	18,4	1	0,029104	41,1	41,1
2	0,29261	10,5	22,5	2	0,08562	13,9	32,2	2	0,013401	18,9	60
3	0,26184	9,4	31,8	3	0,06856	11,1	43,4	3	0,005976	8,4	68,4
4	0,23319	8,3	40,10	4	0,05438	8,8	52,2	4	0,001721	2,4	70,9
5	0,22896	8,2	48,3	5	0,05242	8,5	60,7	5	0,001311	1,9	72,7
6	0,21244	7,6	55,9	6	0,04513	7,3	68	6	0,000242	0,3	73
7	0,19413	6,9	62,8	7	0,03769	6,1	74,1				
8	0,18846	6,7	69,6	8	0,03552	5,8	79,9				
9	0,16578	5,9	75,5	9	0,02748	4,5	84,3				
10	0,16362	5,8	81,3	10	0,02677	4,3	88,7				
11	0,15566	5,6	86,9	11	0,02423	3,9	92,6				
12	0,13624	4,9	91,8	12	0,01856	3	95,6				
13	0,13134	4,7	96,5	13	0,01725	2,8	98,4				
14	0,09926	3,5	100	14	0,00985	1,6	100				
Total	2,8	100		Total:	0,61668	100		Total:	0,070851	100	

Conforme a Tabela 43, as coordenadas principais foram formadas pelas seguintes combinações: CP1 (H5, G1, G2, RD4); CP2 (S1, S4, H1, H2, H3, RD2).

Tabela 43 – Resultado da ACM do Grupo 5 por meio da Burt Ajustado (x1000)

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	S1	11	618	71	296	130	32	574	489	265
2	S2	6	414	71	-560	294	63	358	120	56

3	S3	33	1	55	-6	1	0	1	0	0
4	S4	150	576	17	2	1	0	-55	575	34
5	H1	10	500	64	-60	17	1	321	482	75
6	H2	89	577	38	-12	10	0	88	567	52
7	H3	74	702	44	-86	277	19	-107	425	63
8	H4	22	401	61	176	310	23	-95	91	15
9	H5	6	788	68	717	739	104	-184	49	15
10	PC1	46	597	54	-203	588	65	-26	10	2
11	PC2	107	335	32	-45	136	8	-55	200	24
12	PC3	47	744	59	302	595	148	151	149	80
13	G1	12	693	74	758	666	232	-152	27	20
14	G2	188	693	5	-47	666	15	9	27	1
15	RD1	21	428	66	-266	235	50	241	193	89
16	RD2	148	656	19	-43	211	9	-63	445	43
17	RD3	25	650	63	248	338	52	238	312	103
18	RD4	6	705	73	926	626	173	-329	79	48
19	RD5	1	249	66	359	106	4	418	143	13

Na Figura 10, observamos a existência de relações entre as categorias PC3 (Ótimo), RD3 (De 03 a 06 salários mínimos), RD5 (Acima de 10 salários mínimos), S1 (Muito insatisfeito). Isto indica que os alunos pesquisados, que recebem entre 03 a 10 salários mínimos, têm conhecimento ótimo em computador e estão muito insatisfeitos com o curso de graduação à distância.

Verificamos também uma associação entre as categorias RD1 (Até 01 salário mínimo), H1 (Nenhuma, apenas assisto às aulas) e S2 (Insatisfeito). Isto pode indicar que os alunos pesquisados, que recebem até um salário mínimo, apenas assistem às aulas presenciais e estão insatisfeitos com o curso de graduação. Observamos também associações entre as seguintes categorias RD4 (De 06 a 10 salários mínimos), H5 (Acima de 8 horas), G1 (Sim, faz a primeira graduação). Neste caso, pode indicar que os alunos pesquisados, que recebem de 06 a 10 salários mínimos, cursam a primeira graduação e estudam mais de 8 horas semanais.

No centro do eixo, observamos uma forte associação entre as categorias G2 (Não é a primeira graduação), PC1 (pouco), RD2 (De 01 a 03 salários mínimos), H2 (De 1 a 2 horas), H3 (De 3 a 5 horas), H4 (De 6 a 8 horas), S3 (pouco satisfeito) e S4 (Satisfeito). Isto pode indicar que os alunos pesquisados, que recebem entre 1 a 3 salários mínimos, têm pouco conhecimento em computador, dedicam aos estudos 1 a 8 horas por semana, bem como há alunos pouco satisfeitos com o curso de graduação e alunos satisfeitos.

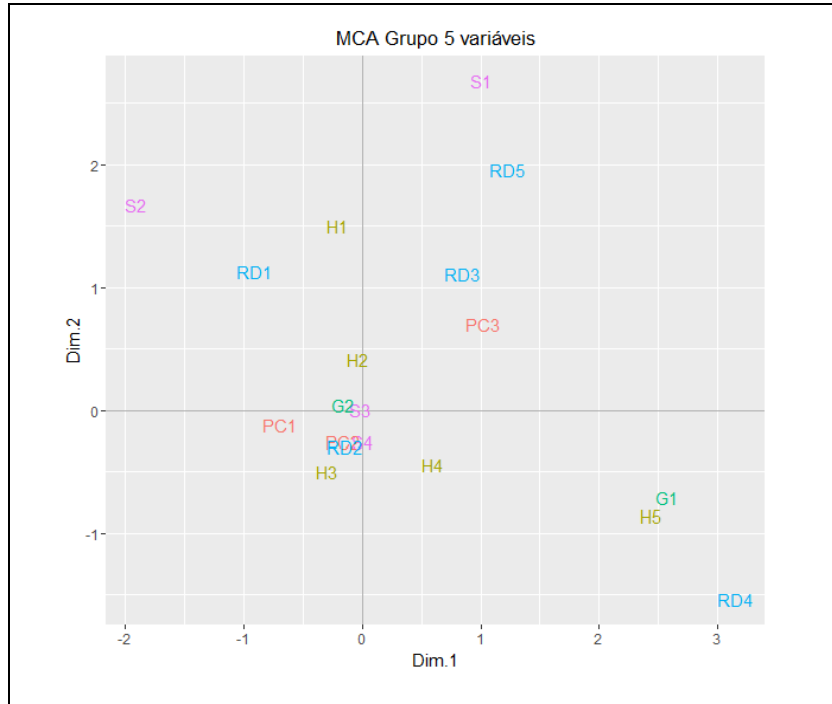


Figura 9 – Mapa de Correspondência do Grupo 5

8 CONCLUSÃO

A ACM é uma poderosa ferramenta de análise de grandes volumes de dados categóricos e permite que o pesquisador analise várias variáveis categóricas simultaneamente. Através da representação gráfica dos dados, ela também facilita a demonstração visual das informações, mostrando mais facilmente a existência ou não de associações entre as categorias das variáveis.

A aplicação dessa técnica de análise se torna importante para atingir os objetivos desta pesquisa devido ao grande número de variáveis categóricas a serem tratadas. Desse modo, esta pesquisa é relevante para a Instituição de Ensino Superior pesquisada, pois auxiliará o gestor do polo de apoio presencial no planejamento de marketing, na divulgação dos cursos de graduação para captação de alunos e também fornecerá informações sobre o perfil dos seus estudantes. Essa última permitirá que o Gestor do Polo de Apoio, junto com a equipe pedagógica, possa elaborar cursos ou outras atividades de complementação de estudo de acordo com as características de seus alunos EAD.

Para atingir os objetivos desta pesquisa, foi formulado um questionário socioeconômico, o qual foi aplicado aos alunos da Instituição de Ensino Superior pesquisada. Os dados coletados foram obtidos de forma completa e verdadeira, sendo utilizados para formação do banco de dados desta pesquisa. Durante a coleta dos dados, foram encontradas algumas dificuldades, como por exemplo a falta de tempo para aplicação de todos os questionários e dificuldades em encontrar todos os alunos

do 2º período dos cursos de graduação nas salas de aulas, uma vez que, no ensino à distância, os alunos podem assistir às aulas em outro local com acesso à *internet*.

Com a base de dados formada, foi aplicada à Análise de Correspondência Múltipla. Com isso, percebe-se que as definições e aplicações dos algoritmos da AC podem ser utilizadas em matrizes multidimensionais (Matriz Indicadora e Matriz de Burt), pois elas fornecem as regras básicas para interpretação dos dados. Além disso, constatou-se que a aplicação da ACM em Matrizes Indicadoras e de Burt é equivalente, diferindo apenas na estimação das coordenadas principais.

Verificamos também que, na ACM, as inércias das coordenadas principais são subestimadas, sendo necessária a aplicação do método do Burt Ajustado para correção das inércias. Dessa forma, o ajuste é melhor para a explicação da variação dos dados da pesquisa, pois aumenta consideravelmente as inércias das duas primeiras coordenadas principais, ou seja, propicia mais qualidade na representação das informações.

A utilização da análise descritiva aplicada aos dados desta pesquisa teve como objetivo descrever o perfil dos alunos pesquisados que estudam na modalidade de ensino à distância. Constatou-se que os alunos entrevistados da Instituição de Ensino Superior pesquisada tem o seguinte perfil: eles têm mais de 30 anos de idade, são casados, têm filhos, trabalham, tem pouco tempo para os estudos, recebem até três salários mínimos, acessam pouco à internet e não têm conhecimento em uma língua estrangeira. Também identificamos um público predominantemente feminino matriculado nos cursos de graduação à distância na Instituição de Ensino Superior pesquisada.

Com esta pesquisa, percebe-se que os alunos que estudam na modalidade à distância buscam a formação profissional e a qualificação, pois veem na EAD uma oportunidade de dar continuidade aos estudos e de se manter competitivo no mercado de trabalho. Nota-se que a EAD é uma opção de estudo para quem não tem disponibilidade de tempo. Essa motivação ocorre, provavelmente, pelo fato de a maioria dos alunos pesquisados terem que conciliar os estudos com o trabalho e a família.

Cabe salientar que esta pesquisa visa contribuir com os estudos sobre o Análise de Correspondência e disseminar as técnicas estatísticas de tratamento de dados, bem como estimular pesquisadores da área educacional e afins a aprofundarem esses estudos.

REFERÊNCIAS

- ABED – **Associação Brasileira de Educação a Distância. Censo EAD**: relatório analítico da aprendizagem a distância no Brasil. 2014. Disponível em: http://www.abed.org.br/censoead2014/CensoEAD2014_portugues.pdf. Acesso em: 20/10/2015.
- ABRAEAD - **Anuário Brasileiro Estatístico de Educação Aberta e a Distância**. 2008. Disponível em: http://www.abraead.com.br/anuario/anuario_2008.pdf. Acesso em: 02/11/2015.
- ALVES, J. R. M. **A história da educação a distância no Brasil**. Carta Mensal Educacional, n.82, jun2007,ISSN 0103-9449. Disponível em: http://www.ipae.com.br/pub/pt/cme/cme_82/index.htm. Acesso em : 10/04/2016
- ALVES, L. **Educação a distância: conceitos e história no Brasil e no mundo**. *Associação Brasileira de Educação a Distância*. v.10. p.83-92.2011. Disponível em: http://www.abed.org.br/revistacientifica/Revista_PDF_Doc/2011/Artigo_07.pdf. Acesso em: 12/10/2015.
- BELLONI, M. L. **Educação a distância**. 4º edição. Campinas, 2006.115p.
- BRASIL. **Decreto nº.5622**, 19 de Dezembro de 2005. Regulamenta o art. 80 da Lei nº 9.394, de 20 de dezembro de 1996 (LDB). Disponível em : http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/decreto/d5622.htm. Acesso em : 20/12/2014.
- BRASIL. **Lei nº.9394**, de 20 de Dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. Disponível em: http://www.planalto.gov.br/CCIVIL_03/leis/L9394.htm. Acesso em : 20/12/2014.
- BOLZAN, R. de F. F. de A. **O conhecimento tecnológico e o paradigma educacional**. 1998. 175f. Dissertação (Mestrado) - Programa de Pós - Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis. 1998. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/158192/137987.pdf>. Acesso em: 22/03/2016
- CRESPO, A. A. **Estatística fácil**. 17º. Ed. São Paulo: Saraiva, 2002.224p.
- DEMARCO, S. R. S.; HONORATO, H. G. **A educação a distância e a tutoria : Alguns olhares**. V Seminário Internacional de Educação a distância. Anais..., 2 e 4 de setembro de 2013.CAED/UFMG. Disponível em: https://www.ufmg.br/ead/seminario/anais/pdf/Eixo_2.pdf. Acesso em: 27 mar. 2014
- CZERMAINSKI, A.B. **Análise de correspondência**. Piracicaba, 2004. Disponível em: <http://www.lce.esalq.usp.br/tadeu/anabeatriz.pdf>. Acesso em: 01 dezembro 2013.
- FÁVERO, L. P.; BELFIORE, P.; SILVA, F.L. da; CHAN; B. L. **Análise de dados**: modelagem multivariada para tomada de decisões. 4ª. Edição. Rio de Janeiro: Elsevier, 2009.646p.

GOMES, S. G. S.; MOTA, J. B; LEONARDO, E. da S. **Reflexão sobre o perfil do aluno como determinante para a motivação e aprendizagem em curso de EAD.** Caderno de Educação, Tecnologia e Sociedade , Inhumas, v. 7, p. 355-363, 2014 DOI 10.14571/cets.v7.355-363. Disponível em: <http://cadernosets.inhumas.ifg.edu.br/index.php/cadernosets/article/download/214/118> . Acesso em : 05/04/2016.

GREENACRE, M. **La práctica del análisis de correspondencias** . Jul. 2008. Fundación BBVA, 2008.384p.ISBN : 978-84-96515710.Disponível em: <http://www.multivariatestatistics.org/practica.html>. Acesso em 25/03/2016

HERMIDA, J. F.; BONFIM, C. R. de S. **A educação à distância : história, concepções e perspectivas.** Revista HISTEDBR On-line, Campinas, n. especial, p.166–181, ago 2006 ISSN: 1676-2584. Disponível em : http://www.histedbr.fe.unicamp.br/revista/edicoes/22e/art11_22e.pdf. Acesso em : 10/04/2016.

JELIHOVSCHI, E. G.; FERRAZ, M. I. F. **Análise do conjunto dos candidatos ao vestibular da UESC no ano de 2008 usando a análise de correspondência.** Revista Brasileira de Biometria [online], 28, 4, p. 117-136, 2010. Disponível em: http://jaguar.fcav.unesp.br/RME/fasciculos/v28/v28_n4/A8_Enio.pdf. Acesso em 20/01/2016.

LANDIM, C. M. das M. P. F.**Educação à distância: algumas considerações.** Rio de Janeiro: s.n.1997.LARSON, R; FARBER, B; Estatística aplicada. 4.ed. Tradução: Luciane Ferreira Pauleti Vianna. São Paulo. Editora: Pearson Prentice Hall, 2010.637p.

MENDES, D. P. C. **Diagnóstico da evasão ocorrida na UFRN para alunos que ingressaram entre 2000 e 2008: uma aplicação de Análise de correspondência.** Natal: Departamento de Estatística, UFRN, 2009. (Monografia-Especialização). Disponível em:http://www.comperve.ufrn.br/conteudo/observatorio/arquivos/monografias/debora_cosme_pereira_mendes.pdf. Acesso em: 02 dezembro 2013.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte: Editora UFMG, 2005.295p.

MORINI, A. M. **Um estudo sobre o perfil do aluno do Ensino a Distância.** Palhoça: Departamento de Educação Educação a Distância,UnisulVirtual, 2006 . (Monografia-Especialização). Disponível em: http://busca.unisul.br/pdf/88702_Ana.pdf. Acesso em: 03/09/2015.

NASCIMENTO, A. **Avaliação de farmácias hospitalares brasileiras utilizando análise de correspondência múltipla.** 2011. 93f. Dissertação (Mestrado) – Programa de Pós-graduação em Engenharia Biomédica, Universidade Federal do Rio de Janeiro,Rio de Janeiro,2011. Disponível em: http://objdig.ufrj.br/60/teses/coppe_m/AlineDoNascimento.pdf . Acesso em: 14/04/2016

OLIVEIRA, O. J de. **Análise de correspondência.** Belém: Faculdade de Estatística, UFPA, 2013. (Monografia – Especialização) Disponível em: http://www.ufpa.br/heliton/arquivos/aplicada/seminarios/M1_06_Odair_Oliveira_Analise_deCorrespondencia.pdf. Acesso em: 05 dezembro 2013

PALLOFF, R. M. e PRATT, K. **O aluno virtual: um guia para trabalhar com estudantes online**. 1ª ed. São Paulo: Artemed, 2004.216p.

PAMPLONA, A. S. **Análise de Correspondência para dados com estrutura de grupo**.1998. 180f. Dissertação (Mestrado) - Instituto de Matemática, Estatística e Computação Cientificada , Universidade Estadual de Campinas, São Paulo, 1998. Disponível em: <http://www.bibliotecadigital.unicamp.br/document/?code=000125777&fd=y>. Acesso em : 5/03/2016.

PRADO, M. V. B. **Métodos de análise de correspondência múltipla: estudo de caso aplicado à avaliação da qualidade do café**. 2012. 78f. Dissertação (Mestrado) – Programa de Pós-graduação em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras, Lavras, 2012. Disponível em: <http://repositorio.ufla.br/handle/1/482>

SCHNITMAN, I.M. **O perfil do aluno virtual e as teorias de estilos de Aprendizagem**. III Simpósio Hipertexto e Tecnologias na Educação. Redes Sociais e Aprendizagem. NEHTE/UFPE. Anais..., 2 e 3 de dezembro de 2010. Recife (PE). Disponível em: <http://www.ufpe.br/nehte/simposio/anais/Anais-Hipertexto-2010/Ivana-Maria-Schnitman.pdf>. Acesso em: 27 mar. 2014

VIDAL, E. M; MAIA, J. E. B; **Educação a distância: rompendo fronteiras**. Introdução à Educação a Distância. RDS Editora, 2010. p. 09-24.

VIDIGAL, B. C. **Análise de correspondência para avaliação de indicadores socioeconômicos, demográficos e de mortalidade**. Juiz de Fora: Universidade Federal de Juiz de Fora, Departamento de Estatística, UFJF, 2010. (Monografia) Disponível em: <http://www.ufjf.br/estatistica/producao-2/tcc/>. Acesso em 30/03/2016.

VILAÇA, M. L. C. **Educação a Distância e Tecnologias: conceitos, termos e um pouco de história**. Revista Magistro [online], v. 1, n. 2, p.89-101, 2010. Disponível em: <http://publicacoes.unigranrio.edu.br/index.php/magistro/article/viewFile/1197/801>. Acesso em: 11/10/2015.

APÊNDICE A - Questionário de caracterização socioeconômico

Qual curso que faz atualmente? _____

- Bacharel Superior de tecnologia
 Licenciatura

1. Você é do sexo:

- a) Masculino
b) Feminino
c) Outros.

2. Qual a sua idade?

- a) De 18 a 22 anos
b) De 23 a 27 anos
c) De 28 a 32 anos
d) De 33 a 37 anos
e) De mais de 38 anos

3. Você se considera:

- a) Branco
b) Negro
c) Pardo
d) Amarelo (de origem oriental)
e) Indígena ou de origem indígena
f) Outra. Qual? _____

4. Estado Civil:

- a) Solteiro(a).
b) Divorciado(a)/ Separado (a).
c) Viúvo(a).
d) Casado(a) ou União estável.

5. Você tem filhos?

- a) Não. b) Sim. Quantos? _____

6. Cidade local da sua residência:

- a) Betim
b) Outra cidade. Qual? _____

7. Qual sua ocupação atual ? Caso a resposta seja negativa passe para a questão nº.09

- a) Não trabalho, apenas estudando.

- b) Profissional Liberal.
- c) Empresário.
- d) Servidor Público Federal/ Estadual/Municipal.
- e) Empregado em Empresa Privada.
- f) Empregado Rural/Agricultor.
- g) Trabalhador Autônomo.
- h) Outras. Qual? _____

08. Qual a sua renda mensal? (Valor do salário mínimo vigente em 2014 = R\$724,00)

- a) Menos de 1 Salário Mínimo.
- b) De 01 a 03 Sal. Mínimos.
- c) De 03 a 06 Sal. Mínimos.
- d) De 06 a 10 Sal. Mínimos.
- e) Mais de 10 Sal. Mínimos.

09. Em que tipo de escola você cursou o ensino médio?

- a) Todo em escola pública.
- b) Todo em escola privada (particular).
- c) Parte em escola pública e parte em escola privada (particular).

10. Que tipo de curso de ensino médio você cursou?

- a) Ensino médio tradicional
- b) Profissionalizante técnico (eletrônica, contabilidade, agrícola, etc)
- c) Educação de Jovens e Adultos – EJA Supletivo
- d) Programa Nacional de Integração da Educação Profissional com a Educação Básica na Modalidade de Jovens e Adultos (PROEJA)
- e) Outros. Especifique: _____

11. Em que período você cursou o Ensino Médio?

- a) Todo diurno.
- b) Maior parte diurno.
- c) Todo noturno.
- d) Maior parte noturno.
- e) Outro. Especifique: _____

12. Você cursa ou já concluiu algum curso de idioma em escola de línguas?

- a) Não
- b) Sim. Qual(is) idioma(s)? _____

13) Este curso é a sua primeira graduação?

- a) Sim.
- b) Não, já sou formado em uma graduação presencial. Qual? _____
- c) Não, já sou formado em uma outra graduação pela EaD. Qual? _____

14. Por que você escolheu um curso na modalidade a distância? Indicar a alternativa que mais se aproxime de sua opção.

- a) Falta de tempo para frequentar diariamente um curso.
- b) Distância das Instituições de Ensino Superior.
- c) Achei que fosse mais fácil que o curso presencial.
- d) Constrangimento em frequentar o curso junto com pessoas mais jovens.
- e) Outra. Qual? _____

15. Qual seu conhecimento em relação à utilização do computador?

- a) Nenhum
- b) Pouco
- c) Bom
- d) Ótimo.

16. Qual é o local onde você mais acessa a internet para desenvolver todas as atividades acadêmicas?

- a) Em casa b) No trabalho c) Em uma LAN house d) Pelo celular
- e) Na instituição de ensino
- f) Outros: especificar _____

17. Quantas horas por semana, aproximadamente, você dedica aos estudos, excetuando as horas de aula na instituição de ensino?

- a) Nenhuma, apenas assisto às aulas.
- b) Uma a duas.
- c) Três a cinco.
- d) Seis a oito.
- e) Mais de oito.

18. Entre as alternativas a seguir, assinale a que melhor expressa sua perspectiva profissional futura:

- a) Já tenho trabalho na área e pretendo continuar nele.
- b) Trabalho em outra área, mas pretendo buscar uma atividade na minha área de graduação.
- c) Vou me dedicar à atividade acadêmica e buscar um curso de pós-graduação.
- d) Vou prestar concurso para atividade em empresa pública.
- e) Pretendo trabalhar em empresa privada.

f) () Ainda não me decidi.

g) () Outros. Especificar: _____

19. Você ficou conhecendo a instituição de ensino superior por meio de:

a) () Alunos da Instituição.

b) () Parentes.

c) () Divulgação em minha escola.

d) () Empresa em que trabalha.

e) () Internet ou outros meios de comunicação.

f) () Outros. Especificar: _____

20. Qual sua satisfação sobre o curso?

a) () Insatisfeito

b) () Pouco Satisfeito

c) () Satisfeito

b) () Muito Insatisfeito.

APÊNDICE B – Probabilidades associadas ao qui-quadrado calculado do conjunto de dados

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21
Curso (X1)	0,000																				
Titulação (X2)	0,000	0,000																			
Sexo (X3)	0,000	0,590	0,000																		
Faixa etária (X4)	0,002	0,587	0,041	0,000																	
Raça (X5)	0,351	0,247	0,872	0,420	0,000																
Estado Civil (X6)	0,027	0,279	0,033	0,000	0,543	0,000															
Nº Filhos (X7)	0,294	0,235	0,033	0,000	0,991	0,000	0,000														
Ocupação (X8)	0,005	0,353	0,521	0,056	0,790	0,737	0,853	0,000													
Renda (X9)	0,028	0,190	0,001	0,098	0,893	0,816	0,201	0,001	0,000												
Escola Ensino Médio (X10)	0,993	0,227	0,396	0,573	0,036	0,210	0,916	0,476	0,701	0,000											
Tipo de curso EM (X11)	0,299	0,287	0,009	0,000	0,220	0,000	0,000	0,568	0,299	0,476	0,000										
Período do Curso (X12)	0,202	0,627	0,429	0,012	0,949	0,450	0,142	0,562	0,108	0,036	0,001	0,000									
Curso de Idioma (X13)	0,974	0,304	0,719	0,535	0,090	0,359	0,942	0,976	0,013	0,000	0,432	0,316	0,000								
Primeira graduação (X14)	0,015	0,958	0,359	0,263	0,968	0,650	0,319	0,988	0,000	0,824	0,439	0,018	0,037	0,000							
Motivo EAD (X15)	0,842	0,628	0,492	0,306	0,999	0,799	0,951	1,000	0,217	1,000	0,296	0,776	0,208	0,035	0,000						
Conhecimento Computador (X16)	0,009	0,054	0,071	0,001	0,789	0,222	0,222	0,576	0,261	0,318	0,088	0,335	0,010	0,013	0,503	0,000					
Local internet (X17)	0,183	0,442	0,946	0,498	0,996	0,684	0,019	0,966	0,427	0,974	0,703	0,227	0,997	0,979	0,000	0,026	0,000				
Horas de estudo (X18)	0,629	0,369	0,568	0,498	0,624	0,215	0,760	0,807	0,000	0,722	0,758	0,745	0,000	0,063	0,026	0,128	0,566	0,000			
Perspectiva profissional (X19)	0,029	0,004	0,568	0,082	0,384	0,773	0,922	0,009	0,000	0,018	0,052	0,000	0,116	0,000	0,000	0,024	0,975	0,003	0,000		
Conhecimento da Instituição de Ensino (X20)	0,318	0,539	0,607	0,676	0,994	0,470	0,003	0,051	0,002	0,819	0,188	0,972	0,220	0,000	0,000	0,592	0,331	0,008	0,000	0,000	
Satisfação (X21)	0,509	0,667	0,353	0,350	0,719	0,867	0,777	0,001	0,002	0,273	0,603	0,000	0,998	0,881	0,000	0,006	0,000	0,871	0,734	0,252	0,000

ANEXO A - Script do software r da análise de correspondência simples

```
##### PACOTES #####
```

```
library(FactoMineR)
library(ca)
library(ade4)
library(factoextra)
```

```
##### ANÁLISE DE CORRESPONDENCIA SIMPLES #####
```

```
data(dado)
head(dado)
```

```
# Soma da LINHAS da Tabela
row.sum <- apply(dado, 1, sum)
head(row.sum)
```

```
# Soma da COLUNAS da Tabela
col.sum <- apply(dado, 2, sum)
head(col.sum)
```

```
# Total de n da tabela
n <- sum(dado)
n
```

```
# Matriz de Correspondência
matriz.corresp <- dado/n
head(matriz.corresp)
matriz.corresp
```

```
# Perfil das linhas
row.profile <- dado/row.sum
head(row.profile)
row.profile
```

```
# Média Perfil de Linhas = Soma das Colunas / total
media.rp <- col.sum/n
media.rp
```

```
# Laundry and Main_meal profiles
Matematica.p <- row.profile["Matematica",]
Portugues.p <- row.profile["Portugues",]
```

```
##### MATRIZ DE DISTÂNCIA #####
```

```
## data: a data frame or matrix;
## average.profile: average profile
```



```

dist.matrix <- function(data, average.profile){
  mat <- as.matrix(t(data))
  n <- ncol(mat)
  dist.mat<- matrix(NA, n, n)
  diag(dist.mat) <- 0
  for (i in 1:(n - 1)) {
    for (j in (i + 1):n) {
      d2 <- sum(((mat[, i] - mat[, j])^2) / average.profile)
      dist.mat[i, j] <- dist.mat[j, i] <- d2
    }
  }
  colnames(dist.mat) <- rownames(dist.mat) <- colnames(mat)
  dist.mat
}
# Distance matrix
dist.mat <- dist.matrix(row.profile, average.rp)
dist.mat <-round(dist.mat, 2)
dist.mat

##### PESO DA LINHA #####

row.sum <- apply(dado, 1, sum) # SOMA DA CADA LINHA
grand.total <- sum(dado) # SOMA GERAL TABELA

row.mass <- row.sum/grand.total # PESO DA LINHA
head(row.mass)
row.mass

##### INERCIA DA LINHA #####
# Distance linha
d2.row <- apply(row.profile, 1,
  function(row.p, av.p){sum(((row.p - av.p)^2)/av.p)},
  average.rp)
as.matrix(round(d2.row,3))

row.inertia <- row.mass * d2.row
head(row.inertia)
row.inertia

### Total inercia
sum(row.inertia)

### Resumo das estatísticas da linha #####
row <- cbind.data.frame(d2 = d2.row, mass = row.mass, inertia = row.inertia)
round(row,3)

##### PERFIL DAS COLUNAS #####

col.profile <- t(dado)/col.sum

```

```

col.profile <- as.data.frame(t(col.profile))
col.profile

# Média Colunas
# average column profile= row sums/grand total
row.sum <- apply(dado, 1, sum)
row.sum
average.cp <- row.sum/n
average.cp

##### MATRIZ DE DISTÂNCIAS #####

d2.col <- apply(col.profile, 2,
               function(col.p, av.p){sum(((col.p - av.p)^2)/av.p)},
               average.cp)
round(d2.col,3)

# Distance matrix
dist.mat <- dist.matrix(t(col.profile), average.cp)
dist.mat <- round(dist.mat, 2)
dist.mat

##### PESO DA COLUNA #####

col.sum <- apply(dado, 2, sum)
col.sum

grand.total <- sum(dado)
grand.total

col.mass <- col.sum/grand.total
col.mass

##### INERCIA COLUNA #####

col.inertia <- col.mass * d2.col
col.inertia

##### RESUMO RESULTADOS DA COLUNA #####

col <- cbind.data.frame(d2 = d2.col, mass = col.mass,
                       inertia = col.inertia)
round(col,3)

##### TESTE QUI-QUADRADO #####

chisq <- chisq.test(dado)
chisq

```

```

chisq$observed
round(chisq$expected,2)
round(chisq$residuals, 3)

# Contribution in percentage (%)
contrib <- 100*chisq$residuals^2/chisq$statistic
round(contrib, 3)

phi2 <- as.numeric(chisq$statistic/sum(dado))
phi2

##### ANALISE DE CORRESPONDÊNCIA #####

# Grand total
n <- sum(dado)
n

# Standardized residuals
residuals <- chisq$residuals/sqrt(n)
residuals

#Number of dimensions
nb.axes <- min(nrow(residuals)-1, ncol(residuals)-1)
nb.axes

Singular value decomposition
res.svd <- svd(residuals, nu = nb.axes, nv = nb.axes)
res.svd

sv <- res.svd$d[1:nb.axes] # singular value
u <- res.svd$u
v <- res.svd$v

# Eigenvalues
eig <- sv^2
eig

# Variances in percentage
variance <- eig*100/sum(eig)
variance

# Cumulative variances
cumvar <- cumsum(variance)
cumvar

eig2<- data.frame(eig = eig, variance = variance,
                  cumvariance = cumvar)
eig2

```

```

barplot(eig[, 2], names.arg=1:nrow(eig),
        main = "Variâncias",
        xlab = "Dimensões",
        ylab = "% variância",
        col = "steelblue")

# Add connected line segments to the plot
lines(x = 1:nrow(eig), eig[, 2],
      type="b", pch=19, col = "red")

##### COORDENADA DAS LINHAS #####
# row sum
row.sum <- apply(dado, 1, sum)
# row mass
row.mass <- row.sum/n

row.sum
row.mass

# row coord = sv * u /sqrt(row.mass)

cc <- t(apply(u, 1, '*', sv)) # each row X sv
row.coord <- apply(cc, 2, '/', sqrt(row.mass))
rownames(row.coord) <- rownames(dado)
colnames(row.coord) <- paste0("Dim.", 1:nb.axes)
round(row.coord,3)

# plot
plot(row.coord, pch=19, col = "blue")
text(row.coord, labels =rownames(row.coord), pos = 3, col ="blue")
abline(v=0, h=0, lty = 2)

##### COORDENADA DAS COLUNAS #####
# Coordinates of columns
col.sum <- apply(dado, 2, sum)
col.mass <- col.sum/n
# coordinates sv * v /sqrt(col.mass)
cc <- t(apply(v, 1, '*', sv))
col.coord <- apply(cc, 2, '/', sqrt(col.mass))
rownames(col.coord) <- colnames(dado)
colnames(col.coord) <- paste0("Dim", 1:nb.axes)
col.coord

# plot
plot(col.coord, pch=17, col = "red")
text(col.coord, labels =rownames(col.coord), pos = 3, col ="red")
abline(v=0, h=0, lty = 2)

```

```
##### BILOT OF ROWS AND COLUMNS TO VIEW THE ASSOCIATION ###
```

```
xlim <- range(c(row.coord[,1], col.coord[,1]))*1.1
ylim <- range(c(row.coord[,2], col.coord[,2]))*1.1
# Plot of rows
plot(row.coord, pch=19, col = "blue", xlim = xlim, ylim = ylim)
text(row.coord, labels =rownames(row.coord), pos = 3, col ="blue")
# plot off columns
points(col.coord, pch=17, col = "red")
text(col.coord, labels =rownames(col.coord), pos = 3, col ="red")
abline(v=0, h=0, lty = 2)
```

```
##### CONTRIBUIÇÃO LINHAS E COLUNAS #####
```

```
# contrib <- row.mass * row.coord^2/eigenvalue
cc <- apply(row.coord^2, 2, "*", row.mass)
row.contrib <- t(apply(cc, 1, "/", eig[1:nb.axes,1])) *100
round(row.contrib, 2)
```

```
# contrib <- col.mass * col.coord^2/eigenvalue
cc <- apply(col.coord^2, 2, "*", col.mass)
col.contrib <- t(apply(cc, 1, "/", eig[1:nb.axes,1])) *100
round(col.contrib, 2)
```

```
### QUALIDADE AJUSTE #####
```

```
#LINHAS#
```

```
row.profile <- dado/row.sum
head(round(row.profile, 3))
```

```
average.profile <- col.sum/n
head(round(average.profile, 3))
```

```
d2.row <- apply(row.profile, 1,
               function(row.p, av.p){sum(((row.p - av.p)^2)/av.p)},
               average.rp)
head(round(d2.row,3))
```

```
row.cos2 <- apply(row.coord^2, 2, "/", d2.row)
round(row.cos2, 3)
```

```
# COLUNAS ##
```

```
col.profile <- t(dado)/col.sum
col.profile <- t(col.profile)
head(round(col.profile, 3))
```

```
average.profile <- row.sum/n  
round(average.profile, 3)
```

```
d2.col <- apply(col.profile, 2,  
  function(col.p, av.p){sum(((col.p - av.p)^2)/av.p)},  
  average.profile)  
round(d2.col,3)
```

```
col.cos2 <- apply(col.coord^2, 2, "/", d2.col)  
round(col.cos2, 3)
```

ANEXO B - Script do software r da análise de correspondência múltipla

```
##### PACOTES #####
```

```
install.packages("ca")  
library(ca)
```

```
install.packages("FactoMineR")  
library(FactoMineR)
```

```
install.packages("ade4")  
library(ade4)
```

```
install.packages("factoextra")  
library(factoextra)
```

```
##### análise de correspondencia múltipla #####
```

```
# Matriz Indicadora #  
mjca(dado, lambda="indicator")  
summary(mjca(dado, lambda="indicator"))  
result.ind <- mjca(dado, lambda="indicator")  
plot(result.ind)
```

```
# Matriz BURT #  
mjca(dado[,1:4], lambda="Burt")  
summary(mjca(dado[,1:4], lambda="Burt"))  
plot(mjca(dado[,1:4], lambda="Burt"))
```

```
# Matriz ADJUSTED #  
mjca(dado[,1:4], lambda="adjusted")  
summary(mjca(dado[,1:4], lambda="adjusted"))  
plot(mjca(dado[,1:4], lambda="adjusted"))
```

```
# Matriz JCA #  
mjca(dado[,1:4], lambda="JCA")  
summary(mjca(dado[,1:4], lambda="JCA"))  
plot(mjca(dado[,1:4], lambda="JCA"))
```