



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS QUIXADÁ
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

CRISLÂNIO DE SOUZA MACÊDO

**ANÁLISE ESTATÍSTICA DA RELAÇÃO ENTRE EVASÃO E AS RESPOSTAS DO
QUESTIONÁRIO PARA INGRESSANTES DA UFC-QUIXADÁ**

QUIXADÁ – CEARÁ

2016

CRISLÂNIO DE SOUZA MACÊDO

ANÁLISE ESTATÍSTICA DA RELAÇÃO ENTRE EVASÃO E AS RESPOSTAS DO
QUESTIONÁRIO PARA INGRESSANTES DA UFC-QUIXADÁ

Monografia apresentada no curso de Ciência da Computação da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharel em Ciência da Computação. Área de concentração: Computação.

Orientador: Críston Pereira de Souza

Co-Orientador: Lucas Ismaily Bezerra
Freitas

QUIXADÁ – CEARÁ

2016

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- M12a Macedo, Crislânio de Souza.
Análise estatística da relação entre evasão e as respostas do questionário para ingressantes da UFC-
Quixadá / Crislânio de Souza Macedo. – 2016.
58 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá,
Curso de Ciência da Computação, Quixadá, 2016.
Orientação: Prof. Dr. Críston Pereira de Souza.
Coorientação: Prof. Me. Lucas Ismaily Bezerra Freitas.
1. Evasão universitária. 2. Análise multivariada. 3. Análise fatorial. I. Título.

CDD 004

CRISLÂNIO DE SOUZA MACÊDO

ANÁLISE ESTATÍSTICA DA RELAÇÃO ENTRE EVASÃO E AS RESPOSTAS DO
QUESTIONÁRIO PARA INGRESSANTES DA UFC-QUIXADÁ

Monografia apresentada no curso de Ciência da
Computação da Universidade Federal do Ceará,
como requisito parcial à obtenção do título de
bacharel em Ciência da Computação. Área de
concentração: Computação.

Aprovada em:

BANCA EXAMINADORA

Críston Pereira de Souza (Orientador)
Campus Quixadá
Universidade Federal do Ceará – UFC

Lucas Ismaily Bezerra Freitas (Co-Orientador)
Campus Quixadá
Universidade Federal do Ceará - UFC

Elvis Miguel Galeas Stancanelli
Campus Quixadá
Universidade Federal do Ceará - UFC

Andreia Libório Sampaio
Campus Quixadá
Universidade Federal do Ceará - UFC

A minha família e amigos que de alguma forma
contribuíram para a construção desse trabalho.
Aos meus pais, José Reis de Macêdo e Aldeni
Artur de Souza Macêdo.

AGRADECIMENTOS

À UFC, por proporcionar uma estrutura que possibilitou a minha estadia para permanecer no curso.

Ao Prof. Dr. Críston Pereira de Souza, pela excelente orientação, paciência, disponibilidade. Por ter acreditado em mim e por ter me ajudado desde o início do curso.

Ao Prof. Ms. Lucas Ismaily Bezerra Freitas, pela excelente orientação, conselhos.

Aos professores participantes da banca examinadora Elvis Miguel Galeas Stancanelli e Andreia Libório Sampaio pela disponibilidade, pelo tempo, pelas valiosas colaborações e sugestões.

Aos meus irmãos Crislene, Marcos Danillo que sempre me apoiaram.

Aos colegas da turma de graduação, David Oliveira, Macílio Pereira, Jhonata Matias, Jonas Costa, Rômulo Costa, André Davys, Wallinson Deives, Sergio Filho, Ana Paula, Airton Filho, pelas críticas e sugestões recebidas.

Aos amigos Elida Correia, Wellington Lucas, Darlan Nascimento, André Barbosa, Douglas Galdino, Maike Bezerra pelo apoio e sinceridade nas críticas nesse estudo e ao longo da trajetória acadêmica.

“Muitas coisas não ousamos empreender por parecerem difíceis; entretanto, são difíceis porque não ousamos empreendê-las.”

(Sêneca)

RESUMO

A evasão ocorre em vários níveis de ensino, sendo que é no primeiro ano seu maior índice, TINTO (2014). Nesse contexto, é importante conhecer os fatores e entender o que poderia levar o aluno a evadir. Esse conhecimento pode ser obtido utilizando os dados da própria universidade. Tais dados podem revelar fatores e indicadores para evasão, e podem ser usados por técnicas multivariadas de dados. Este estudo tem como objetivo analisar fatores sociais, econômicos e motivacionais mais importantes na diferenciação entre alunos evadidos e não evadidos, nos três primeiros semestres dos cursos da UFC-Quixadá, utilizando para esse propósito as respostas do questionário solicitado aos ingressantes de 2015. Nossa amostra contém o total de 237 alunos. Serão apresentadas neste trabalho a Análise de Perfil, que pode ser usada para identificar se dois ou mais grupos do teste possuem perfis significativamente distintos ou similares, e Análise Fatorial, que permite sumarizar os dados. Com essas técnicas chegaremos a um entendimento maior a respeito da evasão na UFC-Quixadá, para que medidas possam ser tomadas com o intuito de diminuir essa prática. O estudo traz contribuições para a compreensão de possíveis motivos que levam o aluno a evadir.

Palavras-chave: Evasão Escolar. Análise multivariada de dados. Análise de Perfil. Teste de Hipótese para Média. Análise Fatorial.

ABSTRACT

Evasion occurs at several levels of education, and in the first year is its highest index. In this context, it is important to know the factors and understand what could lead the student to evade. This knowledge can be obtained using data from the university itself . Such data may reveal factors and indicators for evasion, and may be used by multivariate data techniques. This study aims to analyze the most important social, economic and motivational factors in the differentiation between evaded and non-evaded students, in the first three semesters of UFC-Quixadá courses, using for this purpose, the answers of the questionnaire requested to the newcomers of 2015. Our sample contains the Total of 237 students. In this work, a Profile Analysis will be used, which can be used to identify if two or more target groups have significantly different or similar profiles, and Factor Analysis, which allows the data to be summarized. With these techniques we will arrive at a better understanding of evasion in the UFC-Quixada, so that measures can be taken to reduce this practice. The study brings contributions to the understanding of possible reasons that lead the student to evade.

Keywords: School. Multivariate data analysis. Profile Analysis. Hypothesis Test for Mean. Factor analysis

LISTA DE FIGURAS

Figura 1 – P-Valor para Teste de Hipóteses Bilateral com Rejeição da Hipótese Nula e Sem Rejeição da Hipótese Nula	16
Figura 2 – <i>Critério Scree Plot</i>	21
Figura 3 – Análise de Perfil para os alunos evadidos (vermelho) e não evadidos (azul). .	30
Figura 4 – <i>Scree Plot</i>	36

LISTA DE TABELAS

Tabela 1 – Exemplo de cálculo de comunalidades	23
Tabela 2 – Perguntas com significância estatística	31
Tabela 3 – Teste de Levene para igualdade de variâncias	32
Tabela 4 – Teste t para igualdade de médias	32
Tabela 5 – Comunalidades	35
Tabela 6 – Variância total explicada	36
Tabela 7 – Matriz rotacionada dos fatores	37
Tabela 8 – Questões do estudo 1	47
Tabela 9 – Questões do estudo 2	48
Tabela 10 – Questões do estudo 3	49
Tabela 11 – Matriz rotacionada dos componentes	50
Tabela 12 – Matriz rotacionada dos Fatores	50

LISTA DE ABREVIATURAS E SIGLAS

ACP	Análise de Componentes Principais
AF	Análise Fatorial
CC	Ciência da Computação
DD	Design Digital
EC	Engenharia de Computação
IES	Instituição de Ensino Superior
RC	Redes de Computadores
SI	Sistemas de Informação
SISU	Sistema de Seleção Unificada
SPSS	<i>Statistical Package for the Social Sciences</i>
UFES	Universidade Federal do Espírito Santo
UFC	Universidade Federal do Ceará

SUMÁRIO

1	INTRODUÇÃO	13
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Evasão Escolar	14
2.2	Análise Multivariada de Dados	14
2.3	Teste de Hipótese Para Média	15
2.3.1	<i>Tipos de erro</i>	15
2.3.1.1	<i>P-Valor</i>	15
2.3.1.2	<i>Teste para média</i>	16
2.4	Análise de Perfil	17
2.5	Análise Fatorial	19
2.5.1	<i>Componentes Principais</i>	19
2.5.2	<i>Critério para escolha do número de fatores</i>	20
2.5.3	<i>Rotação dos Fatores</i>	22
2.5.4	<i>Comunalidades</i>	22
3	TRABALHOS RELACIONADOS	24
4	RESULTADOS	25
4.1	Coleta e Tratamento dos Dados	25
4.2	Tratamento das Variáveis no Estudo	25
4.3	Aplicação das Técnicas Estatísticas	28
4.4	Teste de Hipótese para Média	28
4.5	Análise de Perfil	29
4.6	Análise Fatorial	29
4.7	Resultados da Análise de Perfil	29
4.8	Resultados do Teste de Hipótese para Média	30
4.9	Resultados da Análise Fatorial	33
5	CONCLUSÃO	39
6	CONSIDERAÇÕES FINAIS	41
	REFERÊNCIAS	42
	APÊNDICE A – GLOSSÁRIO	44
APÊNDICES	44
	APÊNDICE B – APÊNDICES	47

APÊNDICES	47
B.1 Perguntas do Estudo	47
APÊNDICE C – APÊNDICES	50
APÊNDICES	50
C.1 Fatores	50
ANEXO A – QUESTIONÁRIO	51
A.1 Questionário	51
A.1.1 <i>Levantamento das expectativas e interesses dos alunos em relação ao curso</i>	51
A.1.2 <i>Experiência escolar e profissional.</i>	54
A.1.3 <i>Motivações para o ingresso no curso.</i>	55
A.1.4 <i>Expectativas em relação à conclusão do curso e à inserção no mercado de trabalho.</i>	56

1 INTRODUÇÃO

A evasão no ensino superior é um problema complexo, visto que vários fatores influenciam na decisão do aluno de evadir: questões pessoais, dificuldade financeira, problemas de saúde, abandono involuntário ou abandono voluntário motivado por decisão pessoal, resultados acadêmicos, etc, (MACHADO, 2005). É necessário analisar as relações que envolvem o problema e entender o que poderia levar um aluno ao abandono, para que medidas possam ser tomadas com o intuito de reduzir a evasão. Nesse trabalho utilizamos as respostas do questionário para ingressantes da UFC-Quixadá. Propomos uma análise do perfil dos alunos ingressantes no ano de 2015.

Ao longo dos anos surgiram vários estudos na literatura sobre esse problema, como o de Pintrich et al. (1991) que usa uma visão de motivação regulada na aprendizagem, onde a motivação dos alunos está diretamente ligada à capacidade de auto regular suas atividades de aprendizagem. Em Spady (1970) é descrito o processo de evasão a partir de uma perspectiva do aluno com a instituição.

O modelo de Bean (1980) discute as inter-relações do aluno com o problema da evasão e a integração universitária usando duas técnicas estatísticas: regressão múltipla e análise de caminhos. Na primeira, o autor analisa os efeitos de cada variável e o impacto de uma variável independente em uma variável dependente testada. Na análise de caminhos, que é uma aplicação de regressão múltipla, são estabelecidas as relações de correlação de cada variável, e a regressão múltipla mostra a força dessa relação. Diferente de Bean (1980), usaremos Teste de Hipótese para Média, Análise de Perfil e Análise Fatorial Johnson, Wichern et al. (2002).

No trabalho de Pereira, Brasil e Samohyl (2004) se pretende encontrar fatores que expliquem o problema da evasão, para isso usa-se a Análise Fatorial. Em Junior et al. (2015) é realizada uma comparação de dois grupos de alunos, os formados e os evadidos. O estudo de Machado (2005) busca fatores motivacionais e de contexto para tentar explicar o problema da evasão, para isso ele se baseia em Pintrich et al. (1991).

Este estudo está dividido da seguinte forma: no Capítulo 2 é apresentada a fundamentação teórica, no Capítulo 3 são apresentados os trabalhos relacionados, no Capítulo 4 são apresentados os resultados do estudo, no Capítulo 5 é feita uma conclusão e por fim, no Capítulo 6 são feitas as considerações finais.

2 FUNDAMENTAÇÃO TEÓRICA

Na Seção 2.1 apresentamos a definição de Evasão para o nosso estudo. Na Seção 2.2 definimos a Análise Multivariada de Dados e a ferramenta utilizada para analisar os resultados gerados pelas técnicas estatísticas usadas nesse estudo. Na Seção 2.3 descrevemos o Teste de Hipótese para Média. Na Seção 2.4 apresentamos a Análise de Perfil. Na Seção 2.5 definimos a Análise Fatorial.

2.1 Evasão Escolar

A evasão escolar é um problema que atinge as instituições em todos os níveis de ensino. Em especial, é no primeiro ano de estudo no ensino superior que a maior parte das evasões acontecem. Segundo Filho et al. (2007) a taxa de evasão no primeiro ano é de duas a três vezes maior do que a dos anos seguintes, e esse problema acontece em todo o mundo.

Na literatura o conceito de evasão é tratado de diferentes perspectivas. Em Cardoso (2008), temos o conceito de evasão aparente, que seria a mudança de curso do aluno dentro da própria IES ou para outra IES, e evasão real, que seria o abandono definitivo do sistema por parte do aluno. Já em Filho et al. (2007), temos a evasão anual média, que seria a percentagem de alunos matriculados na IES que não se formaram nem se matricularam no ano seguinte e a evasão total que mede o número de alunos que entraram na IES mas não obtiveram o diploma ao final de um certo número de anos.

Em nosso trabalho faremos um estudo da evasão no primeiro ano dos cursos da UFC-Quixadá, onde será tratado a evasão como sendo o aluno que não se matriculou no semestre subsequente ao seu terceiro semestre de curso.

2.2 Análise Multivariada de Dados

Análise Multivariada se refere a todas as técnicas estatísticas que simultaneamente analisam múltiplas medidas sobre indivíduos ou objetos sob investigação (Hair et al. (2009)). Em nosso estudo, analisaremos um conjunto de dados onde investigaremos suas relações intrínsecas para explicar o problema da evasão.

Para fazer as análises dos dados, usaremos a ferramenta *Statistical Package for the Social Sciences* (SPSS) da IBM com uso de licença para estudante, que oferece suporte as

análises, relatórios estatísticos, suporte a mineração de dados, modelagem preditiva etc, e o programa R que oferece suporte a análises, modelagem preditiva etc. A seguir há uma descrição das técnicas utilizadas nesse trabalho.

2.3 Teste de Hipótese Para Média

Muitas vezes é necessário tirar conclusões sobre parâmetros da população com base em uma amostra. Nesse sentido, a hipótese é uma ideia sobre a população que poderá ser verificada por testes de hipóteses (ou testes de significância). Temos o conceito de hipótese nula (H_0) que é considerada verdadeira durante o teste. Geralmente a hipótese nula representa o contrário da hipótese em análise. Quando o teste de hipótese apresentar uma probabilidade muito pequena (menor do que 5%) em relação à hipótese nula, dizemos que a rejeitamos e aceitamos a hipótese alternativa (H_1) em seu lugar com uma margem de erro (aproximadamente 5%). A hipótese alternativa (H_1) corresponde ao que se quer verificar estatisticamente, ou seja, corresponde à própria hipótese de pesquisa formulada (Barbetta, Reis e Bornia (2004)).

2.3.1 Tipos de erro

Estamos avaliando uma população com base na amostra, portanto não podemos saber com certeza sobre a realidade da população. Estamos avaliando uma hipótese sobre a população, que pode ser aceita ou não.

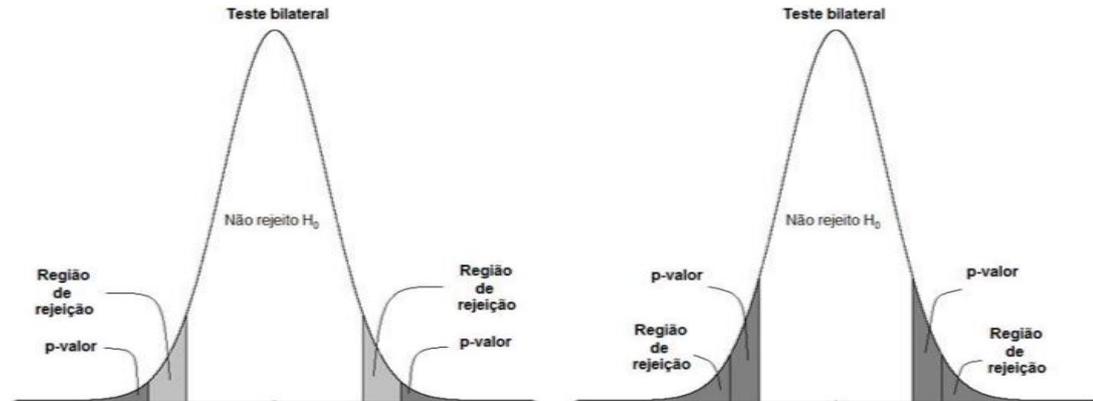
Podemos aceitar uma hipótese sendo ela falsa, ou rejeitar sendo ela verdadeira. Temos os erros I e II que são:

- Tipo I - decisão de rejeitar H_0 quando de fato H_0 é verdadeira.
- Tipo II - decisão de não rejeitar H_0 quando de fato H_0 é falsa.

2.3.1.1 P-Valor

Pressupondo a veracidade de (H_0), podemos calcular a probabilidade de obtermos, por acaso, um resultado tal quanto esperado do que aquele que verificamos na amostra efetivamente observada. A essa probabilidade damos o nome de probabilidade de significância, ou **P-Valor** (Barbetta, Reis e Bornia (2004)). Na Figura 1 temos o **P-Valor** localizado na região de rejeição da hipótese e na região de não rejeição da hipótese.

Figura 1 – P-Valor para Teste de Hipóteses Bilateral com Rejeição da Hipótese Nula e Sem Rejeição da Hipótese Nula



Fonte: O Autor

É importante estimar um valor pelo qual uma hipótese é aceita ou não. O nível de significância do teste é representado pela letra α . Na Figura 2 é estabelecido um nível de significância pela qual a hipótese é rejeitada ou não. É comum adotar um nível de significância de $\alpha = 0,05$, contudo quando se deseja um nível de confiança maior para afirmar (H_1) pode-se adotar $\alpha = 0,01$.

Segundo Barbetta, Reis e Bornia (2004) estabelecido o nível de significância α , que corresponde à probabilidade de ocorrer erro do tipo I no teste, temos a seguinte regra geral de decisão de um teste estatístico:

$$p > \alpha \Rightarrow \text{aceita}(H_0)$$

$$p \leq \alpha \Rightarrow \text{rejeita}(H_0)$$

2.3.1.2 Teste para média

O teste para média é aplicável nas situações em que queremos verificar se uma variável na população pode ser considerada, em média, igual a certo valor (μ_0) (Barbetta, Reis e Bornia (2004)). Será usado o teste **t** para comparar duas amostras independentes com tamanhos diferentes. Nosso principal objetivo com esse teste é testar a hipótese de que as médias são iguais quando consideramos o grupo dos evadidos e o grupo dos não evadidos para cada pergunta no questionário.

2.4 Análise de Perfil

A Análise de Perfil (Johnson, Wichern et al. (2002)), é uma técnica que ajuda a identificar se dois ou mais grupos analisados mostram-se como perfis distintos. Podemos usá-la quando realizamos p tratamentos (ex.: perguntas de um questionário) em dois ou mais grupos (populações). Todas as respostas devem estar na mesma unidade: por exemplo, uma escala de 1 até 5. Além disso, assume-se que as respostas de um grupo são independentes das respostas dos outros grupos. A seguir temos a descrição formal da Análise de Perfil. Todo o conteúdo foi extraído do livro (Johnson, Wichern et al. (2002)).

Sejam $\mu_1 = [\mu_{11}, \mu_{12}, \dots, \mu_{1p}]$ e $\mu_2 = [\mu_{21}, \mu_{22}, \dots, \mu_{2p}]$ as médias dos dois grupos para os p tratamentos. O gráfico destas médias de um grupo, ligadas por segmentos de reta, é chamado perfil.

A análise compara estas médias em 3 etapas:

1. Os perfis são paralelos?

$$H_0^{(1)} : \mu_{1,i} - \mu_{1,i-1} = \mu_{2,i} - \mu_{2,i-1}, \text{ para } i = 2, 3, \dots, p.$$

2. Assumindo que os perfis são paralelos, eles são também coincidentes?

$$H_0^{(2)} : \mu_{1,i} = \mu_{2,i}, \text{ para } i = 1, 2, \dots, p.$$

3. Assumindo que os perfis são coincidentes, todas as médias dos tratamentos são iguais?

$$H_0^{(3)} : \mu_{11} = \mu_{12} = \dots = \mu_{1p} = \mu_{21} = \mu_{22} = \dots = \mu_{2p}.$$

Em forma matricial, podemos escrever $H_0^{(1)} : \mathbf{C}\mu_1 = \mathbf{C}\mu_2$, onde \mathbf{C} vale

$$\mathbf{C}_{(p-1) \times p} = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}.$$

As diferenças tem médias amostrais $\mathbf{C}\bar{\mathbf{x}}_1$ e $\mathbf{C}\bar{\mathbf{x}}_2$. Assumindo variâncias iguais nas duas populações, a variância amostral das diferenças vale $\mathbf{C}\mathbf{S}_p\mathbf{C}'$, onde

$$\begin{aligned} \mathbf{S}_p &= \frac{n_1 - 1}{n_1 + n_2 - 2} \mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2} \mathbf{S}_2, \\ \mathbf{S}_1 &= \frac{\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)'}{n_1 - 1}, \\ \mathbf{S}_2 &= \frac{\sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)'}{n_2 - 1}. \end{aligned}$$

\mathbf{S}_1 e \mathbf{S}_2 são as matrizes de variâncias e covariâncias das populações 1 e 2, respectivamente.

Portanto, o vetor de diferenças da população 1 tem distribuição $N_{p-1}(\mathbf{C}\boldsymbol{\mu}_1, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$ na população 1, e distribuição $N_{p-1}(\mathbf{C}\boldsymbol{\mu}_2, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$ na população 2. T^2 é uma generalização da estatística *T de Student* que é usado em testes de hipóteses multivariada.

Rejeitamos $H_0^{(1)} : \mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2$ (perfis paralelos) com confiança α quando

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{C}' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{C}\mathbf{S}_p\mathbf{C}' \right]^{-1} \mathbf{C}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) > c^2,$$

onde

$$c^2 = \frac{(n_1 + n_2 - 2)(p - 1)}{n_1 + n_2 - p} F_{p-1, n_1 + n_2 - p}(\alpha).$$

Assumindo que os perfis são paralelos, para verificar que são coincidentes basta testar a hipótese de que as somas das médias na população 1 é igual a soma das médias na população 2, ou seja, $\mathbf{1}'\boldsymbol{\mu}_1 = \mu_{11} + \mu_{12} + \dots + \mu_{1p} = \mu_{21} + \mu_{22} + \dots + \mu_{2p} = \mathbf{1}'\boldsymbol{\mu}_2$. Podemos então usar um teste *t* univariado.

Rejeitamos $H_0^{(2)} : \mathbf{1}'\boldsymbol{\mu}_1 = \mathbf{1}'\boldsymbol{\mu}_2$ (perfis coincidentes) com confiança α quando

$$\begin{aligned} T^2 &= \mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{1}'\mathbf{S}_p\mathbf{1} \right]^{-1} \mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \\ &= \left(\frac{\mathbf{1}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{1}'\mathbf{S}_p\mathbf{1}}} \right)^2 > t_{n_1 + n_2 - 2}^2(\alpha/2) = F_{1, n_1 + n_2 - 2}(\alpha). \end{aligned}$$

Se as médias são coincidentes, podemos usar os pontos das duas populações para obter a média amostral

$$\bar{\mathbf{x}} = \frac{\sum_{j=1}^{n_1} \mathbf{x}_{1j} + \sum_{j=1}^{n_2} \mathbf{x}_{2j}}{n_1 + n_2} = \frac{n_1}{n_1 + n_2} \bar{\mathbf{x}}_1 + \frac{n_2}{n_1 + n_2} \bar{\mathbf{x}}_2.$$

Neste caso, resta verificar se as médias dos tratamentos são todas iguais, ou seja, $\mu_1 = \mu_2 = \dots = \mu_p$.

Rejeitamos $H_0^{(3)} : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ (médias dos tratamentos iguais) com confiança α se

$$(n_1 + n_2) \bar{\mathbf{x}}' \mathbf{C}' [\mathbf{C}\mathbf{S}\mathbf{C}']^{-1} \mathbf{C}\bar{\mathbf{x}} > c^2,$$

em que S é a matriz de covariâncias amostrais calculada a partir de todas as $n_1 + n_2$ observações, e

$$c^2 = \frac{(n_1 + n_2 - 1)(p - 1)}{n_1 + n_2 - p + 1} F_{p-1, n_1 + n_2 - p + 1}(\alpha).$$

Para obter os resultados com a Análise de Perfil iremos identificar quais perguntas tem unidades iguais (escalas iguais), assumiremos que os grupos tem variâncias iguais. Para isso faremos um Teste de Levene para identificar quais perguntas não possuem variâncias iguais.

2.5 Análise Fatorial

A Análise Fatorial é uma técnica de análise multivariada que visa representar um conjunto de variáveis em termos de um conjunto subjacente menor de variáveis chamados fatores (Webb (2003)). Os fatores são extraídos através das componentes principais, Johnson, Wichern et al. (2002), uma vez tendo as componentes (fatores) é feito uma rotação nos eixos dos fatores com o intuito de interpretá-los melhor, Corrar, Paulo e Filho (2007). A seguir será descrita essa técnica estatística e seu uso nesse trabalho.

Suponha que temos p variáveis x_1, x_2, \dots, x_p , nós assumimos que existem m variáveis latentes ou fatores ξ_1, \dots, ξ_m , assim que

$$(x_i - \mu_i) = \sum_{k=1}^{m_1} \lambda_{ik} \xi_k + \varepsilon_i, \quad i = 1, \dots, p$$

onde μ é o vetor de médias $(x_1, \dots, x_p)^T$. Sem perda de generalidade, vamos tornar ser zero. Assim, as variáveis ξ_k denotam os fatores, que contribui para todas as variáveis observadas x_i , ε_i é a variação residual das variáveis x_i . λ_{ik} são os pesos das cargas fatoriais.

A Análise Fatorial utilizada em nosso estudo é chamada de Análise Fatorial R (*R-mode factor analysis*) com a qual buscamos analisar um conjunto de variáveis para identificarmos fatores, buscamos desenvolver uma “estrutura subjacente” entre as variáveis, onde essas “estruturas subjacentes” não são explicadas através de uma única variável. Por exemplo, explicar as características de um bom aluno através de variáveis como frequência, nota de prova entre outras variáveis podem explicar melhor as características que distinguem um bom aluno.

2.5.1 Componentes Principais

As componentes principais permitem representar em um espaço de dimensão pequena, as observações de um espaço geral p -dimensional, Peña (2002). As componentes são os fatores, que são primeiro passo para identificar variáveis latentes (não observáveis) que estão gerando a variabilidade nos dados. As componentes permitem transformar as variáveis originais

correlacionadas, em novas não correlacionadas, facilitando a interpretação dos dados.

As componentes descrevem a estrutura de variância e covariância de variáveis correlacionadas x_1, x_2, \dots, x_p em termos de um conjunto de novas variáveis y_1, y_2, \dots, y_p não correlacionadas. Cada y_1, \dots, y_p é uma combinação linear das variáveis x_1, x_2, \dots, x_p , ou seja.

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{ip}x_p$$

Assim, y_1 (1° componente principal) concentra a maior parte da variação dos dados originais entre todas as combinações lineares de x_1, x_2, \dots, x_p . y_2 (2° componente principal) é não correlacionada com y_1 e concentra a maior parte da variação restante. y_3 (3° componente principal) é não correlacionada com y_1 e y_2 e concentra uma parcela ainda menor da variação. y_4 (4° componente principal) é não correlacionada com y_1, y_2 e y_3 e concentra uma parcela ainda menor da variação. Assim sucessivamente.

As componentes tem como propriedade a conservação da variabilidade inicial que é a soma das variâncias das componentes, igual a soma da variância dos dados originais, e a variância geral dos dados é igual a original. A variância da componente h é λ_h , o respectivo autovalor. A soma das variâncias das variáveis originais é $\sum_{i=1}^p \lambda_i$. A proporção de variabilidade total explicada pela componente h é $\lambda_h / \sum \lambda_i$.

2.5.2 Critério para escolha do número de fatores

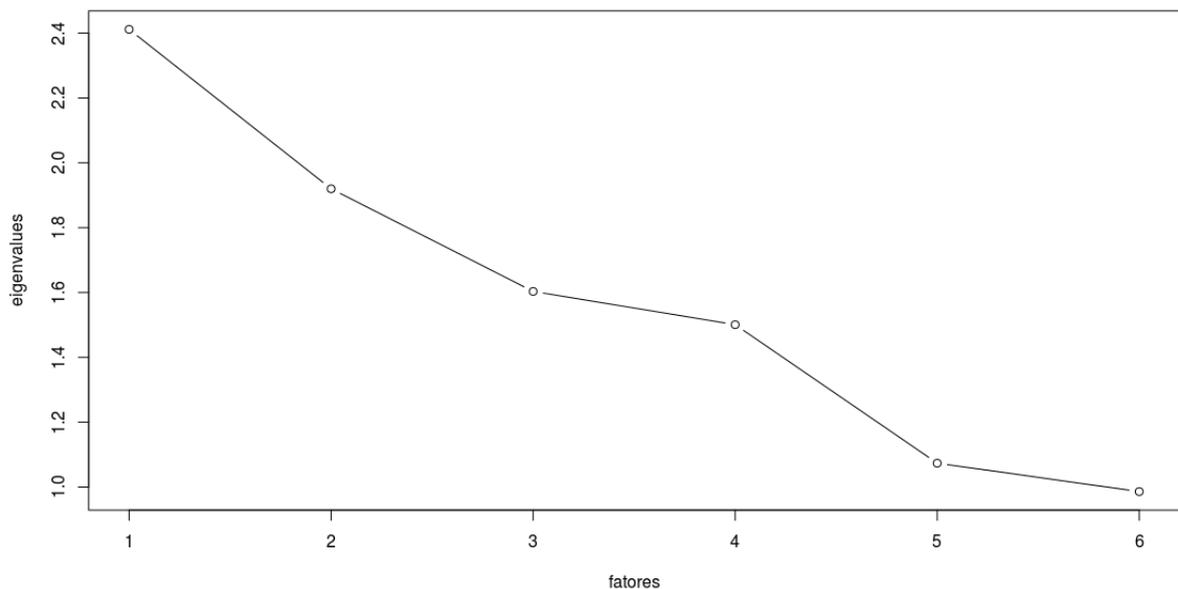
Na Análise Fatorial, determinar o número utilizado de fatores é um ponto chave para a explicação dos dados, pois tendo a escolha de poucos fatores, pode acontecer desses fatores não explicarem os dados bem, ou seja, tais fatores não descrevem os dados originais com uma variação elevada; por outro lado, pode prejudicar a análise. Por exemplo, em um estudo com 40 variáveis, 40 fatores são extraídos, o ponto chave é que os primeiros fatores descrevem melhor os dados originais, ou seja, o primeiro fator é tido como a melhor combinação dos dados, seguido pelo segundo fator e assim por diante. Desse modo, é necessário utilizar algum critério na escolha da quantidade dos fatores.

Um critério para escolha da quantidade de fatores é baseia-se nas observações dos autovalores. O raciocínio segue da seguinte forma: qualquer fator explica a variância de pelo menos uma variável, pois para um conjunto de dados tem que existir pelo menos um número de

fatores menor ou igual ao número de variáveis. Com a análise de componentes principais (ACP), cada variável contribui com um valor 1 do autovalor total. O autovalor é um número que reflete a relevância de um fator, corresponde quanto o fator consegue explicar a variância dos dados, sendo os fatores com o (autovalor) iguais ou acima de 1,0 considerados para a Análise Fatorial e fatores com autovalores abaixo de 1,0 descartados (Corrar, Paulo e Filho (2007)). Segundo Hair et al. (2009) usar o autovalor como critério de escolha do número de fatores é mais confiável quando o número de variáveis está entre 20 e 50.

Outro critério seria observar graficamente os autovalores é o *scree plot* que faz o gráfico dos autovalores em relação aos fatores do estudo. O ponto de corte pelo critério *scree plot* se dá quando os ângulos de inclinação dos autovalores decrescem de forma a ficar como uma reta horizontal. Por exemplo, ao observarmos a Figura 2, pelo critério do autovalor temos um total de 6 fatores. Contudo, o que pode acontecer é que, pelo *scree plot*, enquanto os ângulos de inclinação dos autovalores não se aproximam de uma reta horizontal, haverá mais fatores incluídos. Pelo critério *scree plot* resulta em pelo menos um e, às vezes, dois ou mais fatores incluídos em relação ao critério do autovalor Hair et al. (2009).

Figura 2 – Critério Scree Plot



Fonte: O Autor

2.5.3 Rotação dos Fatores

A rotação dos fatores é uma forma de melhorar a interpretação dos fatores, pois é maximizada a importância, carga fatorial de uma variável em um fator. O modelo fatorial para um vetor de dados é, $X = \mu + Lf + \xi$, é a função da média μ mais a matriz de cargas fatoriais vezes os vetores de fatores comuns mais um fator específico. O que queremos obter é uma rotação apropriada que permita interpretar os dados.

Utilizaremos a rotação *Varimax* que minimiza a ocorrência de uma variável com altas cargas fatoriais em diferentes fatores, permitindo que uma variável seja facilmente identificada em um fator (Corrar, Paulo e Filho (2007)). A rotação *Varimax* envolve escalar as cargas fatoriais dividindo pela comunalidade correspondente University (2004).

2.5.4 Comunalidades

As comunalidades para a i -ésima variável são computadas pela soma ao quadrado das cargas fatoriais.

$$\hat{h}_i = \sum_{j=1}^m \hat{l}_{ij}^2.$$

Em que as cargas fatoriais são $\hat{e}_i \sqrt{\hat{\lambda}_i}$, os elementos do autovetor são multiplicados pelo correspondente autovalor. Ou seja, é a soma das cargas fatoriais associadas a ela elevada ao quadrado, é a proporção da variância de cada variável explicada pelos fatores (Rocha (2014)). As comunalidades representam o quanto a variável tem relação com os fatores. Representam a quantidade média de variação entre as medidas/indicadores e podendo ser calculadas diretamente a partir de cargas de fatores. Pegando a comunalidade total e dividindo pelo total de fatores temos a proporção de variância por fator. $\sum_{i=1}^p \hat{h}_i = \sum_{i=1}^m \lambda_i$.

A variâncias específicas podem ser encontradas subtraindo as comunalidades da variância.

$$\hat{\psi}_i = 1 - \hat{h}_i$$

Como exemplo considere a tabela abaixo:

Computar a comunalidade para matemática discreta é dada por:

$$\hat{h}_1 = 0,313^2 + 0,046^2 + 0,651^2 = 0,523886.$$

A comunalidade total é de 1,457 e a proporção da variância total de explicada por

Tabela 1 – Exemplo de cálculo de comunalidades

variável	fator 1	fator 2	fator 3
matemática discreta	0,313	0,046	0,651
probabilidade e estatística	0,534	-0,222	-0,041
cálculo	-0,230	0,449	0,670

Fonte: O Autor

Nota: fatores com notas dos alunos

cada um dos três fatores é $1,457/3 = 0,485$. Uma variância específica para a disciplina/variável probabilidade e estatística é dado por $\hat{\psi}_i = 1 - 0,336121 = 0,6638$.

3 TRABALHOS RELACIONADOS

A seguir são apresentados os trabalhos que auxiliaram na construção deste estudo. Machado (2005) fez um estudo com o objetivo de investigar o problema da evasão do curso de Ciência da Computação de uma instituição sem fins lucrativos onde os cursos são pagos. Considerando as turmas que concluíram entre 1999 e 2004, a evasão média é aproximadamente 71%. Seu estudo é de natureza quantitativa, descritiva correlacional e exploratória. Em seu estudo Machado (2005) levantou os motivos pelos quais os alunos escolheram o curso, bem como a intenção de permanecer no curso. O estudo foi restrito à investigação de fatores relacionados à motivação e ao uso de estratégias baseadas no modelo de Pintrich et al. (1991), comparando os resultados obtidos com outros dois cursos que não eram de tecnologia da informação e que não apresentavam alto histórico de evasão no primeiro ano de curso.

A nossa abordagem se assemelha a de Machado (2005) por investigar os fatores motivacionais que expliquem a evasão no primeiro ano do curso, e por sua natureza exploratória, quantitativa e correlacional. Nosso estudo diferencia-se por buscar identificar quais as variáveis são mais significativas na detecção de alunos que ingressem em um dos seis cursos em estudo e que venham a desistir do curso após terem cursado os três primeiros semestres.

No trabalho de Junior et al. (2015) é feita uma análise de dados da Universidade Federal do Espírito Santo (UFES) no período de 2006 a 2012. Para entendimento do problema da evasão, foram utilizadas técnicas de análise multivariada de dados: Análise Fatorial e Regressão Logística. Como referencial teórico, foi adotado um modelo que relaciona aspectos como a estrutura da instituição, o esforço do aluno e sua aprendizagem. Nosso trabalho assemelha-se pelo uso da técnica Análise Fatorial para encontrar os fatores que influenciam na evasão. Entretanto, nosso trabalho acrescenta o fato que identificaremos quais perguntas são significantes para diferenciar os perfis de alunos.

Em Silva et al. (2014) foi feito um estudo do ensino superior no Brasil, investigando os fatores que estão relacionados à evasão escolar e às políticas tanto do Estado quanto da Instituição de Ensino Superior (IES) para a permanência do aluno na IES.

Assim como Silva et al. (2014), iremos analisar alunos que estão se beneficiando das políticas do Estado como o Sistema de Seleção Unificada (SISU), e políticas estabelecidas pela IES como concessão de bolsas, analisando a relação com a evasão.

4 RESULTADOS

Os procedimentos adotados neste estudo são divididos em duas etapas: (i) coleta e tratamento dos dados; (ii) aplicação das técnicas estatísticas;

4.1 Coleta e Tratamento dos Dados

Os dados coletados para o estudo foram obtidos através de um questionário aplicado no ato da matrícula em fevereiro de 2015. Em maio de 2016, a secretaria acadêmica do Campus fez uma segunda coleta de dados para identificar o *status* do aluno (cancelado ou não), utilizando um sistema da IES para gerar o relatório. Para preservar o anonimato dos alunos, recebemos os dados de todos os alunos matriculados em 2015 sem o nome e a matrícula.

Para poder aplicar a Análise Fatorial é necessário realizar um tratamento prévio nos dados. Realizamos normalização nos dados numéricos (mantendo-os entre 0 e 1) para que todas as variáveis fiquem na mesma escala, igualando assim as importâncias das variáveis no modelo gerado. Nenhum valor discrepante e nenhum valor perdido (*missing value*) foi detectado na amostra.

4.2 Tratamento das Variáveis no Estudo

No questionário (ver anexo A) temos as perguntas que foram consideradas no estudo. Tivemos um total de 51 perguntas no estudo. Algumas perguntas foram representadas de forma binária, outras de forma intervalar. A seguir descrevemos brevemente o que foi feito para algumas perguntas.

A pergunta 11, sobre a renda, transformamos a resposta “renda até 1000 reais” em 0,5, a resposta “renda de 1001 até 2000 reais” em 1,5, a resposta “renda de 2001 até 3000 reais” em 2,5, a resposta “renda de 3001 até 4000 reais” em 3,5, a resposta “renda de 4001 até 5000 reais” em 4,5, a resposta “renda de 5001 até 6000 reais” em 5,5, e a resposta “renda acima de 6000 reais” em 6,5.

Na pergunta 19, a respeito da distância do local de moradia atual para a Universidade, foi tirado o ponto médio das respostas. Transformamos a resposta “até 10 km” em 5, a resposta “entre 10 km e 25 km” em 17,5, a resposta “entre 25 km e 50 km” em 37,5, a resposta “entre 50 km e 100 km” em 75 e a resposta “acima de 100 km” em 125.

A pergunta 20, sobre o meio de transporte usado para chegar a Universidade, é

composta por cinco alternativas de respostas: “Automóvel particular”, “Transporte alternativo (táxi, mototáxi, etc.)”, “Ônibus universitário”, “Bicicleta”, “Outro”. Estas respostas foram agrupadas em duas categorias: “transporte privado”, composta apenas pela resposta “Automóvel particular”, e “outros”, composta pelas respostas restantes.

A pergunta a respeito da existência de necessidades especiais foi dividida nas perguntas 21 e 22, em que nas perguntas queríamos identificar os alunos com dificuldade visual ou de atenção, na outra queríamos saber se o aluno tinha alguma dificuldade visual ou não. O motivo dessa separação é a necessidade de diferenciação entre dificuldade visual e de atenção para as outras dificuldades.

As perguntas 23, 24, e 25, sobre a escolaridade, foram dispostas da seguinte forma: na pergunta sobre o ensino fundamental, transformamos a resposta “pública” em 0, a resposta “privada” em 1, “pública-privada” em 0,5, e a resposta “estrangeira” em 0,5. Na pergunta sobre o ensino médio transformamos a resposta “pública” em 0, a resposta “pública-profissional” em 0, a resposta “privada” em 1, a resposta “pública-privada” em 0,5, e a resposta “estrangeira” em 0,5.

A pergunta 26, sobre a forma de ingresso na Universidade, foi apresentada como ampla concorrência e outros. As perguntas 27 e 28 são relacionadas ao trabalho durante a graduação. Queremos saber se o aluno está trabalhando na área de TI. As perguntas 29 a 32 estão relacionadas às atividades extracurriculares. As perguntas 33 a 36 estão associadas a fatores motivacionais que determinaram a escolha do curso.

As perguntas 37 a 51 são todas escalares, normalizadas para valores entre 0 a 1, todas têm um nível de mensuração em que podemos fazer uma quantificação das diferenças entre elas. Nas perguntas de 37 a 40 transformamos as respostas: “pouquíssima afinidade” igual a 0, “alguma afinidade” igual a 0,3, “boa afinidade” igual a 0,7 e “muita afinidade” igual a 1. Na pergunta 41, a respeito das horas de estudo, transformamos a resposta “até 2 horas de estudo” em 0, a resposta “de 2 a 3 horas de estudo” em 2,5, a resposta “de 3 a 4 horas de estudo” em 3,5, a resposta “de 4 a 6 horas de estudo” em 5, e a resposta “acima de 7 horas de estudo” em 7.

A pergunta 42 que indica o nível de motivação em relação ao curso, transformamos a resposta “totalmente desmotivado” em 0, “pouco motivado” em 0,3, “motivado” em 0,7, e a resposta “muito motivado” em 1. As perguntas 43 a 47, a respeito da medida que o aluno conhece o curso, transformamos a resposta “pouquíssimo conhecimento” em 0, a resposta “algum conhecimento” em 0,3, a resposta “conheço bem” em 0,7, e a resposta “conheço muito bem” em 1. E finalmente as perguntas 48 a 51, sobre a percepção acerca da vida acadêmica, transformamos

as resposta “nunca se verifica” em 0, a resposta “poucas vezes se verifica” em 0,3, a resposta “verifica-se com frequência” em 0,7, e a resposta “verifica-se sempre” em 1.

Para aplicar a Análise de Perfil tivemos um total de 17 perguntas, todas em uma mesma unidade, normalizadas para valores entre 0 a 1. Para que todas as perguntas na Análise de Perfil ficassem na mesma unidade, dividimos em cada resposta da pergunta pelo maior valor, exceto nas perguntas em que já se tinha uma escala de 0 até 1.

As seguintes perguntas fizeram parte do estudo na Análise de Perfil, são elas: pergunta 11 a respeito da renda, 19 sobre a distância para a Universidade, 37 a 40 sobre a afinidade em áreas no ENEM, 41 a respeito do tempo de estudo, 42 sobre a motivação em relação ao curso que escolheu, 43 até 47 a cerca da medida que o aluno conhece o curso e finalmente as perguntas 48 até 51 a respeito da percepção sobre a vida acadêmica no início do curso.

A seguir mostramos as transformações feitas para cada pergunta no estudo a respeito da Análise de Perfil.

Na pergunta 11, “Qual a renda mensal de sua família?”, transformamos a resposta “renda até 1000 reais” em 0,0769, a resposta “renda de 1001 até 2000 reais” em 0,2307, a resposta “renda de 2001 até 3000 reais” em 0,3846, a resposta “renda de 3001 até 4000 reais” em 0,5384, a resposta “renda de 4001 até 5000 reais” em 0,6923, a resposta “renda de 5001 até 6000 reais” em 0,8461, e a resposta “renda acima de 6000 reais” em 1.

As perguntas 37 a 40 não sofreram qualquer transformação, pois seus valores já estavam normalizados para valores entre 0 a 1. Assim seus valores permaneceram como: “pouquíssima afinidade” igual a 0, “alguma afinidade” igual a 0,3, “boa afinidade” igual a 0,7, “muita afinidade” igual a 1.

Na pergunta 41, “Quantas horas diárias você dedica aos estudos fora do horário de aula?”, transformamos a resposta “até 2 horas de estudo” em 0, a resposta “de 2 a 3 horas de estudo” em 0,3846, a resposta “de 3 a 4 horas de estudo” em 0,5384, a resposta “de 4 a 6 horas de estudo” em 0,7142, e a resposta “acima de 7 horas de estudo” em 1.

Na pergunta 42, “Como você avalia sua motivação em relação ao curso que escolheu?”, transformamos a resposta “totalmente desmotivado” em 0, a resposta “pouco motivado” em 0,3, a resposta “motivado” em 0,7, e a resposta “muito motivado” em 1. As perguntas 43 a 47 não sofreram nenhuma transformação pois seus valores já estavam normalizados para valores entre 0 a 1. Seus valores permaneceram como: “pouquíssimo

conhecimento” igual a 0, “algum conhecimento” igual a 0,3, “conheço bem” igual a 0,7, “conheço muito bem” igual a 1. As perguntas 48 a 51, sobre a percepção sobre a vida acadêmica, permaneceram com valores: “nunca se verifica” igual a 0, “poucas vezes se verifica” igual a 0,3, “verifica-se com frequência” igual a 0,7, “verifica-se sempre” igual a 1.

Todas essas transformações têm o intuito de melhorar a forma de interpretação nos dados, para que se possa ao longo do estudo explicar da forma mais estatisticamente significativa os dados apresentados, ou seja, que a explanação do estudo reflita a veracidade dos dados a respeito da evasão.

4.3 Aplicação das Técnicas Estatísticas

Nessa etapa aplicaremos nos dados as seguintes técnicas: Análise Fatorial, Teste de Hipótese para Média e Análise de Perfil. Iremos reduzir os dados em fatores, os quais explicam a variação dos dados utilizando a Análise Fatorial.

Com a Análise Fatorial teremos uma explicação melhor dos dados, os dados serão representados através de fatores em que analisaremos as relações dos fatores com a evasão escolar.

Com o Teste de Hipótese para Média saberemos quais variáveis diferenciam entre os grupos de alunos evadidos e não evadidos com significância estatística. Com a Análise de Perfil iremos verificar se os grupos de alunos respondem da mesma forma ou não.

Para realizar as técnicas apresentadas, usaremos para fazer as análises o programa R (2016) em que será o Análise de Perfil em que queremos saber se existe uma diferença entre os grupos de alunos evadidos, não evadidos quanto as respostas do questionário.

4.4 Teste de Hipótese para Média

Nosso objetivo com esse teste é verificar se há evidências de diferenças entre os alunos evadidos e alunos não evadidos. Desejamos verificar se os grupos de alunos evadidos e não evadidos têm comportamentos diferentes para alguma variável em estudo. As hipóteses são:

H_0 : em média, os dois grupos de alunos são *iguais* para uma determinada variável no estudo; e

H_1 : em média, os dois grupos de alunos são *diferentes* para uma determinada variável no estudo.

4.5 Análise de Perfil

A Análise de Perfil é útil para identificar se dois ou mais grupos do teste possuem perfis significativamente distintos ou similares (Parmigiani et al. (2003)).

Iremos responder uma questão importante na Análise de Perfil: os grupos têm perfis semelhantes no conjunto de dados?

Para alcançar esse propósito iremos verificar as três etapas descritas anteriormente. Se os perfis são paralelos, ou seja, há uma relação entre as perguntas respondidas entre os grupos de forma que a diferença entre a média de cada pergunta seja proporcional, assumindo que os perfis são paralelos, elas são coincidentes, assumindo que são coincidentes as médias dos tratamentos são iguais, ou seja, se os grupos respondem da mesma forma para cada pergunta.

4.6 Análise Fatorial

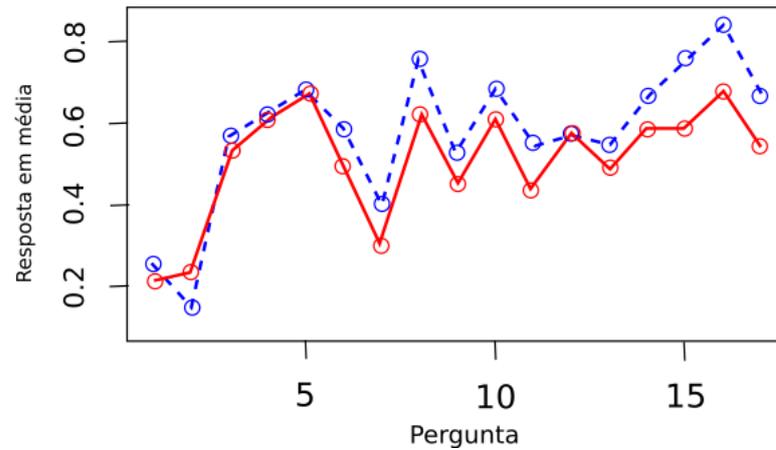
De acordo com as recomendações na literatura (HAIR et al., 2009), realizamos os seguintes passos: (1) verificação de existência de pelo menos 30 observações na amostra; (2) existência de correlações entre variáveis acima de 0,3; (3) determinação do número de fatores, utilizando *scree plot* e autovalores acima de 1; (4) rotação *varimax* (KAISER, 1958) dos fatores, para facilitar a interpretação; (5) identificação das variáveis com coeficiente acima de 70% em cada fator; e (6) interpretação do relacionamento entre variáveis e fatores.

4.7 Resultados da Análise de Perfil

Nesta Seção são apresentados os resultados obtidos dos dados dos ingressantes da UFC-Quixadá do ano de 2015, com a aplicação do Teste de Hipótese para Média, Análise de Perfil e da Análise Fatorial.

Para obter os resultados primeiro testamos a hipótese de que os perfis são paralelos, descrito na Seção 2.4. O resultado do teste mostrou que os perfis são paralelos. Como o resultado mostrou que os perfis são paralelos, as hipóteses de que os perfis são coincidentes ou que são iguais é refutado. A seguir temos na Figura 3 os perfis .

Figura 3 – Análise de Perfil para os alunos evadidos (vermelho) e não evadidos (azul).



Fonte: O Autor

Com o resultado do teste concluímos que os grupos de alunos evadidos e não evadidos respondem de forma diferente para as perguntas analisadas. A próxima etapa consiste em verificar para quais perguntas os alunos respondem diferentemente com significância estatística. Para esse propósito usaremos o Teste de Hipótese para Média, que é um teste univariado.

4.8 Resultados do Teste de Hipótese para Média

No Teste de Hipótese foi utilizado o teste t para grupos independentes com o intuito de avaliar a diferença entre as médias dos grupos de alunos evadidos e não evadidos.

Na Tabela 2 temos os dados das perguntas com significância estatística, e nas Tabelas 3 e 4 são mostrados os resultados do Teste de Hipótese para Média. Através do teste de hipótese queremos saber se existe diferença, com significância estatística, nas seguintes:

- na distância da moradia atual, entre os alunos evadidos e não evadidos?
- nos principais fatores que influenciam na escolha do curso, entre os alunos evadidos e não evadidos?
- na quantidade de horas de estudo dedicadas, entre os alunos evadidos e não evadidos?
- na motivação em relação à escolha do curso, entre os alunos evadidos e não evadidos?
- na percepção sobre a vida acadêmica, entre os alunos evadidos e não evadidos?

Tabela 2 – Perguntas com significância estatística

Núm. Pergunta	Nome da Variável	status	Tamanho da Amostra	Média
19	dist_para_universidade	Não evadido	206	0,1837
		Evadido	30	0,2858
33	esc_curso_fat_identificacao	Não evadido	206	0,78
		Evadido	30	0,43
35	esc_curso_fat_inf_a_fam	Não evadido	206	0,12
		Evadido	30	0,30
41	tempo_estudo	Não evadido	206	2,7791
		Evadido	30	2,1167
42	motivacao_curso	Não evadido	206	0,7519
		Evadido	30	0,6200
49	vida_acad_realizacao_prof	Não evadido	206	0,7597
		Evadido	30	0,5833
51	vida_acad_n_mudaria_curso	Não evadido	206	0,8413
		Evadido	30	0,6700

Como observamos na Tabela 2 sobre as estatísticas descritivas dos grupos, a média dos alunos evadidos em relação à distância para Universidade é aproximadamente 55,57% maior do que os não evadidos. Sobre a escolha do curso, a afinidade dos alunos evadidos é em média 44,48% menor do que os não evadidos, e a influência dos pais, amigos, familiares e mercado promissor é em média menor para alunos não evadidos do que evadidos, com um total de 12% alunos não evadidos e 30% evadidos, ou seja, a influência desses fatores para a escolha do curso em alunos evadidos é 150% maior em média do que em alunos não evadidos.

Em relação ao número de horas de estudo e à motivação em relação ao curso que escolheu, a média dos alunos não evadidos é maior que a média dos alunos evadidos.

Percebemos pela média da amostra que o grupo evadido dedica menos horas de estudo: 23,83% menor do que os alunos não evadidos. A motivação em relação ao curso que escolheu é menor entre os alunos evadidos: 17,54% menor do que os alunos não evadidos.

Desse modo, no primeiro passo para fazermos o teste de hipótese analisamos o teste de igualdade de variância descrito na Tabela 3. Consideramos a significância estatística de 5%, Härdle e Simar (2007). Ao observamos os valores de significância para as variáveis na Tabela 3 em que consideramos um intervalo de confiança de 95%, somente a variável tempo_estudo tem um nível de significância maior que 5%, mostra que existe diferenças significativas entre as variâncias dos dois grupos. Portanto, somente para essa variável usamos a parte da saída chamada de *equal variances assumed* (suposição de variâncias iguais). Para o restante das variáveis foi assumida variâncias diferentes.

Os resultados da Tabela 4 revelaram que com 95% de confiança temos as variáveis descritas na Tabela 2 com respostas significativas entre os grupos de alunos. Os valores de

Tabela 3 – Teste de Levene para igualdade de variâncias

	Teste de Levene para Igualdade de variâncias	
	F	Nível de significância
dist_para_universidade:		
Igualdade de variância assumida/	5,137	0,024
Igualdade de variância não assumida		
esc_curso_fat_identificacao:		
Igualdade de variância assumida/	12,163	0,001
Igualdade de variância não assumida		
esc_curso_fat_inf_a_fam:		
Igualdade de variância assumida/	19,250	0,0001
Igualdade de variância não assumida		
tempo_estudo:		
Igualdade de variância assumida/	0,019	0,890
Igualdade de variância não assumida		
motivacao_curso:		
Igualdade de variância assumida/	0,915	0,340
Igualdade de variância não assumida		
vida_acad_realizacao_prof:		
Igualdade de variância assumida/	11,044	0,001
Igualdade de variância não assumida		
vida_acad_n_mudaria_curso:		
Igualdade de variância assumida/	10,211	0,002
Igualdade de variância não assumida		

Tabela 4 – Teste t para igualdade de médias

	Teste t de igualdade de médias			
	t	Sig.(2-caudal)	Mais baixo	Mais alto
dist_para_universidade:				
Igualdade de variância assumida/	-2,026	0,044	-0,20138	-0,00281
Igualdade de variância não assumida		0,088	-0,022011	0,01592
esc_curso_fat_identificacao:				
Igualdade de variância assumida/	4,179	0,0001	0,184	0,0512
Igualdade de variância não assumida	3,611	0,001	0,152	0,544
esc_curso_fat_inf_a_fam:				
Igualdade de variância assumida/	-2,630	0,009	-0,312	-0,045
Igualdade de variância não assumida	-2,028	0,051	-0,358	0,001
tempo_estudo:				
Igualdade de variância assumida/	2,434	0,016	0,12617	1,19875
Igualdade de variância não assumida	2,728	0,009	0,17213	1,15279
motivacao_curso:				
Igualdade de variância assumida/	3,078	0,002	0,04748	0,21640
Igualdade de variância não assumida	2,895	0,006	0,03956	0,22432
vida_acad_realizacao_prof:				
Igualdade de variância assumida/	3,772	0,0001	0,08425	0,26850
Igualdade de variância não assumida	2,956	0,006	0,05503	0,29772
vida_acad_n_mudaria_curso:				
Igualdade de variância assumida/	3,482	0,001	0,07436	0,26816
Igualdade de variância não assumida	2,620	0,013	0,03829	0,30424

significância são menores que 5%, logo existem evidências de diferenças estatisticamente significativas entre essas duas médias.

A implicação disso é que rejeitamos a hipótese nula (H_0) e aceitamos a hipótese alternativa (H_1).

Assim podemos afirmar que:

- A distância do local de moradia atual para a Universidade é menor em média para alunos evadidos do que alunos não evadidos.
- Em relação ao fator que determinou a escolha pelo seu curso. “É uma área com a qual me identifico. Por influências dos pais, dos amigos, de familiares e por conta de um mercado de trabalho promissor”, todos esses fatores têm valores em média menores para alunos evadidos do que para alunos não evadidos.
- Em relação à percepção acadêmica sobre o início do curso os itens: “Julgo que o meu curso me permitirá me realizar profissionalmente e pretendo concluir meu curso nesta instituição”, são menores em média para alunos evadidos do que alunos não evadidos.
- A dedicação de horas de estudo para os alunos evadidos é inferior aos alunos não evadidos.
- A motivação em relação ao curso que escolheu para os alunos evadidos é inferior ao dos alunos não evadidos.

A próxima etapa consiste em verificar com a Análise Fatorial os relacionamentos entre as variáveis, e quais são os fatores que melhor descrevem os dados.

4.9 Resultados da Análise Fatorial

Na Tabela 5 temos a proporção da variância de cada variável explicada pelos fatores comuns, as comunalidades, que corresponde respectivamente às perguntas das Tabelas 8, 9, 10 do Apêndice A, exceto pelo *status*. Na 1^o coluna temos as variáveis do estudo, 2^o coluna representa os autovalores e na 3^o coluna temos o percentual de cada variável explicada pelo modelo fatorial. A comunalidade para uma determinada variável pode ser interpretada como a proporção de variação naquela variável explicada por todos os fatores. Cerca de 70,2% da variação da variável *forma_ingresso* é explicada pelo modelo fatorial.

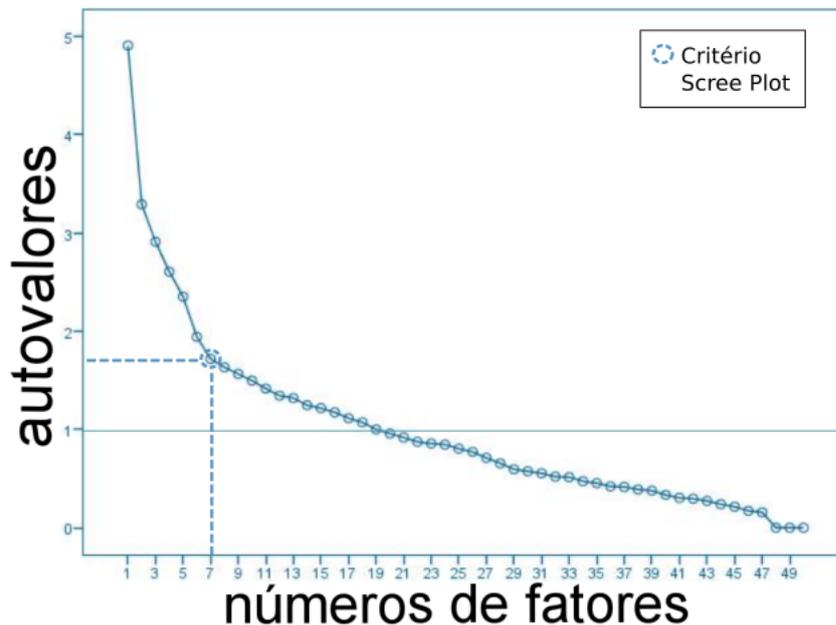
Observa-se na Tabela 6 a porcentagem de variância explicada por cada fator. Tivemos um total de 70,649% de variabilidade das variáveis originais, ou seja, 70,649% da variância total é explicada por 19 fatores.

Tabela 5 – Comunalidades

Variáveis do estudo	Inicial	Extração
renda	1,000	0,530
dist_pra_universidade	1,000	0,716
afinidade_ling_cod_tec	1,000	0,611
afinidade_ciencias_hum_tec	1,000	0,710
afinidade_matematica_tec	1,000	0,704
afinidade_ciencias_nat_tec	1,000	0,634
tempo_estudo	1,000	0,618
motivacao_escolha_curso	1,000	0,593
conhec_curso_atv_exerc	1,000	0,675
conhec_curso_nivel_dedic	1,000	0,663
conhec_curso_disc_ofert	1,000	0,648
conhec_curso_comp_hab_req	1,000	0,703
conhec_curso_mercado	1,000	0,599
vida_acad_boa_comp_para_ti	1,000	0,633
vida_acad_realizacao_prof	1,000	0,661
vida_acad_concluir_curso	1,000	0,685
vida_acad_n_mudaria_curso	1,000	0,681
status	1,000	0,540
sexo	1,000	0,506
estado_civil	1,000	0,841
tem_filhos	1,000	0,691
moradia_familia	1,000	0,879
moradia_amigos	1,000	0,877
moradia_sozinho	1,000	0,759
moradia_parentes	1,000	0,713
moradia_outros	1,000	0,842
residindo_zona_urbana	1,000	0,972
residindo_zona_rural	1,000	0,972
transporte_pra_universidade	1,000	0,676
necessidade_especial_va	1,000	0,607
necessidade_especial	1,000	0,714
escola_fundamental	1,000	0,686
escola_medio	1,000	0,797
escola_medio_prof	1,000	0,604
forma_ingresso	1,000	0,702
trabalha_na_area	1,000	0,649
trabalha_em_outras_areas	1,000	0,786
atv_extrac_religioso	1,000	0,543
atv_extrac_art_cultural	1,000	0,718
atv_extrac_poli_social	1,000	0,505
atv_extrac_esportivo	1,000	0,636
esc_curso_fat_identificacao	1,000	0,794
esc_curso_fat_falta_opcao	1,000	0,806
esc_curso_fat_uni_publica	1,000	0,696
curso_es	1,000	0,808
curso_si	1,000	0,805
curso_rc	1,000	0,790
curso_cc	1,000	0,821
curso_ec	1,000	0,764
curso_dd	1,000	0,761

Tabela 6 – Variância total explicada

Fator	Autovalores iniciais			Soma das cargas ao quadrado da rotação	
	Total	% da variância	% acumulado	% da variância	% acumulado
1	4,907	9,813	9,813	6,010	6,010
2	3,290	6,580	16,394	5,673	11,683
3	2,911	5,822	22,215	5,593	17,276
4	2,606	5,211	27,426	5,496	22,772
5	2,351	4,703	27,426	5,103	27,875
6	1,942	3,885	36,014	4,268	32,143
7	1,718	3,436	39,450	4,174	36,318
8	1,632	3,265	42,715	3,147	39,465
9	1,566	3,131	45,846	3,052	42,517
10	1,497	2,994	45,846	3,010	45,528
11	1,416	2,832	51,672	2,961	48,488
12	1,344	2,688	54,359	2,887	51,375
13	1,320	2,639	56,999	2,843	54,218
14	1,246	2,639	59,491	2,824	57,043
15	1,217	2,434	61,925	2,818	59,861
16	1,174	2,348	64,273	2,805	62,666
17	1,115	2,229	66,502	2,761	65,427
18	1,072	2,145	68,647	2,708	68,135
19	1,001	2,002	70,649	2,514	70,649

Figura 4 – *Scree Plot*

Fonte: O Autor

Observamos na Tabela 6 que a variância explicada por cada fator está na ordem do maior para o menor. Na Figura 4 observamos essa representação, em que no eixo x temos os fatores, e no eixo y os autovalores. O critério de autovalor acima de 1 resulta em 19 fatores, com variância total explicada de 70,649%. Entretanto, adotamos o critério do *scree plot*, conforme

Tabela 7 – Matriz rotacionada dos fatores

	Fator						
	1	2	3	4	5	6	7
conhec_curso_atv_exerc	0,751						
conhec_curso_nivel_dedic	0,645						
conhec_curso_disc_ofert	0,751						
conhec_curso_comp_hab_req	0,784						
conhec_curso_mercado	0,394						
vida_acad_boa_comp_para_ti	0,341		0,515				
renda		0,615					
moradia_sozinho		0,548					
escola_fundamental		0,723					
escola_medio		0,861					
escola_medio_prof		-0,419					
forma_ingresso		0,636					
tempo_estudo			0,357				
motivacao_escolha_curso			0,515				
vida_acad_realizacao_prof			0,735				
vida_acad_concluir_curso			0,739				
vida_acad_n_mudaria_curso			0,726				
<i>status</i>			-0,383				
estado_civil				0,904			
tem_filhos				0,786			
moradia_outros				0,905			
curso_rc				0,337	0,370		
dist_pra_universidade					0,719		
moradia_familia					0,871		
moradia_amigos					-0,793		
residindo_zona_urbana						0,963	
residindo_zona_rural						-0,963	
afinidade_ling_cod_tec							0,462
afinidade_ciencias_nat_tec							-0,432
sexo							0,580
curso_dd							0,663

observado na Figura 4, resultando em apenas 7 fatores decritos na Tabela 7, devido à mudança brusca de variância explicada entre o 7^o e 8^o fatores, com uma variância total explicada de 36,3%.

Para uma melhor explicação dos fatores utilizamos o método de rotação Varimax sugerido por (KAISER, 1958), que consiste em concentrar os coeficientes significantes em poucas variáveis, facilitando a interpretação dos fatores. A Tabela 7 representa os 7 primeiros fatores extraídos após a rotação *varimax*, que juntos representam 36,318% da variação dos dados. Para denominar os fatores foi utilizado um dos seguintes critérios: o fator foi denominado de acordo com uma das variáveis que apresentam maior coeficiente, ou o fator foi denominado segundo a sua característica geral (combinação de várias variáveis). A seguir uma descrição de cada fator.

O Fator 1 (chamado “Conhecimento da área e do curso”), é formado por 7 variáveis,

e agrupou variáveis que representam o conhecimento em relação ao curso escolhido e a percepção sobre a vida acadêmica no início do curso, como: as atividades e tarefas exercidas durante o curso, o nível de dedicação exigido, as disciplinas ofertadas, as competências e habilidades requeridas, as condições do mercado de trabalho e as competências para a área de tecnologia da informação.

O Fator 2 (chamado “Poder aquisitivo”), dá maior peso para: morar sozinho, ter estudado em escola particular, não ter feito profissionalizante, e ter ingressado na Universidade por ampla concorrência. Ou seja, características de alunos com maior poder aquisitivo.

O Fator 3 (chamado “Motivação em fazer o curso”), é formado pelos indicadores: ter boas competências para a área de tecnologia da informação, tempo de estudo maior, motivação em relação ao curso escolhido, boa percepção sobre a vida acadêmica no início do curso, tais como, “o curso permitir realização profissional ao aluno, conclusão do curso e não mudança de curso”, além do *status* do aluno (evadido ou não) marcado como não evadido.

O Fator 4 (chamado “Casado e/ou com filhos”), é composto por: estado civil (casado), ter filhos, morando com o marido/esposa ou companheiro/companheira, e ser do curso de redes de computadores.

No Fator 5 (chamado “Família distante de Quixadá”), temos alunos que moram com a família e moram longe da Universidade, e com maior chance de ser do curso de Redes de Computadores.

O Fator 6 (chamado “Zona Urbana x Zona Rural”), diferencia quem mora em zona urbana de quem mora em zona rural.

O Fator 7 (chamado “Alunos do curso de Design Digital”), permite diferenciar alunos de Design Digital dos outros cursos. São alunos que em geral são do sexo feminino, tem menos afinidade com ciências da natureza e suas tecnologias, e mais afinidade com linguagens e códigos e suas tecnologias, além de serem de alunos do curso de Design Digital.

Os demais fatores estão nas Tabelas 11 e 12 do Apêndice B.

5 CONCLUSÃO

Nesse trabalho foi aplicado técnicas estatísticas para tentar descrever os dados do questionário aplicado aos alunos ingressantes no ano de 2015. No estudo foi encontrado 7 fatores que descrevem os dados e explicam 36,31% da variância total. Através da Análise de Perfil foi comprovado que os alunos evadidos e não evadidos respondem o questionário de forma diferente. Após obtermos o resultado da Análise de Perfil o próximo passo foi verificar para quais perguntas os alunos evadidos e não evadidos respondem de forma diferente. Foi feito um Teste de Hipótese para Média, para verificar quais perguntas os grupos de alunos respondem diferentemente. Com o teste foi obtido um total de 7 perguntas que diferenciam os perfis de alunos, ou seja, essas perguntas tem significância estatística.

Com o resultado do teste foi obtido que os alunos evadidos moram em média 55,57% mais distantes que os alunos não evadidos. Foi identificado que os alunos evadidos: se identificam no curso em média 44,48% menos que os alunos não evadidos. Estudam em média 23,83% menos. Acreditam que o curso o permitirá ter uma realização profissional em média 23,21% menos que alunos não evadidos. Tem em média uma motivação de 17,54% a menos que os alunos não evadidos. Pretende concluir o curso 20,36% em média menos que alunos não evadidos. Contudo os alunos evadidos escolheram o curso por influência dos pais, familiares, amigos, por conta do mercado promissor em média a 150% mais que alunos não evadidos. Dentre as 7 perguntas, na Tabela 2, cinco delas estão nos fatores mais importantes encontrados (em termos de variância explicada), ou seja, as 2 perguntas não inclusas nos 7 primeiros fatores não tiveram significância estatística.

A pergunta 35 não se encontra em nenhum dos fatores, e a pergunta 19, (“Tomando como base seu local de moradia atual, qual a distância de onde reside para a universidade?”) encontra-se no Fator 5 (chamado “Família distante de Quixadá”). Já as perguntas 41 (“Quantas horas diárias você dedica aos estudos fora do horário de aula?”), 42 (“Como você avalia sua motivação em relação ao curso que escolheu?”), 49 (“De acordo com sua percepção sobre sua vida acadêmica neste início de curso, assinale os itens a seguir: [Julgo que o meu curso me permitirá me realizar profissionalmente.]”) e 50 (“De acordo com sua percepção sobre sua vida acadêmica neste início de curso, assinale os itens a seguir: [Pretendo concluir meu curso nesta instituição.]”) encontram-se no Fator 3 (chamado “Motivação em fazer o curso”).

Com a Análise Fatorial foi agrupado perguntas mais correlacionadas em fatores, ou seja, um fator agrupa um conjunto de perguntas que melhor se relaciona. Tivemos um total de 7

fatores que se mostraram mais importante no estudo, sendo que o Fator 3 foi o que se destacou, pois esse Fator agrupou um número maior de perguntas significantes obtidas através do Teste de Hipótese para Média. Assim que o Fator 3 foi o que se mostrou mais significante para tentar prever o aluno que irá evadir. Para tentar prever qual aluno irá evadir, o mais importante como vimos na análise é ver como o aluno marcou a motivação dele em fazer o curso, confirmando assim a relação entre o Fator 3 e o *status*, mostrando que tais fatores possuem as respostas significantes para diferenciar os perfis de alunos e portanto podem ser indicadores da evasão escolar. Em especial o Fator 3, que tem o *status* do aluno, é um grande indicador para a evasão, sugerindo que o aluno se que declara mais motivado tende a evadir menos.

6 CONSIDERAÇÕES FINAIS

Várias conclusões adicionais poderiam ser extraídas dos dados. Como trabalhos futuros, sugere-se repetir as análises feitas nesse trabalho considerando os dados de anos posteriores. Para esse propósito é necessário que, por exemplo, o questionário esteja o mais padronizado possível para que análises já feitas possam ser consideradas para estudos posteriores. Ficou em aberto fazer uma análise confirmatória dos dados, por exemplo, com a aplicação da Análise Fatorial Confirmatória, (HAIR et al., 2009). Assim seria possível verificarmos o quão próximo a especificação dos fatores combina com a realidade (os dados verdadeiros).

No estudo de SILVA (2014), foi feito um questionamento em relação à falta de estudos relacionados à evasão na UFC-Quixadá. No presente trabalho buscamos entender melhor o problema através dos dados da UFC-Quixadá. Assim como SILVA (2014), concordamos com a criação de um sistema integrado aos sistemas já existentes na universidade para mostrar as análises geradas através de estudos como este.

REFERÊNCIAS

- BARBETTA, P. A.; REIS, M. M.; BORNIA, A. C. **Estatística: para cursos de engenharia e informática**. [S.l.]: Atlas São Paulo, 2004. v. 3.
- BEAN, J. P. Dropouts and turnover: The synthesis and test of a causal model of student attrition. **Research in higher education**, Kluwer Academic Publishers, v. 12, n. 2, p. 155–187, 1980.
- CARDOSO, C. B. Efeitos da política de cotas na universidade de Brasília: uma análise do rendimento e da evasão. 2008.
- CORRAR, L. J.; PAULO, E.; FILHO, J. M. D. Análise multivariada: para os cursos de administração, ciências contábeis e economia. **São Paulo: Atlas**, p. 280–323, 2007.
- FILHO, R. L. L. S.; MOTEJUNAS, P. R.; HIPÓLITO, O.; LOBO, M. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, SciELO Brasil, v. 37, n. 132, p. 641–659, 2007.
- HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. [S.l.]: Bookman Editora, 2009.
- HÄRDLE, W.; SIMAR, L. **Applied multivariate statistical analysis**. [S.l.]: Springer, 2007. v. 22007.
- JOHNSON, R. A.; WICHERN, D. W. et al. **Applied multivariate statistical analysis**. [S.l.]: Prentice hall Upper Saddle River, NJ, 2002. v. 5.
- JUNIOR, J. S. S.; BRASIL, G. H.; CARNEIRO, T. C. J.; CORASSA, M. A. de C. Análise estatística da evasão na universidade federal do Espírito Santo e uma avaliação de seus determinantes. 2015.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, Springer, v. 23, n. 3, p. 187–200, 1958.
- MACHADO, O. A. Evasão de alunos de cursos superiores: fatores motivacionais e de contexto. **Evasão de alunos de cursos superiores: fatores motivacionais e de contexto**, 2005.
- OLKIN, I. **Contributions to probability and statistics: essays in honor of Harold Hotelling**. [S.l.]: Stanford University Press, 1960.
- PARMIGIANI, G.; GARRETT, E. S.; IRIZARRY, R. A.; ZEGER, S. L. The analysis of gene expression data: an overview of methods and software. In: **The analysis of gene expression data**. [S.l.]: Springer, 2003. p. 1–45.
- PEÑA, D. **Análisis de datos multivariantes**. [S.l.]: McGraw-Hill Madrid, 2002. v. 24.
- PEREIRA, F. C. B.; BRASIL, G. H.; SAMOBYL, R. W. Análise fatorial e a evasão de alunos no ensino superior. 2004.
- PINTRICH, P. R. et al. A manual for the use of the motivated strategies for learning questionnaire (mslq). ERIC, 1991.
- R, T. R. F. The R project for statistical computing, R version 3.3.1. **The R Foundation**, 2016.
- ROCHA, M. V. G. Análise fatorial com o método de extração análise de componentes principais aplicados ao consumo de combustível. 2014.

SILVA, A. S. d. et al. Retenção ou evasão: a grande questão social das instituições de ensino superior. Universidade Presbiteriana Mackenzie, 2014.

SILVA, T. J. D. Extração de conhecimento nos dados da universidade federal do ceará via mineração de dados: Descoberta e análise dos perfis dos alunos. 2014.

SPADY, W. G. Dropouts from higher education: An interdisciplinary review and synthesis. **Interchange**, Springer, v. 1, n. 1, p. 64–85, 1970.

TINTO, V. **Student success and the construction of inclusive educational communities.[SI]: American Association of State Colleges and Universities-AASCU, 2005.** 2014.

UNIVERSITY, T. P. S. **Factor Analysis.** 2004. Disponível em:
<http://sites.stat.psu.edu/~ajw13/stat505/fa06/17_factor/13_factor_varimax.html>.

WEBB, A. R. **Statistical pattern recognition.** [S.l.]: John Wiley & Sons, 2003.

APÊNDICE A – GLOSSÁRIO

Análise Multivariada Análise de múltiplas variáveis de forma simultânea, observando suas relações intrínsecas.

Autovalor Soma em coluna de cargas fatoriais ao quadrado para um fator. Representa a quantia de variância explicada por um fator, (Hair et al. (2009)).

Cargas Fatoriais São as correlações entre as variáveis originais e os fatores (Hair et al. (2009)). Quanto maior a carga fatorial, maior a correlação com determinado fator seja ela negativamente ou positivamente.

Estatística F Teste estatístico que avalia a hipótese nula de médias iguais entre duas amostras independentes. Para determinar se a estatística F é suficientemente grande para justificar a hipótese nula, ou seja, aceitação de diferenças entre as médias dos grupos, basta verificar se o correspondente nível de significância é aceitável (Hair et al. (2009)).

Estatística t Teste estatístico que avalia a significância estatística entre dois grupos em uma única variável dependente. O quão grande deve ser o valor do teste *t* para ser considerada estatística das médias dos grupos.

Matriz de Correlação Tabela que mostra as intercorrelações entre todas as variáveis (Hair et al. (2009)). É uma matriz quadrada em que os elementos são as correlações entre as variáveis analisadas.

Significância estatística É uma medida estimada do grau em que este resultado é verdadeiro. Está relacionada ao nível de confiança ao rejeitar uma hipótese quando esta na verdade é verdadeira.

Teste t É um teste de hipótese que segue uma distribuição *T de Student*. O Teste t avalia a significância estatística da diferença entre duas médias amostrais independentes para uma única variável dependente (Hair et al. (2009)).

Teste de Levene para igualdade de variâncias O teste de Levene é utilizado para testar se *k* amostras tem variâncias iguais, se são homogêneas ou não (Olkin (1960)).

Varimax Um método de rotação ortogonal, concentrando-se na simplificação das colunas em uma matriz fatorial. Geralmente considerado superior a outros métodos de rotação fatorial ortogonal para conseguir uma estrutura fatorial simplificada (Hair et al. (2009)).

Variância A variância de uma variável aleatória é uma medida da variabilidade, indicando "o quão longe" em geral os seus valores se encontram do valor esperado.

Variabilidade É uma medida de dispersão que mostra o quão esticada ou espremida é uma distribuição. Exemplos comuns de medidas de dispersão estatística são a variância, o desvio padrão.

APÊNDICE B – APÊNDICES

B.1 Perguntas do Estudo

Tabela 8 – Questões do estudo 1

Núm.	Nome da Variável / Pergunta
1	sexo - Qual seu sexo ?
2	estado_civil - Qual o seu estado civil? *
3	curso_es - Você está matriculado no curso de Engenharia de Software ?
4	curso_si - Você está matriculado no curso de Sistema de Informação ?
5	curso_rc - Você está matriculado no curso de Redes de Computadores ?
6	curso_cc - Você está matriculado no curso de Ciência da Computação ?
7	curso_ec - Você está matriculado no curso de Engenharia da Computação ?
8	curso_dd - Você está matriculado no curso de Design Digital ?
9	opc_enem - Quais foram a 1e a 2opção de cursos que você escolheu no ENEM/Sisu, respectivamente?
10	tem_filhos - Tem filhos?
11	renda - Qual a renda mensal de sua família?
12	moradia_familia - Você está morando atualmente com a família?
13	moradia_amigos - Você está morando atualmente com amigos?
14	moradia_sozinho - Você está morando atualmente sozinho?
15	moradia_parentes - Você está morando atualmente com parentes?
16	moradia_companheiro - Você está morando atualmente com sua esposa/marido ou companheiro/companheira ?
17	zona_urbana - Você está residindo na zona urbana ?
18	zona_rural - Você está residindo na zona rural ?
19	dist_pra_universidade - Tomando como base seu local de moradia atual, qual a distância de onde reside para a universidade?
20	meio_de_transporte - Qual meio de transporte você utiliza para chegar à universidade?
21	necessidade_esp_v_a - Você apresenta alguma necessidade educacional especial, visual/atenção?
22	necessidade_esp_v - Você apresenta alguma necessidade educacional especial visual?
23	ensino_fundamental - Que tipo de escola você cursou o ensino fundamental?
24	ensino_medio - Que tipo de escola você cursou o ensino médio?

Tabela 9 – Questões do estudo 2

Núm.	Nome da Variável / Pergunta
25	ensino_medio_prof - Você cursou o o ensino médio em escola publico profissional na área de informática ?
26	forma_ingresso - Qual sua forma de ingresso na UFC-Quixadá?
27	trabalha_na_area - Você trabalha ou exerce alguma atividade remunerada (emprego) na área de TI?
28	trabalha_em_outras_areas - Você trabalha ou exerce alguma atividade remunerada (emprego) em outras áreas que não seja a de TI?
29	atv_extrac_art_cultural - Fora da universidade, você participa de algum tipo de atividade de cunho religioso ?
30	atv_extrac_religioso - Fora da universidade, você participa de algum tipo de atividade de cunho artístico/cultural ?
31	atv_extrac_poli_social - Fora da universidade, você participa de algum tipo de atividade de cunho político/social ?
32	atv_extrac_esportivo - Fora da universidade, você participa de algum tipo de atividade de cunho recreativo/esportivo ?
33	esc_curso_fat_identificacao - O principal fator que determinou sua escolha por este curso: é uma área com a qual me identificação ?
34	esc_curso_fat_falta_opcao - O principal fator que determinou sua escolha por este curso: não havia na cidade onde eu moro o curso do meu interesse/ foi o curso que consegui aprovação com minha nota do ENEM ?
35	esc_curso_fat_uni_publica - O principal fator que determinou sua escolha por este curso: por influência dos pais, dos amigos e de familiares/ por conta de um mercado de trabalho promissor ?
36	esc_curso_fat_inf_a_fam - O principal fator que determinou sua escolha por este curso: por influência dos pais, dos amigos e de familiares ?
37	afinidade_ling_cod_tec - Indique em que medida você tem afinidade com as áreas a seguir: [Linguagens, códigos e suas tecnologias.]
38	afinidade_ciencias_hum_tec - Indique em que medida você tem afinidade com as áreas a seguir: [Ciências humanas e suas tecnologias.]
39	afinidade_matematica_tec - Indique em que medida você tem afinidade com as áreas a seguir: [Matemática e suas tecnologias.]
40	afinidade_ciencias_nat_tec - Indique em que medida você tem afinidade com as áreas a seguir: [Ciências da natureza e suas tecnologias.]

Tabela 10 – Questões do estudo 3

Núm.	Nome da Variável / Pergunta
41	tempo_estudo - Quantas horas diárias você dedica aos estudos fora do horário de aula?
42	motivacao_escolha_curso - Como você avalia sua motivação em relação ao curso que escolheu?
43	conhec_curso_atv_exerc - Indique em que medida você conhece sobre o curso de sua escolha: [As atividades e tarefas exercidas durante o curso.]
44	conhec_curso_nivel_dedic - Indique em que medida você conhece sobre o curso de sua escolha: [O nível de dedicação exigido.]
45	conhec_curso_disc_ofert - Indique em que medida você conhece sobre o curso de sua escolha: [As disciplinas ofertadas.]
46	conhec_curso_comp_hab_req - Indique em que medida você conhece sobre o curso de sua escolha: [As competências e habilidades requeridas.]
47	conhec_curso_mercado - Indique em que medida você conhece sobre o curso de sua escolha: [As condições do mercado de trabalho (salário, empregabilidade, etc).]
48	vida_acad_boa_comp_para_ti - De acordo com sua percepção sobre sua vida acadêmica neste início de curso, assinale os itens a seguir: [Tenho boas competências para a área de tecnologia da informação.]
49	vida_acad_realizacao_prof - De acordo com sua percepção sobre sua vida acadêmica neste início de curso, assinale os itens a seguir: [Julgo que o meu curso me permitirá me realizar profissionalmente.]
50	vida_acad_concluir_curso - De acordo com sua percepção sobre sua vida acadêmica neste início de curso, assinale os itens a seguir: [Pretendo concluir meu curso nesta instituição.]
51	vida_acad_n_mudaria_curso - De acordo com sua percepção sobre sua vida acadêmica neste início de curso, assinale os itens a seguir: [Mesmo que pudesse, não mudaria de curso.]

APÊNDICE C – APÊNDICES

C.1 Fatores

A seguir na Tabela 11 temos os fatores de 6 a 12, já na Tabela 12 temos os fatores de 13 a 19.

Tabela 11 – Matriz rotacionada dos componentes

	Componente						
	6	7	8	9	10	11	12
conhecimento_curso_mercado							-0,542
tempo_estudo					0,304		
residindo_zona_urbana	0,963						
residindo_zona_rural	-0,963						
afinidade_ling_cod_tec		-0,462		0,451			
afinidade_ciencias_nat_tec		0,432				0,386	
sexo		-0,580					
curso_dd		-0,663					
escolha_curso_fator_falta_opcao			-0,86				
afinidade_ciencias_hum_tec				0,803			
transporte_pra_universidade				-0,370			
atividade_extracurricular_religioso					-0,488		
curso_es					-0,821		
curso_ec					0,358		
moradia_parentes						-0,814	
atividade_extracurricular_politico_social							0,628

Tabela 12 – Matriz rotacionada dos Fatores

	Fator						
	13	14	15	16	17	18	19
forma_ingresso	-0,237						
trabalha_em_outras_areas	0,825						
transporte_pra_universidade	-0,385					0,405	
necessidade_especial_va							-0,749
necessidade_especial						0,804	
atividade_extracurricular_art_cultural		0,792					
atividade_extracurricular_esportivo			0,534	0,333			
escolha_curso_fator_identificacao				0,356			
escolha_curso_fator_uni_publica				-0,764			
curso_si		-0,836					
curso_cc					0,786		
curso_ec					-0,539		

ANEXO A – QUESTIONÁRIO

A.1 Questionário

Caro(a) aluno(a), a UFC do Campus Quixadá tem o interesse em conhecer o perfil dos alunos que ingressaram em seus cursos, para fins de otimizarmos nossas ações no sentido de melhor atender às demandas dos estudantes. Sua colaboração é de fundamental importância para o aprimoramento do curso e da UFC Campus Quixadá.

*Obrigatório

A.1.1 *Levantamento das expectativas e interesses dos alunos em relação ao curso*

1. Número de Matrícula: *
2. Cite a cidade que você morava antes de estudar na UFC Campus Quixadá:
3. Endereço completo ATUAL (Onde você está morando? RUA, NÚMERO, BAIRRO, CIDADE):
4. Telefone:
5. Email:
6. Curso em que você está matriculado: *
 - Engenharia de software
 - Redes de computadores
 - Ciência da computação
 - Design digital
 - Sistemas da informação
 - Engenharia de computação
7. Quais foram a primeira e a segunda opção de cursos que você escolheu no ENEM/Sisu, respectivamente? *
8. Como você tomou conhecimento do curso de sua escolha aqui na UFC-Quixadá?
*
 - Em sites especializados.
 - No site da universidade.

- Soube no momento da minha inscrição no SISU.
- Jornal/TV
- Feira das profissões
- Escola e cursinhos
- Obtive conhecimento através de amigos e familiares.
- Outro:

9. Sexo: *

- Masculino
- Feminino

10. Qual o seu estado civil? *

- Solteiro (a)
- Casado(a) ou União estável
- Separado(a) ou Divorciado
- Viúvo(a)

11. Tem filhos? *

- Não
- Sim

12. Qual a renda mensal de sua família? * Considere todo tipo de renda (seja com aluguel, benefícios, salários, etc) de todas as pessoas que moram em sua casa.

- Até 1000 reais
- De 1001 até 2000 reais
- De 2001 até 3000 reais
- De 3001 até 4000 reais
- De 4001 até 5000 reais
- De 5001 até 6000 reais
- Acima de 6000 reais

13. Com quem você está morando atualmente? *

- Família (pai, mãe, irmãos...)

- Amigos
- Parentes
- Sozinho
- Com sua esposa/marido ou companheiro(a)

14. Você está residindo em: *

- Zona urbana
- Zona rural

15. Tomando como base seu local de moradia atual, qual a distância de onde reside para a universidade? *

- Até 10 km
- Entre 10 km e 25 km
- Entre 25 km e 50 km
- Entre 50 km e 100 km
- Acima de 100 km

16. Qual meio de transporte você utiliza para chegar à universidade? *

- Automóvel particular
- Transporte alternativo (táxi, mototáxi, etc)
- Ônibus universitário
- Bicicleta
- Outro

17. Você apresenta alguma necessidade educacional especial? *

- Não
- Sim. Auditiva
- Sim. Visual
- Sim. Física
- Outro

A.1.2 Experiência escolar e profissional.

1. Que tipo de escola você cursou o ensino fundamental? *

- Pública
- Particular
- Pública e particular
- Estrangeira

2. Que tipo de escola você cursou o ensino médio? *

- Pública Tradicional
- Pública - Profissionalizante na Área de Informática
- Particular
- Particular e pública
- Estrangeira

3. Qual sua forma de ingresso na UFC Campus Quixadá? *

- Cota L1 (Candidatos que cursaram todo o ensino médio em escolas públicas, com renda familiar bruta per capita igual ou inferior a 1,5 salário mínimo)
- Cota L2 (Candidatos que cursaram todo o ensino médio em escolas públicas, com renda familiar bruta per capita igual ou inferior a 1,5 salário mínimo, autodeclarados pretos, pardos ou indígenas)
- Cota L3 (Candidatos que cursaram todo o ensino médio em escolas públicas, INDEPENDENTE DA RENDA)
- Cota L4 (Candidatos que cursaram todo o ensino médio em escolas públicas, INDEPENDENTE DA RENDA, autodeclarados pretos, pardos ou indígenas)
- ENEM/Sisu - Cota para pessoa com deficiência
- Ampla concorrência
- Transferência
- Admissão de graduado

- Aluno estrangeiro
4. Você trabalha ou exerce alguma atividade remunerada (emprego)? * (Bolsa NÃO conta).
- Sim, na área de tecnologia da informação ou afins.
 - Sim, mas NÃO na área de tecnologia da informação ou afins.
 - Não exerço nenhuma atividade remunerada.
5. Fora da universidade, você participa de algum tipo de atividade, organização e associação? * Pode assinalar mais de uma opção.
- Não participo de nenhuma atividade.
 - Grupo de bairro ou associação comunitária.
 - Igreja ou grupo religioso.
 - Partido político ou movimento estudantil.
 - ONG ou movimento social.
 - Clube recreativo ou associação esportiva.
 - Grupo cultural (dança, teatro, música, coral, etc).
 - Outro

A.1.3 Motivações para o ingresso no curso.

1. Indique o principal fator que determinou sua escolha por este curso: *
- É uma área com a qual me identifico.
 - Não havia na cidade onde moro o curso de meu real interesse.
 - Por influência dos pais, dos amigos e de familiares.
 - Foi o curso que consegui aprovação com minha nota do ENEM.
 - Por conta de um mercado de trabalho promissor.
 - Pelo fato de a instituição de ensino ser uma universidade pública.
 - Outro
2. Indique em que medida você tem afinidade com as áreas a seguir: * (Pouquíssima afinidade, Alguma afinidade, Boa afinidade, Muita afinidade)
- Linguagens, códigos e suas tecnologias.

- Ciências humanas e suas tecnologias.
 - Matemática e suas tecnologias.
 - Ciências da natureza e suas tecnologias.
3. Quantas horas diárias você dedica aos estudos fora do horário de aula? *
- Até 2 horas
 - De 2 a 3 horas
 - De 3 a 4 horas
 - De 4 a 6 horas
 - Acima de 6 horas
 - Não tenho tempo disponível além dos horários das aulas.
4. Como você avalia sua motivação em relação ao curso que escolheu? *
- Totalmente desmotivado.
 - Pouco motivado.
 - Motivado.
 - Muito motivado.

A.1.4 Expectativas em relação à conclusão do curso e à inserção no mercado de trabalho.

1. Indique em que medida você conhece sobre o curso de sua escolha: *
(Pouquíssimo conhecimento, Algum conhecimento, Conheço bem, Conheço muito bem)
- As atividades e tarefas exercidas durante o curso.
 - O nível de dedicação exigido.
 - As disciplinas ofertadas.
 - As competências e habilidades requeridas.
 - As condições do mercado de trabalho (salário, empregabilidade, etc).
2. De acordo com sua percepção sobre sua vida acadêmica neste início de curso, assinale os itens a seguir: * (Nunca se verifica, Poucas vezes se verifica, Verifica-se com frequência, Verifica-se sempre)

- Tenho boas competências para a área de tecnologia da informação.
- Julgo que o meu curso me permitirá me realizar profissionalmente.
- Pretendo concluir meu curso nesta instituição.
- Mesmo que pudesse, não mudaria de curso.

3. Assinale DOIS fatores que você acredita que o ajudarão a concluir o curso com êxito: *

- O conhecimento que os professores têm das matérias e a habilidade em ensiná-las.
- O apoio da família.
- Auxílio moradia (benefício oferecido pela UFC aos alunos de baixa renda para custearem sua moradia na cidade).
- Alimentação no refeitório universitário gratuita.
- Esforço e dedicação pessoal para com os estudos e a realização dos compromissos escolares.
- Envolvimento em atividades de ensino, pesquisa e extensão.
- Experiência de estágio na área.
- Auxílio creche.
- Outro:

4. Assinale as DUAS principais dificuldades que você percebe influenciar a conclusão do curso: *

- Dificuldade de Transporte
- Problemas Financeiros
- Problemas Familiares
- Problemas Pessoais
- Dificuldade de Socialização
- Dificuldade de Organizar Horários de Estudos
- Dificuldade de Concentração
- Dificuldade com o aprendizado da matemática
- Dificuldade com o aprendizado de programação

- Outro

5. Quais suas expectativas com a conclusão de seu curso? *

- Gostaria de conseguir um bom emprego.
- Gostaria de prestar um concurso e trabalhar no setor público.
- Gostaria de ganhar dinheiro com o meu próprio negócio.
- Gostaria de seguir a vida acadêmica (mestrado, doutorado, etc).
- Outro

6. O que você espera do curso que está iniciando? Quais suas expectativas? *