



Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Damião Flávio dos Santos

**Análise de Correspondência: Um estudo sobre a evasão escolar
no curso de Bacharelado em Estatística - UEPB/Campus I**

Campina Grande
Dezembro de 2015

Damião Flávio dos Santos

**Análise de Correspondência: Um estudo sobre a evasão escolar
no curso de Bacharelado em Estatística - UEPB/Campus I**

Trabalho de conclusão de curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador: Tiago Almeida de Oliveira

Campina Grande
Dezembro de 2015

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

S237a Santos, Damião Flávio dos.

Análise de correspondência [manuscrito] : um estudo sobre a evasão escolar no curso de Bacharelado em Estatística - UEPB/Campus I / Damião Flávio dos Santos. - 2015.
45 p. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2015.

"Orientação: Prof. Dr. Tiago Almeida de Oliveira, Departamento de Estatística".

1. Análise multivariada. 2. Evasão escolar. 3. Mapas perceptuais. 4. Estatística. I. Título.

21. ed. CDD 519.535

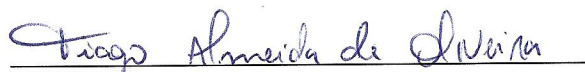
Damião Flávio dos Santos

**Análise de Correspondência: Um estudo sobre a evasão escolar
no curso de Bacharelado em Estatística - UEPB/Campus I**

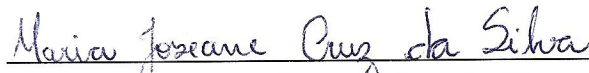
Trabalho de conclusão de curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 11/12/2015

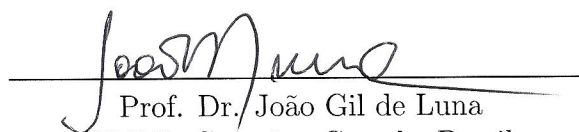
Banca Examinadora:



Prof. Dr. Tiago Almeida de Oliveira
Orientador



Prof. Msc. Maria Joseane Cruz da Silva
UEPB, Campina Grande, Brasil



Prof. Dr. João Gil de Luna
UEPB, Campina Grande, Brasil

Dedicatória

Dedico este trabalho a Deus, aos meus familiares e amigos que tanto torceram e acreditaram na minha formação acadêmica.

Agradecimentos

A realização desse trabalho tornou-se possível graças ao imenso amor de Deus e Sua presença em minha vida, dando assim, esperança e fé em dias melhores e ir em busca de um futuro que tanto almejei.

Aos meus pais José Araújo dos Santos e Maria Lopes dos Santos que sempre estiveram do meu lado me apoiando nas mais difíceis decisões, em especial minha mãe que sempre acreditou em minha capacidade e com seu imenso amor de mãe, pode transmitir toda sua experiência de vida, baseada em respeito ao próximo, na perseverança, na educação, e principalmente na fé em Deus.

Aos meus irmãos Francisca, Francisco e Fabiana, cunhados Cláudio José e Salomão Mendes e aos meus lindos sobrinhos Sthefani Mendes e Kauê Victor, primos Dirceu, Daiana, Diego e Daila, que mesmo distantes, me mostram o quão sou um homem feliz por tê-los em minha vida.

À minha namorada Anna Hemyli Antunes Pamplona, por me ajudar na coleta dos dados, pelo companheirismo, e acima de tudo por está sempre caminhando comigo e me fazendo mais feliz do que já sou.

Aos meus tios Nivaldo e Toinha, que me deram acolhida em Campina Grande e sempre estiveram me aconselhando e apoiando.

À família Trajano, pela acolhida e pelo carinho durante todos esses anos.

Aos meus amigos do Grupo 6 Sigma que de maneira direta contribuíram para meu crescimento e formação.

Aos meus amigos Aline, Arnete, Deise, Pablo, Regina, Sônia, Waleska e Wanessa pelos momentos agradáveis dentro e fora da Universidade.

Ao professor Tiago Almeida de Oliveira, pela orientação e aprendizado por esses meses que trabalhamos juntos.

Aos professores Maria Joseane e João Gil por participarem da banca e todos os professores do departamento de Estatística pelas aulas ministradas e conhecimento repassado.

Enfim, a todos que contribuíram de alguma forma para este trabalho. Muito obrigado.

Resumo

Neste trabalho é apresentado uma abordagem sobre análise de correspondência, sendo de forma teórica e prática, em que a mesma é uma técnica gráfica projetada para representar as associações entre as categorias das variáveis em um espaço com pouca dimensionalidade. Ela pode ser considerada como método de escalas, e pode ser vista como um complemento para outros métodos, como escalonamento multidimensional em um gráfico bidimensional. A importância deste método é permitir que as relações entre as variáveis categóricas possam ser observadas através de gráficos com poucas dimensões. Neste trabalho procedeu-se um estudo com os alunos que evadiram do curso de bacharelado em Estatística da UEPB entre 2010 à 2014 e são apresentados os resultados descritivos da pesquisa e a aplicação da análise de correspondência, com as variáveis relacionadas aos motivos que levaram os alunos a optarem pelo curso, motivos de evadir e de reingresso no curso. A análise de correspondência, mostrou-se válida para analisar dados categóricos e sendo assim, propícia para nossa pesquisa, em que, percebeu-se dos que disseram não saber qual curso optar estavam fortemente relacionado a evadirem do curso por dificuldade em cálculo, assim como para os indicados a optarem pelo curso por amigos ou parentes, evadiram por mudança de curso.

Palavras-chave: Análise multivariada; Mapas perceptuais; Desistência acadêmica.

Abstract

This paper presents an approach to correspondence analysis, and theoretical and practical aspects, in that it is a graphical technique designed to represent the associations between categories of variables in a space with low dimensionality. It can be considered as scale method, and can be seen as a complement to other methods, such as multidimensional scaling in a two-dimensional graph. The importance of this method is to allow the relationships between categorical variables can be observed through graphs with few dimensions. In this work we proceeded to a study of students who dropped out of the bachelor's degree in Statistics from UEPB from 2010 to 2014 and shows the descriptive results of the research and application of correspondence analysis with the variables related to the reasons that led students to opt for the course, reasons for escape and rejoin the course. Correspondence analysis, proved to be valid to analyze categorical data and therefore, conducive to our research, in which he realized in those who said they did not know which course to opt were strongly related to elope the course by difficulty in calculus, as well as for your search to opt for the course by friends or relatives, evade by change to course.

Keywords: Multivariate analysis; perceptual maps; academic waiver.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
2	Fundamentação Teórica	p. 12
2.1	A evasão no curso de Estatística	p. 12
2.2	Marco Histórico da Análise de Correspondência	p. 14
2.3	Análise de Correspondência	p. 15
2.3.1	Tabelas de contingência	p. 16
2.3.2	Desenvolvimento Algébrico da AC	p. 17
2.3.3	Nuvens de pontos	p. 20
2.3.4	Centro de gravidade ou centroide	p. 21
2.3.5	Inércia	p. 23
3	Materiais e métodos	p. 24
4	Resultados e discussões	p. 27
4.1	Análise descritiva	p. 27
4.1.1	Perfil dos alunos	p. 27
4.2	Aplicação da Análise de Correspondência	p. 32
5	Conclusões	p. 36

Referências	p. 37
Apêndice A - Formulário	p. 39
Apêndice B - Rotina no R	p. 40

Lista de Figuras

1	Número de vagas e formandos no curso de Estatística no Brasil	p. 12
2	Imagem de Jean-Paul Benzécri	p. 14
3	Número de publicações sobre “correspondence analysis” de 2000 a 2011 de acordo com ISI-Web of Science (2012)	p. 15
4	Representação tri-dimensional dos perfis de linha sob dois ângulos de visão. Adaptado de (SILVA, 2012)	p. 21
5	Representação geométrica dos perfis linha e do centróide (c) no espaço Euclidiano tridimensional. Adaptado de (SILVA, 2012)	p. 22
6	Gráfico da porcentagem de alunos entrevistados por gênero	p. 27
7	Gráfico da porcentagem de alunos entrevistados por estado civil	p. 28
8	Gráfico da porcentagem de alunos entrevistados por faixa etária.	p. 28
9	Gráfico da porcentagem de alunos entrevistados por renda familiar	p. 29
10	Gráfico da porcentagem de alunos entrevistados por tipo de escola que cursou o 2º grau e turno que cursou o 2º grau	p. 30
11	Gráfico da porcentagem de alunos entrevistados por motivos de optar pelo curso e relacionado a participação em eventos científicos	p. 31
12	Gráfico da porcentagem de alunos entrevistados por motivos de evasão no curso	p. 31
13	Gráfico da porcentagem de alunos entrevistados por motivadores para reingresso no curso	p. 32
14	Mapa perceptual entre motivo de optar pelo curso e motivo de evadir	p. 33
15	Mapa perceptual entre motivo de evasão com motivadores para reingresso no curso	p. 34

Lista de Tabelas

1	Representação tabular de contingência com I linhas e J colunas.	p. 16
2	Descrição das variáveis coletadas na pesquisa	p. 26
3	Organização tabular dos dados para obtenção da média da idade dos entrevistados	p. 28
4	Organização tabular dos dados para obtenção da média da renda familiar dos entrevistados	p. 29
5	Representação tabular de contingência dos motivos de optar pelo curso e motivos de evadir	p. 32
6	Decomposição da inércia sobre os eixos principais	p. 33
7	Decomposição da inércia sobre os eixos principais	p. 34

1 Introdução

A evasão de alunos no ensino superior é, certamente, um problema que atinge de maneira direta as instituições, seja pública ou privada no Brasil. As perdas de estudantes que começam o curso e não conclui, afeta todo o sistema educacional, principalmente de forma social e econômica. Porém, pouco se fala sobre algum programa de controle a evasão nas universidades brasileiras, ou ao menos que existe de fato um considerável percentual de desistência dos discentes, apenas fala-se em números de ingresso em ensino superior.

Diante disso, é importante obter informações acerca dos motivos da desistência e possíveis soluções, de modo que leve a entender de maneira geral ou específica, o que está relacionado ao problema. Desta forma, foi realizado uma pesquisa com os alunos evadidos do curso de Bacharelado em Estatística na Universidade Estadual da Paraíba, Campus I, onde foi possível coletar várias variáveis sobre o tema, para que assim, tivesse informações suficientes para compor este trabalho. Para que isso fosse possível, aplicou-se uma técnica de análise multivariada chamada Análise de Correspondência (AC), com a finalidade de identificar através de um mapa perceptual, as relações existentes entre as categorias das variáveis estudadas e neste caso, verificar possíveis relações entre os motivadores de evasão e outras variáveis.

Este trabalho teve como objetivo realizar uma Análise de Correspondência aos dados de alunos evadidos do Bacharelado de Estatística da UEPB - Campus I, além de realizar uma análise do perfil dos alunos que evadiram do curso de Bacharelado em Estatística da UEPB - Campus I, apresentar o desenvolvimento algébrico de cada etapa da Análise de correspondência, aplicar a AC entre os motivos de cursar e os motivos de evasão. aplicar a AC entre os motivos de evasão e os motivos de reingresso.

2 Fundamentação Teórica

Nesta seção serão apresentados os elementos textuais para melhor compreensão do trabalho de conclusão de curso no qual será mostrado o problema da evasão escolar no ensino superior no Brasil, a evasão no curso de Estatística no Brasil, a origem da Análise de Correspondência, os principais aspectos da técnica.

2.1 A evasão no curso de Estatística

Inicia-se este trabalho apresentado-se a Figura 1, em que tem-se o número de alunos egressos do curso de estatística conjuntamente com o número de vagas no Brasil.

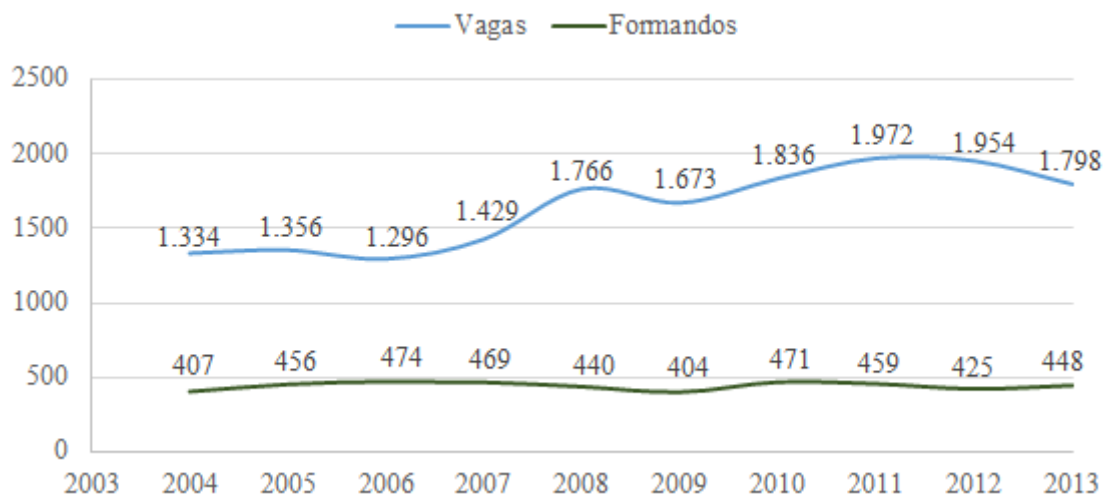


Figura 1: Número de vagas e formandos no curso de Estatística no Brasil

Fonte: INEP

O aumento do número de vagas ao longo dos anos devido ao acréscimo de novos cursos de estatística pelo Brasil e o número de concluintes manteve-se constante o que demonstra o problema da evasão no curso de Estatística no Brasil, levando-se em conta que nem todos os alunos evadiram (há o problema da retenção no curso de Estatística

nas séries ou disciplinas iniciais), mas pode-se afirmar que a maioria dos alunos que não formaram no tempo foi devido a evasão.

Segundo Costa (1979), retenção refere-se aos alunos de uma mesma geração que não conseguiram, dentro um mesmo ciclo normal de estudos, o número mínimo de créditos exigidos para a conclusão de um determinado curso. Rendimento configura-se quando temos alunos de uma mesma geração que iniciaram um curso e o concluíram ao final. O baixo rendimento do aluno pode acabar levando à retenção e este à evasão. Pode ocorrer evasão por vários motivos: trabalho, doença grave, morte ou transferência.

De acordo com Nunes e Biurrun (2010) que referenciou um estudo sobre evasão, realizado pela Pró-Reitoria de Planejamento (PROPLAN) da UFRGS em 1991, temos os seguintes conceitos para Evasão:

- Evasão definitiva: é a saída definitiva do sistema de ensino. Dentro da universidade pode ser caracterizada por abandono de curso, desistência definitiva ou transferência para outra Instituição de Ensino
- Evasão temporária: é toda e qualquer saída temporária da Universidade tais como trancamento voluntário e trancamento “ex-offício”.
- Evasão de curso: é a transferência interna de curso dentro da mesma Instituição.

De acordo com Teles (1999) a evasão escolar no ensino superior compreende como toda forma de saída do curso de graduação que não seja a própria graduação do aluno no curso. Nesse conceito incluem-se as desistências, mudanças de curso, abandono, transferência e jubramento (desligamento compulsório por decurso do tempo máximo de graduação).

A evasão escolar, de qualquer maneira, gera um grande prejuízo ao sistema educacional, por se tratar de um problema que atinge de maneira direta a instituição, assim como o próprio aluno evadido, pois, os motivadores da evasão não estão exclusivamente ligados aos discentes, mas também, a própria instituição por falta de informação e consequentemente, por inexistência de um programa de combate a evasão.

2.2 Marco Histórico da Análise de Correspondência

De acordo com Greenacre e Blasius (2006) a história da análise de correspondência (AC) pode ser rastreada até Hirsfeld (1935) (mais tarde mudando seu nome para Hartley), que deu uma formulação algébrica da correlação entre linhas e colunas de uma Tabela de contingência. Fisher (1940) usou as mesmas ideias no âmbito da análise discriminante e também é considerado um dos “pais fundadores” da técnica.



Figura 2: Imagem de Jean-Paul Benzécri

Segundo Fox (2010), a análise de correspondência se popularizou na década de 60 e 70 por meio do estatístico francês Jean-Paul Benzécri, primeiramente na França e se estendendo pela Europa, porém, teve suas origens em trabalhos do início do século XX feitos por Pearson e Fisher, dois dos mais importantes nomes da estatística. A análise de correspondência é uma técnica de análise multivariada, que permite verificar graficamente a relação existente entre as categorias das variáveis qualitativas, podendo assim, verificar tanto entre duas variáveis, no caso da Análise de Correspondência Simples (AC), assim como, mais de duas, no caso Análise de Correspondência Múltipla (ACM). A AC pode ser entendida, segundo Greenacre e Blasius (2006), como um caso particular da análise de componentes principais (PCA) para dados categóricos, uma vez que PCA trabalha com dados quantitativos.

Segundo Prado (2012), se realizarmos uma busca pelo tema “correspondence analysis”, por meio da plataforma de pesquisa ISI - Web of Science (2012), é possível observar o aumento das publicações sobre o tema na última década, sendo apresentado na Figura 3.

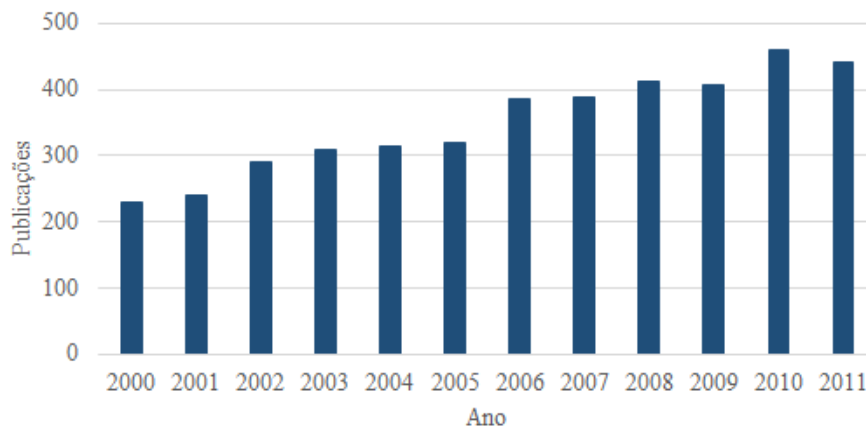


Figura 3: Número de publicações sobre “correspondence analysis” de 2000 a 2011 de acordo com ISI-Web of Science (2012)

Fonte: (PRADO, 2012)

Neste trabalho é utilizada a AC, apresentando o desenvolvimento algébrico da técnica, assim como a aplicação nos dados obtidos da pesquisa com os alunos evadidos do curso de Bacharelado em Estatística da Universidade Estadual da Paraíba, Campus I.

2.3 Análise de Correspondência

De acordo com Cunha (1997), a análise de correspondência estuda a associação entre variáveis categóricas de forma bivariada por meio da AC ou multivariada (ACM), por meio de um conjunto de mapas (gráficos) perceptuais/intuitivos, em que nestes mapas verifica-se a existência de proximidade, chamada de similaridade ou dissimilaridade entre os objetos propostos, ou seja, entre as categorias das variáveis em um específico estudo.

De acordo com Prado (2012), a associação entre as categorias é observada graficamente por meio da localização de cada uma delas em relação aos eixos. Pontos de categorias, localizados próximo a origem apresentam associações pequenas com as demais, enquanto que, quanto mais afastadas da origem, e próximas umas das outras, maior a associação entre as categorias. De forma geral, categorias presentes em um mesmo quadrante, apresentam características comuns. Embora o conceito da técnica seja de fácil entendimento, assim como a aplicação em pacotes computacionais seja facilitada, é importante o entendimento dos métodos matemáticos para uma compreensão adequada da técnica, da sua aplicabilidade e também da interpretação dos resultados obtidos, tornando assim, uma análise mais confiável e eficiente.

2.3.1 Tabelas de contingência

Tabelas de contingência são tabelas com classificação cruzada de n objetos por duas ou mais variáveis, normalmente categóricas, sendo as categorias mutuamente exclusivas e exaustivas:

Tabela 1: Representação tabular de contingência com I linhas e J colunas.

Linhas	Colunas				Totais das Linhas
	1	2	...	J	
1	n_{11}	n_{12}	...	n_{1J}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2J}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮
I	n_{I1}	n_{I2}	...	n_{IJ}	$n_{I.}$
Totais das colunas	$n_{.1}$	$n_{.2}$...	$n_{.J}$	$n_{..} = n$

Em que:

- n_{ij} é a frequência observada da categoria i da variável de linha e j da variável de coluna, onde $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, J$.
- $n_{i.}$ é a frequência total da i -ésima categoria da variável de linha, ou seja, o total marginal da i -ésima linha, dado por $n_{i.} = \sum_{j=1}^J n_{ij}$.
- $n_{.j}$ é a frequência total da j -ésima categoria da variável de coluna, ou seja, o total marginal da j -ésima coluna, dado por $n_{.j} = \sum_{i=1}^I n_{ij}$.
- $n_{..}$ ou n é o total geral da tabela ou o total de indivíduos, dado por $n_{..} = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$.

Alguns autores como Fávero (2009), Mingoti (2005), Goncalves (2009) afirmam que uma das formas de se iniciar a AC é por meio da construção da tabela de contingência e em seguida realizar o teste qui-quadrado, no intuito de testar a hipótese nula (H_0) de independência entre as linhas e colunas da tabela, construída a partir do cruzamento das categorias das variáveis estudadas. Para o teste qui-quadrado iremos utilizar a seguinte expressão:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

em que o_{ij} representa a frequência observada na i -ésima linha e j -ésima coluna, e_{ij} representa a frequência esperada na i -ésima linha e j -ésima coluna, $i = 1, \dots, I$ representa o número de linhas e $j = 1, \dots, J$ o número de colunas na tabela de contingência. A frequência esperada é calculada por

$$e_{ij} = \frac{n_{i.}n_{.j}}{n_{..}}$$

com $n_{i.}$, $n_{.j}$ e $n_{..}$ representando as frequências marginais e o total da tabela de contingência a ser analisada.

2.3.2 Desenvolvimento Algébrico da AC

Segundo Mingoti (2005), a matriz de correspondência $\mathbf{P}_{p \times q}$ é constituída das proporções $p_{ij} = \frac{n_{ij}}{n_{..}}$, ou seja, o número de elementos que pertencem à categoria i da variável X e à categoria j da variável Y, dividido pelo total de indivíduos e sendo assim, a soma dos elementos de $\mathbf{P}_{p \times q}$ passa a ser, portanto, igual a 1. Assim, temos a seguinte matriz:

$$\mathbf{P}_{p \times q} = \begin{bmatrix} \frac{n_{11}}{n} & \frac{n_{12}}{n} & \dots & \frac{n_{1J}}{n} \\ \frac{n_{21}}{n} & \frac{n_{22}}{n} & \dots & \frac{n_{2J}}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n_{J1}}{n} & \frac{n_{J2}}{n} & \dots & \frac{n_{JJ}}{n} \end{bmatrix}$$

A partir da matriz \mathbf{P} , calcula-se a matriz de perfil das linhas e das colunas, em que precisa-se obter os vetores de frequências relativas marginais \mathbf{r} e \mathbf{c} , vetores linhas e colunas respectivamente, também chamada de massa, para cada célula da tabela de contingência, podendo ser construído a partir dos elementos dos vetores \mathbf{r} e \mathbf{c} das seguintes formas:

$$r_i = \sum_{j=1}^J p_{ij} = \sum_{j=1}^J \frac{n_{ij}}{n}$$

$$c_j = \sum_{i=1}^I p_{ij} = \sum_{i=1}^I \frac{n_{ij}}{n}$$

De acordo com Fávero (2009), massa é uma medida que indica a influência de um

objeto com base em sua frequência marginal, ou seja, representa o quanto do total de cada categoria está presente no estudo (vertical e horizontal). A massa afeta o centroide, que é uma média ponderada do perfil linha e da coluna. Além disso, os pontos com maior massa puxam o centroide linha (coluna) fortemente para sua posição.

Com a criação dos vetores de perfil de linhas e colunas, segue para a construção das matrizes de perfil de linhas e de colunas, onde a mesma é composta da matriz diagonal dos elementos de \mathbf{r} e de \mathbf{c} , respectivamente. Assim, temos as matrizes \mathbf{D}_r e \mathbf{D}_c :

$$\mathbf{D}_r = \begin{bmatrix} r_1 & 0 & 0 & 0 \\ 0 & r_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & r_J \end{bmatrix}$$

$$\mathbf{D}_c = \begin{bmatrix} c_1 & 0 & 0 & 0 \\ 0 & c_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & c_I \end{bmatrix}$$

Agora, precisa-se fazer a decomposição em coordenadas principais como proposto por Mingoti (2005). Para isso, considere a matriz $\tilde{\mathbf{P}}$, dada por:

$$\tilde{\mathbf{P}} = \mathbf{P} - \mathbf{r}\mathbf{c}'$$

Essa matriz tem entradas do tipo: $(p_{ij} - \frac{n_{i.}n_{.j}}{n})$, ou seja, os termos representam uma comparação da proporção observada no interior de cada casela da tabela, com aquela esperada sob um modelo no qual as variáveis X (Linhas) e Y (Colunas) são independentes.

Segundo Infantosi *et al.* (2014, p. 475) pré-multiplicando e pós-multiplando a matriz $\tilde{\mathbf{P}}$ por $\mathbf{D}_c^{-1/2}$ e $\mathbf{D}_r^{-1/2}$ obtêm-se um nova matriz conforme equação (2.1) que é a padronização da matriz de perfis. A ideia por trás da construção desta matriz está na necessidade da disposição gráfica de vetores multivariados em utilizar a origem das próprias variáveis, ou determinar uma nova origem por meio do método de mínimos quadrados, no qual a soma dos quadrados das distâncias dos pontos até os eixos é minimizada. O método computacional mais utilizado para tal minimização é o algoritmo de Decomposição por Valores Singulares (DVS), em que a matriz de perfis é fatorada em três matrizes (INFANTOSI *et al.*, 2014).

$$\mathbf{S} = \mathbf{D}_c^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}')\mathbf{D}_r^{-1/2} \quad (2.1)$$

que também pode ser escrito da seguinte maneira:

$$s_{ij} = (p_{ij} - r_i c_j) / \sqrt{r_i c_j}$$

A partir da matriz \mathbf{S} , será feita a Decomposição por Valores Singulares - DVS, onde de acordo com Johnson e Wichern (2007) pode ser escrito como

$$\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}',$$

em que $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_k\}$, com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$, \mathbf{U} é uma matriz ortogonal de ordem $n \times k$, isto é, $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$. O conjunto de valores $\{\lambda_i\}$ são chamados de valores singulares de \mathbf{S} . Se \mathbf{U} e \mathbf{V} são escritos em termos de seus vetores coluna, $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_k]$ e $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_k]$, então $\{\mathbf{u}_i\}$ são os vetores singulares à esquerda de \mathbf{S} e $\{\mathbf{v}_i\}$ são os vetores singulares à direita de \mathbf{S} . A matriz \mathbf{S} pode então ser escrita como

$$\mathbf{S} = \sum_{i=1}^k \lambda_i \mathbf{u}_i \mathbf{v}_i'$$

Segue-se então, a partir das matrizes \mathbf{U} , \mathbf{V} e $\mathbf{\Lambda}$, pode-se construir as coordenadas principais das linhas e das colunas.

As coordenadas principais das linhas são definidas como:

$$\begin{aligned} \mathbf{F} &= \mathbf{D}_r^{-1/2} \mathbf{U}\mathbf{\Lambda} = \\ \mathbf{F} &= \mathbf{D}_r^{-1/2} \mathbf{S}\mathbf{V} \end{aligned}$$

E as coordenadas principais das colunas são:

$$\begin{aligned} \mathbf{G} &= \mathbf{D}_c^{-1/2} \mathbf{V}\mathbf{\Lambda} = \\ \mathbf{G} &= \mathbf{D}_c^{-1/2} \mathbf{S}'\mathbf{U} \end{aligned}$$

Por meio das matrizes \mathbf{F} e \mathbf{G} , constroem-se os gráficos de dispersão das coordenadas principais linhas e colunas, para que se possa analisar a associação entre as categorias de suas variáveis. Este é o gráfico de correspondência, pode-se também traçar os gráficos somente das coordenadas principais das linhas e/ou colunas (MINGOTI, 2005).

2.3.3 Nuvens de pontos

Devido ao conceito básico da Análise de Correspondência ser a verificação de associação as categorias das variáveis, é necessário representar tanto as categorias de linha, quanto de colunas em um mesmo gráfico, para o entendimento completo desta técnica é necessário maiores esclarecimentos teóricos geométricos acerca do assunto que podem ser visto em Naito (2007), porém não deixa-se aqui de explicitar a ideia da construção gráfica neste texto. Sendo assim, será apresentado um exemplo numérico que facilitará o entendimento geométrico.

Admite-se que as componentes de cada vetor seguem uma distribuição multinomial condicional ao total da linha ou coluna e obviamente sua soma é 1, o que gera uma dependência linear entre tais componentes.

Assim sendo, cada um desses pontos vetoriais (perfil de linha ou coluna) podem ser representados em uma dimensão menor do que originalmente foram projetados, ou seja, dimensão $(J-1)$ para representar os perfis de linha e dimensão $(I-1)$ para representar os perfis de coluna. (SILVA, 2012)

Para maior entendimento, é utilizado um exemplo de Silva (2012), em que tem-se a seguinte matriz:

$$\mathbf{W} = \begin{bmatrix} 50 & 50 & 50 \\ 100 & 300 & 100 \\ 120 & 0 & 120 \\ 100 & 10 & 0 \end{bmatrix}$$

No qual, obtém-se a matriz \mathbf{P} .

$$\mathbf{P} = \begin{bmatrix} 0,05 & 0,05 & 0,05 \\ 0,10 & 0,30 & 0,10 \\ 0,12 & 0,0 & 0,12 \\ 0,10 & 0,01 & 0,0 \end{bmatrix}$$

Além da matriz \mathbf{P} , constrói-se a matriz de perfil de linha \mathbf{D}_r e de coluna \mathbf{D}_c .

$$D_r = \begin{bmatrix} 0,15 & 0,0 & 0,0 & 0,0 \\ 0,0 & 0,50 & 0,0 & 0,0 \\ 0,0 & 0,0 & 0,24 & 0,0 \\ 0,0 & 0,0 & 0,0 & 0,11 \end{bmatrix}$$

$$D_c = \begin{bmatrix} 0,37 & 0,0 & 0,0 \\ 0,0 & 0,36 & 0,0 \\ 0,0 & 0,0 & 0,27 \end{bmatrix}$$

A partir dos perfis de linha, é representado geometricamente no espaço Euclidiano tridimensional.

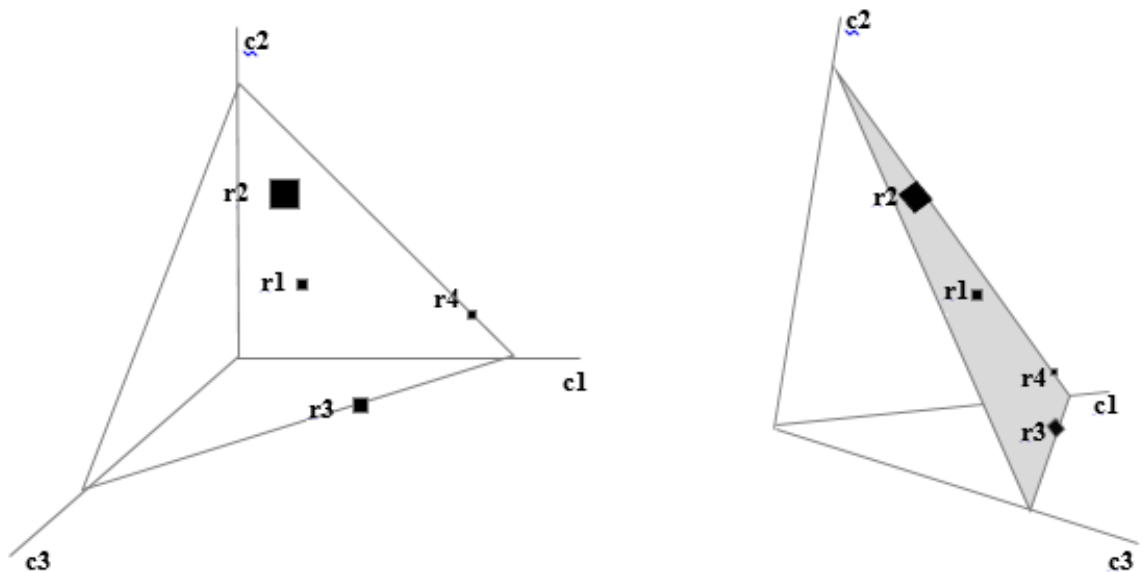


Figura 4: Representação tri-dimensional dos perfis de linha sob dois ângulos de visão. Adaptado de (SILVA, 2012)

2.3.4 Centro de gravidade ou centroide

Segundo Benzecri (1992, p. 28) o centro de gravidade de um sistema de pontos com massas que lhe são atribuídas (números positivos ou zeros) é uma generalização espacial da média aritmética. Sendo assim, para se determinar o centro de gravidade ou centroide,

deve-se fazer a média das componentes (coordenadas) referentes a cada eixo de base, ponderadas pela sua respectiva massa.

O centroide linha de uma tabela de contingência indica geometricamente a posição média dos perfis linha, como se fosse o centro de gravidade ou o ponto de equilíbrio da matriz de dados (CZERMAINSKI, 2004). A média ponderada dos perfis linha ou perfil médio de linhas ou centroide, será dado por:

$$g_r = A' r,$$

com

$$A = D_r^{-1} P.$$

Utilizando o exemplo de anteriormente, obtém-se $g_r = [0,37 \ 0,36 \ 0,27]$, que será projetado juntamente com os perfis de linha:

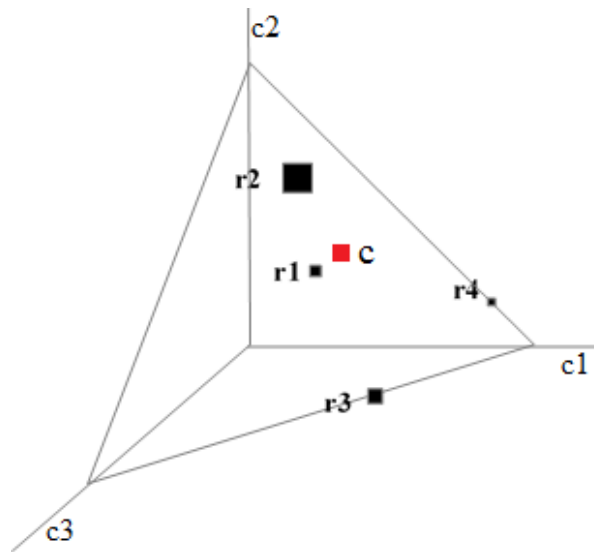


Figura 5: Representação geométrica dos perfis linha e do centróide (c) no espaço Euclidiano tridimensional. Adaptado de (SILVA, 2012)

Da mesma forma, pode-se obter a média ponderada dos perfis coluna ou perfil médio de coluna, da seguinte forma:

$$g_c = B' c,$$

com

$$B = D_c^{-1} P'.$$

2.3.5 Inércia

De acordo com Silva (2012), com o intuito de medir a dispersão dos pontos em relação ao centro de gravidade, calcula-se a inércia, que é um conceito da física, semelhante ao conceito estatístico de variância. Segundo Hair (2010) a inércia total de uma tabela de tabulação cruzada é calculada como o qui-quadrado total dividido pela frequência total (soma de linhas ou colunas). Segundo Johnson e Wichern (2002) o total de inércia está relacionada com a estatística qui-quadrado podendo ser escrito como χ^2/n .

A inércia pode então ser calculada para cada categoria de linha ou coluna e assim quantificar sua contribuição ao total da inércia, essa é também definida como a soma dos quadrados de todos os valores singulares ou autovalores não nulos (FÁVERO, 2009). Assim:

$$\sum_{i=1}^k \lambda_i^2.$$

Em que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$ são os elementos não nulos da matriz Λ . A proporção de explicação da i -ésima coordenada principal em relação a inércia total é dada por:

$$\frac{\lambda_i^2}{\sum_{i=1}^k \lambda_i^2}.$$

Assim, para cada dimensão, a inércia é igual ao quadrado do valor singular ou autovalor. Portanto, de acordo com Silva (2012), pode-se medir a importância relativa de cada eixo (dimensão) para reter a inércia total da nuvem sobre si mesmo e esta medida é dada pela seguinte proporção:

$$\omega_\alpha = \frac{\lambda_\alpha}{\sum_{\alpha=1}^k \lambda_\alpha} \cdot 100\%$$

Para redução de dimensionalidade selecionam-se os principais eixos que captam a inércia da nuvem de pontos, e que ao mesmo tempo possam ser representados de forma bi ou tridimensional. Sendo assim, pode-se medir o quão boa é esta representação, através do somatório dos ω_α dos eixos envolvidos na representação. Deve-se procurar manter um número de eixos tal que a porcentagem de inércia explicada seja superior a 50% (NAITO, 2007, p. 49)

3 Materiais e métodos

Os dados utilizados para a aplicação da Análise de Correspondência provém de uma pesquisa realizada com os alunos que evadiram com curso de Bacharelado em Estatística na Universidade Estadual da Paraíba, Campus I, situada em Campina Grande, estado da Paraíba, em que foram considerados apenas para o estudo os alunos que iniciaram o curso de Bacharelado em Estatística após 2010.1, pois neste período houve uma reformulação do curso e deu-se início a grade semestral do curso.

Foi aplicado um questionário semiestruturado com 46 alunos, no período de 19 de Novembro de 2015 à 05 de Dezembro de 2015. Ao todo entre 2010.1 à 2014.1 houveram mais de 250 alunos que haviam evadido. A escolha dos indivíduos para compor a amostra, deu-se por conveniência, em que o universo a ser pesquisado se deu a partir de uma lista repassada pela coordenação do curso, com os alunos que haviam evadido do curso, a partir desta lista foi possível obter o contato do aluno e desta forma aplicar o questionário. Para obter informações mais apuradas, definiu-se que os indivíduos deveriam ter carga horária maior do que zero, pois alguns ingressam no curso e não persistem até o término do primeiro período e portanto, não poderiam com clareza enumerar as causas para evasão.

Antes da aplicação do questionário oficial, foi feito um pré-teste, com a finalidade de deixá-lo o mais eficiente possível, assim como também, foi utilizado a base de um questionário construído em Novembro de 2012, na disciplina de Métodos e Técnicas de Pesquisa, onde naquela mesma ocasião foi realizado uma pesquisa piloto com 9 alunos evadidos. O questionário desta pesquisa foi aplicado de 3 maneiras distintas:

- Pessoalmente: Onde o entrevistado foi abordado pessoalmente e o entrevistador realizou as perguntas e anotou-se as respostas;
- Telefone: Através da lista de contatos, foi realizado a pesquisa com os alunos por telefone, de modo que, aqueles que não haviam o contato, tentou-se a terceira maneira;

- Mala Direta: Foi construído o questionário na plataforma do Google, onde era gerado um link e o mesmo foi enviado para os e-mails daqueles que na lista não havia telefone de contato. O entrevistado ao enviar o questionário, as respostas eram enviadas para outra plataforma no Google Drive, e em seguida transpassadas para o questionário físico.

Após a aplicação dos questionários, foi construído a plataforma de recebimento dos dados em Excel (2013), de modo que também foi possível realizar as primeiras análises descritivas dos dados, com a geração de tabelas e gráficos. A aplicação da análise de correspondência deu-se em tabelas de contingência, sendo essas, tabelas de duas entradas, onde nas linhas são apresentadas as categorias de uma variável, nas colunas as categorias de outra variável e nas células as frequências observadas de objetos ou indivíduos. A representação gráfica se dá para as categorias das variáveis e por cada variável (GREENACRE, 2007).

As variáveis analisadas na análise de correspondência foram motivo de cursar, motivo de evadir e motivo de reingressar. Para a realização da Análise de Correspondência, utilizamos o Software livre R em sua versão 3.2.2 (TEAM, 2013) e para buscar o banco de dados foi preciso salvá-lo em formato de csv. A descrição das variáveis estudadas tanto na AC como apresentadas na parte descritiva é apresentada na Tabela 2, em que tem-se na primeira coluna da tabela as variáveis, na segunda coluna a sigla das categorias das variáveis e na terceira coluna, a descrição das categorias.

Tabela 2: Descrição das variáveis coletadas na pesquisa

Variáveis	Sigla	Categorias
A1 : Sexo	A1: 1	A1: Masculino
	A1: 2	A1: Feminino
A2: Estado civil	A2: 1	A2: Solteiro
	A2: 2	A2: Casado
	A2: 3	A2: Separado
	A2: 4	A2: Divorciado
	A2: 5	A2: Viúvo
	A2: 6	A2: União estável
	A2: 7	A2: Não Informa
A3: Você tem filhos?	A3: 1	A3: Sim
	A3: 2	A3: Não
A3.1: Quantos Filhos		Espontânea
A4: Idade	A4: 1	A4: 18 a 24 anos
	A4: 2	A4: 25 a 34 anos
	A4: 3	A4: 35 a 44 anos
	A4: 4	A4: 45 a 54 anos
	A4: 5	A4: 55 a 64 anos
	A4: 6	A4: Mais de 64 anos
	A4: 7	A4: Não Informa
A5: Renda Familiar	A5: 1	A5: Até 1 SM
	A5: 2	A5: De 1 a 3 SM
	A5: 3	A5: De 3 a 6 SM
	A5: 4	A5: De 6 a 8 SM
	A5: 5	A5: Mais de 8 SM
B1: Tipo de escola cursou o 2º grau	B1: 1	B1: Somente em escola pública
	B1: 2	B1: Maior parte em escola pública
	B1: 3	B1: Somente em escola particular
	B1: 4	B1: Maior parte em escola particular
B2: Turno que cursou o 2º grau	B2: 1	B2: Somente no turno diurno
	B2: 2	B2: Maior parte no turno diurno
	B2: 3	B2: Somente no turno noturno
	B2: 4	B2: Maior parte no turno noturno
B3: Principal motivo de optar pelo curso	B3: 1	B3: Baixa concorrência
	B3: 2	B3: Não sabia qual curso optar
	B3: 3	B3: Indicação de um/a amigo/a ou parente
	B3: 4	B3: Possibilidade de atuar em várias áreas
	B3: 5	B3: Afinidade com a área/ curso
	B3: 6	B3: Facilidade de emprego após conclusão
B4: Participou de algum evento científico	B4: 1	B4: Sim
	B4: 2	B4: Não
B5: Principal motivo de evadir	B5: 1	B5: Por causa do trabalho
	B5: 2	B5: Optou por outro curso
	B5: 3	B5: Dificuldade em cálculo
	B5: 4	B5: Assuntos pessoais (Gravidez, separação etc.)
B6: Principal motivação para reingresso	B6: 1	B6: Melhoria na infraestrutura
	B6: 2	B6: Mais eventos relacionados ao curso
	B6: 3	B6: Mais oportunidades de bolsas
	B6: 4	B6: Melhoria nas condições financeiras
	B6: 5	B6: Melhoria por parte do mercado de trabalho (Estágio)
	B6: 6	B6: Outros motivadores
	B6: 7	B6: Não voltaria

4 Resultados e discussões

4.1 Análise descritiva

Uma análise descritiva é de suma importância, antes da aplicação da Análise de Correspondência, de modo que, pode-se entender o comportamento do fenômeno estudado e o auxilie para a interpretação dos resultados obtidos pela técnica.

4.1.1 Perfil dos alunos

Neste item apresenta-se o perfil dos alunos evadidos entrevistados, onde alguns resultados se refletiu da própria natureza do perfil dos alunos de Estatística, como por exemplo, haver mais pessoas do sexo masculino do que feminino.

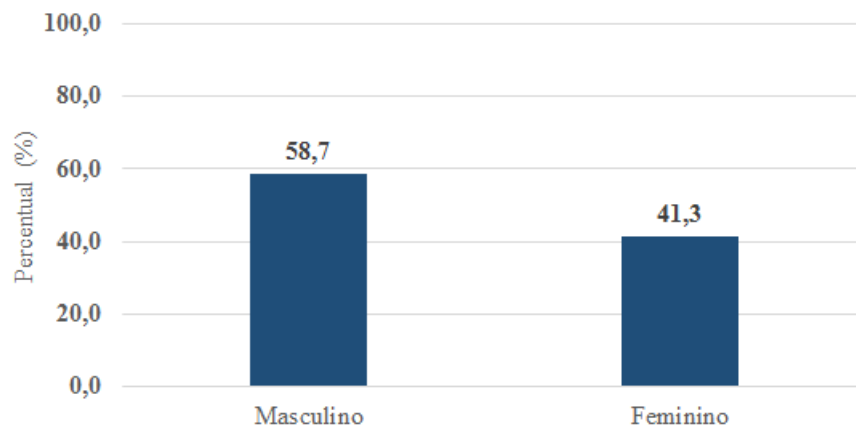


Figura 6: Gráfico da porcentagem de alunos entrevistados por gênero

Dentre os estudantes evadidos que foram entrevistados, 56,5% eram solteiros, 39,1% casados e 4,3% divorciados.

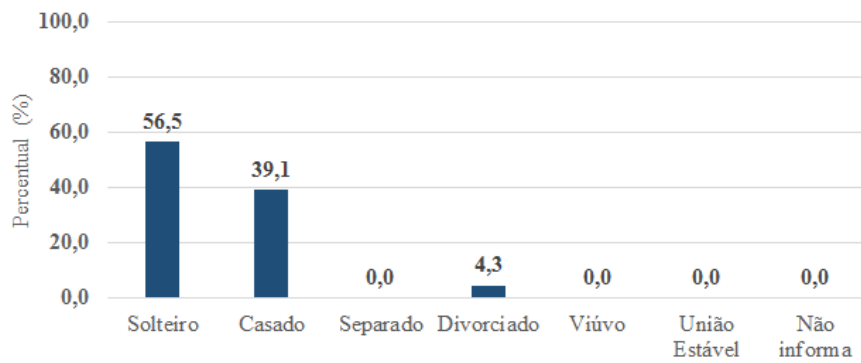


Figura 7: Gráfico da porcentagem de alunos entrevistados por estado civil

De acordo com a Figura 8, a maioria dos alunos (50,0%) são pessoas de idade de 25 a 34 anos, 37,0% com idade de 18 a 24 anos e 13,0% são estudantes de 35 a 44 anos de idade.

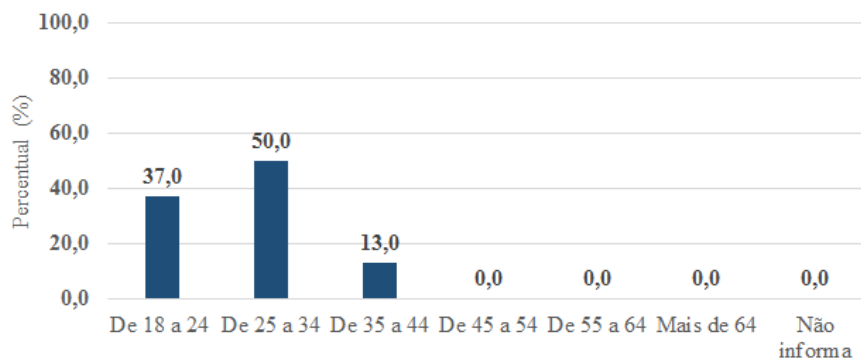


Figura 8: Gráfico da porcentagem de alunos entrevistados por faixa etária.

Sendo assim, pode-se também calcular a média de idade com intervalo, onde segundo Pommer (2013), a média é dada pela expressão do somatório do percentual da classe vezes a média do intervalo de classe, em que x_m representa o ponto médio de cada classe e f a frequência da classe.

Tabela 3: Organização tabular dos dados para obtenção da média da idade dos entrevistados

Classe (R\$)	f	x_m	$f \cdot x_m$
18 a 24	17	21	357
25 a 34	23	29,5	678,5
35 a 44	6	39,5	237,5
Total	46	-	1.272,5

$$\bar{x} = \frac{\sum f * x_m}{\sum f} = \frac{1.272,5}{46} = 27,663 \cong 28.$$

Portanto, a média de idade dos alunos que evadiram que foram entrevistados é aproximadamente 28 anos. No que refere a renda familiar dos entrevistados, ou seja, a soma da renda de cada indivíduo de sua residência, 69,6% tem renda familiar de 1 a 3 salários mínimos (SM), ou seja, de R\$ 788,00 a R\$ 2.364,00. Além disso, 13,0% tem renda familiar até 1 salário mínimo, 15,2% de 3 a 6 SM, e 2,2% de 6 a 8 SM.

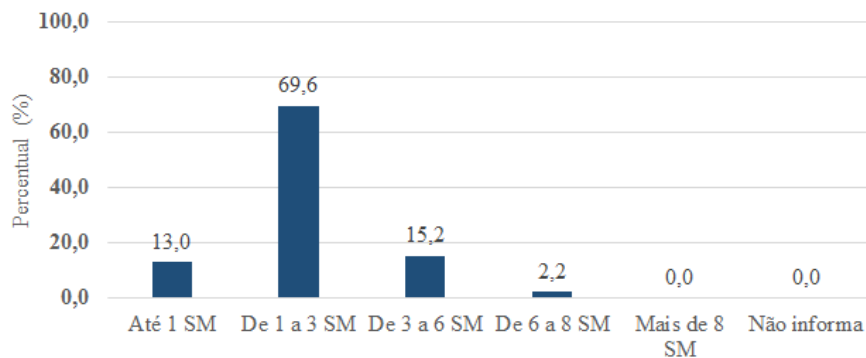


Figura 9: Gráfico da porcentagem de alunos entrevistados por renda familiar

Da mesma maneira que foi calculado para encontrar a média de idade, faz-se para a média da renda familiar, como mostra a Tabela 4 e em média, a renda familiar dos alunos entrevistados evadidos foi R\$ 1.807,26.

Tabela 4: Organização tabular dos dados para obtenção da média da renda familiar dos entrevistados

Classe (R\$)	f	x_m	$f * x_m$
0,00 a 788,00	6	394	2.364
788,00 a 2.364,00	32	1.576	50.432
2.364,00 a 4.728,00	7	3.546	24.822
4.728,00 a 6.304,00	1	5.516	5.516
Total	46	-	83.134

$$\bar{x} = \frac{\sum f * x_m}{\sum f} = \frac{83.134}{46} = 1.807,26$$

Foi questionado também, se os entrevistados tem filhos e 32,6% informaram que tem

e em média são $1,86 \cong 2$ filhos. Referente ao tipo de escola que os entrevistados cursaram o ensino médio (2º Grau), 78,3% cursaram somente em escola pública e referente ao turno que cursou 65,2% informaram que cursaram somente em turno diurno.

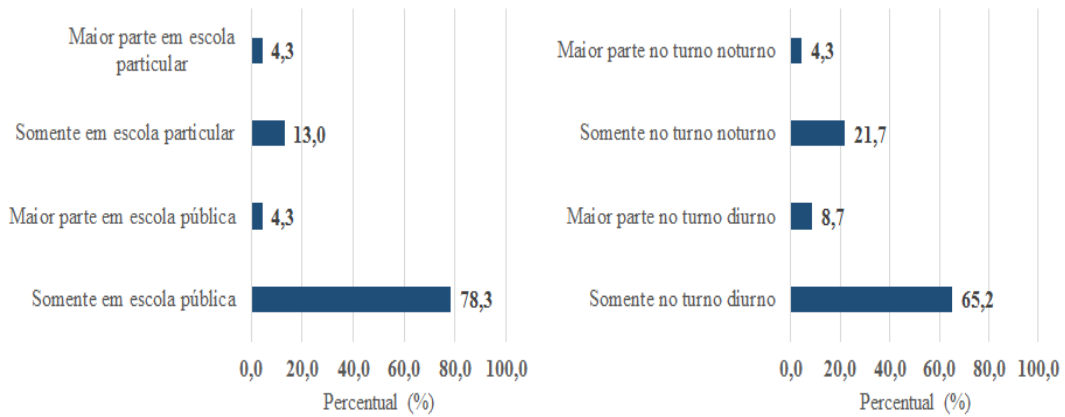


Figura 10: Gráfico da porcentagem de alunos entrevistados por tipo de escola que cursou o 2º grau e turno que cursou o 2º grau

Buscando compreender o motivo para a evasão dos alunos, foi questionado alguns pontos que podem influenciar na escolha de evadir do curso, como por exemplo, os motivos que o fez escolher pelo curso. Como pode-se observar na Figura 11, o principal motivador para optar pelo curso, foi devido a baixa concorrência (26,1%), seguido de afinidade com a área ou com o curso.

É importante destacar que também foi questionado se o aluno havia participado de algum evento científico, como congresso, Simpósio, Semana da Estatística ou congresso de iniciação científica, e como mostra na Figura 11, 41,3% informaram que participaram de pelo menos um evento desses citados.

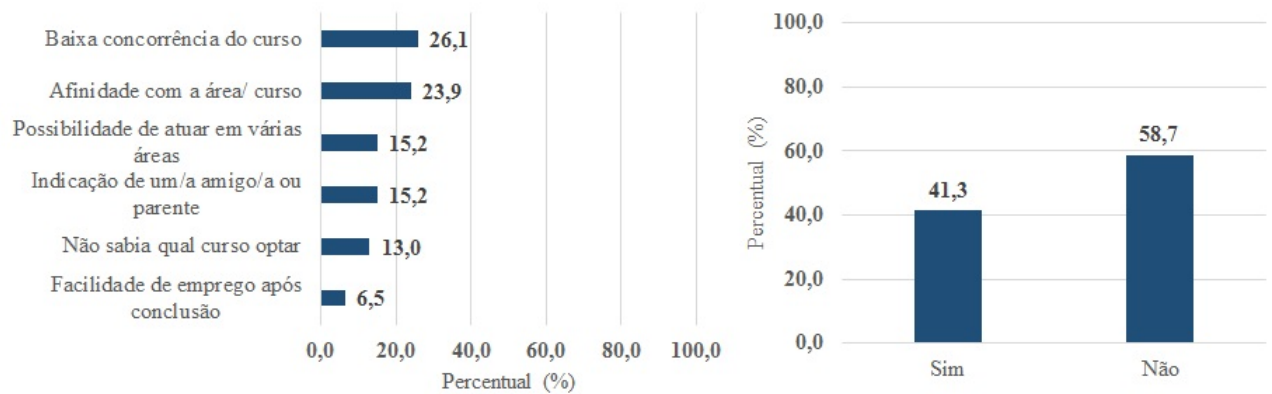


Figura 11: Gráfico da porcentagem de alunos entrevistados por motivos de optar pelo curso e relacionado a participação em eventos científicos

O gráfico abaixo refere-se ao motivo de evasão do aluno, em que 37,0% evadiu devido a ter optado por outro curso, 34,8% por assuntos pessoais como: Gravidez (17,4%), Deslocamento (2,2%), Mudança de endereço (4,3%), dificuldades financeiras (6,5%), Separação (2,2%) e Doença na família (2,2%). Além disso, 21,7% evadiu devido ao trabalho e 6,5% por dificuldade em cálculo.

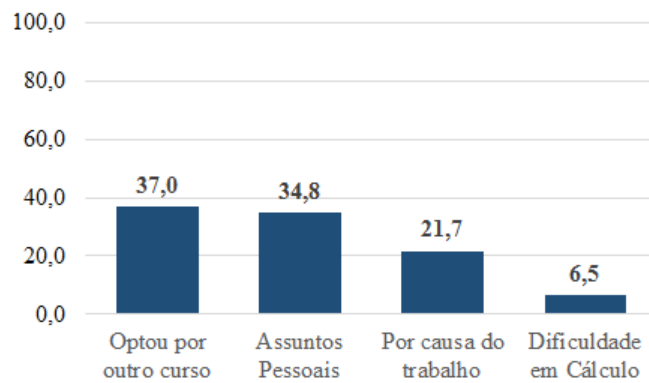


Figura 12: Gráfico da porcentagem de alunos entrevistados por motivos de evasão no curso

Foi questionado de forma estimulada, os motivadores para reingresso dos alunos, e 23,9% informaram que dentre os citados, o principal motivador seria se houvesse melhoria por parte do mercado de trabalho (Estágio), 15,2% citaram outros (Ter mais tempo (13,0%) e morar em Campina Grande (2,2%)), se houvesse mais incentivo por parte dos professores e melhoria nas condições financeiras foi citado por 15,2% e mais oportunidade de bolsas, melhoria na infraestrutura e mais eventos relacionados ao curso foram citados por 13,0%, 6,5% e 4,3%, respectivamente.



Figura 13: Gráfico da porcentagem de alunos entrevistados por motivadores para regresso no curso

4.2 Aplicação da Análise de Correspondência

Para facilitar a visualização na tabela, buscou-se uma nomenclatura para as categorias das variáveis, utilizando as abreviações apresentadas na Tabela 2. A aplicação da técnica primeiramente será realizada em função dos motivos de cursar com os motivos de evadir.

Tabela 5: Representação tabular de contingência dos motivos de optar pelo curso e motivos de evadir

Motivo de evadir	Motivos de optar pelo curso						Totais das linhas
	B3.1	B3.2	B3.3	B3.4	B3.5	B3.6	
Mudou de curso	2	2	4	2	6	1	17
Assuntos pessoais	6	1	2	2	4	1	16
Trabalho	3	2	1	2	1	1	10
Dificuldade em Cálculo	1	1	0	1	0	0	3
Total	12	6	7	7	11	3	46

Pela Tabela 6, verificou-se que o eixo 1 é responsável por 69,0% da inércia total e o eixo 2 por 26,2%, que conjuntamente somam 95,1% da inércia total, sendo assim, existe boa representatividade dos perfis no espaço \mathfrak{R}^2 .

Tabela 6: Decomposição da inércia sobre os eixos principais

Dimensão	Valor	%	Cum%
1	0,142431	69,0	69,0
2	0,054047	26,2	95,1
3	0,010093	4,9	100,0
4	00000000	0,0	100,0
Total	0,206571	100,0	

Como pode-se observar na Figura 14, sugere relação entre o motivo de cursar (B3:3) (Indicação de um/a amigo/a ou parente) e (B3:5) (Afinidade com a área do curso) e ter evadido porque mudou de curso. Aqueles que disseram que não sabia qual curso optar (B3:2), está associado a evadir do curso por causa de dificuldade em cálculo, e por estarem distantes da origem, implica que existe maior associação entre elas. Dos que informaram ter desistido por causa do trabalho está associado a ter optado pelo curso por ter possibilidade de atuar em várias áreas (B3:4). Por fim, daqueles que optou pelo curso por causa da baixa concorrência (B3:1) e por ter facilidade de emprego após conclusão (B3:6) se associa a ter evadido por assuntos pessoais.

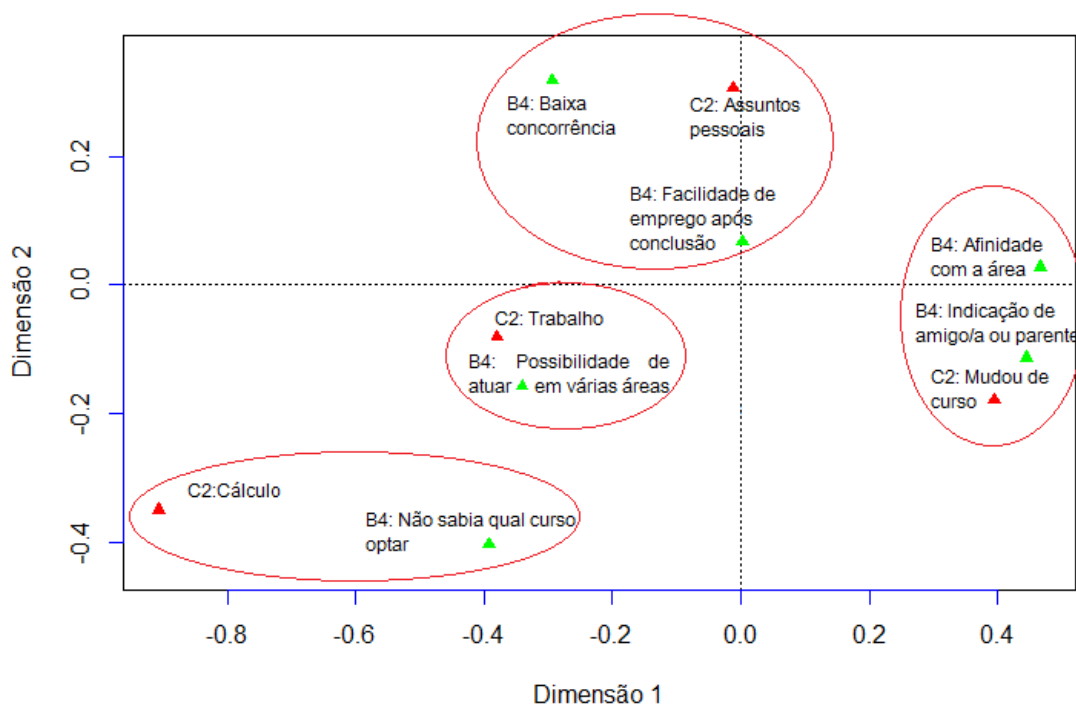


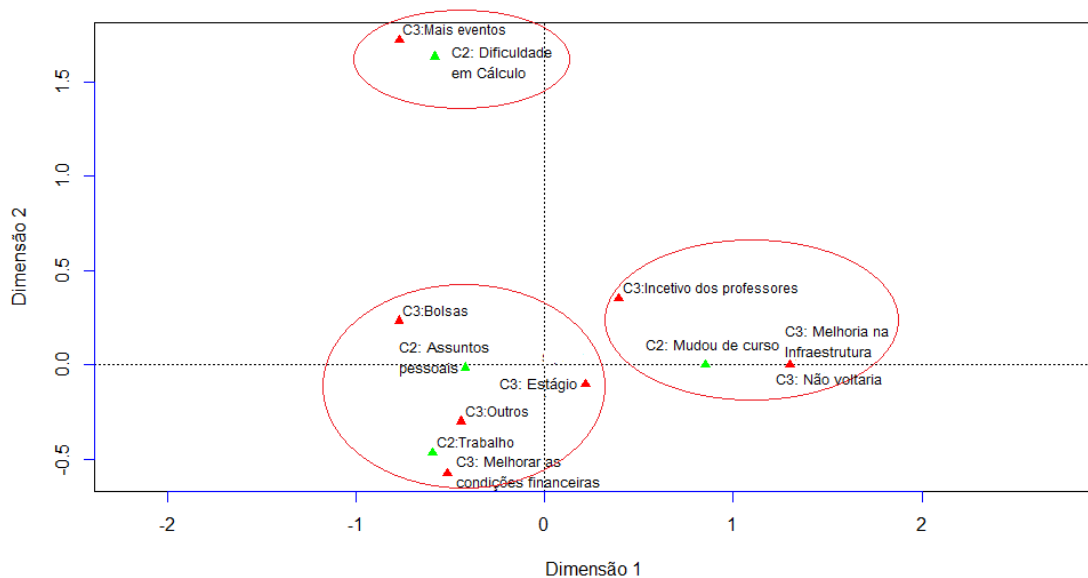
Figura 14: Mapa perceptual entre motivo de optar pelo curso e motivo de evadir

Para os motivos de evasão e os motivadores de reingresso, foi realizada a técnica, em que acordo com a Tabela 7, verificou-se que o eixo 1 é responsável por 55,1% da inércia total e o eixo 2 por 25,6%, que conjuntamente somam 80,7% da inércia total, e assim como para os eixos da primeira análise, também existe boa representatividade dos perfis no espaço \mathbb{R}^2 .

Tabela 7: Decomposição da inércia sobre os eixos principais

Dimensão	Valor	%	Cum%
1	0,502312	55,1	55,1
2	0,233851	25,6	80,7
3	0,127377	14,0	94,7
4	0,048293	5,3	100,0
5	0,000000	0,0	100,0
6	0,000000	0,0	100,0
7	0,000000	0,0	100,0
Total	0,911833	100,0	

Figura 15: Mapa perceptual entre motivo de evasão com motivadores para reingresso no curso



Pode-se perceber que, daqueles que informaram ter evadido por causa de dificuldade em cálculo está fortemente associado a voltaria se no curso houvesse mais eventos rela-

cionados a Estatística, onde essa relação pode ser compreendida tanto pela proximidade dos pontos, quanto pela distância ao eixo, e por estarem em um mesmo quadrante.

Percebe-se também, que formou-se um conjunto de categorias que estão associadas e o que leva a entender que são categorias relacionadas ao meio profissional do aluno evadido, pois, dos que disseram que evadiram por causa do trabalho e por assuntos pessoais, associa-se com os que informaram que voltaria se melhorasse as condições financeiras e outros (ter mais tempo e morar em Campina Grande), além da possibilidade de voltar se houvesse mais bolsas e estágios.

Partindo para o lado institucional, verifica-se associações entre o motivo de ter evadido porque mudou de curso, com voltaria para o curso de houvesse mais incentivo por parte dos professores, melhoria na infraestrutura ou até mesmo não voltaria. Desta forma, com essa representação e interpretação, sugere que há fenômenos que envolve não somente ao aluno, mas também por parte da Universidade.

5 Conclusões

A partir desse estudo, podemos levantar algumas considerações a cerca dos resultados descritivos da pesquisa sobre os motivos da evasão dos alunos no curso de Estatística na UEPB, Campus I, onde nos trouxe reflexões importantes acerca do tema, tanto pela ótica da causa, como também sobre um olhar de possível solução. Partindo dos resultados da aplicação da AC entre as variáveis motivos de optar pelo curso e motivos de evadir, as principais considerações é acerca daqueles que disseram não saber qual curso optar, está relacionado com os que informaram terem evadido por dificuldade em cálculo, assim como há relação entre ter optado por indicação de amigo/a ou parente e evadir por ter mudado de curso.

Além disso, aplicando a análise de correspondência entre os motivos de evadir e motivação para reingresso, percebeu-se relação entre os que disseram ter evadido por dificuldade em cálculo e reingressariam no curso se houvesse mais eventos relacionados ao curso. É importante destacar a relação entre os motivadores de reingresso que envolve a instituição (Incentivo dos professores e melhoria na infraestrutura) com ter evadido do curso por ter mudado de curso, sugerindo assim, que alguns motivadores de evasão não se caracteriza na responsabilidade apenas do aluno, mas também da instituição de ensino.

A análise de correspondência, mostrou ser uma técnica de bastante valia para análise de dados categóricos e sendo assim, propícia para nossa pesquisa. Além disso, por ser uma técnica de análise multivariada, onde o foco normalmente é análise de dados métricos, materiais sobre o tema publicado no Brasil ainda é escasso, mesmo que seja uma técnica que pode ser válida para várias áreas do conhecimento, no que se refere a pesquisas exploratórias. Sua aplicação e interpretação, apesar de ser simples, requer um bom embasamento teórico para sua construção.

Aos interessados em mais informações sobre a técnica, recomenda-se o leitura de Benzecri (1992), Greenacre (1993), Johnson e Wichern (2007), Silva (2012),

Referências

- BENZECRI, J. P. *Correspondence Analysis Handbook*. New York: Marcel Dekker, 1992.
- COSTA, V. *Evasão, Retenção e Rendimento em Relação à Ordem de Opção Atendida nos Cursos de Graduação da UFRGS*. Dissertação (Mestrado) — UFRGS, 1979.
- CUNHA, J. *Análise Multidimensional de Dados Categóricos: A aplicação das Análises de Correspondência Simples e Múltipla em Marketing e sua Integração com Técnicas de Análise de Dados Quantitativos*. Rio Grande do Sul: UFRGS, 1997.
- CZERMAINSKI, A. B. C. *Análise de Correspondência*. Piracicaba-SP, 2004.
- EXCEL, M. Microsoft. *Redmond, Washington: Computer Software.*, 2013.
- FÁVERO, L. P. *Análise de dados: modelagem multivariada para tomada de decisões*. Rio de Janeiro: Elsevier, 2009.
- FISHER, R. A. The precision of discriminant functions. *Annals of Eugenics*, 1940.
- FOX, J. *Multiple Correspondence Analysis*. Londres: SAGE, 2010.
- GONCALVES, M. T. Aplicação da análise de correspondência à avaliação institucional da fecilcam. In: *IV Encontro de Produção Científica e Tecnológica*. [S.l.: s.n.], 2009.
- GREENACRE, M.; BLASIUS, J. *Multiple Correspondence Analysis and Related Methods*. New York: Chapman & Hall/CRC, 2006.
- GREENACRE, M. J. *Theory and applications of correspondence analysis*. London: Academic Press, 1993.
- GREENACRE, M. J. *Correspondence analysis in practice*. 2. ed. Boca Raton: Chapman & Hall/CRC, 2007.
- HAIR, J. F. *Multivariate Data Analysis*. Seventh. New Jersey: Prentice Hall, 2010.
- HIRSFELD, H. O. A connection between correlation and contingency. *Proceedings of the Cambridge Philosophical Society*, p. 520, 524, 1935.
- INFANTOSI, A. F. C.; GAMA DIAS COSTA, J. C. da; ALMEIDA, R. M. V. R. de. Análise de correspondência: bases teóricas na interpretação de dados categóricos em ciências da saúde. *Cadernos de Saude Publica*, n. 30(3), p. 475 – 486, Março 2014.
- JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall, 2002.
- JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall, 2007.

- MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: UFMG, 2005.
- NAITO, S. D. N. P. *Análise de Correspondências Generalizada*. Dissertação (Mestrado em Bioestatística) — Universidade de Lisboa, 2007.
- NUNES, L. N.; BIURRUM, A. L. Perfil do aluno evadido do curso de estatística da ufmg. In: *19º SINAPE*. [S.l.: s.n.], 2010.
- POMMER, W. M. *Conceitos e Aplicações de Estatística para cursos de Ciências Gerenciais: Uma abordagem introdutória*. São Paulo: [s.n.], 2013.
- PRADO, M. V. B. *Métodos de Análise de Correspondência Multipla: Estudo de caso aplicado à avaliação da qualidade do café*. Dissertação (Mestrado) — Universidade Federal de Lavras, Lavras - MG, 2012.
- SILVA, Y. V. da. *Análise de Correspondência: Uma abordagem geométrica*. Dissertação (Mestrado em Estatística Aplicada e Biometria) — Universidade Federal de Viçosa, 2012.
- TEAM, R. C. *R Language Definition*. [S.l.]: Available from CRAN sites, 2013.
- TELES, A. R. T. F. O estudo da evasão como um dos elementos de subsídio às reformas curriculares. In: *Anais do XIII Congresso Brasileiro de Ensino de Engenharia COBENGE 95*. Recife: [s.n.], 1999. p. 1199–1208.

APÊNDICE A – Formulário



Pesquisa Acadêmica – Alunos Evadidos

A evasão dos alunos do Curso de Bacharelado em Estatística na UEPB

Prezada/o Aluno,

As informações dadas pelo discente neste formulário são de grande importância para nossa pesquisa. Pedimos que responda às questões com a maior sinceridade possível, pois suas respostas não serão identificadas individualmente, até porque o objetivo deste trabalho está centrado nas informações coletivas, dadas por todos os discentes. Desde já agradecemos por sua colaboração.

As questões deverão ser respondidas colocando no parêntese, o número correspondente à resposta. Não será necessário utilizar o “X” para as respostas desse questionário.

A - Perfil do entrevistado:

- A1. Sexo: () 1 – Masculino 2 – Feminino
- A2. Estado Civil:
() 1 - Solteiro 2 - Casado 3 - Separado 4 - Divorciado 5 - Viúvo 6 - União estável 7 - Não Informa
- A3. Você tem filho(s)? () 1 – Sim. Quantos: _____ 2 – Não
- A4. Idade:
() 1 - 18 a 24 2 - 25 a 34 3 - 35 a 44 4 - 45 a 54 5 - 55 a 64 6 - +64 7 – Não informa
- A5. Renda Familiar:
() 1 – Menos de 1 SM 2 - De 1 a 3 SM 3 - De 3 a 6 SM 4 - De 6 a 8 SM 5 - + 8 SM
6 - Não Sabe/Não Informa

B – Questões Gerais:

- B1. Em que tipo de escola você cursou o ensino médio (2º grau)?
() 1 – Somente em escola pública 2 - Maior parte em escola pública. 3 - Somente em escola particular
4 - Maior parte em escola particular.
- B2. Em que turno você cursou o ensino médio (2º grau)?
() 1 – Somente no turno diurno 2 – Maior parte no turno diurno 3 – Somente no turno noturno
4 – Maior parte no turno noturno
- B3. Qual foi o principal motivo que levou você a optar pelo Curso de Bacharelado em Estatística?
() 1 – Baixa concorrência do curso 2 - Não sabia qual curso optar 3 - Indicação de um/a amigo/a ou parente
4 - Possibilidade de atuar em várias áreas 5 - Afinidade com a Área/ Curso 6 – Facilidade de emprego após conclusão
- B4. Você participou de algum evento científico (Congresso, Semana da Estatística, Congresso de iniciação Científica)?
() 1 – Sim 2 – Não
- B5. Qual foi o principal motivo que levou você a desistência do Curso de Bacharelado em Estatística?
() 1 – Por causa do trabalho 2 – Optou por outro curso 3 – Dificuldade em cálculo
4 – Assuntos Pessoais (Gravidez, separação, mudou de endereço, dificuldades financeiras)
- B6. Qual seria a principal motivação para reingresso no Curso de Bacharelado em Estatística?
() 1 – Melhoria na infraestrutura 2 – Mais eventos relacionados ao curso 3 – Mais oportunidade de bolsas
4 – Mais incentivo por parte dos professores 5 – Melhoria nas condições financeiras 6 – Melhoria por parte do mercado de trabalho (Estágio) 7 – Outros motivadores 8 – Não voltaria de maneira nenhuma

APÊNDICE B – Rotina no R

Análise de correspondência de forma algébrica entre motivo de optar pelo curso e o motivo de evadir.

```
#####Matriz de contingência:
```

```
B3.1=c(2,6,3,1)
```

```
B3.2=c(2,1,2,1)
```

```
B3.3=c(4,2,1,0)
```

```
B3.4=c(2,2,2,1)
```

```
B3.5=c(6,4,1,0)
```

```
B3.6=c(1,1,1,0)
```

```
A=cbind(B3.1,B3.2,B3.3,B3.4,B3.5,B3.6)
```

```
##### Matriz de Correspondência
```

```
X=A*(1/sum(A))
```

```
##### Massas das linhas e colunas
```

```
r=c(A%*%rep(1,6)*(1/sum(A)))
```

```
c=c(t(A)%*%rep(1,4)*(1/sum(A)))
```

```
##### Matriz de perfil das linhas e colunas
```

```
Dr=diag(r)
```

```
Dc=diag(c)
```

```
##### Matriz
```

```
Pp=X-r%*%t(c)
```

```
##### Matriz S:
```

```
S=(sqrt(solve(Dr)))%*%Pp%*%(sqrt(solve(Dc)))
```

```
##### Decomposição da Matriz S:
```

```
SS=svd(S)
```

```
##### Coordenadas principais das Linhas:
```

```
F=(sqrt(solve(Dr)))%*%SS$u%*%diag(SS$d)
```

```
##### Coordenadas principais das colunas:
```

```
G=(sqrt(solve(Dc)))%*%SS$v%*%diag(SS$d)
```

```
##### Plotando as duas coordenadas:
```

```
Dimensão1=c(F[,1],G[,1])
```

```
Dimensão2=c(F[,2],G[,2])
```

```
plot(Dimensão1,Dimensão2)
```

```
##### Inércia Total
```

```
In.T=sum((SS$d)^2)
```

Análise de correspondência de forma direta, utilizando os pacotes do R, entre motivo de cursar e motivo de evadir.

```
dados=read.csv("AC1.csv",sep=";") ### listado no fim do apendice

summary(dados)

##### pacotes de interesse

install.packages('ca',depend=T)
require(ca)

suppressPackageStartupMessages(library(ca))

suppressPackageStartupMessages(library(FactoMineR))

suppressPackageStartupMessages(library(vcd))

require(FactoMineR)

require(ggplot2)

plot(mjca(dados),col=c("blue","red","green","black","brown1","brown4",
"darkgoldenrod","cornflowerblue","darkslateblue"),xlab="Dimensão 1",
ylab="Dimensão 2")

summary(mjca(dados))
```

Arquivo de dados AC1.csv após ser lido no R

B3	B5
1	4 Assuntos pessoais
2	6 Trabalho
3	5 Trabalho
4	4 Cálculo
5	4 Trabalho
6	1 Assuntos pessoais
7	5 Assuntos pessoais
8	5 Mudou de curso
9	5 Mudou de curso
10	4 Assuntos pessoais
11	2 Trabalho
12	1 Assuntos pessoais
13	5 Mudou de curso
14	1 Assuntos pessoais
15	1 Mudou de curso
16	5 Assuntos pessoais
17	5 Assuntos pessoais
18	2 Mudou de curso
19	6 Mudou de curso
20	3 Mudou de curso
21	3 Mudou de curso
22	3 Mudou de curso
23	1 Trabalho
24	1 Assuntos pessoais
25	2 Cálculo
26	6 Assuntos pessoais
27	5 Mudou de curso
28	1 Assuntos pessoais
29	4 Trabalho
30	2 Trabalho
31	1 Mudou de curso
32	3 Trabalho
33	2 Mudou de curso

34 2 Assuntos pessoais
35 3 Mudou de curso
36 4 Mudou de curso
37 1 Assuntos pessoais
38 1 Trabalho
39 5 Mudou de curso
40 1 Trabalho
41 4 Mudou de curso
42 5 Mudou de curso
43 3 Assuntos pessoais
44 3 Assuntos pessoais
45 1 Cálculo
46 5 Assuntos pessoais